Ann. Inst. Statist. Math.
Vol. 45, No. 1, 159–171 (1993)

ROBUST TESTS IN GROUP SEQUENTIAL ANALYSIS: ONE- AND TWO-SIDED HYPOTHESES IN THE LINEAR MODEL

MERVYN J. SILVAPULLE[1] AND PRANAB K. SEN[2]

[1]*School of Agriculture, La Trobe University, Bundoora, Victoria 3083, Australia*
[2]*Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7400, U.S.A.*

(Received August 5, 1991; revised December 24, 1991)

Abstract. Consider the linear model $Y = X\theta + E$ in the usual matrix notation where the errors are independent and identically distributed. We develop robust tests for a large class of one- and two-sided hypotheses about θ when the data are obtained and tests are carried out according to a group sequential design. To illustrate the nature of the main results, let $\hat{\theta}$ and $\tilde{\theta}$ be an M- and the least squares estimator of θ respectively which are asymptotically normal about θ with covariance matrices $\sigma^2(X^tX)^{-1}$ and $\tau^2(X^tX)^{-1}$ respectively. Let the Wald-type statistics based on $\hat{\theta}$ and $\tilde{\theta}$ be denoted by RW and W respectively. It is shown that RW and W have the same asymptotic null distributions; here the limit is taken with the number of groups fixed but the numbers of observations in the groups increase proportionately. Our main result is that the asymptotic Pitman efficiency of RW relative to W is (σ^2/τ^2). Thus, the asymptotic efficiency-robustness properties of $\hat{\theta}$ relative to $\tilde{\theta}$ translate to asymptotic power-robustness of RW relative to W. Clearly, this is an attractive result since we already have a large literature which shows that $\hat{\theta}$ is efficiency-robust compared to $\tilde{\theta}$. The results of a simulation study show that with realistic sample sizes, RW is likely to have almost as much power as W for normal errors, and substantially more power if the errors have long tails. The simulation results also illustrate the advantages of group sequential designs compared to a fixed sample design, in terms of sample size requirements to achieve a specified power.

Key words and phrases: Clinical trial, comparison of two treatments, composite hypothesis, inequality tests, interim analysis, long tailed distribution, M-estimator, Pitman efficiency, power-robustness, repeated tests, Wald-type statistics.

1. Introduction

A distinguishing feature of group sequential procedures is that they enable us to carry out several repeated analyses at various stages as the data accrue. Usually in group sequential analysis the number of such interim analyses is small compared to the number of observations. For the comparison of two means, it has been demonstrated that a properly designed group sequential procedure could be expected to require smaller total sample size than the corresponding traditional non-sequential procedure to detect a departure from the null hypothesis when such a departure is "large". These are the main attractive features of group sequential procedures compared to the fixed-sample procedures. For a review of group sequential analyses, see Jennison and Turnbull (1991).

It appears that the group sequential analysis has been applied mainly in clinical trials where ethical issues play a major role. The ethical issues arise due to the necessity to avoid subjecting patients to inferior treatments by detecting such treatments early in the study using interim analyses. Thus, group sequential designs with multiple treatments has not attracted much attention (see Whitehead (1983), Chapter 8). However, in many areas of research, other issues such as total cost and resource limitations (space, instruments, personnel etc.) play the major role. In such situations, group sequential designs may be more attractive than fixed sample designs. This paper is concerned about group sequential analysis involving linear models and tests of various hypotheses about the regression parameter.

Tests of a few different types of hypotheses in group sequential designs have been studied in the literature; for example, two-sided hypothesis with two means (Pocock (1977), O'Brien and Fleming (1979)), one-sided hypothesis with two means (DeMets and Ware (1980, 1982)), and equality of k means (Siegmund (1980)). However, tests of multivariate one-sided hypothesis and inequality constraints on the regression parameter in the linear model have not been investigated although, this has been studied extensively for fixed sample designs (see, for example, Robertson *et al.* (1988) and the many references therein). In this paper, we consider robust tests of such hypotheses in group sequential designs; here robustness is interpreted with respect to departures of the error distribution from normality. Our formulation is quite general. It does incorporate the simple but important two- and k-sample problems mentioned above. It also allows repeated analysis of covariance to accommodate concomitant variables.

The general class of hypotheses that we consider here can be tested using a Wald-type statistic based on the least squares estimator of the regression parameter although, as indicated above, this has not been investigated. Since the least squares estimator is sensitive to outliers, we propose to replace its role by a more robust estimator, for example by an M-estimator. It turns out that the efficiency-robustness of an M-estimator with respect to the least squares estimator translates to power-robustness of the corresponding test statistic in terms of Pitman efficiency. In other words, as one would expect, the test statistic based on an M-estimator is itself less sensitive to outliers than is the one based on the least squares estimator. A Monte Carlo study illustrates this large sample result for reasonably small samples and moderately long tailed error distributions. The illustration of the power-robustness of the test statistic based on a robust estimator compared to that based on the least squares estimator is the main theme of this paper.

2. Preliminaries

Let us write the linear model corresponding to a set of n independent observations as $Y = X\theta + E$ where Y is $n \times 1$, X is $n \times p$ and $\theta = (\theta_1, \ldots, \theta_p)^t$. We shall assume that every element of the first column of X is one so that θ_1 is the intercept. Let the null and alternative hypotheses be

$$(2.1) \qquad H_0 : \theta \in K \quad \text{and} \quad H_1 : \theta \in C$$

respectively, where K and C are closed, convex and positively homogeneous subsets of the p-dimensional Euclidean space, and $K \subset C$; a set P is said to be positively homogeneous if $\lambda x \in P$ whenever $x \in P$ and $\lambda \geq 0$. For various technical reasons, let us assume that the linear space spanned by K is contained in C. This formulation is quite general, and it does incorporate $H_0 : R\theta = 0$ against $H_1 : R_1\theta \geq 0$, $R_2\theta \neq 0$, where $R = [R_1^t R_2^t]^t$ is a full row-rank matrix and R_1 or R_2 could be R itself. Obviously, this special case incorporates the familiar one- and two-sided two sample problems, the usual analysis of variance problems and the analysis of covariance problems.

