The Annals of Statistics
1998, Vol. 26, No. 4, 1279–1305

INTERACTIONS AND OUTLIERS IN THE TWO-WAY ANALYSIS OF VARIANCE

BY WOLFGANG TERBECK AND P. LAURIE DAVIES

University of Essen

The two-way analysis of variance with interactions is a well established and integral part of statistics. In spite of its long standing, it is shown that the standard definition of interactions is counterintuitive and obfuscates rather than clarifies. A different definition of interaction is given which among other advantages allows the detection of interactions even in the case of one observation per cell. A characterization of unconditionally identifiable interaction patterns is given and it is proved that such patterns can be identified by the L^1 functional. The unconditionally identifiable interaction patterns describe the optimal breakdown behavior of any equivariant location functional from which it follows that the L^1 functional has optimal breakdown behavior. Possible lack of uniqueness of the L^1 functional can be overcome using an M functional with an external scale derived independently from the observations. The resulting procedures are applied to some data sets including one describing the results of an interlaboratory test.

1. Introduction.

1.1. *A simple example.* The standard model of the two-way layout with interactions is often written in the form

$$X_{ijk} = \mu + a_i + b_j + c_{ij} + \varepsilon_{ijk}, \qquad 1 \le k \le K_{ij},\ 1 \le i \le I,\ 1 \le j \le J. \tag{1}$$

The model (1) is overparameterized and to avoid this, the following restrictions are conventionally placed on the row and column effects and the interactions, respectively:

$$\sum_i a_i = \sum_j b_j = 0 \tag{2}$$

and

$$\sum_i c_{ij} = \sum_j c_{ij} = 0 \tag{3}$$

There is no reason to accept any of these restrictions. In this paper, the restrictions (2) play no role. Only (3) is of interest and we claim it is counterintuitive as may be seen in the following example. For the sake of simplicity, we set $K_{ij} = 1$ and the noise ε to zero. Three different therapies are given to three

Received November 1996; revised January 1998.

AMS 1991 *subject classifications*. Primary 62J10; secondary 62F35.

Key words and phrases. Analysis of variance, interactions, outliers, breakdown patterns, robust statistics, L^1 functional, M functional.

different groups of patients. Therapy A causes an improvement in Group 1, but there are no other effects. The results are summarized by the table

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{4}$$

The obvious interpretation of these results is that there is an interaction between Therapy A and Group 1 and that is all. However, if we write (4) in the form (1) subject to (2) and (3), we obtain a main effect, row and column effects and the interaction pattern

$$\begin{pmatrix} 4/9 & -2/9 & -2/9 \\ -2/9 & 1/9 & 1/9 \\ -2/9 & 1/9 & 1/9 \end{pmatrix}. \tag{5}$$

We now have interactions everywhere and the original clear and simple interpretation has been lost. This example was also known to Daniel [7] and is of itself sufficient to discredit the usual definition of interaction. It is also discussed in [20], pages 178–180. A practical example with the same structure is Table 5.6 of Cochran and Cox ([6], page 164). It has been analyzed by Daniel ([7], Sections 4.3 and 4.7) and Hampel, Ronchetti, Rousseeuw and Stahel ([18], Section 1.1d).

We note that in the context of this paper there is no difference between an interaction and an outlier, and we shall use both terminologies. Tukey [36] calls such observations exotic. Failure to detect interactions is equivalent to failure to detect outliers, and in the context of outliers, such a failure is referred to as breakdown. When we therefore refer to optimal breakdown behavior, we also mean optimal detection of interactions.

We restrict attention throughout to the case of one observation per cell, that is, $K_{ij} = 1$. The reason for this is that it is the most difficult case and that the general case can be reduced to it by the simple expedient of replacing the observations in each cell by their median.

The example (4) is given in [11] as are (without proof) Corollary 2.4 and Theorem 2.11.

1.2. *Group invariance and equivariance.* The model (1) is clearly invariant with respect to the following group of operations:

PR	permute the rows
PC	permute the columns
AR	add fixed numbers to the rows
AC	add fixed numbers to the columns
IRC	interchange rows and columns
M	multiply all observations by a nonzero number

It seems reasonable to demand that any method of analyzing a two-way table should be equivariant with respect to these group operations. Most but not all methods suggested in the literature are equivariant. An exception is Tukey's median Polish [35, 20], which is not equivariant with respect to AR, AC and IRC. Terbeck [34] contains more information on this topic.

1.3. *Previous work.* The definition of interaction we give in Section 2.1 is not new and may be found in [8]. Daniel showed how the least-squares residuals may be used to detect certain patterns of outliers. In many cases Tukey's median polish correctly identifies outliers in the two-way table, but it cannot always be relied upon to do this [20]. Tukey [36] pointed out the questionability of the side conditions (3) and he also considered the problem of splitting up the residuals into a noise and an interaction part. Tukey's median polish can be shown to detect all interaction patterns described by Corollary 2.7, but it does not detect all those described by Corollary 2.8. Methods based on the differences $\Delta_{kj} = \text{med}_i\{x_{ij} - x_{ik}\}$ for the differences $b_j - b_k$, and $\delta_{st} = \text{med}_j\{x_{tj} - x_{sj}\}$ for the differences $a_t - a_s$ have been developed by Hoaglin, Mosteller and Tukey ([21], page 45). In general, these methods find all interaction patterns that satisfy either Corollary 2.7 or Corollary 2.8. They do not, however, find all unconditionally identifiable interaction patterns as defined below. More detailed information is given in [34]. Bradu [3], Bradu and Hawkins [5] and Gentleman and Wilk [15, 16] have also considered the problem of identifying multiple outliers in the two-way analysis of variance. Hubert [24] has treated the corresponding problem for two-way contingency tables. She shows that in this situation the L^1 functional has the highest possible breakdown point. A discussion of the problem of outliers in the analysis of variance can be found in [18]. He, Jurečková, Koenker and Portnoy [19], Bradu [4] and Ellis and Morgenthaler [14] considered the breakdown behavior of the L^1 functional for fixed regressors. Their work leads to necessary and sufficient conditions for a subset of the regressors to be safe for the L^1 functional. The condition is not easy to check, but it applies to any linear regression model. In the particular case of the two-way table, we are able to give a simple necessary and sufficient condition for a subset to be safe (Theorem 2.3) and, moreover, we are able to show that an unsafe subset is unsafe for any functional which is equivariant with respect to the allowable group of transformations of a two-way table as described above.

An additive structure is not the only possible structure for the two-way table. The results we give may be applied to the multiplicative model by taking logarithms. Many other structures can be constructed such as the row- and column-linear models developed by Mandel [26, 27]. They face the same problems and as yet have not been robustified. The linear model has the great advantage of simplicity, and even if it is not an adequate model, the residuals from a robust fit provide a good starting point for developing an improved model.

Huber [23] states "Embarrassingly, the robustification of the statistics of two-way tables is still wide open." We hope this paper reduces the embarrassment.

1.4. *Contents.* In Section 2 we consider the no-noise model and introduce a different definition of interaction. It is also shown that the L^1 functional has certain optimality properties with respect to breakdown or the identification of interactions. Noise is included in Section 3, where it is shown that the L^1 functional is no longer in general uniquely defined. This problem can be overcome by using an M functional which is shown to inherit the optimality properties of the L^1 functional. Section 4 contains some remarks on breakdown in the two-way table. Section 5 contains several data sets which are analyzed according to the procedures developed in this paper. Proofs are relegated to the Appendix.

2. The "no-noise" case.

2.1. *Minimizing the interactions.* The no-noise model is given by

$$x_{ij} = a_i + b_j + c_{ij}. \tag{6}$$

The definition of interaction which we propose is the following. Given a data set $X = (x_{ij})$, $1 \le i \le I$ and $1 \le j \le J$, find effects a_i and b_j so that

$$\#\{(i, j)\colon c_{ij} \neq 0\} \stackrel{!}{=} \min. \tag{7}$$

Clearly a solution always exists, so the question of interest is that of uniqueness. An example of nonuniqueness is the following, also known to Hampel: if we add -1 to the first row and 1 to the second column of the table

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{8}$$

we obtain the table

$$\begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

both of which have two interaction terms. It is not possible to reduce the number of interactions further.

If we replace (8) by

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{9}$$

then any nonzero choice of the main, column and row effects leads to an increase in the number of nonzero interactions. Thus the interactions given by

(9) represent the unique solution. These two examples show that, in general, uniqueness depends on the position and the values of the interaction terms. In some cases, however, the question of uniqueness can be reduced to that of the position alone. The simplest case is the following. The table

$$\begin{pmatrix} \alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{10}$$

is clearly the unique solution to our problem regardless of the value of α. We call such interaction patterns *unconditionally identifiable* and they are our main object of study. The reasons for this are that such patterns can be characterized, they can be detected by the L^1 functional, the unconditionally identifiable interaction patterns are linked to optimal breakdown behavior and, finally, they form a sufficiently rich class to be of use in practical problems. Indeed if there are too many interactions, then the additive model may not be adequate.

