

of the Department of Statistics
at Queen's College University

ROBUST WALD-TYPE TESTS OF ONE-SIDED HYPOTHESES
IN THE LINEAR MODEL

by

Mervyn J. Silvapulle

School of Agriculture, La Trobe University
Bundoora, Australia

(Prepared while a Visiting School at the Department of
Biostatistics, Univ. of N.C., Chapel Hill, NC)

Institute of Mimeo Series No. 1878

May 1990

ROBUST WALD-TYPE TESTS OF ONE-SIDED HYPOTHESES
IN THE LINEAR MODEL

Mervyn J. Silvapulle
University of North Carolina at Chapel Hill and
La Trobe University

ABSTRACT

Let us consider the linear model $Y = X\theta + E$ in the usual matrix notation, where the errors are independent and identically distributed. Our main objective is to develop robust Wald-type tests for a large class of hypotheses on θ ; by 'robust' we mean robustness in terms of size and power against long tailed error distributions.

First, assume that the error distribution is symmetric about the origin. Let $\hat{\theta}_L$ and $\hat{\theta}$ be the least squares and a robust estimator of θ . Assume that they are asymptotically normal about θ with covariance matrices $\sigma^2(X^tX)^{-1}$ and $\tau^2(X^tX)^{-1}$ respectively. So, $\hat{\theta}$ could be an M- or a High Breakdown Point estimator. Robust Wald-type tests based on $\hat{\theta}$ (denoted by RW) are studied here for testing a large class of one-sided hypotheses on θ . It is shown that the asymptotic null distribution of RW and that of the usual Wald-type statistic based on $\hat{\theta}_L$ (denoted by W) are the same. This is a useful result since the critical values and procedures for computing the p-values for W are directly applicable to RW as well. A more important result is that the Pitman asymptotic efficiency of RW relative to W is (σ^2/τ^2) which is precisely the asymptotic efficiency of $\hat{\theta}$ relative $\hat{\theta}_L$. In other words, the efficiency-robustness properties of $\hat{\theta}$ relative $\hat{\theta}_L$ translate to power-robustness of RW relative to W.

The above results hold for asymmetric error distributions as well with a minor modification. The general theory presented incorporates Wald-type statistics based on a large class of estimators which includes M-, Bounded Influence and High Breakdown Point estimators; the main requirement is that the estimator under consideration be asymptotically normal about θ and hence, it does not explicitly require the errors to be iid or symmetric. The results of a simulation study show that in realistic situations, RW based on an M-estimator is likely to have at least as much power as W, and more if the errors have long tails. A simple example illustrates the application of RW, and its advantages over W.

Key Words : asymmetric errors; asymptotic efficiency; bounded influence; breakdown point; composite hypothesis; inequality constraints; M-estimator; Pitman efficiency; R-estimator.

Abbreviated title : Robust test of one-sided hypotheses.

Acknowledgements : The author wishes to thank Professors P. K. Sen and R.J. Carroll for some helpful discussions.

1. INTRODUCTION

Let us consider the linear model $y_i = x_i^t \theta + \epsilon_i$, $i = 1, \dots, n$, where $\theta = (\theta_1, \dots, \theta_p)^t$, $x_i = (x_{i1}, \dots, x_{ip})^t$, $x_{i1} = 1$, and x_1, \dots, x_n are non-stochastic. We may express this in the familiar form $Y = X\theta + E$. For the time being, assume that the errors are independent with a common distribution function which has mean zero and finite variance. It is well known that various M- and Bounded Influence estimators of θ are substantially more efficiency-robust than is the least squares estimator (see Hampel *et al* (1986), Giltinan *et al* (1986) and the references therein). Our objective is to show that, for testing a large class of one- and two-sided hypotheses on θ , such robust estimators may be used for obtaining Wald-type tests which are more robust than are the usual Wald-type tests based on the least squares estimator; robustness of tests will be interpreted as robustness of size and power against long tailed errors.

In what follows, σ^2 , $\hat{\theta}_L$ and S^2 denote the error variance which may be ∞ , the least squares estimator of θ and the corresponding error mean square respectively.