Let $\hat{\theta}_n$ be a robust estimator of θ, for example an M-estimator or a bounded influence estimator (see Yohai and Zamar (1988), Giltinan *et al.* (1986) and Krasker and Welsch (1983)) based on a sample of size n; in what follows we shall drop the suffix n. Our objective is to use $\hat{\theta}$ for the testing problem (2.1). The conjecture is that tests based on robust estimators rather than the least squares estimator would also be robust in terms of power.

Now let us define some regularity conditions:

CONDITION A.

(i) $n^{1/2}(\hat{\theta}-\theta) \xrightarrow{d} N(0, \tau^2 B^{-1})$ as $n \to \infty$ for some $\tau > 0$ and a positive definite matrix B; B does not depend on θ, and the scale for τ is fixed by some condition to avoid indeterminacy.

(ii) $\hat{\tau}$ and $\hat{B}$ are consistent estimators of τ and B respectively; $\hat{B}$ does not depend on θ.

(iii) $\mathcal{L}_\theta(\hat{\theta} - \theta, \hat{\tau}) = \mathcal{L}_0(\hat{\theta}, \hat{\tau})$, where $\mathcal{L}_\theta$ denotes the distribution at θ.

Let us define

$$(2.2) \quad T_n = n\hat{\tau}^{-2}[\inf\{(\hat{\theta} - b)^t\hat{B}(\hat{\theta} - b) : b \in K\} - \inf\{(\hat{\theta} - b)^t\hat{B}(\hat{\theta} - b) : b \in C\}].$$

Note that T_n is a measure of how closer is $\hat{\theta}$ to the alternative parameter space than to the null parameter space. Since $\hat{\theta}$ is expected to be close to θ, we would reject H_0 when T_n is large. If $\hat{\theta}$ is the least squares estimator then for testing $H_0 : R\theta = 0$ against $H_1 : R\theta \neq 0$ where R is a full row-rank matrix, T_n is the usual Wald statistic $S^{-2}(R\hat{\theta})^t\{R(X^tX)^{-1}R^t\}^t(R\hat{\theta})$ with the usual choices, $\hat{B} = n^{-1}(X^tX)$ and $\hat{\tau}^2 = S^2$ where S^2 is the variance of the least squares residuals. For a discussion on T_n when the sampling procedure is not sequential, see Silvapulle (1992).

We assume that the group sequential procedure is designed to have a maximum of M interim analyses. Let n_j denote the number of observations from the j-th group, $j = 1, \ldots, M$. To obtain asymptotic results, we allow the total maximum possible sample size $n = (n_1 + \cdots + n_M)$ to increase while the proportions of observations in the groups are fixed. That is, we interpret the group sequential procedure as one that involves interim analyses at the fixed time points $t_1, \ldots, t_M$ where $t_j = (n_1 + \cdots + n_j)/n$.

For any sequence S_n and $0 \leq t \leq 1$, we shall write $S_n(t)$ for $S_{[nt]}$ where $[\cdot]$ denotes the integer part. Let $W_n = n^{1/2}\tau^{-1}B^{1/2}(\hat{\theta} - \theta)$ and $W(t)$ be a vector of p independent copies of the standard Wiener process on $[0, 1]$. Now, we define another condition.

CONDITION B. As $n \to \infty$, the joint distribution of $\{W_n(t_1), \ldots, W_n(t_M)\}$ converges to that of $\{t_1^{-1/2}W(t_1), \ldots, t_M^{-1/2}W(t_M))\}$.

This condition says that a particular finite dimensional distribution of $W_n(t)$ converges to that of $t^{-1/2}W(t)$ (see, Billingsley (1968)); it is weaker than the requirement that $W_n(\cdot)$ converges weakly in the usual Skorokhod topology of $D[0, 1]^p$ (see, Sen (1981)). Let us make another remark about Condition B. Often we can obtain a representation of the form $n^{1/2}(\hat{\theta} - \theta) = n^{-1/2}\sum \xi_i + o_p(1)$, where $\xi_1, \ldots, \xi_n$ are independent with mean zero and finite covariances which are bounded in some sense. This representation for the least squares estimator is obvious once we replace $n(X^tX)^{-1}$ by its limit in the expression for that estimator; for such a representation of an M-estimator with possibly *asymmetric* errors, see Silvapulle (1985); and for a class of one-step generalized M-estimators with high break down point see Simpson *et al.* (1990). Once this representation is obtained, we can show that Condition B is satisfied by an argument similar to that in Billingsley ((1968), p. 69).

In what follows, we shall assume that Conditions A and B are satisfied unless the contrary is obvious.