2.2. *Unconditional identifiability.* Given a data set $X = (x_{ij})$, we call the set

$$\mathscr{C}(X) = \{(c_{ij})\text{: there exist } a_i, b_j \text{ s.t. } x_{ij} = a_i + b_j + c_{ij}\}$$

the set of *possible residual matrices*. For each residual matrix $C = (c_{ij})$ we write

$$R(C) = \#\{(i, j)\text{: } c_{ij} \neq 0\}.$$

The minimization problem can be formulated as finding a matrix $C \in \mathscr{C}(X)$ with

$$R(C) = \min\{R(C')\text{: } C' \in \mathscr{C}(X)\}.$$

DEFINITION 2.1. A residual matrix C is called *identifiable* if there exists no matrix $C' \in \mathscr{C}(C)$ with $C \neq C'$ and $R(C') \leq R(C)$.

We restrict attention to those interaction matrices which are identifiable for any choice of the nonzero interactions. The pattern of the nonzero residuals is described by an $(I \times J)$ matrix with entries 0 and 1, where a 1 represents a nonzero interaction.

DEFINITION 2.2. An interaction pattern $P = (p_{ij})$ is called *unconditionally identifiable* if, for every choice $c_{ij} \neq 0$ in the cells with $p_{ij} = 1$ and $c_{ij} = 0$, whenever $p_{ij} = 0$, we get an identifiable residual matrix $C = (c_{ij})$.

2.3. *Characterization of unconditionally identifiable interaction patterns.* The main result on unconditionally identifiable interaction patterns is the following theorem.

THEOREM 2.3. *An interaction pattern P is not unconditionally identifiable if and only if either of the following statements holds.*

(a) *There exists a row or column of P which contains at least as many* 1*'s as* 0*'s.*

(b) *We can permute the rows and columns of P to obtain a matrix P'*

$$P' = \begin{pmatrix} P'_{11} & P'_{12} \\ P'_{21} & P'_{22} \end{pmatrix}, \tag{11}$$

where none of the matrices is empty and where P'_{12} and P'_{21} together contain at least as many 1*'s as* 0*'s.*

The proof is presented in the Appendix.

COROLLARY 2.4. *Let P be an interaction pattern and consider the following operations: to each row or column of P may be added a row or column of* 1*'s with addition modulus* 2*, that is, $1 + 1 = 0$. Then P is unconditionally identifiable if and only if there is no such sequence of operations which results in an interaction pattern $P' \neq P$ with $R(P') \leq R(P)$.*

COROLLARY 2.5. *If P is an unconditionally identifiable interaction pattern and P' is another interaction pattern satisfying $p'_{ij} = 0$ whenever $p_{ij} = 0$, then P' is also unconditionally identifiable.*

COROLLARY 2.6. *An $I \times J$ unconditionally identifiable interaction pattern has at most*

$$\min\left\{\left(J - \left\lfloor \frac{J-1}{2} \right\rfloor\right)\left\lfloor \frac{I-2}{2} \right\rfloor, \left(I - \left\lfloor \frac{I-1}{2} \right\rfloor\right)\left\lfloor \frac{J-2}{2} \right\rfloor\right\} + \left\lfloor \frac{I-1}{2} \right\rfloor\left\lfloor \frac{J-1}{2} \right\rfloor$$

interactions.

The upper bound of Corollary 2.6 is sharp for $I = J = 3$. If $I = J = 5$, then the maximum number of interactions is six whereas Corollary 2.6 gives seven. We refer to [34] for more information on the number of interactions in unconditionally identifiable patterns. Using Corollary 2.4 we see that we require $\min\{J2^{I-1}, I2^{J-1}\}$ operations to check unconditionally identifiability. The two following special cases are sometimes useful.

COROLLARY 2.7. *If P is an interaction pattern such that the majority of rows and the majority of columns do not contain any interactions, then P is unconditionally identifiable.*

COROLLARY 2.8. *If P is an interaction pattern such that in each row and each column less than a quarter of the cells contain an interaction, then P is unconditionally identifiable.*

The interaction patterns described in Corollary 2.7 are precisely those considered by Daniel [9], who showed how to detect them on the basis of the least-squares residuals.

2.4. L^1 *optimality.* It turns out that data arising from unconditionally identifiable patterns can be correctly analyzed using the L^1 functional. On writing

$$c_{ij} = c_{ij}(a_i, b_j) = x_{ij} - (a_i + b_j), \tag{12}$$

the L^1 or least absolute deviation functional is defined as the arg min of the function

$$F(a_1, \ldots, a_I, b_1, \ldots, b_J) = \sum_{i,\,j} \left| c_{ij}(a_i, b_j) \right|. \tag{13}$$

A characterization of solutions of such a minimization problem was given by El-Attar, Vidyasagar and Dutta ([13], Lemma 2.1). In our situation all the functions c_{ij} are linear and the gradient of c_{ij} is an $(I + J)$ vector with ith and $(I + j)$th component -1 and 0 elsewhere. This leads to the following proposition.

PROPOSITION 2.9. *The effects a_i and b_j minimize* (13) *if and only if there exist $\alpha_{ij} \in [-1; 1]$ for every cell (i, j) with $c_{ij} = 0$ such that*

$$\sum_{j:\ c_{ij} \neq 0} \operatorname{sgn} c_{ij} + \sum_{j:\ c_{ij} = 0} \alpha_{ij} = 0, \qquad 1 \leq i \leq I, \tag{14}$$

$$\sum_{i:\ c_{ij} \neq 0} \operatorname{sgn} c_{ij} + \sum_{i:\ c_{ij} = 0} \alpha_{ij} = 0, \qquad 1 \leq j \leq J. \tag{15}$$

A residual matrix $C = (c_{ij})$ minimizes the least absolute deviation if and only if we can find an $(I \times J)$ matrix $A = (\alpha_{ij})$ for which the following hold:

1. If $c_{ij} \neq 0$, then $\alpha_{ij} = \operatorname{sgn}(c_{ij}) \neq 0$.
2. If $c_{ij} = 0$, then $\alpha_{ij} \in [-1; 1]$.
3. All row sums and column sums of A equal 0.

Using this representation of L^1 optimality, we can prove the following lemma.

LEMMA 2.10. *If the residual matrix C has an unconditionally identifiable interaction pattern P, then C is a solution of the least absolute deviation problem.*

In general, solutions to the L^1 problem in the two-way layout are not unique [1]. The following theorem is therefore stronger than Lemma 2.10:

THEOREM 2.11. *Let X be a matrix and C and let C' be two residual matrices in $\mathscr{C}(X)$ such that C has an unconditionally identifiable interaction pattern and C' minimizes the least absolute deviation. Then $C = C'$.*

An adaptation of general simplex methods to the special case of a two-way table was given by Armstrong and Frome [1, 2].

Theorem 2.11 is no longer valid for a matrix C whose interaction pattern is not unconditionally identifiable as shown in the example

$$\begin{pmatrix} 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

whose unique L^1 solution is

$$\begin{pmatrix} -1 & -1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

2.5. *A general solution to the parsimony criterion.* An algorithm which calculates the minimum number of nonzero interactions for any given data set is described in [34]. It is, however, slow and can only be used for small data sets. Furthermore, we have not succeeded in extending it to cover the case of noisy data, so we do not pursue this topic further.

3. Interactions and noise.

3.1. *The inclusion of noise.* The inclusion of noise leads to the model

$$x_{ij} = a_i + b_j + c_{ij} + \varepsilon_{ij}, \tag{16}$$

where the ε_{ij} are usually taken to be independently and identically distributed random variables. At first sight the noise (ε_{ij}) and the interactions (c_{ij}) are confounded. However, if we accept that the noise (ε_{ij}) is small, then we can tentatively identify large residuals as interactions or outliers rather than noise. This is the path we shall pursue.

3.2. *Location functionals and noise.* In Section 2 we saw that in the no-noise case the L^1 functional detects all interactions which form an unconditionally identifiable interaction pattern. We now show that the influence of such interactions on the values for the row and column effects is bounded even for noisy data. More generally, we can prove optimal breakdown behavior for a class of M functionals defined by minimization of a strictly convex ρ function with a given scale.

Another possibility is to define location and scale as the 0's of the ψ and χ functions (see [22]). Although the empirical evidence is good in that we have not found any data sets where this fails, there are no proofs of uniqueness or theoretical results on breakdown behavior. This is worthy of further investigation.

In order to define the M functional, we need an initial or external scale functional. We construct this in Section 3.3, but for now we assume its existence, denote it by s_e and assume it satisfies

$$\sup\{s_e(X+C)\colon C \in \mathscr{C}_{UI}\} < \infty \tag{17}$$

for all matrices X, where $\mathscr{C}_{\mathrm{UI}}$ is the set of all $(I \times J)$ matrices with an unconditionally identifiable interaction pattern.

The M functional for a data set (x_{ij}) is defined as a minimizer of

$$\sum_{i,j} \rho\left(\frac{x_{ij} - (a_i + b_j)}{s_e(X)}\right) \tag{18}$$

if $s_e(X) > 0$ and as an L^1 solution if $s_e(X) = 0$. By fixing $a_1 = 0$ and choosing a strictly convex ρ function, we can guarantee that (18) always has a unique minimum. In the two-way layout we can prove robustness if $|\rho(x) - k_0|x||$ is bounded for some $k_0 \neq 0$.