Suppose that we wish to test $H_0 : R\theta = 0$ against $H_1 : R\theta \neq 0$, where R is a $q \times p$ matrix and $\text{rank}(R) = q \leq p$. The Wald statistic W based on $\hat{\theta}_L$ is given by

$$W = S^{-2} (R\hat{\theta}_L)^t \{R(X^t X)^{-1} R^t\}^{-1} (R\hat{\theta}_L). \quad (1.1)$$

Since $\hat{\theta}_L$ and S^2 are sensitive to outliers in the y -variable, particularly those appearing at "large" x -values, one would expect that the above Wald statistic would also be sensitive to such outliers. To overcome the non-robustness of W , Wald-type statistics based on robust estimators have been proposed (see Hampel *et al* (1986, p. 363)). Let $\hat{\theta}$ be an M-estimator of θ , and assume that it is approximately $N\{\theta, \tau^2(X^t X)^{-1}\}$ for some $\tau > 0$; this does require some symmetry-like condition on the errors. For the same testing problem as that relates to (1.1), Schrader and Hettmansperger (1980) obtained interesting results for the robust Wald-type statistic

$$\hat{\tau}^{-2} (R\hat{\theta})^t \{R(X^t X)^{-1} R^t\}^{-1} (R\hat{\theta}) \quad (1.2)$$

where $\hat{\tau}$ is a consistent estimator of τ . For some related discussions see also Sen (1982).

Here, we investigate Wald-type statistics based on robust estimators of θ for testing rather general hypotheses on θ ; particular attention is given to one-sided hypotheses on θ . First, assume that the error distribution is symmetric about the origin. Let RW denote our robust Wald-type statistic (precise definition is given in the next section) based on an estimator $\hat{\theta}$ which is asymptotically $N\{\theta, \tau^2(X^tX)^{-1}\}$; thus, $\hat{\theta}$ could be an M- or a High Breakdown Point estimator (see Yohai and Zamar (1988)). Let W denote the corresponding Wald-type statistic based on $\hat{\theta}_L$ which is asymptotically $N\{\theta, \sigma^2(X^tX)^{-1}\}$. It turns out that W and RW have the same asymptotic null distributions. Hence the critical values and procedures for computing the p-values for W are directly applicable to RW as well. This is an important result since, in some situations, computation of critical and p-values for W and RW may be a non-trivial task. For testing a large class of equality and inequality constraints on θ , the Pitman asymptotic efficiency of RW relative to W turns out to be (σ^2/τ^2) . To put it simply, if m and n are large positive integers such that $(\sigma^2/\tau^2) \simeq (m/n)$ then the local asymptotic power of RW with sample size n is approximately equal to that of W with sample size m. Since (σ^2/τ^2) is also the asymptotic efficiency of $\hat{\theta}$ relative to $\hat{\theta}_L$, we may say that RW is robust compared to the corresponding W, and that the efficiency-robustness properties of $\hat{\theta}$ relative to $\hat{\theta}_L$ translate to robustness of their corresponding Wald-type statistics.

Now, let us relax the symmetry assumption on the error distribution. In this case, the intercept and location of the error distribution are confounded, and hence we consider hypotheses involving the slope component β of θ only, where $\beta = (\theta_2, \dots, \theta_p)^t$. Let us write the linear model as $Y = 1\alpha + Z\beta + E$, where 1 is a column of ones and the columns of Z are centered so that $1^tZ = 0$. Let $\hat{\beta}$ be an estimator of β such that $(\hat{\beta} - \beta)$ is asymptotically $N\{0, \tau^2(Z^tZ)^{-1}\}$; so, $\hat{\beta}$ could be an M- or R-estimator (see Silvapulle (1985), and Carroll and Welsh (1988)). Again, it turns out that the Pitman asymptotic efficiency of RW based on $\hat{\beta}$ relative to W is (σ^2/τ^2) , and the asymptotic null distributions of RW and W are the same. Clearly, the implications of similar results discussed in the above paragraph are applicable here as well. Since the main theoretical results are asymptotic in nature, we carried out a simulation study to evaluate the behaviour of RW compared to W; the results, reported in Section 5, are very encouraging. In Section 6, a simple numerical example is worked to illustrate the application of RW corresponding to an M-estimator.