Since the probability of type I error may depend on the particular point in the null parameter space, we need to ensure that the *maximum* probability of type I error over the null parameter space does not exceed a nominal level, say 0.05. The "spending function" (Lan and DeMets (1983)) of a group sequential design specifies how the total risk (i.e. the maximum probability of type I error over the null parameter space) is to be spent during the M stages of interim analyses. Let α be the total risk and let α_j be the maximum probability over the null parameter space of rejecting the null hypothesis at least by stage j when the null hypothesis is true. So, $\alpha_1 \leq \cdots \leq \alpha_M = \alpha$. *In what follows we shall assume that $\alpha_1, \ldots, \alpha_M$ and hence α are given.*

To define another process, let $K(\gamma) = \bigcup_{j=1}^{\infty} K_j(\gamma)$ where $K_j(\gamma) = \{b - j\gamma : b \in K\}$ and $\gamma \in K$; since K is a convex cone, $K(\gamma)$ is the *tangent cone* of K at γ. Now, define

$$(2.3)\quad \begin{aligned} T(t;\gamma) = {} & \inf[\{t^{-1/2}W(t) - B^{1/2}b\}^t\{t^{-1/2}W(t) - B^{1/2}b\} : b \in K(\gamma)] \\ & - \inf[\{t^{-1/2}W(t) - B^{1/2}b\}^t\{t^{-1/2}W(t) - B^{1/2}b\} : b \in C]. \end{aligned}$$

In what follows, we shall write $T_n(\cdot)$ for $[T_n(t_1), \ldots, T_n(t_M)]$ and $T(\cdot;\gamma)$ for $[T(t_1;\gamma), \ldots, T(t_M;\gamma)]$. This is consistent with the usual notation since we are interested in the values of the function $T_n(t)$ and $T(t;\gamma)$ only at $t_1, \dots, t_M$. Note that $T_n(\cdot)$ is a function of $\hat{\theta}$ and hence its distribution may depend on θ. Therefore, in probability statements involving $T_n(\cdot)$ we shall write "pr_θ". However, $T(\cdot;\gamma)$ is a function of $t^{-1/2}W(t)$ which does not depend on any unknown parameters; therefore, we do not need a suffix for "pr".

3. The main results

Let us first state a theorem relating to the null distribution of $T_n(\cdot)$.

THEOREM 3.1. *Assume that Conditions A and B are satisfied and that the null hypothesis in* (2.1) *holds. Then* $T_n(\cdot)$ *converges in distribution to* $T(\cdot;\theta)$, $\theta \in K$.

(i) *Suppose that* K *is a linear space. Then the distribution of* $T_n(\cdot)$ *is the same for any* $\theta \in K$, *and hence* $T_n(\cdot)$ *converges in distribution to* $T(\cdot;0)$ *as* $n \to \infty$.

(ii) *Suppose that* K *is not necessarily a linear space. Then for any positive integer* $k \leq M$ *and any given constants* $c_1, \dots, c_M > 0$, *we have*

$$\begin{aligned}\sup[\mathrm{pr}_b\{T_n(t_i) > c_i \text{ for some } i = 1, \dots, k\} : b \in K] \\ = \mathrm{pr}_0\{T_n(t_i) > c_i \text{ for some } i = 1, \dots, k\} \\ \to \mathrm{pr}\{T(t_i;0) > c_i \text{ for some } i = 1, \dots, k\} \quad \text{as} \quad n \to \infty.\end{aligned}$$

The proof of the above theorem is given in the Appendix, and its implications are discussed in the next section. The main use of Theorem 3.1 is that it gives some results relating to the null distribution of the test statistics and enables us to define a set of M critical values for the M stages of interim analyses.

Now, we wish to show that the efficiency-robustness properties of some robust estimators translate to power robustness of the corresponding test statistics. So, let us consider the sequence of local hypotheses $H_{0n} : \theta = \theta^*$ and $H_{1n} : \theta = (\theta^* + n^{-1/2}\Delta)$ where θ^* is a boundary point of K and Δ is a fixed point such that $(\theta^* + n^{-1/2}\Delta) \in C$ for $n \geq 1$. Let us suppose that $\hat{\theta}_1$ and $\hat{\theta}_2$ are two estimators satisfying the regularity conditions in Section 2 with τ_1 and τ_2 respectively. Note that we are assuming that the matrix B is the same for both estimators but the values of τ may be different. We wish to compare the power of $T_{n1}(\cdot)$ corresponding to $\hat{\theta}_1$ with that of $T_{n2}(\cdot)$ corresponding to $\hat{\theta}_2$. So, let $m = m(n)$ be the integer part of $n\lambda$ for some fixed $0 < \lambda < \infty$. Suppose that $T_{n1}(\cdot)$ and $T_{m2}(\cdot)$ have the same joint asymptotic distribution under H_{1n}. Then we may refer to λ as the Pitman asymptotic efficiency of $T_{n1}(\cdot)$ relative to $T_{n2}(\cdot)$. As an example, note that if $\lambda = 2$ then $T_{m2}(\cdot)$ corresponds to a group sequential procedure with twice the sample size as that corresponding to $T_{n1}(\cdot)$, but both procedures have equal asymptotic power. Now, we have the following:

THEOREM 3.2. *Let* $\hat{\theta}_1$ *and* $\hat{\theta}_2$ *be two estimators of* θ. *Suppose that, Conditions A and B are satisfied with* τ_1 *and* τ_2 *for* $\hat{\theta}_1$ *and* $\hat{\theta}_2$ *respectively; the B matrix in Condition A*(i) *is assumed to be the same for* $\hat{\theta}_1$ *and* $\hat{\theta}_2$. *Let* T_{nj} *be defined as in* (2.2) *with* $\hat{\theta}_j$, $\hat{\tau}_j$ *and* $\hat{B}_j$, *for* $j = 1, 2$. *Then the Pitman asymptotic efficiency of* $T_{n1}(\cdot)$ *relative to* $T_{n2}(\cdot)$ *is* $(\tau_2/\tau_1)^2$.

4. Discussion

4.1 *Critical values*

First assume that the null hypothesis holds with K being a linear space. Then by part (i) of Theorem 3.1, the distribution of $T_n(\cdot)$ and its limiting distribution do not depend on the particular value of θ in K. Hence, it suffices to compute the finite sample or asymptotic critical values corresponding to $\theta = 0$; for these critical values, the probability of rejecting the null hypothesis at or prior to stage j is α_j for $j = 1, \ldots, M$ and for every $\theta \in K$.

