Discussiones Mathematicae
Probability and Statistics 24 (2004) 183–195

SOME INFERENTIAL QUESTIONS IN REGARD TO ANALYSING TWO-WAY LAYOUTS AND ASSOCIATED LINEAR MODEL THEORY AND PRACTICE

Brenton R. Clarke

Mathematics and Statistics
Division of Science and Engineering
Murdoch University, Murdoch, W.A., 6150, Australia

and

Antony G. Monaco

Department of Health, Government of Western Australia
W.A. 6004, Australia

Abstract

In analysing a well known data set from the literature which can be thought of as a two-way layout it transpires that a robust adaptive regression approach for identifying outliers fails to be sensitive enough to detect the possible interchange of two observations. On the other hand if one takes the classical approach of diagnostic checking one may also stop too early and be satisfied with a model that falls short of a more detailed analysis that takes account of heteroscedasticity in the data. An exact F-test for heteroscedasticity in the two way layout is compared with various more general tests proposed by Shukla. In conclusion it is noted that when modelling the particular form of heteroscedasticity countenanced here, the estimated column effects are unchanged from those estimated from the model assuming homogeneous error variance structure. It is only the estimated variances of these column effects that changes.

Keywords: outlier; two-way layout; adaptive estimation; heteroscedaticity.

2000 Mathematics Subject Classification: 62J10, 2F05.

1. Introduction

In this paper we discuss questions relating to the analysis of the two-way layout with one observation per cell. The linear model is formulated for example by letting the observation in the (i, j)th cell be represented by a random variable $Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$ $(i = 1, ..., r; j = 1, ..., s)$. Here μ, α_i and β_j are the general mean and the effects of the ith row and the jth column, respectively, e_{ij} represents the error associated with the (i, j)th cell, and we assume for the present that the e_{ij} are independent and normally distributed, $e_{ij} \sim N(0, \sigma^2)$. The row effects and column effects are assumed to sum to zero respectively. By analysing a well known data set of Immer et al. summarized in R.A. Fisher's book "Design of Experiments" (1949, p. 66) we illustrate that the robust outlier detection method in regression discussed of Clarke (2000) may not be sensitive enough to identify potential outliers in the two-way layout, even though this method works extremely well on other known data sets in regression. For example Daniel (1976, p. 159) identifies two observations as possibly being interchanged for years. While one of these observations is highlighted by the method in Clarke (2000) it is not identified as an outlier. The reader who follows up the data laid out in Fisher's book may see that there are observations on yield for each of two years at six different locations for five varieties of barley. Allowing for a fixed effects design and possible interactions between locations and years gives a 12×5 two-way layout. The interest here is primarily in yield for the five varieties of barley and thus we consider a fixed effects model.

Assuming Daniel is correct and interchanging the two observations, a half-normal plot of the residuals indicates the two-way fixed effects model is a good fit. However, this also is a fallacy, for we note that four of the five largest absolute residuals belong to one variety *Trebi* The consequent analysis showing that variety *Trebi* is indeed having a larger error variance is discussed in Clarke and Godolphin (1992, pp. 2520–2522). There it is also noted that the associated F-statistic for testing heteroscedasticity, in this case where variety *Trebi* has increased variance and the other varieties have a common error structure, is in fact optimal, as it is equivalent to an exact likelihood ratio test based on the error contrasts (see Clarke and Godolphin 1992, §7.).

A more general F-statistic can be written down for testing where normal errors for say the first $s - l$ columns of the two-way layout have error variance σ^2 and the remaining normal errors, say in the last l columns have error variance λ^2. It is unknown whether this statistic is optimal when

$2 \leq l \leq s-2$, but all indications from the empirical analysis of power given in this paper suggest this is so. We compare the given statistic with more general tests of heteroscedasticity illustrating superior performance of the former when the particular error structure we have described here holds. Various tests of heteroscedaticity given in Shukla (1972, 1982) are compared for type I error and power.