2. PRELIMINARIES

Let the null and alternative hypotheses be

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

respectively, where $K \subset C \subset \mathbb{R}^p$ and \mathbb{R}^p is the p -dimensional Euclidean space. Strictly speaking we should define H_1 as " $\theta \in C, \theta \notin K$ "; however, $\theta \notin K$ is implicit. The general approach adopted here for defining Wald-type statistics for (2.1) is not new. However, it is instructive to briefly mention it for a simple situation. So, let us temporarily assume that the errors are independent and identically distributed (iid) with mean zero and finite variance. The Wald-type statistic in (1.1) is the same as (see, Wolak (1987))

$$S^{-2} \left[\inf \{ (\hat{\theta}_L - b)^t X^t X (\hat{\theta}_L - b) : Rb = 0 \} - \inf \{ (\hat{\theta}_L - b)^t X^t X (\hat{\theta}_L - b) : b \in \mathbb{R}^p \} \right]$$

Note that this statistic makes use of only the information that $S^{-1}(\hat{\theta}_L - \theta)$ is approximately $N\{0, (X^t X)^{-1}\}$ to assess if $\hat{\theta}_L$ is "closer" to the alternative parameter space than to the null parameter space (see, Robertson *et al* (1988, p224)). We shall adopt this approach here. So, for testing H_0 against H_1 in (2.1) we define a Wald-type statistic W based on $\hat{\theta}_L$ as

$$W = S^{-2} \left[\inf \{ (\hat{\theta}_L - b)^t X^t X (\hat{\theta}_L - b) : b \in K \} - \inf \{ (\hat{\theta}_L - b)^t X^t X (\hat{\theta}_L - b) : b \in C \} \right]. \quad (2.2)$$

To define a robust Wald-type statistic, let $\hat{\theta}$ be a robust estimator of θ , for example an M- or a Bounded Influence estimator (see Yohai and Zamar (1988), Giltinan *et al* (1986) and Krasker and Welsch (1983)). Now, let us introduce the following definitions : A set P in an Euclidean space is said to be *positively homogeneous* if $\lambda x \in P$ whenever $x \in P$ and $\lambda \geq 0$.

CONDITION A : (i) K and C are closed, convex and positively homogeneous; the linear space spanned by K is contained in C . (ii) $n^{-1} X^t X \rightarrow V$ as $n \rightarrow \infty$ where V is positive definite.

CONDITION B : (i) $n^{1/2}(\hat{\theta} - \theta) \rightarrow_d N(0, \tau^2 B^{-1})$ as $n \rightarrow \infty$, where $\tau > 0$ and the matrix B is positive definite and non-stochastic. (ii) $\hat{\tau}$ and B_n are consistent estimators of τ and B respectively; B_n and B do not depend on θ . (iii) Either (a) K is a linear space or (b) $\mathcal{L}_\theta(\hat{\theta} - \theta, \hat{\tau}) = \mathcal{L}_\theta(\hat{\theta}, \hat{\tau})$ where \mathcal{L}_θ denotes the distribution at θ .

Throughout, we shall assume that Condition A is satisfied, and that Condition B is also satisfied unless the contrary is obvious. Now, for testing H_0 against H_1 in (2.1), we define a robust Wald-type

statistic RW as

$$RW = n\hat{\tau}^{-2} \left[\inf\{(\hat{\theta}-b)^t B_n(\hat{\theta}-b) : b \in K\} - \inf\{(\hat{\theta}-b)^t B_n(\hat{\theta}-b) : b \in C\} \right]. \quad (2.3)$$

3. THE MAIN RESULTS

In this section, we state the main results for the hypothesis testing problem in (2.1). The proofs are given in the Appendix. First, it is convenient to state a general result.