THEOREM 3.1. *Let (x_{ij}) be an arbitrary matrix, s_e be a scale functional satisfying* (17) *and ρ be a strictly convex function such that $|\rho(u) - k_0|u||$ is bounded for some constant $k_0 \neq 0$. For each $C \in \mathscr{C}_{\mathrm{UI}}$ we minimize*

$$\sum_{i,j} |x_{ij} - a_i - b_j| \quad \text{if } s_e(X+C) = 0,$$

$$\sum_{i,j} \rho\left(\frac{x_{ij} - a_i - b_j}{s_e(X+C)}\right) \quad \text{if } s_e(X+C) > 0.$$

If the solutions are a_i and b_j with $a_1 = 0$, then

$$\sup\{|a_i^C|\colon C \in \mathscr{C}_{\mathrm{UI}}\} < \infty$$

and

$$\sup\{|b_j^C|\colon C \in \mathscr{C}_{\mathrm{UI}}\} < \infty.$$

The proof is given in the Appendix.

Theorem 3.1 implies that functionals defined by the theorem have the best possible breakdown behavior in the two-way layout.

In the examples below we use the ρ function

$$\rho(u, A) = \frac{u^2}{(1 + A|u|)} \tag{19}$$

for some tuning constant $A > 0$. For small A, the $\rho(u)$ function behaves like u^2 and for large A, like $|u|$. Simulations with standard Gaussian noise and scale functional $s_e = 1$ show that for $A = 10$ the ability to detect outliers is comparable to that of the L^1 functional and larger values of A have no further advantage. For values of A as low as 5 the ability to detect outliers is impaired. In the examples below we therefore set $A = 10$ in (19). Unfortunately this choice of A leads to numerical problems. Steepest descent turns out to be very unstable because of inaccuracies in the calculation of the gradient. The Nelder–Meade algorithm [28] does work, but may have to be restarted several times due to degeneracy of the simplex. The following method proved satisfactory.

STEP 1. Iterate median polish 10 times.
STEP 2. Calculate robust scale functional s (see Section 3.3).
STEP 3. Calculate direction of steepest descent (a^*, b^*).
STEP 4. Minimize in direction (a^*, b^*).
STEP 5. Repeat Steps 3 and 4 to convergence.

Step 1 is included because median polish is very fast and the final result is often very close to the solution of the minimization problem [20]. A robust scale functional as in Step 2 is described in the next section. This is calculated only once. The calculation of (a^*, b^*) in Step 3 is simple. The minimization problem in Step 4 is now a one-dimensional problem which may be solved by bisection.

3.3. *Initial robust scale functionals.* In order to calculate the M functional of the last section we require an initial scale functional s_e which satisfies (17). If the scale is "known," then s can simply be taken to be this known value. If there is more than one observation per cell, one simple method is to take the median of the cells' MADs. This is a case of "borrowing strength" (Tukey). The most difficult situation is that of one observation per cell, where the scale itself has to be based on the observations while allowing for the possibility of interactions. We address this problem.

One possibility is to calculate the "tetrad differences" $(x_{ij} + x_{kl}) - (x_{il} + x_{kj})$ because these terms do not depend on the row and column effects ([5]; [18], Section 8.4). Using Theorem 3.2 below, it can be shown that in the case of an unconditionally identifiable interaction pattern the proportion of tetrad differences not effected by interactions is at least $1/(16(I-1))$, where, without loss of generality, we assume $I \leq J$ [34]. This is the only universal bound we have, but it can be very poor. In the case of a 5×5 table it can be shown directly that at least 16 of the possible 100 tetrad differences are not effected, whereas the general bound gives only 1. A second attempt is to calculate differences of the form $x_{i'j} - x_{ij}$, $i' \neq i, 1 \leq j \leq J$. These terms do not depend on any column effects. It turns out that for unconditionally identifiable interaction patterns a sufficient number of them do not contain any interaction terms for it to be possible to make use of them.

THEOREM 3.2. *Let P be an unconditionally identifiable interaction pattern. Then for each row i there exists another row $i' \neq i$ such that*

$$\#\{j\colon p_{ij} = 0 = p_{i'j}\} > \frac{J+1}{4}.$$

The proof is given in the Appendix.

For each pair i, i' of rows, calculate the length of the shortest interval containing more than $(J+1)/4$ of the differences $x_{ij} - x_{i'j}$ and denote it by $s(i, i')$. For each row i we set

$$s_{\min}(i) = \min_{i' \neq i} s(i, i')/\sqrt{2}.$$

It follows from the Theorem 3.2 that $s_{\min}(i)$ does not explode for patterns of unconditionally identifiable outliers. If the number of rows is very large, then $s_{\min}(i)$ may be very small. Although this is theoretically of no consequence, it may in practice lead to considerable numerical problems when minimizing the ρ function. To avoid this, we set

$$s^*(i) = \exp(A(J) + B(J)\log(I))s_{\min}(i), \tag{20}$$

where $A(J)$ and $B(J)$ are as follows. If $3 \leq J \leq 6$, then $B(J) = 1.0$ and

$$A(3) = 0.9, \qquad A(4) = 1.6, \qquad A(5) = 2.1, \qquad A(6) = 2.5.$$

For $J = 7$ we set $A(7) = 1.1$ and $B(7) = 0.5$. Finally if $J \geq 8$, then

$$\begin{aligned}
A(J) &= 2.7J^{-0.3}, \qquad B(J) = 2.8J^{-0.8}, \qquad \operatorname{mod}(J, 4) = 0,\\
A(J) &= 4.0J^{-0.4}, \qquad B(J) = 3.0J^{-0.8}, \qquad \operatorname{mod}(J, 4) = 1,\\
A(J) &= 4.3J^{-0.4}, \qquad B(J) = 3.1J^{-0.8}, \qquad \operatorname{mod}(J, 4) = 2,\\
A(J) &= 2.1J^{-0.2}, \qquad B(J) = 1.5J^{-0.6}, \qquad \operatorname{mod}(J, 4) = 3.
\end{aligned}$$

The constants given above were determined by simulation and make the $s^*(i)$ approximately median consistent for Gaussian noise. We repeat the process for the columns to obtain $s^*(j)$ and define the initial robust scale functional s_e by

$$s_e = \frac{1}{I+J}\left(\sum_{i=1}^{I} s^*(i) + \sum_{j=1}^{J} s^*(j)\right). \tag{21}$$

3.4. *Final robust scale functionals.* Simulations show that the functional s_e has both a large variability and a large bias in the presence of interac-

tions [34]. This has little effect on the result of the minimization problem, where scale is effectively a nuisance parameter. It does, however, make s_e unsuitable as a measure of the scale of the noise. An improved scale functional for the noise may be obtained from the residuals as described in Theorem 3.3. For the statement of the theorem, we require the definition of the finite sample breakdown point [12]. Given a scale functional S and a data set X we define

$$\varepsilon^*(S, X) = \min\left\{\frac{k}{IJ}: \sup_{Y \in \mathscr{Y}_k}\{S(Y)\} = \infty\right\},$$

where

$$\mathscr{Y}_k = \{(y_{ij}): \#\{(i, j): y_{ij} \neq x_{ij}\} = k\}.$$

THEOREM 3.3. *For a given data set X and for each $C \in \mathscr{C}_{\mathrm{UI}}$ define $R^C = (r_{ij}^C)$ by*

$$x_{ij} + c_{ij} = a_i^C + b_j^C + r_{ij}^C,$$

where $a_1^C = 0$ and

$$\sup_{i, j, C \in \mathscr{C}_{\mathrm{UI}}} \{|a_i^C|, |b_j^C|\} < \infty.$$

Let S be a Lipschitz-continuous (with respect to the L^1 norm) scale functional such that

$$\varepsilon^*(S, X') > \left(\min\left\{\left(J - \left\lfloor \frac{J-1}{2} \right\rfloor\right)\left\lfloor \frac{I-2}{2} \right\rfloor, \left(I - \left\lfloor \frac{I-1}{2} \right\rfloor\right)\left\lfloor \frac{J-2}{2} \right\rfloor\right\} + \left\lfloor \frac{I-1}{2} \right\rfloor\left\lfloor \frac{J-1}{2} \right\rfloor\right)\Big/ IJ \tag{22}$$

for all $X' = (x'_{ij})$. Then

$$\sup\{S(R^C): C \in \mathscr{C}_{\mathrm{UI}}\} < \infty.$$

The term in (22) comes from Corollary 2.6. It is always less than $\frac{1}{2}$ and thus the MAD satisfies the conditions of Theorem 3.3. The functional we shall adopt here is the following M functional for scale. If the residuals are r_{ij}, then we define the scale functional s_0 to be the unique solution of

$$\sum_{ij} \chi\left(\frac{r_{ij}}{s_0}\right) = 1 - 2\varepsilon, \tag{23}$$

where

$$\chi(u) = \frac{u^4 - 1}{u^4 + 1} \tag{24}$$

and

$$\varepsilon = \min\left\{\left(J - \left\lfloor \frac{J-1}{2} \right\rfloor\right)\left\lfloor \frac{I-2}{2} \right\rfloor, \left(I - \left\lfloor \frac{I-1}{2} \right\rfloor\right)\left\lfloor \frac{J-2}{2} \right\rfloor\right\} + \left\lfloor \frac{I-1}{2} \right\rfloor\left\lfloor \frac{J-1}{2} \right\rfloor \Big/ IJ. \tag{25}$$