Finally, it is a curious phenomenon that when considering the alternative error variance structure given above, that the estimated column effects are unchanged, while the row effects do change. While the estimated column effects do not change under the alternative covariance structure their estimated variances do change and this was demonstrated in Table II in Clarke and Godolphin (1992). The theory of considering when least squares estimates (which assume a covariance matrix of σ^2 times the identity matrix), are equal to the generalized least squares estimates (assuming an alternative covariance structure) is discussed in Puntanen and Styan (1989). In our paper we have an example where only a partition of the least squares estimator is unchanged when going from the least squares estimator to the generalized least squares estimator under a more general covariance matrix. Visualise, the column effects are unchanged whereas the row effects alter.

2. The robust adaptive approach

If Cuthbert Daniel is correct then in ordering the observations column by column from the two-way layout, then observations 33 and 34 in the vector of observations $\mathbf{Y}$ should appear as outliers. The approach taken in Clarke (2000) to identify outliers in regression was to minimize an estimated asymptotic variance of the trimmed likelihood estimator. Letting $V_n(g)$ be the objective function given in Clarke (2000) and $\tilde{J}(g)$ the potential outlying observations we obtain Table 2.1. The value g refers to the number of observations to be trimmed from the sample when evaluating the trimmed likelihood with a trimming proportion $\alpha = g/n$.

Table 2.1. Fisher's practical example

	g	$V_n(g)$	$\tilde{J}(g)$
$\tilde{g} =$	0	170.6	-
	1	196.8	34
	2	222.1	34,53
	3	240.5	20,34,44

For this data it appears that the ATLA algorithm in Clarke (2000) does not identify any outlying observations, viz. the preferred value of trimming is $\tilde{g} = 0$ observations, though observation 34 is flagged as the first potential outlier should one exist. However observation 53 is highlighted as the next most potentially aberrant, seemingly disqualifying observation 33 of being so.

Behaving as if Daniel were correct and swapping the observations for location 5 and variety *Velvet* leads to Table 2.2 which again highlights no outliers. Using our approach of outlier identification we would stop here and proceed no further in the modelling of these data.

Table 2.2. Fisher's modified practical example

	g	$V_n(g)$	$\tilde{J}(g)$
$\tilde{g} =$	0	139.8	-
	1	164.4	57
	2	178.7	57,58
	3	194.2	20,44,53
	4	197.2	20,44,51,53
	5	202.7	20.44,51,52,53
	6	Computationally out of range	

3. The classical approach to modelling the data

On fitting a fixed effects two-way layout design the accepted approach to modelling is to do a diagnostic check of the residuals, for instance to see if there is any departure from the normality assumption for the unobserved errors. A half-normal probability plot of the estimated residuals is a typical diagnostic tool where ordered absolute residuals are plotted against the expected values from a normal distribution. The plot below shows no sign of significant deviation from the model normal assumption, it appearing to not deviate greatly from a straight line. On the other hand it can be seen that four of the five largest residuals belong to variety *Trebi*. For instance variety *Trebi* residuals are plotted in circles. It is also noted in Clarke and Godolphin (1992) that *Trebi* has the largest variety effect estimate.

In keeping with the well known fact that increased mean values often lead to increased variances it is then sensible to test to see if variety *Trebi* has a different error variance associated with it, and then model the data accordingly as given in Clarke and Godolphin (1992).