Now, assume that the null hypothesis holds but K may not be a linear space. Then by part (ii) of Theorem 3.1, the distribution of $T_n(\cdot)$ and its limiting distribution depend on the particular value of θ in K. However, $\theta = 0$ is the least favourable point in K for $T_n(\cdot)$ and its limiting distribution. To illustrate this for the finite sample case, suppose that the vector of critical values $[c_1, \ldots, c_M]$ correspond to the spending function $[\alpha_1, \ldots, \alpha_M]$ for $T_n(\cdot)$ at $\theta = 0$. Then for any $\theta \in K$ and $j = 1, \ldots, M$, we have

$$\begin{aligned} \text{pr}_\theta[\text{reject } H_0\text{, at least, by stage } j] &= \text{pr}_\theta[T_n(t_i) > c_i \text{ for some } i \leq j] \\ &\leq \text{pr}_0[T_n(t_i) > c_i \text{ for some } i \leq j] \\ &= \text{pr}_0[\text{reject } H_0\text{, at least, by stage } j] = \alpha_j. \end{aligned}$$

The same holds even if the role of $T_n(\cdot)$ is replaced by its asymptotic distribution $T(\cdot;\theta)$.

Thus, whether K is a linear space or not, and whether we are dealing with the finite sample or asymptotic case, we need the critical values corresponding to $\theta = 0$ only. *So, in what follows whenever we refer to critical values we would implicitly assume that they correspond to* $\theta = 0$.

Let $[d_1, \ldots, d_M]$ denote the vector of *asymptotic* critical values for $T_n(\cdot)$. Then $\alpha_1 = \text{pr}[T(t_1; 0) > d_1]$ and $\alpha_j = \text{pr}[T(t_i; 0) > d_i$ for some $i \leq j]$ for $j = 2, \ldots, M$. It is well-known that $\text{pr}[T(t; 0) > d_1]$ can be written as a weighted sum of χ^2 probabilities (known as chi-bar squared distribution). Unfortunately, it appears that the probabilities for α_j $(j \geq 2)$ do not reduce to such expressions, and we do not know if it is at all possible to develop a convenient algorithm for computing $d_2, \ldots, d_M$. Since the evaluation of $T(t; 0)$ involves constrained minimization of only quadratic functions of $W(t)$, we can generate pseudo random observations on $T(\cdot; 0)$ rather easily; so, a solution is to estimate $[d_1, \ldots, d_M]$ by Monte Carlo. However, one possible difficulty is that the validity of $[d_1, \ldots, d_M]$ for finite samples depend on the rate of convergence of $T_n(\cdot)$; note that in the early stages of a group sequential analysis, we are likely to have only small samples. Hence, it would be desirable to use the *exact* critical values rather than the asymptotic critical values $[d_1, \ldots, d_M]$.

Let $[c_1, \ldots, c_M]$ denote the *exact* critical values for $T_n(\cdot)$. Then, as above, we have $\alpha_1 = \text{pr}_0\{T_n(t_1) > c_1]$ and $\alpha_j = \text{pr}_0\{T_n(t_i) > c_i$ for some $i \leq j\}$. Since we can generate pseudo observations on $T_n(\cdot)$ with $\theta = 0$ rather easily, we can estimate $c_1, \ldots, c_M$ by Monte Carlo. It is easier to estimate c_1 first, and then c_2 and so on; this is the procedure adopted in our simulations.

4.2 *Size function*

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators of θ satisfying Conditions A and B. Suppose that K is not a linear space and that the true value $\theta \neq 0$. Let $T_{n1}(\cdot)$ and $T_{n2}(\cdot)$ be two statistics. Then the probabilities of type I error for $T_{n1}(\cdot)$ and $T_{n2}(\cdot)$ depend on the particular value of θ in K (see, Theorem 3.1, part (ii)). In an application, if $T_{n1}(\cdot)$ and $T_{n2}(\cdot)$ were used and their size functions were different then it would be difficult to make a meaningful comparison between the p-values for $T_{n1}(\cdot)$ and $T_{n2}(\cdot)$. This difficulty may arize even if the critical values for the two test statistics are exact for the least favourable point, the origin. However, for $\hat{\theta}_1$ and $\hat{\theta}_2$ with a common B matrix, this difficulty is overcome in large samples since, by Theorem 3.1, $T_{n1}(\cdot)$ and $T_{n2}(\cdot)$ have the same limiting size functions. (Recall that there are well-known estimators with a common B matrix; for example, least squares and M-estimators.) If K is a linear space, the difficulty does not arise (see below).

4.3 *Symmetric errors*

Let us assume that the error distribution is symmetric about the origin. Let $\hat{\theta}$ be an estimator of θ such that $(\hat{\theta} - \theta)$ is asymptotically $N\{0, \tau^2(X^tX)^{-1}\}$ for some $\tau > 0$. So, $\hat{\theta}$ could be an M-estimator (see Hampel *et al.* (1986)) or a high break down point estimator (see Yohai and Zamar (1988), Simpson *et al.* (1990)). Let $\tilde{\theta}$ be the least squares estimator. Note that the B matrix for $\hat{\theta}$ and $\tilde{\theta}$ is $\lim n^{-1}X^tX$; hence, by Theorem 3.1 the asymptotic critical values for the two corresponding test statistics $\hat{T}_n(\cdot)$ and $\tilde{T}_n(\cdot)$ are also the same. By Theorem 3.2, the Pitman asymptotic efficiency of $\hat{T}_n(\cdot)$ relative to $\tilde{T}_n(\cdot)$ is $\{\text{var(error)}/\tau^2\}$ which is precisely the asymptotic efficiency of $\hat{\theta}$ relative to $\tilde{\theta}$. In other words, the asymptotic efficiency-robustness of $\hat{\theta}$ relative to $\tilde{\theta}$ translates to power-robustness of $\hat{T}_n(\cdot)$ relative to $\tilde{T}_n(\cdot)$. This is a very useful result since a large literature already exists which shows that some M-estimators satisfying the above conditions are asymptotically more efficient than the least squares estimator when the errors have long tails and that they are almost equally good when the errors are normally distributed.