THEOREM 1. Suppose that Conditions A and B are satisfied. For $\alpha \in K$, let $K(\alpha) = \lim\{b - n\alpha : b \in K\}$ as $n \rightarrow \infty$, T_0 be a random variable distributed as $N(0, B^{-1})$, θ be a fixed point in K and let

$$Q_\alpha = \inf\{(T_0 - b)^t B(T_0 - b) : b \in K(\alpha)\} - \inf\{(T_0 - b)^t B(T_0 - b) : b \in C\}. \quad (3.1)$$

(A). Suppose that B(iii)(a) is satisfied. Then, for $c > 0$, $pr_\theta(RW \geq c) \rightarrow pr(Q_0 \geq c)$ as $n \rightarrow \infty$.

(B) Suppose that B(iii)(b) is satisfied. Then, for $c > 0$,

$$\sup\{pr_b(RW \geq c) : b \in K\} = pr_\theta(RW \geq c); \quad (3.2)$$

further, $pr_\theta(RW \geq c) \rightarrow pr(Q_\theta \geq c)$ as $n \rightarrow \infty$.

It is clear from the above theorem that to compute the asymptotic critical and p-values for RW, we need to use the distribution of Q_0 which turns out to be chi-bar squared (see Shapiro (1988)) for a large class of hypothesis testing problems of practical interest. Since hypotheses involving linear functions of θ are more common than those involving non-linear functions, let us apply the above theorem to such hypotheses; the results take simpler forms in this case. To state the results, let us write $w(p, i, A)$ for the probability that exactly i components of a p -dimensional $N(0, A)$ random vector are positive. For a discussion on the computation of $w(p, i, A)$, see Shapiro (1988). The next corollary follows from Theorem 1 and the results in Wolak (1987); it includes all the hypotheses considered in Wolak (1987).

COROLLARY 1. Suppose that Conditions A, B(i) and B(ii) are satisfied. Let R be a $q \times p$ matrix, $rank(R) = q \leq p$, and let R be partitioned as $R^t = [R_1^t, R_2^t]$ where R_1 is $r \times p$.

(A) Let the null and alternative hypotheses be respectively $H_0 : R\theta = 0$ and $H_1 : R_1\theta \geq 0$, and let

$\Pi = R_1 B_n R_1^t$. Then, under H_0 , for any $c > 0$, we have

$$pr_{\theta}(RW \geq c) - \sum_{i=0}^r w(r, i, \Pi) pr\{\chi^2(q-r+i) \geq c\} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

(B) Let the null and alternative hypotheses be respectively $H_0 : R_1\theta \geq 0, R_2\theta = 0$ and $H_1 : \theta$ is unrestricted,

and let $\Pi = [R_1 B_n R_1^t - R_1 B_n R_2^t \{R_2 B_n R_2^t\}^{-1} R_2 B_n R_1^t]$. Suppose that Condition (B) (iii) (b) is satisfied.

Then $\sup\{pr_b(RW \geq c) : b \in K\} = pr_{\theta}(RW \geq c)$, and

$$pr_{\theta}(RW \geq c) - \sum_{i=0}^r w(r, r-i, \Pi) pr\{\chi^2(q-r+i) \geq c\} \rightarrow 0 \text{ as } n \rightarrow \infty, \text{ for any } c > 0.$$

In the above results, we considered only the asymptotic null distribution of the test statistic. To study the asymptotic power of RW, we adopt an approach based on Hajek and Sidak (1967) and Hannan (1956). Let us define a sequence of local hypotheses as $H_{0n} : \theta = \theta^*$ 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 $(\theta^* + n^{-1/2}\Delta)$ lies in C for any $n \geq 1$. Let RW_1 and RW_2 be defined as in (2.3) for two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ respectively. We are interested to compare the asymptotic power of RW_1 relative to RW_2 . In general, this is difficult to do. However, if the asymptotic covariance matrices of $\hat{\theta}_1$ and $\hat{\theta}_2$ are proportional then we can obtain an interesting result. To discuss this, let us suppose that $m = m(n)$ is a function of n such that $(m/n) \rightarrow \lambda$ as $n \rightarrow \infty$ for some $0 < \lambda < \infty$, and $0 < \lim pr(RW_{1n} \geq c | H_{1n}) = \lim pr(RW_{2m} \geq c | H_{1n}) < 1$, where the suffices n and m in RW_1 and RW_2 respectively, are the sample sizes. Then, we may refer to λ as the *Pitman asymptotic relative efficiency* (ARE) of RW_1 with respect to RW_2 . Now, we have the following :