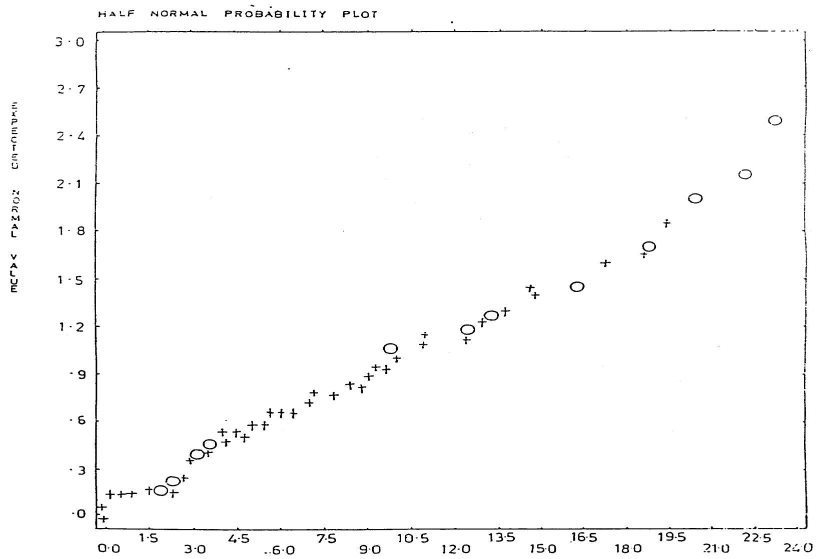

Figure 1. Half-normal probability plot of residuals

4. Tests of Heteroscedasticity

4.1. An exact test of heteroscedasticity

An elegant description of the contrasts that make up the sums of squares in a two-way design is given in Clarke (2002). The error contrasts in a two-way layout with r rows and s columns can be conveniently written in the form

$$\mathbf{w} = (\mathbf{B_s} \otimes \mathbf{B_r})\mathbf{Y},$$

where $\otimes$ represents kronecker product and the matrices $\mathbf{B_r}$ and $\mathbf{B_s}$ are respectively $(r-1) \times r$ and $(s-1) \times s$ partitions from the Helmert matrices of order r and s, respectively. For example the rows of $\mathbf{B_r}$ and $\mathbf{B_s}$ are orthogonal to the unit vectors of length r and s respectively.

See Clarke (2002) and Clarke and Godolphin (1992) for example. When the first $s-l$ columns of the two-way layout have error variance σ^2 and the remaining columns have error variance λ^2 using Lemma 6.1 of Clarke and Godolphin (1992) it is possible to set up an exact F-test for the null hypothesis $H_0 : \sigma^2 = \lambda^2$ versus either a one tailed or two tailed alternative. In fact, $\mathbf{w}' = (\mathbf{w}'_{\nu_1}, \mathbf{w}'_{\nu_2})$ and

$$\mathbf{F} = \frac{\mathbf{w}'_{\nu_2}\mathbf{w}_{\nu_2}/\nu_2}{\mathbf{w}'_{\nu_1}\mathbf{w}_{\nu_1}/\nu_1}, \tag{4.1}$$

where $\nu_1 = (s-l-1)(r-1)$ and $\nu_2 = l(r-1)$. From Theorem 6.1 of Clarke and Godolphin (1992) the statistic $\mathbf{F}$ is compared with the appropriate critical point(s) of the Fisher F-distribution. The special case of this F-test is when only the last column of the two way layout has a differing error variance, where $l = 1$, and then the test corresponds to Test 2 of Russell and Bradley (1958) and is known to be optimal as was shown in Clarke and Godolphin (1992, §7).

4.2. Some tests of heteroscedasticity of Shukla and further variants

Shukla (1972, 1982) considered a null hypothesis of equal columnon error variances, considered in the form of

$$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_s^2$$

The alternative hypothesis is that the column variances are not all equal (non-homogeneous) whence

$$H_a : \sigma_j^2 \neq \sigma_{j'}^2 \text{ for at least one pair } j \neq j' \text{ and } j, j' = 1, 2, ..., s.$$

This more general alternative encompasses the particular alternative considered in the previous subsection.