The scale functional s_0 can be made median consistent (for Gaussian noise) as follows. Without loss of generality, we assume that $I \le J$ and set

$$s = \frac{s_0}{E(I) - F(I, J)J^{-1}}, \tag{26}$$

where

I	3	4	5	6	7	8	9	10	11	≥ 12
$E(I)$	1.27	1.00	0.90	0.85	0.80	0.78	0.77	0.75	0.74	$0.65 + 0.90I^{-1}$
$F(I, \text{even})$	−1.24	0.84	0.25	0.84	0.50	0.84	0.60	0.84	0.60	0.78
$F(I, \text{odd})$	−0.12	0.84	0.56	0.84	0.70	0.84	0.70	0.84	0.70	0.78

3.5. *Identifying interactions or outliers.* To identify outliers in the two-way table, we propose the following procedure. The row and column effects are calculated using the ρ function defined by (19) with tuning constant $A = 10$. The initial scale functional s_0 we use is that defined by (20). The residuals r_{ij} are then calculated and divided by the scale function (26) to give standardized residuals $r^*_{ij} = r_{ij}/s$. All cells (i, j) for which

$$|r^*_{ij}| > \mathrm{OF}(I, J) \tag{27}$$

are identified as outliers. The factor $\mathrm{OF}(I, J)$ is chosen so that the probability of identifying some cell as an outlier is 0.05 under Gaussian noise, that is,

$$\mathrm{P}\Big(\max_{ij}\{|r^*_{ij}|\} > \mathrm{OF}(I, J)\Big) = 0.05.$$

Simulations give the following approximation for $\mathrm{OF}(I, J)$. We set

$$z(I, J) = \Phi^{-1}(\alpha_N),$$

where Φ denotes the distribution function of the standard normal distribution, $N = IJ$ and

$$\alpha_N = (1 + 0.95^{1/N})/2.$$

Without loss of generality we may suppose that $I \le J$. For $I = 3$ we have

$$\mathrm{OF}(3, 3) = 2.7 \ \mathrm{OF}(3, 4) = 2.7 \ \mathrm{OF}(3, J) = z(3, J) + 0.45, \qquad J \ge 5.$$

For $4 \le I \le 8$ we set

$$O(I, J) = z(I, J) + G(I) + H(I)/J,$$

where

I	4	5	6	7	8
$G(I)$	0.25	0.30	0.20	0.20	0.20
$H(I)$	2.8	2.5	2.8	2.4	2.7

Finally for $I \geq 9$ we set

$$\mathrm{OF}(I, J) = z(I, J)\exp(\exp(0.5 - 0.02I)/J).$$

The following approximation for $z(I, J)$ may be used:

$$z(I, J) = \sqrt{2\log N} - \left(\tfrac{1}{2}\log(\log N) - 2.4\right)\Big/\sqrt{2\log N} - 1.48/\log N. \tag{28}$$

4. Breakdown patterns. In this section we shall interpret interactions as outliers to conform with the usage in robust statistics. We distinguish between breakdown points and breakdown behavior. The breakdown point of a functional represents the minimum number of outliers which can cause functionals to break down. In the two-way table the minimum number is $\min\{\lfloor (I+1)/2 \rfloor, \lfloor (J+1)/2 \rfloor\}$ and occurs when all the outliers are in one row or one column. As we have seen, it is possible for a functional to withstand more outliers than this if they are spread out over the table. The best we can hope for is the identification of arbitrary outliers which form an unconditionally identifiable pattern. If the pattern is not unconditionally identifiable, then there is a choice of outliers such that there is another pattern with at most the same number of outliers and which is, in the context of the model, indistinguishable from the first. Furthermore, the outliers may be chosen to be arbitrarily large with the result that any procedure which is equivariant with respect to the group of allowable transformations (Section 1.2) will break down. From this it follows that any method which eventually detects arbitrarily large outliers forming an unconditionally identifiable pattern has the optimal breakdown behavior. Theorem 3.1 shows that the L^1 functional has the optimal breakdown behavior.

Let us examine the breakdown behavior of the Hampel–Rousseeuw least median of squares [17, 30]. First, we note that the optimal breakdown point is not obtained by minimizing the median of the absolute residuals, but rather by minimizing the hth order statistic where $h = [n/2] + [(p+1)/2]$ and $p-1$ denotes the maximum number of points on a lower dimensional plane. We refer to [33], page 125, and [10], page 1851, with the correction that the maximum number of points on a lower dimensional plane is $p-1$ and not p. In the case of the 5×5 table "least median of squares," therefore, has the highest breakdown point if we minimize the size of the 23rd order statistic of the absolute residuals. Furthermore, if we minimize any smaller order statistic, it is clear that we may then fail to identify two outliers in any row or column. The following modification gives optimal breakdown behavior. Let k be such that the pattern formed by the largest k absolute residuals is unconditionally identifiable, but that formed by the largest $k+1$ absolute residuals is not. Choose the row and column effects to minimize the $(IJ-k)$th absolute resid-

ual. Note that k is not constant, but depends on the data. This modification is not trivial because it requires the concept of unconditionally identifiable patterns.

5. Examples.

5.1. *A constructed example.* The first example is one we have constructed so that the size and position of each interaction is known. The data are

$$\begin{pmatrix} 24.12 & -4.17 & -2.69 & 11.73 & 19.85 & 40.54 & 5.96 & 13.61 & 14.88 \\ -8.82 & -12.16 & -24.99 & 3.16 & 20.19 & 19.16 & 9.58 & 3.43 & 3.15 \\ 0.12 & 10.13 & -0.58 & 25.05 & 33.35 & 30.60 & 7.02 & 15.29 & 15.95 \\ 26.45 & 9.57 & 10.40 & 27.92 & 46.07 & 43.49 & 21.01 & 28.52 & 29.42 \\ -12.19 & -15.05 & -14.62 & -13.41 & 6.67 & 6.15 & -18.06 & -10.36 & -10.31 \\ 22.03 & 20.62 & 7.29 & 13.31 & 41.76 & 38.66 & 16.03 & 24.06 & 23.38 \\ 10.35 & 7.62 & -4.43 & 10.51 & 28.12 & 28.25 & 15.18 & 26.76 & 23.11 \\ 0.79 & -1.22 & -10.94 & 3.39 & 20.48 & 19.80 & 9.27 & -8.89 & 16.67 \\ 3.16 & 0.75 & 4.05 & 19.39 & 34.13 & 31.00 & -2.86 & 4.52 & 6.86 \end{pmatrix}$$

and were generated as follows. The matrix

$$C = \begin{pmatrix} 12.3 & -13.1 & 0.0 & 0.0 & -12.9 & 11.8 & 0.0 & 0.0 & 0.0 \\ -11.9 & -12.0 & -14.2 & 0.0 & 0.0 & 0.0 & 13.9 & 0.0 & 0.0 \\ -13.9 & 0.0 & 0.0 & 10.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & -13.8 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 10.9 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & -10.3 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 12.9 & 14.2 & 12.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 13.7 & -12.9 & 12.9 \\ 0.0 & 0.0 & 14.2 & 15.1 & 12.9 & 11.1 & 0.0 & 0.0 & 0.0 \end{pmatrix},$$

has an unconditionally identifiable interaction pattern. To it were added the row effects 12.0, 2.0, 13.2, 26.8, −12.9, 22.1, 9.8, 2.1, 3.4 and the column effects 1.0, −2.9, −13.8, 0.9, 18.5, 17.2, −6.5, 2.1, 2.1. Finally, using a pseudo-random number generator Gaussian white noise with 0 mean and unit variance was added to all cells.

The standardized (division by the standard deviation) least-squares residuals are

$$\begin{pmatrix} 2.59 & -1.44 & -0.14 & -0.30 & -1.81 & 1.73 & -0.57 & 0.12 & -0.17 \\ -1.06 & -0.67 & -1.91 & 0.37 & 0.45 & 0.12 & 2.26 & 0.50 & -0.06 \\ -1.92 & 0.84 & -0.03 & 1.81 & 0.33 & -0.30 & -0.64 & 0.16 & -0.24 \\ 0.66 & -1.35 & -0.17 & 0.23 & 0.50 & -0.10 & -0.25 & 0.42 & 0.06 \\ 0.19 & 0.67 & 1.78 & -0.72 & -0.10 & -0.33 & -0.79 & -0.09 & -0.60 \\ 0.58 & 1.31 & -0.02 & -1.66 & 0.44 & -0.25 & -0.43 & 0.33 & -0.31 \\ -0.28 & 0.22 & -0.88 & -0.94 & -0.76 & -0.88 & 0.64 & 2.02 & 0.86 \\ -0.08 & 0.55 & -0.14 & -0.31 & -0.22 & -0.48 & 1.48 & -2.41 & 1.61 \\ -0.68 & -0.12 & 1.50 & 1.51 & 1.18 & 0.49 & -1.69 & -1.05 & -1.16 \end{pmatrix}.$$

and give no hint of any interaction. Indeed, the largest absolute standardized residual is 2.59 in cell (1, 1). Simulations with the Gaussian model give a 0.95 quantile of 3.276 for the largest absolute standardized residual.