4.4 *Asymmetric errors*

Now, assume that the error distribution is not necessarily symmetric. Let us write the linear model as $Y = 1\beta_0 + Z\beta + E$ where 1 is a column of ones, and the columns of Z are centered so that $1^tZ = 0$. So, we have $\beta = (\theta_2, \dots, \theta_p)^t$. Since the error distribution is not necessarily symmetric, there is an arbitrariness in the definition of β_0 (see Carrol and Welsh (1988), Silvapulle (1985)). So, assume that the null and alternative hypotheses do not involve the intercept.

Let $\hat{B}$ be a robust estimator of β such that $(\hat{\beta} - \beta)$ is asymptotically $N\{0, \tau^2(Z^tZ)^{-1}\}$ for some $\tau > 0$. So, $\hat{\beta}$ could be an M-estimator (see Silvapulle (1985), Carrol and Welsh (1988)). Now, instead of T_n in (2.2), let us define

$$(4.1)\quad \hat{U}_n = \hat{\tau}^{-2}[\inf\{(\hat{\beta}-b)^tZ^tZ(\hat{\beta}-b) : b \in K_0\} - \inf\{(\hat{\beta}-b)^tZ^tZ(\hat{\beta}-b) : b \in C_0\}]$$

where K_0 and C_0 are the null and alternative parameter spaces for the slope component β. Let $\tilde{\beta}$ be the least squares estimator of β. Since $(\tilde{\beta} - \beta)$ is asymptotically

$N\{0, \sigma^2(Z^tZ)^{-1}\}$ where $\sigma^2 = \text{var(error)}$, let us define

(4.2) $\tilde{U}_n = S^{-2}[\inf\{(\tilde{\beta}-b)^t Z^t Z(\tilde{\beta}-b) : b \in K_0\} - \inf\{(\tilde{\beta}-b)^t Z^t Z(\tilde{\beta}-b) : b \in C_0\}]$

where S^2 is the residual variance corresponding to least squares. By Theorem 3.1, $\tilde{U}_n(\cdot)$ and $\hat{U}_n(\cdot)$ have the same asymptotic size functions, and hence the same asymptotic critical values. Further by Theorem 3.2, the Pitman asymptotic efficiency of $\hat{U}_n(\cdot)$ relative to $\tilde{U}_n(\cdot)$ is (σ^2/τ^2) which is the asymptotic efficiency of $\hat{\beta}$ relative to $\tilde{\beta}$. So, again we can make comments about the asymptotic critical values and the Pitman efficiency similar to those in Subsection 4.3.

A drawback of Huber-type M-estimators is that they have low break down points when the number of regression parameters, p is large. So, recently there has been some interest on high break down point estimators (for example, see Giltinan *et al.* (1986)). However a major draw back of many such estimators is that they require the error distribution to be symmetric. One exception is the so called Mallows-type estimator or one-step version of it (see Mallows (1975), Simpson *et al.* (1990), Giltinan *et al.* (1986)). These are essentially weighted M-estimators with appropriately chosen weights which depend on X. Let $\hat{\beta}$ be one such estimator. Assume that $(\hat{\beta}-\beta)$ is asymptotically $N(0, \tau^2 Q_n^{-1})$ where $\tau > 0$ and Q_n depends on X. Let $\tilde{\beta}$ be the corresponding weighted least squares estimator with the weights being the same as those for $\hat{\beta}$. Then, $(\tilde{\beta}-\beta)$ is asymptotically $N(0, \sigma^2 Q_n^{-1})$ where $\sigma^2 = \text{var(error)}$. Let $\hat{U}_n$ and $\tilde{U}_n$ be the test statistics in (4.1) and (4.2) respectively with Z^tZ replaced by Q_n. Then by Theorem 3.1, $\hat{U}_n(\cdot)$ and $\tilde{U}_n(\cdot)$ have the same asymptotic size function. Further, by Theorem 3.2, the asymptotic Pitman efficiency of $\hat{U}_n(\cdot)$ relative to $\tilde{U}_n(\cdot)$ is (σ^2/τ^2) which is the asymptotic efficiency of $\hat{\beta}$ relative to $\tilde{\beta}$. So, again we can make comments similar to those in Subsection 4.3.

5. Simulation results

A simulation study was carried out to investigate some of the main results presented above. In particular, our objectives were (i) to compare a robust statistic based on an M-estimator with that based on the least squares estimator, and (ii) to illustrate that substantial reductions in total sample size required are possible with a group sequential design compared to a fixed sample design.

5.1 *Design of the study*

We consider the two-way analysis of variance model: $y_{ijk} = \mu + \alpha_i + \gamma_j + \epsilon_{ijk}$, $i = 1, 2$, $j = 1, 2, 3$ and k is the index for an observation within a given cell. This model may be written as $Y = X\theta + E$ where $\theta^t = (\theta_1, \theta_2, \theta_3, \theta_4) = (\mu, \alpha_1, \gamma_1, \gamma_2)$ and the columns of X corresponding to $(\alpha_1, \gamma_1, \gamma_2)$ are centered.