In his 1972 paper Shukla considers two statistics for testing these more general hypotheses. Then in 1982 Shukla proposes a Bartlett type test statistic for testing the above hypotheses and considers several approximations to the null distribution of the statistic. We shall not go into great detail about these tests as the discussion of the statistics is found in Shukla's papers. However we list here the tests of Shukla and the tests resulting from the above $\mathbf{F}$ in (4.1). In addition we consider a statistic $\mathbf{F}^{\sim}$ that can be thought of as being formed from say splitting the $r \times s$ two-way layout into two two-way layouts of orders $r \times (s-l)$ and $r \times l$, evaluating the error variances from the two two-way layouts and then letting $\mathbf{F}^{\sim}$ be the resulting Fisher F-statistic for testing equality of variances. In using the above F-statistics we have the added option of doing a oneway or two way test, e.g. we may consider either $H_a: \ \lambda^2 \neq \sigma^2$ or $H_a: \ \lambda^2 > \sigma^2$. Clearly a one-tailed test is more powerful at the same significance level as the two tailed test. We list the tests below.

Table 4.1. Tests considered in later Comparisons of Power & Type I error

Test Number	Test	Description
1.	Sh72	Shukla's 72 test based on Mauchly's 1940 test of sphericity
2.	Shlrt	Likelihood ratio test(asymptotic) based on error contrasts
3.	χ_I	First approximation to Shukla's Bartlett type statistic
4.	χ_{II}	Second approximation to Shukla's Bartlett type statistic
5.	ShF	Third approximation of Shukla's Bartlett type statistic
6.	χ_J	Johnson approximation as described by Shukla
7.	χ_U	Final approximation variant of Shukla 1982
8.	$\mathbf{F}^{\neq}$	Two tailed F test from equation (4.1)
9.	$\mathbf{F}^{\sim\neq}$	Two tailed F-test using separate two-way ANOVA's
10.	$\mathbf{F}^{>}$	One tailed F-test from equation (4.1)
11.	$\mathbf{F}^{\sim>}$	One tailed F-test using separate two-way ANOVA's

For each of these tests we carry out Type I Error and Power Comparisons at a nominal significance level of 0.05. Naturally the Type I error of the

F-statistics is exactly 0.05 and these therefore do not need to be included. In the subsequent simulations we generate 4,275 samples to achieve a two decimal point accuracy in the estimates of Type I error. The proportion of significant test statistics is recorded in Table 4.2. It can be noted that test Sh72 performs poorly in terms of Type I error, while the asymptotic likelihood ratio test based on the error contrasts in Shlrt has questionable Type I error for smaller designs. The first three approximations to Shukla's Bartlett type statistic perform reasonably well in terms of Type I error. The last two approximations also lead to questionable Type I errors for small designs.

Table 4.2. Comparison of Type I errors as measured by simulation

Design Size		Sh72	Shlrt	χ_I	χ_{II}	ShF	χ_J	χ_U
r	s							
3	3	0.118	0.110	0.036	0.037	0.037	0.063	0.018
	5	0.615	0.0387	0.047	0.052	0.065	0.124	0.035
	10	0.803	0.495	0.053	0.058	0.058	0.270	0.052
8	3	0.072	0.059	0.051	0.059	0.053	0.061	0.023
	5	0.110	0.086	0.045	0.046	0.046	0.066	0.029
	10	1.000	0.578	0.045	0.045	0.045	0.079	0.043
20	3	0.056	0.052	0.038	0.052	0.019	0.043	0.017
	5	0.057	0.052	0.053	0.055	0.055	0.059	0.040
	10	0.0772	0.061	0.050	0.050	0.050	0.064	0.045

We now go on to look at a selection of designs plotting empirically determined power curves for a set of values l (the parameter that determines the number of columns having error variance LAMBDA=λ^2 under the alternative hypothesis). The first $s-l$ columns are simulated to have error variance one. Figure 2 considers the power curves for a 3×5 design with $l=2$ for which tests with adequate Type I error (in the range (0.04, 0.06)) are compared. All the F-statistics excel in terms of power curves. Shukla's approximations come a poor second.