The M functional described in Section 3.2 gives $s_1 = 2.09$ and the following standardized residuals:

$$\begin{pmatrix} 6.37 & -5.77 & -0.01 & -0.39 & -5.22 & 5.45 & -0.02 & 0.04 & 0.34 \\ -4.31 & -4.52 & -5.60 & 0.59 & 0.02 & 0.30 & 6.79 & 0.25 & -0.19 \\ -5.99 & 0.19 & 0.11 & 5.10 & 0.35 & -0.19 & -0.39 & -0.04 & -0.03 \\ 0.30 & -6.37 & -0.93 & 0.18 & 0.14 & -0.31 & 0.00 & -0.01 & 0.12 \\ 0.37 & 0.39 & 5.65 & -1.04 & -0.16 & 0.37 & -0.14 & -0.06 & -0.34 \\ 0.53 & 1.25 & -0.08 & -4.47 & 0.42 & -0.29 & -0.04 & 0.19 & -0.43 \\ 0.01 & 0.10 & -0.62 & -0.74 & -1.04 & -0.20 & 4.62 & 6.55 & 4.51 \\ 0.40 & 0.03 & 0.43 & 0.02 & -0.53 & -0.08 & 5.96 & -6.33 & 5.59 \\ -0.32 & -0.07 & 6.55 & 6.62 & 4.95 & 4.23 & -0.89 & -0.97 & -0.15 \end{pmatrix}.$$

The 0.95 quantile is 3.98 and all interactions are correctly identified. The large value of s_1 is a bias effect caused by the outliers.

5.2. *Daniel's example.* The second example is taken from [9]. The data are the results of ear tests and were originally published by Roberts and Corssen [29]. They have also been analyzed by Bradu [3] and by Bradu and

Hawkins [5]. The rows correspond to sound frequencies and the columns represent different occupational groups. The data are

$$\begin{pmatrix} 2.1 & 6.8 & 8.4 & 1.4 & 14.6 & 7.9 & 4.8 \\ 1.7 & 8.1 & 8.4 & 1.4 & 12.0 & 3.7 & 4.5 \\ 14.4 & 14.8 & 27.0 & 30.9 & 36.5 & 36.4 & 31.4 \\ 57.4 & 62.4 & 37.4 & 63.3 & 65.5 & 65.6 & 59.8 \\ 66.2 & 81.7 & 53.3 & 80.7 & 79.7 & 80.8 & 82.4 \\ 75.2 & 94.0 & 74.3 & 87.9 & 93.3 & 87.8 & 80.5 \\ 4.1 & 10.2 & 10.7 & 5.5 & 18.1 & 11.4 & 6.1 \end{pmatrix}.$$

Using the M-functional procedure, we obtain a scale value of $s_1 = 5.08$ and the standardized residuals are

$$\begin{pmatrix} 0.13 & -0.18 & 0.99 & -1.14 & 0.34 & -0.19 & -0.00 \\ 0.12 & 0.14 & 1.06 & -1.08 & -0.11 & -0.96 & 0.00 \\ -2.10 & -3.26 & 0.00 & 0.01 & -0.01 & 0.76 & 0.58 \\ 0.19 & -0.07 & -4.13 & 0.21 & -0.47 & 0.33 & -0.01 \\ -1.40 & 0.41 & -4.32 & 0.31 & -1.00 & 0.00 & 1.12 \\ -1.01 & 1.46 & -1.56 & 0.35 & 0.30 & 0.00 & -0.63 \\ 0.02 & -0.01 & 0.94 & -0.84 & 0.52 & -0.01 & -0.25 \end{pmatrix}.$$

The 0.95 quantile under the Gaussian model is 3.85, leading to interactions in the cells $(4, 3)$ and $(5, 3)$. Bradu [3] finds the following six interactions or outliers: $(3, 1)$, $(3, 2)$, $(3, 3)$, $(4, 3)$, $(5, 3)$ and $(6, 3)$. Daniel [9] finds five of these, but not $(3, 3)$. Bradu and Hawkins [5] identify the cells $(3, 2)$, $(4, 3)$ and $(5, 3)$.

5.3. *An interlaboratory test.* We indicate briefly how the above method may be used to analyze a certain form of interlaboratory test. The rows now represent laboratories and the columns represent samples whose concentrations are to be determined. Such tests often have to be analyzed in a routine manner for data sets with up to 300 laboratories. The challenge is to provide an automatic analysis which can withstand and identify outlying observations and laboratories. For this purpose we prefer to use a simple model, such as a multiplicative one, rather than more complicated ones, such as row-linear models [26, 27], which attempt to force an additional and possibly unjustified structure on the data. For example, for the two data sets considered in Chapter 10 and Section 13.4 of [26], a row-linear relationship without outliers is the exception and there is no strategy for coping with nonlinearities or outliers.

To account for laboratory variability a random effects model may be used, but there are often additional sources of variation such as the error scale being dependent on the level of concentration. More work remains to be done on this topic and so we do not attempt to give a full analysis of the data below. It consists of 10 samples of sewage sludge which were sent to 21 laboratories, each of which had to report the lead concentration of each sample. The data were first analyzed by Lischer [25]. The data are

$$\begin{pmatrix}
151 & 300 & 259 & 164 & 146 & 223 & 279 & 232 & 218 & 146 \\
136 & 283 & 240 & 146 & 114 & 211 & 270 & 204 & 196 & 132 \\
130 & 280 & 230 & 140 & 110 & 200 & 250 & 200 & 200 & 133 \\
149 & 263 & 249 & 147 & 120 & 213 & 251 & 210 & 198 & 149 \\
147 & 291 & 254 & 160 & 127 & 221 & 276 & 223 & 212 & 152 \\
145 & 281 & 261 & 155 & 130 & 210 & 284 & 238 & 249 & 116 \\
138 & 282 & 233 & 160 & 125 & 209 & 257 & 215 & 200 & 111 \\
158 & 310 & 263 & 174 & 135 & 235 & 282 & 239 & 220 & 153 \\
141 & 292 & 246 & 159 & 116 & 219 & 272 & 221 & 207 & 136 \\
165 & 285 & 236 & 181 & 135 & 243 & 308 & 251 & 216 & 172 \\
145 & 301 & 359 & 195 & 175 & 257 & 315 & 268 & 242 & 139 \\
115 & 275 & 181 & 136 & 115 & 210 & 254 & 214 & 215 & 122 \\
120 & 260 & 226 & 132 & 105 & 186 & 238 & 197 & 182 & 117 \\
143 & 285 & 248 & 166 & 127 & 187 & 260 & 224 & 208 & 116 \\
142 & 271 & 227 & 154 & 116 & 201 & 258 & 230 & 207 & 121 \\
132 & 287 & 245 & 140 & 106 & 209 & 266 & 210 & 202 & 124 \\
142 & 304 & 249 & 171 & 141 & 269 & 287 & 199 & 240 & 130 \\
127 & 274 & 233 & 150 & 122 & 201 & 244 & 206 & 195 & 130 \\
110 & 240 & 200 & 120 & 110 & 180 & 200 & 210 & 150 & 120 \\
168 & 290 & 261 & 172 & 143 & 225 & 270 & 230 & 217 & 162 \\
242 & 333 & 152 & 121 & 192 & 240 & 308 & 274 & 155 & 142
\end{pmatrix}.$$

To facilitate comparison with the analysis given by Lischer [25], we use an additive model and report the residuals from the model. It could be argued

that the deviations of each laboratory reading from the estimated sample concentrations may be more appropriate, but this requires more work based on a random effects model. The normalized residuals based on the M functional ($s_1 = 9.04$) are as follows:

$$\begin{pmatrix}
-0.23 & 0.25 & 0.22 & -0.30 & 0.86 & -0.41 & 0.01 & -0.00 & 0.07 & -0.05 \\
0.00 & 0.25 & 0.01 & -0.40 & -0.79 & 0.15 & 0.90 & -1.21 & -0.48 & 0.29 \\
0.39 & 0.97 & -0.05 & -0.02 & -0.18 & -0.02 & -0.26 & -0.61 & 1.01 & 1.45 \\
1.61 & -1.79 & 1.17 & -0.12 & 0.04 & 0.54 & -1.03 & -0.38 & -0.09 & 2.34 \\
-0.02 & -0.10 & 0.32 & -0.09 & -0.59 & 0.02 & 0.32 & -0.35 & 0.05 & 1.26 \\
0.01 & -0.96 & 1.34 & -0.40 & -0.01 & -0.95 & 1.45 & 1.56 & \underline{4.39} & -2.47 \\
0.23 & 0.15 & -0.76 & 1.16 & 0.44 & -0.06 & -0.53 & 0.01 & -0.03 & -2.02 \\
-0.13 & 0.67 & -0.02 & 0.13 & -1.04 & 0.24 & -0.34 & 0.09 & -0.40 & 0.04 \\
-0.19 & 0.50 & -0.08 & 0.29 & -1.32 & 0.29 & 0.37 & -0.08 & -0.01 & -0.02 \\
-0.12 & -2.85 & -3.76 & 0.14 & -1.79 & 0.37 & 1.78 & 0.66 & -1.60 & 1.39 \\
\underline{-4.13} & -2.88 & \underline{8.04} & -0.11 & 0.83 & 0.11 & 0.75 & 0.74 & -0.52 & \underline{-4.07} \\
-1.57 & 0.12 & \underline{-5.77} & -0.76 & 0.07 & 0.79 & -0.12 & 0.64 & 2.37 & -0.07 \\
0.19 & -0.33 & 0.42 & 0.01 & 0.18 & -0.65 & -0.68 & -0.02 & -0.07 & 0.59 \\
0.08 & -0.22 & 0.20 & 1.12 & -0.04 & -3.20 & -0.90 & 0.31 & 0.15 & -2.17 \\
1.16 & -0.58 & -0.94 & 0.97 & -0.08 & -0.46 & 0.06 & 2.15 & 1.23 & -0.44 \\
-0.21 & 0.92 & 0.79 & -0.84 & -1.45 & 0.16 & 0.68 & -0.32 & 0.41 & -0.37 \\
-1.61 & 0.31 & -1.27 & 0.09 & -0.07 & \underline{4.30} & 0.51 & \underline{-4.04} & 2.12 & -2.20 \\
-0.24 & 0.01 & -0.02 & 0.79 & 0.85 & -0.20 & -1.23 & -0.24 & 0.15 & 0.82 \\
0.40 & -1.23 & -1.14 & -0.00 & 2.05 & 0.00 & -3.57 & 2.73 & -2.29 & 2.24 \\
1.45 & -1.06 & 0.24 & 0.39 & 0.33 & -0.39 & -1.19 & -0.42 & -0.25 & 1.52 \\
\underline{7.47} & 1.53 & \underline{-13.98} & \underline{-7.42} & 3.58 & -0.89 & 0.85 & 2.28 & \underline{-9.27} & -2.86
\end{pmatrix}.$$

All standardized residuals larger than 3.91 are declared as outlying.

APPENDIX

PROOF OF THEOREM 2.3. We start with a matrix C of interactions whose locations are given by an interaction pattern P. If P is not unconditionally identifiable, then we can find row effects a_i and column effects b_j and nonzero values of the interactions with the following property. When we add the row and column effects to C to obtain a matrix C', then C' either has fewer nonzero elements or the the same number of nonzero elements, but at other locations. By permuting rows and columns if necessary, we obtain a matrix T of the

square block form

$$T = \left(T_{kl}\right)_{\substack{1\le k\le (r+1)\\ 1\le l\le (r+1)}} = \begin{pmatrix} \mathbf{0} & \mathbf{1} & \cdots & \mathbf{1} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{1} & \vdots \\ \mathbf{1} & \cdots & \mathbf{1} & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \cdots & \cdots & \mathbf{1} & \mathbf{1} \end{pmatrix}. \tag{29}$$

The $\mathbf{0}$ denotes that the corresponding row and column effects sum to zero and $\mathbf{1}$ denotes that the sum is nonzero. We call such a matrix an addition matrix. We can reduce the number of nonzero elements of the matrix C' by choosing those interactions located at the 1's of the matrix T to be $c_{ij} = -a_i - b_j \neq 0$. The number of nonzero interactions is therefore minimized by choosing T to maximize the number of interactions minus the number of 0's of C at these locations. From this it follows that an interaction pattern is not unconditionally identifiable if and only if the number of interactions located at the 1's is at least equal to the number of 0's of C located at the 1's. If the last row of blocks is present, then replacing a block $T_{(r+1)l} = \mathbf{1}$ by $T_{(r+1)l} = \mathbf{0}$ gives another addition matrix. From the definition of T it follows that this block contains at least as many interactions as 0's. Summing over the last row of blocks of T, we see that there must be at least one row of C which contains as many interactions as 0's. The same argument applies to the last column of blocks if present.

Suppose now that the last row and column of blocks are not present. If we replace two nondiagonal blocks $T_{ij} = \mathbf{1}$ and $T_{ji} = \mathbf{1}$ by $T_{ij} = \mathbf{0}$ and $T_{ji} = \mathbf{0}$, then we obtain another addition matrix. Again we can conclude that the number of interactions in the two blocks together is at least equal to the number of 0's of C. From this it follows that the partition

$$P = \left(\begin{array}{c|ccc} P_{11} & P_{12} & \cdots & P_{1r} \\ \hline P_{21} & P_{22} & \cdots & P_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ P_{r1} & P_{r2} & \cdots & P_{rr} \end{array}\right)$$

leads to a representation of P as given in the theorem. We have therefore proved the "only if" part of the theorem. The "if" part is obvious. □

PROOF OF COROLLARIES 2.7 AND 2.8. Under the conditions of each corollary, P is not unconditionally identifiable if and only if we can find a permutation of rows and columns leading to a partition of the form (11). We can assume that P_{12} is a $(k\times l)$ submatrix and P_{21} is a $((I-k)\times(J-l))$ submatrix.

If $k \le I/2$ and $l \le J/2$, then P_{12} and P_{21} contain at least $IJ/2$ cells. Each of the conditions in Corollary 2.7 and 2.8 guarantees that less than a quarter

of all cells of P contains an interaction, implying that the majority of the cells of P_{12} and P_{21} have to be 0.

If $k \geq I/2$ and $l \geq J/2$ the same arguments hold by symmetry, so without loss of generality we need only consider the case $k \leq I/2$ and $l > J/2$. If P satisfies the condition of Corollary 2.8, then clearly the majority of cells in these two submatrices is 0 as required. To prove Corollary 2.7, we can (after an appropriate permutation of rows and columns) partition each submatrix P_{ij} into four subsubmatrices

$$P' = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} = \left(\begin{array}{cc|cc} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ \hline A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{array}\right)$$

such that:

(a) The rows which do not contain any interaction are divided into the submatrices A_{2j} and A_{3j}, $1 \leq j \leq 4$.

(b) The columns which do not contain any interaction are divided into the submatrices A_{i2} and A_{i3}, $1 \leq i \leq 4$.

If we compare this partition with

$$P' = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} = \left(\begin{array}{c|ccc} A_{11} & A_{12} & A_{13} & A_{14} \\ \hline A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{array}\right),$$

we see that the total number of interactions contained in submatrices Q_{12} and Q_{21} equals that of P_{12} and P_{21} (as the submatrices A_{2j}, A_{3j}, A_{i2} and A_{i3} do not contain any interactions). However, from the choice of k and l it follows that the submatrices A_{12} and A_{21} together contain no more cells than the submatrices A_{23}, A_{24}, A_{32} and A_{42}. Thus the number of 0's contained in the submatrices Q_{12} and Q_{21} can be at most as large as that of submatrices P_{12} and P_{21}. Therefore, in order to prove that the majority of the cells of P_{12} and P_{21} are 0, it suffices to prove this for Q_{12} and Q_{21}. This follows, however, from the condition of Corollary 2.7 because each column of submatrix Q_{21} and each row of submatrix Q_{12} contains more exact values 0 than interactions 1. □

PROOFS OF LEMMA 2.10. Suppose there exists a residual matrix C with an unconditionally identifiable interaction pattern P which is not L^1 optimal. From the arguments of Section 2.4 it follows that each $(I \times J)$ matrix $A = (\alpha_{ij})$ satisfying $\alpha_{ij} \in [-1; 1]$ for all (i, j) and $\alpha_{ij} = \operatorname{sgn} c_{ij}$ whenever $c_{ij} \neq 0$ contains a row or column which does not sum up to zero. We define a function

$$T(A) = \sum_i \left|\sum_j \alpha_{ij}\right| + \sum_j \left|\sum_i \alpha_{ij}\right|$$

for all matrices A which satisfy the given conditions. Among all matrices which minimize T, we select one with the minimum number of rows and columns summing up to 0. After sorting the rows of A by descending row sum and the columns by ascending column sum we can divide A into nine submatrices:

$$\begin{array}{ccccc} A_{11} & | & A_{12} & | & A_{13} \\ \hline A_{21} & | & A_{22} & | & A_{23} \\ \hline A_{31} & | & A_{32} & | & A_{33} \\ \underbrace{\qquad} & & \underbrace{\qquad} & & \underbrace{\qquad} \\ \sum_i \alpha_{ij} < 0 & & \sum_i \alpha_{ij} = 0 & & \sum_i \alpha_{ij} > 0 \end{array} \quad \begin{array}{l} \Big\} \ \sum_j \alpha_{ij} > 0 \\ \\ \Big\} \ \sum_j \alpha_{ij} = 0 \\ \\ \Big\} \ \sum_j \alpha_{ij} < 0 \end{array} \tag{30}$$

As A minimizes T, it follows that for all cells in A_{13}, $\alpha_{ij} = -1$ whenever $c_{ij} = 0$. Similarly $\alpha_{ij} = 1$ for all cells of A_{31} with $c_{ij} = 1$. Moreover, the definition of A implies that $\alpha_{ij} = -1$ for the submatrices A_{12} and A_{23} whenever $c_{ij} = 0$. Similarly, $\alpha_{ij} = 1$ in the submatrices A_{21} and A_{32} whenever $c_{ij} = 0$.