Let us define $H_0 : \theta_3 = \theta_4 = 0$, $H_1 : \theta_3, \theta_4 \geq 0$ and $H_2 : \theta$ is unrestricted. The hypothesis testing problems investigated are (i) H_0 against H_2 (this is a two-sided problem), (ii) H_0 against H_1 (this is a one-sided problem) and (iii) H_1 against H_2 (this is an inequality constraint problem and hence the origin is least favourable). The error distributions considered are $\Phi(t)$, $0.8\Phi(t) + 0.2\Phi(t/3)$,

$0.8\Phi(t)+0.2\Phi\{(t-2)/3\}$ and $0.8\Phi(t)+0.2\Phi(t/5)$, where Φ is the standard normal distribution function. These four error distributions provide a reasonable neighbourhood of symmetric and asymmetric distributions around Φ. Let $\hat{\theta}$ be the M-estimator corresponding to Huber's Proposal 2 (Huber (1977, 1981)) with the kinks in the ψ function being at ± 1.5. Thus, for a set of n observations, the estimate $\hat{\theta}$ and a scale estimate $\hat{\sigma}$ were obtained by solving the estimating equations $\sum \psi(r_i)x_i = 0$ and $\sum \chi(r_i) = (n-p)A$, where x_i is the i-th row of X, $r_i = \{(y_i - x_i^t\theta)/\sigma\}$, $\psi(t) = (d/dt)\rho(t)$,

$$\rho(t) = \begin{cases} 2^{-1}t^2 & |t| < c \\ c|t| - 2^{-1}c^2 & |t| \geq c \end{cases}$$

with $c = 1.5$, $\chi(t) = t\psi(t) - \rho(t)$ and $A = E\{\chi(U)\}$ with U being a standard normal random variable. It may be verified that Conditions A and B are satisfied for this estimator. Since H_0, H_1 and H_2 do not involve the intercept term, the results hold even though one of the error distributions is not symmetric. The M-estimators were computed using the Huber-Dutter algorithm (Huber and Dutter (1974)). The inequality constrained estimators were computed using the subroutine BCOAH in the IMSL. All the programs were written in double precision FORTRAN.

We considered a group sequential procedure with 3 groups, 18 observations in each group and the spending function specified by $[\alpha_1, \alpha_2, \alpha_3] = [0.01, 0.025, 0.05]$. So, we have $M = 3$ and $n_1 = n_2 = n_3 = 18$.

While it is possible to estimate the asymptotic critical values by Monte Carlo, we prefer to use the exact ones. We observed that the exact critical values for normal errors were slightly larger than those for the other long tailed contaminated normals that we considered. This is consistent with similar observations for the fixed sample case (see Schrader and Hettmansperger (1980)). So, to adopt a conservative approach, we used the exact critical values for the normal errors, which were estimated by Monte Carlo. We estimated the Average Number of Observations to Termination (ANT) as $\{18p_1 + 36p_2 + 54(1 - p_1 - p_2)\}$ where p_1 and p_2 are the probabilities of rejecting the null hypothesis precisely at stages 1 and 2 respectively. Note that ANT is proportional to the "average number of groups to termination" (see, Pocock (1977), Table 3), and "expected number of tests performed before termination" (see DeMets and Ware (1980), Section 3.2). ANT corresponds to "Average Sample Number" used in fully sequential procedures (see, Sen (1981), p. 236). We also estimated the number N^* which we define as follows: suppose that for a given model and a test statistic, the group sequential design, with overall size 5%, yielded a power of p%; then the corresponding fixed sample test at 5% level requires N^* observations to yield a power of p%. Thus, $(N^* - \text{ANT})$ is the reduction in sample size requirement achieved by the group sequential design compared to the fixed sample design. Hence, it is a measure of the efficiency of the former compared to the latter.

5.2 *Results*

We shall use the following abreviations for the test statistics: $LS02$, $LS01$ and $LS12$ denote $T_n(\cdot)$ based on the least squares estimator for H_0 against H_2, H_0 against H_1 and H_1 against H_2 respectively. Similarly, $M02$, $M01$ and $M12$

Table 1. Simulation estimates[1] of size (%), power (%), ANT and N^*.

True θ[2]			Test statistic					
θ_3	θ_4		$LS02$	$M02$	$LS01$	$M01$	$LS12$	$M12$
			Error distribution: $\Phi(t)$					
0.0	0.0	Size:	4	4	5	5	5	5
1.0	1.0	Power:	85	83	91	91	80	78
		ANT:	43	44	39	39	43	44
		N^*:	52	50	49	51	47	45
			Error distribution: $0.8\Phi(t) + 0.3\Phi(t/3)$					
0.0	0.0	Size:	3	4	5	5	4	4
1.5	1.5	Power:	81	91	88	96	75	87
		ANT:	42	40	38	35	42	41
		N^*:	56	51	55	54	46	45
			Error distribution: $0.8\Phi(t) + 0.2\Phi\{(t-2)/3\}$					
0.0	0.0	Size:	4	5	6	5	4	5
1.5	1.5	Power:	72	88	81	93	67	84
		ANT:	44	41	40	37	45	42
		N^*:	55	51	55	52	48	46
			Error distribution: $0.8\Phi(t) + 0.2\Phi(t/5)$					
0.0	0.0	Size:	3	4	5	4	4	3
1.75	1.75	Power:	63	92	73	96	59	89
		ANT:	46	40	42	35	46	39
		N^*:	57	51	57	53	51	47

[1]The estimates are based on 1000 replications. Size and power are the probabilities of rejecting the null hypothesis at least by the last stage, under the null and alternative hypothesis respectively.

[2]Since the test statistics do not depend on (θ_1, θ_2) it was fixed at $(0.0, 0.0)$. The true values of (θ_3, θ_4) given in the table are applicable to the first four test statistics only. To obtain the true values for the last two test statistics, change the sign of the value of θ_4.

denote the corresponding ones based on the M-estimator. Some of the simulation estimates are given in Table 1. The main observations are the following:

(i) *The errors are normal*: Since the exact critical values were estimated for normal errors, the estimated sizes of LS- and M-statistics are not different from the nominal levels. Note that the probabilities of rejecting the null hypothesis by LS- and M-statistics are essentially the same; in fact, this was the case for each of the three stages. So, there is hardly any difference between LS- and M-statistics in terms of size and power.