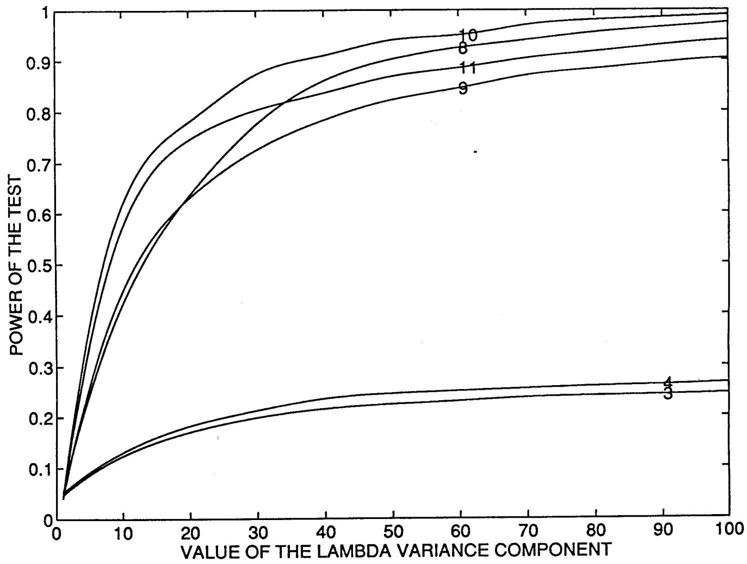

Figure 2. Estimated power curves for the 3×5 design with $l = 2$

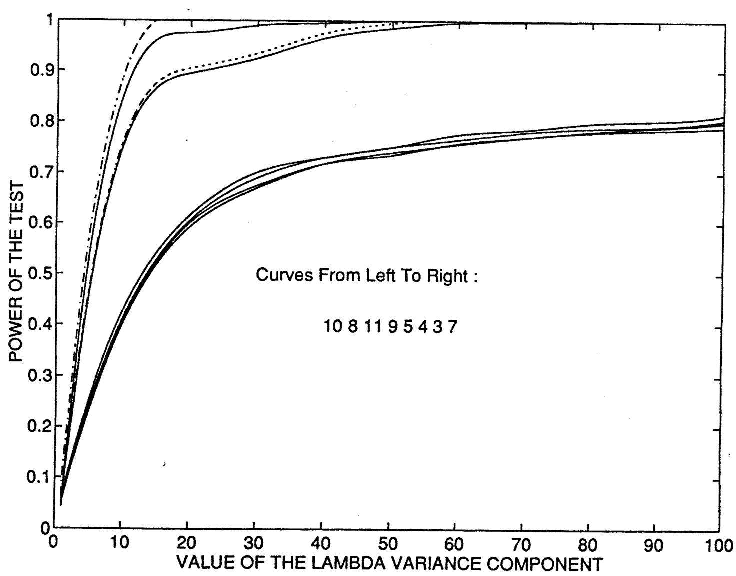

Figure 3. Estimated power curves for the 3×10 design with $l = 3$

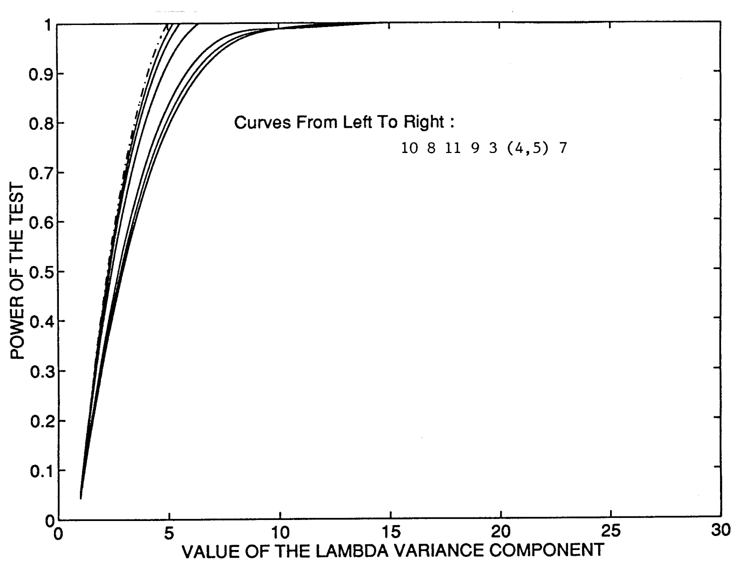

Figure 4. Estimated power curves for the 8 × 10 design with $l = 3$

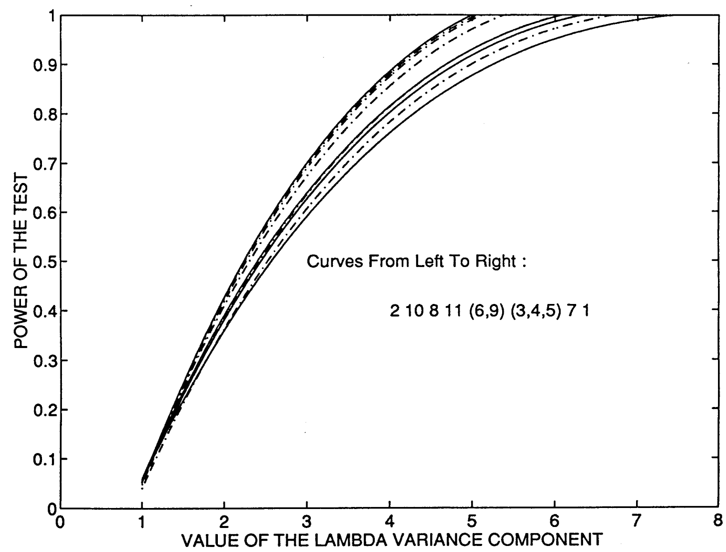

Figure 5. Estimated power curves for the 20 × 5 design with $l = 2$

In Figure 3 a 3×10 design with $l = 3$ again showed the F-tests are much more powerful than the statistics based on Shukla's approximations of the sistribution of the Bartlett type statistic. The test statistics formulated from the construction first proposed in Clarke and Godolphin (1992) are then the most powerful. Considering an 8×10 design with $l = 3$ Figure 4 shows that the F-tests are better in terms of power. Finally in Figure 5 when considering a 20×5 design with $l = 2$, since all the tests are adequate in terms of Type I error all are included in the graph. Note that the asymptotic likelihood ratio test performs the best but is indistinguishable from the F-tests formulated from Clarke and Godolphin (1992).

This last figure begs the question as to whether or not the F-test of Clarke and Godolphin (1992) is equivalent to the exact likelihood ratio test based on the error contrasts when $2 \leq l \leq s-2$. That is, is it the most powerful test? It has already been shown to be the case when $l = 1$. For the more general case it is an open question.

5. Equivalence of least squares and generalized least squares estimates