If in the representation (30) the submatrix A_{11} is empty, then either $\sum_j \alpha_{ij} \le 0$ for all i or $\sum_i \alpha_{ij} \ge 0$ for all j. Both cases cannot occur simultaneously, because otherwise all row and column sums would be 0, contradicting the fact that C is not an L^1 solution. Hence, either there exists a row i of A with strictly positive row sum such that $\alpha_{ij'} = -1$ whenever $c_{ij'} = 0$ or there exists a column j with strictly negative column sum such that $\alpha_{i'j} = 1$ whenever $c_{i'j} = 0$. In this case there exists either a row i or a column j of C which contains more interactions than exact values 0. As this contradicts the unconditional identifiability of P, A_{11} cannot be empty. Similarly A_{33} cannot be empty.

We have shown that

$$A = \left(\begin{array}{ccc} A_{11} & A_{12} & A_{13} \\ \hline A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{array}\right) = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

is a decomposition of A into four nonempty matrices. If we decompose the unconditionally identifiable interaction pattern P in the same manner, we see that in B_{12} and B_{21} together more than half of all α_{ij} can be chosen. The restrictions on these "free" α_{ij} imply

$$\sum_{(i,\,j)\in B_{21}} \alpha_{ij} - \sum_{(i,\,j)\in B_{12}} \alpha_{ij} > 0.$$

On the other hand, we have

$$\sum_{(i,\,j)\in B_{11}\cup B_{21}} \alpha_{ij} - \sum_{(i,\,j)\in B_{11}\cup B_{12}} \alpha_{ij} < 0$$

because of (30). These two inequalities contradict each other. □

PROOF OF THEOREM 2.11 AND COROLLARY 2.6. We assume there exists a residual matrix C with an unconditionally identifiable interaction pattern and an L^1 solution $C' \in \mathscr{C}(C)$ with $C \neq C'$ and $c'_{ij} = c_{ij} + a_i + b_j$ for all (i, j). After permuting the rows and columns, we can assume the the a_i are increasing and that the b_j are decreasing. This implies that $a_1 + b_1 = 0$.

Next we show that not all a_i are equal. If they all are equal, then not all the b_j can be equal as $C' \neq C$. This implies $a_i + b_J < 0$ for all i and hence 0 is not a median of the last column of C', contradicting the L^1 optimality. Similarly, not all the b_j are equal.

The corresponding addition matrix S therefore has the decomposition

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} = \left(\begin{array}{c|c} a_i + b_j = 0 & a_i + b_j > 0 \\ \hline a_i + b_j < 0 & a_i + b_j = \varnothing \end{array}\right), \tag{31}$$

where no submatrix S_{ij} is empty.

As the set of all L^1 solutions is convex, it follows that

$$D(\lambda) = (1 - \lambda)C + \lambda C' = C + \lambda S$$

is L^1 optimal. We define

$$\lambda_0 := \min\left\{\left|\frac{c_{ij}}{a_i + b_j}\right| : (i, j) \text{ such that } c_{ij} \neq 0 \text{ and } a_i + b_j \neq 0\right\}, \tag{32}$$

where the minimum of the empty set is ∞. For all $\lambda \in (0;\ \min\{\lambda_0, 1\})$ we know

$$\begin{aligned} 0 &= \sum_{i, j} |d_{ij}(\lambda)| - \sum_{i, j} |c_{ij}| \\ &= \lambda\Bigg(\sum_{i, j:\ c_{ij} \neq 0} \operatorname{sgn}(c_{ij}) \cdot (a_i + b_j) + \sum_{i, j:\ c_{ij} = 0} |a_i + b_j| \Bigg). \end{aligned} \tag{33}$$

As C is an L^1 solution, there exists a matrix $A = (\alpha_{ij})$ with coefficients $\alpha_{ij} \in [-1; 1]$ and $\alpha_{ij} = \operatorname{sgn} c_{ij}$ whenever $c_{ij} \neq 0$ and such that all row and column sums are 0. Using A we can write (33) as

$$0 = \sum_{i, j:\ c_{ij} = 0} \big(|a_i + b_j| - \alpha_{ij}(a_i + b_j)\big).$$

Thus whenever $c_{ij} = 0 \neq a_i + b_j$ we have $\alpha_{ij} = \operatorname{sgn}(a_i + b_j)$.

If we decompose C and A in the same manner into four submatrices as in S [see (31)], it follows from the unconditional identifiablity of C that

$$\begin{aligned} 0 &< \sum_{(i, j) \in A_{12}} \alpha_{ij} - \sum_{(i, j) \in A_{21}} \alpha_{ij} \\ &= \Bigg(\sum_{(i, j) \in A_{11}} \alpha_{ij} + \sum_{(i, j) \in A_{12}} \alpha_{ij} \Bigg) - \Bigg(\sum_{(i, j) \in A_{11}} \alpha_{ij} + \sum_{(i, j) \in A_{21}} \alpha_{ij} \Bigg). \end{aligned}$$

This contradicts the fact that all row and column sums of A are 0. □

From (33) we obtain the following lemma which will be useful for the proof of Lemma A.2.

LEMMA A.1. *Let C be a residual matrix with an unconditionally identifiable interaction pattern and let $a_1, \dots, a_I$ and $b_1, \dots, b_J$ be such that $a_i + b_j \neq 0$ for at least one cell (i, j). Then we have*

$$\sum_{i,\, j:\ c_{ij} \neq 0} \operatorname{sgn}(c_{ij})(a_i + b_j) + \sum_{i,\, j:\ c_{ij}=0} |a_i + b_j| > 0.$$

PROOF OF THEOREM 3.1. The key to the proof of Theorem 3.1 is the following lemma.

LEMMA A.2. *Let $S \subset \mathbb{R}$ and k_0, $K > 0$. For each $s \in S$ let $\rho_s\colon \mathbb{R} \to \mathbb{R}$ be a function such that $|\rho_s(x) - k_0|x|| < K$ for all x. Consider a fixed data matrix $X = (x_{ij})$. For all $C \in \mathscr{C}_{\mathrm{UI}}$ and $s \in S$ define*

$$r_{ij}(C, s) = \arg\min\left\{\sum_{i,\, j} \rho_s(t_{ij})\colon (t_{ij}) \in \mathscr{C}(X + C)\right\}$$

and let $a_i(C, s)$ and $b_j(C, s)$ be the row and column effects, respectively, with $a_1(C, s) = 0$ and satisfying

$$x_{ij} + c_{ij} = a_i(C, s) + b_j(C, s) + r_{ij}(C, s).$$

Then we have

$$\sup\{|a_i(C, s)|\colon C \in \mathscr{C}_{\mathrm{UI}} \text{ and } s \in S\} < \infty$$

and

$$\sup\{|b_j(C, s)|\colon C \in \mathscr{C}_{\mathrm{UI}} \text{ and } s \in S\} < \infty.$$

PROOF. Consider a matrix X and sequences $(C_k)_k \subset \mathscr{C}_{\mathrm{UI}}$, $(s_k)_k \subset S$. We write $a_{ki} = a_i(C_k, s_k)$, $b_{kj} = b_j(C_k, s_k)$ and $r_{kij} = r_{ij}(C_k, s_k)$. Define λ_k by

$$\lambda_k = \max_{ij}\{|a_{ki}|, |b_{kj}|\}.$$

If the lemma is false, then there exists a subsequence of $(\lambda_k)_k$ which tends to infinity. Without loss of generality we may assume that the sequence (λ_k) itself tends to infinity and that the following limits exist:

(a) $\alpha_i = \lim_{k\to\infty}(a_{ki}/\lambda_k) \in [-1; 1]$ exists for all i.
(b) $\beta_j = \lim_{k\to\infty}(b_{kj}/\lambda_k) \in [-1; 1]$ exists for all j.
(c) $\operatorname{sgn}(c_{kij}) = \operatorname{sgn}(c_{1ij})$ for all (i, j).
(d) $\gamma_{ij} = \lim_{k\to\infty}(c_{kij}/\lambda_k) \in [-\infty; \infty]$ exists for all (i, j).

As $C_1 \in \mathscr{C}_{\mathrm{UI}}$ and $\operatorname{sgn}(\gamma_{ij}) = 0$ whenever $c_{1ij} = 0$, it follows that the matrix $H = (\eta_{ij})$ defined by $\eta_{ij} = \operatorname{sgn}(\gamma_{ij})$ has an unconditionally identifiable interaction pattern. From Lemma A.1 we may conclude that

$$\sum_{(i,j):\ \eta_{ij}=0} |\alpha_i + \beta_j| - \sum_{(i,j):\ \eta_{ij}\neq 0} \operatorname{sgn}(\eta_{ij})(\alpha_i + \beta_j) > 0.$$

We have

$$\begin{aligned}
&\liminf_{k\to\infty} \frac{1}{\lambda_k}\left(\sum_{i,j}\left|(x_{ij} + c_{kij}) - (a_{ki} + b_{kj})\right| - \sum_{i,j}\left|x_{ij} + c_{kij}\right|\right) \\
&\quad = \liminf_{k\to\infty} \sum_{i,j}\left(\left|\frac{c_{kij} - (a_{ki} + b_{kj})}{\lambda_k}\right| - \left|\frac{c_{kij}}{\lambda_k}\right|\right) \qquad (34) \\
&\quad \geq \sum_{(i,j):\ \eta_{ij}=0} |\alpha_i + \beta_j| - \sum_{(i,j):\ \eta_{ij}\neq 0} \operatorname{sgn}(\eta_{ij})(\alpha_i + \beta_j) > 0,
\end{aligned}$$

which implies that the sequence

$$\left(\sum_{i,j} |r_{kij}| - \sum_{i,j} |x_{ij} + c_{kij}|\right)_k$$

is not bounded above. As

$$\begin{aligned}
0 &\geq \sum_{i,j} \rho_{s_k}(r_{kij}) - \sum_{i,j} \rho_{s_k}(x_{ij} + c_{kij}) \\
&\geq k_0 \cdot \left(\sum_{i,j} |r_{kij}| - \sum_{i,j} |x_{ij} + c_{kij}|\right) - 2IJK,
\end{aligned}$$

this leads to a contradiction. □

To prove Theorem 3.1 we define $s^* = \sup\{s(X + C)\colon C \in \mathscr{C}_{\mathrm{UI}}\}$. Then $s^* < \infty$ and on defining $S = [0; s^*]$, we have $s(X + C) \in S$ for all $C \in \mathscr{C}_{\mathrm{UI}}$.