(ii) *The error is contaminated normal*: Some of the estimated sizes are slightly smaller but still close to the corresponding nominal level. As expected, the probability of detecting a departure from the null hypothesis is higher with the M-statistic compared with the LS-statistic. Further, the ANT required for such a

Table 2. Monte Carlo estimates of the exact critical values for normal errors.[1)]

	$LS02$	$M02$	$LS01$	$M01$	$LS12$	$M12$
Stage 1	13.02	14.55	10.20	10.95	8.53	9.70
Stage 2	9.66	10.09	7.35	7.59	6.57	6.90
Stage 3	6.95	7.30	5.45	5.51	4.60	4.73

[1)]The estimates are based on 5000 replications.

higher power is smallef for the M-statistic compared to the LS-statistic. A closer inspection of the intermediate simulation results reveals that this is mainly due to the fact that the probability of early detection of the departure from H_0 is higher with the M-statistic. Another important, but expected observation is that as the tails of the error distribution become heavier, the performance of the M-statistic compared to the LS-statistic becomes substantially better. Finally, the advantage of the group sequential design compared to the fixed sample design, in terms of sample size requirements, is clear from the observation that ANT is smaller than N^* in all the cases.

We also did consider the spending function $[\alpha_1, \alpha_2, \alpha_3] = [0.01, 0.03, 0.05]$ in the simulation study. For each of the four error distributions, each of the two spending functions and each of the three hypothesis testing problems, four values of θ in the alternative parameter space were considered. These four values were chosen to give a wide range for the estimated power, for example from 0.5 to 0.9. The two observations made above also emerge in all these cases.

Thus, in summary, we could say that the overall performance of the M-statistic was better than that of the LS-statistic within the class of error distributions that we considered, and the group sequential design required fewer observations on average than the fixed sample design.

Acknowledgements

We are grateful to the referees for their valuable suggestions.

Appendix

PROOF OF THEOREM 3.1. Let $\theta \in K$ be fixed. If we replace $\hat{\tau}$ and $\hat{B}$ in T_n (see (2.2)) by their probability limits, we have

$$T_n = n\tau^{-2}[\inf\{(\hat{\theta} - b)^t B(\hat{\theta} - b) : b \in K\} - \inf\{(\hat{\theta} - b)^t B(\hat{\theta} - b) : b \in C\}] + o_p(1). \tag{A.1}$$

Now, we have

$$T_n(t) = \inf\{\|W_n(t) - B^{1/2}b\|^2 : b \in K - n^{1/2}\tau^{-1}\theta\} - \inf\{\|W_n(t) - B^{1/2}b\|^2 : b \in C\} + o_p(1). \tag{A.2}$$

Note that $T_n(\cdot) = [T_n(t_1), \ldots, T_n(t_M)]$ is a continuous function of $W_n(\cdot) = [W_n(t_1), \ldots, W_n(t_M)]$ which converges weakly to $[t_1^{-1/2}W(t_1), \ldots, t_M^{-1/2}W(t_M)]$. Hence $T_n(\cdot)$ converges weakly to $T(\cdot;\theta) = [T(t_1;\theta), \ldots, T(t_M;\theta)]$. Now, part (i) of the theorem follows since $T(\cdot;\theta) = T(\cdot;0)$ when K is a linear space.

To prove part (ii), note that $K+\theta \subset K$ and $C+\theta = C$. Let $S(K, C, n, \hat{\tau}, \hat{B}, \hat{\theta})$ denote the T_n in (2.2) and let $\hat{\tau}^i$, $\hat{B}^i$, and $\hat{\theta}^i$ denote the estimates based on a sample of size $[nt_i]$, $i = 1, \ldots, M$. Let k be a positive integer such that $k \leq M$, and let $c_1, \ldots, c_M > 0$ be fixed constants. Then

$$\begin{aligned} \text{(A.3)}\quad \text{pr}_\theta\{T_n(t_i) > c_i \text{ for some } i = 1, \ldots, k\} \\ = \text{pr}_\theta\{S(K, C, [nt_i], \hat{\tau}^i, \hat{B}^i, \hat{\theta}^i) > c_i, \text{ for some } i = 1, \ldots, k\} \\ \leq \text{pr}_\theta\{S(K+\theta, C, [nt_i], \hat{\tau}^i, \hat{B}^i, \hat{\theta}^i) > c_i, \text{ for some } i = 1, \ldots, k\} \\ = \text{pr}_0\{T_n(t_i) > c_i \text{ for some } i = 1, \ldots, k\}. \end{aligned}$$

The rest of the proof follows from (A.1), (A.2) and (A.3).

PROOF OF THEOREM 3.2. Let θ^* be a boundary point of K, and Δ be fixed such that $(\theta^* + n^{-1/2}\Delta)$ lies in C for every n. Let $H_{0n} : \theta = \theta^*$ and $H_{1n} : \theta = (\theta^* + n^{-1/2}\Delta)$. Then $\{H_{1n}\}$ is contiguous to $\{H_{0n}\}$ (see Hajek and Sidak (1967), Sen (1981)). Let $m = m(n)$ be the integer part of $n(\tau_2/\tau_1)^2$. Let $(\hat{\tau}_1, \hat{B}_1, \hat{\theta}_1, T_{n1})$ and $(\hat{\tau}_2, \hat{B}_2, \hat{\theta}_2, T_{m2})$ be based on samples of sizes n and m respectively. Let $U_n = n^{1/2}\tau_1^{-1}(\hat{\theta}_1 - \theta^* - n^{-1/2}\Delta)$ and $V_m = m^{1/2}\tau_2^{-1}(\hat{\theta}_2 - \theta^* - n^{-1/2}\Delta)$.