The usual approach to fitting the fixed effect parameters under the alternative error structure is to first find estimates of the error variance parameters, say $\tilde{\sigma}^2$ and $\tilde{\lambda}^2$, based on the error contrasts. This leads to an estimated covariance matrix of the errors $\tilde{\boldsymbol{\Sigma}}$ from which the fixed effects parameters are then estimated via say generalized least squares where

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{GLS}} = (\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Y},$$

where $\boldsymbol{X}$ is the design matrix and $\tilde{\boldsymbol{\Sigma}}$ is the estimated covariance matrix. It can be shown analytically in the case $l = 1$ that none of the column fixed effect parameters are changed from the least squares counterparts. The only change is in the estimated error variances of the column effects. Using numerical examples this also appears to be the case for $2 \leq l \leq s-2$. The row fixed effects do change under the alternative error structure. This curious result may be reflected more widely, e.g. it may be $\boldsymbol{\Sigma} = \boldsymbol{\sigma^2 V}$ and the particular generalized least squares estimator

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{GLS}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_{\boldsymbol{1,GLS}} \\ \hat{\boldsymbol{\beta}}_{\boldsymbol{2,GLS}} \end{pmatrix} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X} + \boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Y},$$

where $\boldsymbol{H\beta} = \mathbf{0}$ are restraints on the parameters, so that