Using the function ρ we define $\rho_s(x) = s\rho(x/s)$ for $s > 0$ and $\rho_0(x) = k_0 \cdot |x|$. Because of the conditions on ρ,

$$\left|\rho_s(x) - k_0|x|\right| = s\left|\rho\left(\frac{x}{s}\right) - k_0\left|\frac{x}{s}\right|\right| < sK \leq s^*K < \infty.$$

As the terms $\sum_{i,j} \rho_0(t_{ij})$ and $\sum_{i,j} \rho_s(t_{ij})$ have the same minimizers as $\sum_{i,j} |t_{ij}|$ and $\sum_{i,j} \rho(t_{ij}/s)$, respectively, Theorem 3.1 follows from Lemma A.2. □

PROOF OF THEOREM 3.2. We need only consider the first line. After permuting the columns if necessary, we can write C in the form

$$C = \left(\begin{array}{c|c} * \cdots * & 0 \cdots 0 \\ \hline C_{12} & C_{22} \end{array}\right),$$

where $*$ denotes an interaction. Let $l \leq (I-1)/2$ be the number of interactions in the first line. As C is unconditionally identifiable, the submatrix C_{22} must

contain more 0's than interactions. From this it follows that there exists a line $i > 1$ such that

$$\bigl|\{j > l\colon c_{ij} = 0\}\bigr| > \frac{I-l}{2} \geq \frac{I+1}{4}. \qquad \square$$

PROOF OF THEOREM 3.3. As S is Lipschitz continuous, we can find a constant Λ such that

$$\bigl|S(R^C) - S(X+C)\bigr| \leq \Lambda \sum_{i,\,j} \bigl|r_{ij}^C - (x_{ij} + c_{ij})\bigr| = \Lambda \sum_{i,\,j} \bigl|a_i^C + b_j^C\bigr|.$$

This implies

$$\begin{aligned}&\sup\bigl\{S(R^C)\colon C \in \mathscr{C}_{\mathrm{UI}}\bigr\}\\ &\qquad \leq \sup\bigl\{S(X+C)\colon C \in \mathscr{C}_{\mathrm{UI}}\bigr\} + \Lambda \sup\biggl\{\sum_{i,\,j}\bigl|a_i^C + b_j^C\bigr|\colon C \in \mathscr{C}_{\mathrm{UI}}\biggr\}.\end{aligned}$$

The second term on the right-hand side is bounded because of the robustness of the location estimator. The first term on the right-hand side is also bounded because in an unconditionally identifiable interaction pattern, each row and each column contains more 0's than interactions so the scale functional cannot break down because of (22). $\square$

Acknowledgments. We acknowledge the help of an Editor and two referees, whose comments helped improve the paper.

The results of the paper are part of the first author's Ph.D. dissertation at the University of Essen.

REFERENCES

[1] ARMSTRONG, R. D. and FROME, E. L. (1976). The calculation of least absolute value estimators for two-way-tables. In *Proceedings of the Statistical Computing Section* 101–106. Amer. Statist. Assoc., Washington DC.

[2] ARMSTRONG, R. D. and FROME, E. L. (1979). Least-absolute-values-estimators for one-way and two-way tables. *Naval Res. Logist.* **26** 79–96.

[3] BRADU, D. (1975). *E.D.V. in Biologie und Medizin* **6**, Heft 4. Fischer, Stuttgart.

[4] BRADU, D. (1997). Identification of outliers by means of L^1 regression: Safe and unsafe configurations. *Comput. Statist. Data Anal.* **24** 271–281.

[5] BRADU, D. and HAWKINS, D. M. (1982). Location of multiple outliers in two-way tables, using tetrads. *Technometrics* **24** 103–108.

[6] COCHRAN, W. G. and COX, F. M. (1957). *Experimental Designs*, 2nd ed. Wiley, New York.

[7] DANIEL, C. (1960). Locating outliers in factorial experiments. *Technometrics* **2** 149–156.

[8] DANIEL, C. (1976). *Applications of Statistics to Industrial Experimentation.* Wiley, New York.

[9] DANIEL, C. (1978). Patterns in residuals in the two-way-layout. *Technometrics* **20** 385–395.

[10] DAVIES, P. L. (1993). Aspects of robust linear regression. *Ann. Statist.* **21** 1843–1899.

[11] DAVIES, P. L. (1995). Data features. *Statist. Neerlandica* **49** 185–245.

[12] DONOHO, D. L. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, CA.

[13] EL-ATTAR, R. A., VIDYASAGAR, M. and DUTTA, S. R. K. (1979). An algorithm for l_1-norm minimization with application to nonlinear l_1-approximation. *SIAM J. Numer. Anal.* **16** 70–86.

[14] ELLIS, S. P. and MORGENTHALER, S. (1992). Leverage and breakdown in L_1 regression. *J. Amer. Statist. Assoc.* **87** 143–148.
[15] GENTLEMAN, J. F. and WILK, M. B. (1975a). Detecting outliers in a two-way table. I. Statistical behavior of residuals. *Technometrics* **17** 1–14.
[16] GENTLEMAN, J. F. and WILK, M. B. (1975b). Detecting outliers in a two-way table. II. Supplementing the direct analysis of residuals. *Biometrics* **31** 387–410.
[17] HAMPEL, F. R. (1975). Beyond location parameters: robust concepts and methods. In *Proceedings of the 40th Session of the ISI* **46** 375–391.
[18] HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics*. Wiley, New York.
[19] HE, X., JURECKOVA, R., KOENKER, R. and PORTNOY, S. (1990). Tail behavior of regression estimators and their breakdown points. *Econometrica* **58** 1195–1214.
[20] HOAGLIN, D. D., MOSTELLER F. and TUKEY, J. W. (1983). *Understanding Robust and Exploratory Data Analysis*. Wiley, New York.
[21] HOAGLIN, D. C., MOSTELLER, F. and TUKEY J. W. (1985). *Exploring Data Tables, Trends, and Shapes*. Wiley, New York.
[22] HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.
[23] HUBER, P. J. (1995). Robustness: Where are we now? *Student* **1** 75–86.
[24] HUBERT, M. (1996). The breakdown value of the L_1 estimator in contingency tables. *Statist. Probab. Lett.* **33** 419–425.
[25] LISCHER, P. (1993). Ringversuche zur bestimmung des qualitätsstandards von laboratorien. In *Seminar der Region Oesterreich-Schweiz der Internationalen Biometrischen Gesellschaft, Innsbruck.*
[26] MANDEL, J. (1991). *Evaluation and Control of Measurements*. Dekker, New York.
[27] MANDEL, J. (1995). *Analysis of Two-Way Layouts*. Chapman and Hall, New York.
[28] NELDER, J. A., MEAD, R. (1965). A simplex method for function minimization. *Computer Journal* **7** 308–313.
[29] ROBERTS, J. and COHRSSEN, J. (1968). Hearing levels of adults. U.S. National Center for Health Statistics Publications, Series 11, No. 31, p. 36, Table 4, Rockland, MD.
[30] ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880.
[31] ROUSSEEUW, P. J. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications* **B**. *Proceedings of the 4th Pannonian Symp. Math. Statist.* (W. Grossmann, C. G. Pflug, I. Vincze and W. Wertz, eds.) Akadémiai Kiadó, Budapest.
[32] ROUSSEEUW, P. J. and CROUX, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc.* **88** 1273–1283.
[33] ROUSSEEUW, P. J. and LEROY, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.
[34] TERBECK, W. (1996). Interaktionen in der zwei-faktoren-varianzanalyse, Ph.D. dissertation, Univ. Essen.
[35] TUKEY, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.
[36] TUKEY, J. W. (1993). Exploratory analysis of variance as providing examples of strategic choices. In *New Directions in Statistical Data Analysis and Robustness* (S. Morgenthaler, E. Ronchetti and W. A. Stahel, eds.). Birkhäuser, Basel.

FACHBEREICH MATHEMATIK UND INFORMATIK
UNIVERSITÄT GH ESSEN
45117 ESSEN
GERMANY
E-MAIL: laurie.davies@uni-essen.de