Assume that H_{0n} holds. Since $(\hat{\tau}_1, \hat{B}_1) \xrightarrow{p} (\tau_1, B)$, we have $Z_{0n} = Z_{1n} + o_p(1)$, where $Z_{0n} = n\hat{\tau}_1^{-2}\inf\{(\hat{\theta}_1 - b)^t\hat{B}_1(\hat{\theta}_1 - b) : b \in K\}$ and $Z_{1n} = n\tau_1^{-2}\inf\{(\hat{\theta}_1 - b)^t B(\hat{\theta}_1 - b) : b \in K\}$. Now, since $K(\theta^*) = \lim\{K - n^{1/2}\tau_1^{-1}\theta^*\}$, we have $Z_{1n} = Z_{2n} + o_p(1)$, and hence

$$\text{(A.4)}\qquad Z_{0n} - Z_{2n} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty,$$

where $Z_{2n} = \inf\{(U_n + \tau_1^{-1}\Delta - b)^t B(U_n + \tau_1^{-1}\Delta - b) : b \in K(\theta^*)\}$. Similarly

$$\text{(A.5)}\qquad W_{0m} - W_{2m} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty,$$

where $W_{0m} = m\hat{\tau}_2^{-2}\inf\{(\hat{\theta}_2 - b)^t\hat{B}_2(\hat{\theta}_2 - b) : b \in K\}$ and

$$W_{2m} = \inf\{(V_m + \tau_1^{-1}\Delta - b)^t B(V_m + \tau_1^{-1}\Delta - b) : b \in K(\theta^*)\}.$$

Now, by contiguity, we have from (A.4) and (A.5),

$$\text{(A.6)}\qquad Z_{0n} - Z_{2n} \xrightarrow{p} 0 \quad \text{and} \quad W_{0m} - W_{2m} \xrightarrow{p} 0 \quad \text{under} \quad \{H_{1n}\}.$$

The limiting results in (A.6) also hold with K and $K(\theta^*)$ replaced by C. Since $\hat{\theta}_1$ and $\hat{\theta}_2$ satisfy Conditions A(i) and (iii), U_n and $V_m \xrightarrow{d} N(0, B^{-1})$ under $\{H_{1n}\}$. This, together with (A.6) imply that $T_{n1}(\cdot)$ and $T_{m2}(\cdot)$ have the same limiting distribution under $\{H_{1n}\}$. Therefore, the Pitman asymptotic efficiency of $T_{n1}(\cdot)$ relative to $T_{n2}(\cdot)$ is $\lim(mn^{-1}) = (\tau_2/\tau_1)^2$.

References

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

Carroll, R. J. and Welsh, A. H. (1988). A note on the asymmetry and robustness in linear regression, *Amer. Statist.*, **42**, 285–287.

DeMets, D. and Ware, J. H. (1980). Group sequential methods for clinical trials with a one-sided hypothesis, *Biometrika*, **78**, 651–660.

DeMets, D. and Ware, J. H. (1982). Asymmetric group sequential boundaries for monitoring clinical trials, *Biometrika*, **69**, 661–663.

Giltinan, D. M., Carroll, R. J. and Ruppert, D. (1986). Some new estimation methods for weighted regression when there are possible outliers, *Technometrics*, **28**, 219–230.

Hajek, J. and Sidak, Z. (1967). *Theory of Rank Tests*, Academic Press, New York.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: the Approach Based on Influence Functions*, Wiley, New York.

Huber, P. J. (1977). Robust methods of estimating regression coefficients *Mathematische Operationsforschung und Statistik, Series Statistics*, **8**, 41–53.

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Huber, P. J. and Dutter, R. (1974). Numerical solutions of robust regression problems, *Compstat* 1974 (ed. G. Bruckmann), 165–172, Physika Verlag, Wien.

Jennison, C. and Turnbull, B. W. (1991). Group sequential tests and repeated confidence intervals, *Handbook of Sequential Analysis* (eds. B. K. Ghosh and P. K. Sen), Chapter 12, Marcel Dekker, New York.

Krasker, W. S. and Welsch, R. E. (1983). Efficient bounded-influence regression estimation, *J. Amer. Statist. Assoc.*, **77**, 595–604.

Lan, K. K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials, *Biometrika*, **70**, 659–663.

Mallows, C. L. (1975). On some topics in robustness, Memorandum, Bell Telephone Laboratories, Murray Hill, New Jersey (unpublished).

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials, *Biometrics*, **35**, 549–556.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials, *Biometrika*, **64**, 191–199.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference*, Wiley, New York.

Schrader, R. M. and Hettmansperger, T. P. (1980). Robust analysis of variance based upon a likelihood ratio criterion, *Biometrika*, **67**, 93–102.

Sen, P. K. (1981). *Sequential Nonparametrics*, Wiley, New York.

Siegmund, D. (1980). Sequential χ^2 and F tests and the related confidence intervals, *Biometrika*, **67**, 389–402.

Silvapulle, M. J. (1985). Asymptotic behaviour of robust estimators of regression and scale parameters with fixed carriers, *Ann. Statist.*, **13**, 1490–1497.

Silvapulle, M. J. (1992). Robust Wald-type tests of one-sided hypotheses in the linear model, *J. Amer. Statist. Assoc.*, **87**, 156–161.

Simpson, D. G., Ruppert, D. and Carroll, R. J. (1990). One-step *GM*-estimates for regression with bounded influence and high breakdown point (unpublished).

Whitehead, J. (1983). *The Design and Analysis of Sequential Clinical Trials*, Ellis Horwood, Chichester.

Yohai, V. J. and Zamar, R. H. (1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale, *J. Amer. Statist. Assoc.*, **83**, 406–413.