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{LSE}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_{\mathbf{1},\boldsymbol{LSE}} \\ \hat{\boldsymbol{\beta}}_{\mathbf{2},\boldsymbol{LSE}} \end{pmatrix} = (\boldsymbol{X'X} + \boldsymbol{H'H})^{-1}\boldsymbol{X'Y}.$$

See Scheffé (1959 p. 19) for example. We may ask when is $\hat{\boldsymbol{\beta}}_{\mathbf{2},\boldsymbol{LSE}} \equiv \hat{\boldsymbol{\beta}}_{\mathbf{2},\boldsymbol{GLS}}$ for an appropriate partition of the fixed effect parameters? See Puntanen and Styan (1989) for related discussions of the general case of when $\hat{\boldsymbol{\beta}}_{\boldsymbol{GLS}} = \hat{\boldsymbol{\beta}}_{\boldsymbol{LSE}}$.

6. Conclusion

We began by noting that a robust adaptive outlier detection procedure used in Clarke (2000) may not be sensitive enough to show up outliers in the two-way layout. This is despite the procedure showing efficacious usefulness in detecting outliers in some well known data sets in that paper. We should not be disappointed however since we can observe a quote from Huber (1996, p. 62) on Future directions “somewhat embarrassingly, the robustification of the statistics of two-way tables still is wide open” “Typically there are so few degrees of freedom per cell that the customary asymptotic approaches are out of the question”. In our application it may be that the overall size of the adaptive outlier detection procedure is too small to achieve enough sensitivity to outliers for the data of Immer *et al.*

On the other hand we have the classical approach of using half-normal probability plots, which shows no departure from the initial linear model with assumptions of normal errors and homogeneous variance structure. But on further inspection there is definitely a departure from the assumption of homogeneous error variance in that variety *Trebi* has a larger variance than the other four varieties. This is illustrated in Clarke and Godolphin (1992). This resulted in a discussion of tests of heteroscedasticity in the two-way layout more generally, showing that the F-test of Clarke and Godolphin (1992) is in fact quite powerful and the question is posed as to whether it corresponds to the exact likelihood ratio test based on the error contrasts for the case where $2 \leq l \leq s - 2$.

Finally we posed the question or gave an example where a partition of the least squares estimated effects are the same as the corresponding partition of the generalized least squares estimates assuming a different covariance structure.

References

[1] B.R. Clarke, *An adaptive method of estimation and outlier detection in regression applicable for small to moderate sample sizes*, Discussiones Mathematicae Probability and Statistics **20** (2000), 25–50.

[2] B.R. Clarke, *A representation of orthogonal components in analysis of variance*, International Mathematical Journal **1** (2002), 133–147.

[3] B.R. Clarke and E.J. Godolphin, *Uncorrelated residuals and an exact test for two variance components in experimental design*, Communications in Statistics, Part A - Theory and Methods **21** (1992), 2501–2526.

[4] C. Daniel, *Applications of Statistics to Industrial Experimentation*, Wiley, New York 1976.

[5] R.A. Fisher, *Design of Experiments*, 5th ed., Oliver and Boyd, Edinburgh 1949.

[6] P.J. Huber, *Robust Statistical Procedures*, 2nd ed., Society for Industrial and Applied Mathematics, Philadelphia 1996.

[7] S. Puntanen and G.P.H. Styan, *The equality of the ordinary least squares estimator and the best linear unbiased estimator*, (with discussion) The American Statistician **43** (1989), 153–164.

[8] T.S. Russell and R.A. Bradley, *One-way variances in a two-way classification*, Biometrika **45** (1958), 111–129.

[9] H. Scheffé, *The Analysis of Variance*, Wiley, New York 1959.

[10] G.K. Shukla, *An invariant test for the homogeneity of variances in a two-way classification*, Biometrics **28** (1972), 1063–1072.

[11] G.K. Shukla, *Testing the homogeneity of variances in a two-way classification*, Biometrika **69** (1982), 411–416.

Received 17 December 2003
Revised 28 June 2004