THEOREM 2. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators of θ . Suppose that, for $j = 1$ and 2 , $n^{1/2}(\hat{\theta}_j - \theta) \rightarrow_d N(0, \tau_j^2 B^{-1})$ as $n \rightarrow \infty$ for some $\tau_j > 0$, and that Condition B is satisfied for $\hat{\theta}_j$ with τ_j and $\hat{\tau}_j$ substituted for τ and $\hat{\tau}$ respectively. Let RW_j be defined as in (2.3) with $\hat{\theta}_j$, $\hat{\tau}_j$ and B_n , for $j = 1, 2$. Then the Pitman ARE of RW_1 with respect to RW_2 is $(\tau_2/\tau_1)^2$.

In fact, The above results hold in a somewhat more general form. Suppose that θ is replaced by a subvector of it in (2.1), and then (2.2), (2.3) and Condition B are modified appropriately with B_n^{-1} , B^{-1} and $(X^t X)^{-1}$ replaced by their corresponding submatrices. Then, Theorems 1 and 2, and Corollary 1 hold with

only minor modifications to take account of the reduction in the dimension of θ . At the cost of making some discussions somewhat messy, we can allow B and B_n to be functions of θ , but for the situations that we have in mind, they are independent of θ .

4. DISCUSSION OF THE MAIN RESULTS.

Throughout this section we will assume that the errors are iid; note that this is not an assumption in Theorems 1, 2 or Corollary 1. Then, provided that $\sigma^2 < \infty$, Condition B (namely, B (i), (ii) and (iii b)) is satisfied for $\hat{\theta}_L$ with $B_n = n^{-1}X^tX$, $B = V$ and $\hat{\tau} = S$; hence Theorems 1 and 2 and Corollary 1 are applicable to W in (2.2) as well. However, in general, we do not require σ^2 or other moments of the error term to be finite to ensure that Condition B is satisfied for robust estimators such as M- and Bounded Influence estimators. Thus, in contrast to W , RW corresponding to such robust estimators are applicable even if the errors do not have finite moments.

4.1 Symmetric Errors

Suppose 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 Breakdown Point estimator (see Yohai and Zamar (1988), Simpson *et al* (1990)); let RW in (2.3) correspond to this $\hat{\theta}$.

By Theorem 1, the asymptotic null distributions of RW and W are the same, provided $\sigma^2 < \infty$. So, for large n , the critical values and procedures for computing p-values for W are applicable to RW as well. Thus, once $\hat{\theta}$ and $\hat{\tau}$ have been computed, the rest of the computational procedures required for RW are the same as those for W . This is an important result since in some one-sided hypothesis testing problems, computation of critical and p-values for W and RW may be a non-trivial task.

By Theorem 2, the Pitman ARE of RW with respect to W is (σ^2/τ^2) which is precisely the asymptotic efficiency of $\hat{\theta}$ relative to $\hat{\theta}_L$. This is an important observation since we may conclude that the efficiency-robustness properties of $\hat{\theta}$ relative to $\hat{\theta}_L$, which is well documented for many choices of $\hat{\theta}$, carry over to their corresponding Wald-type statistics. We can say the same in comparing an M-estimator with the High

Breakdown Point estimator of Yohai and Zamar (1988) since the asymptotic covariance matrices of these estimators are proportional.

4.2 Asymmetric Errors

Now, we consider the situation when the error distribution may be asymmetric. Let us write the linear model as $Y = 1\alpha + Z\beta + E$, where 1 is a column of ones and β is the slope component of θ . Suppose that the columns of Z are centered so that $1^t Z = 0$. Since the error distribution may be asymmetric, the intercept terms estimated by different estimators are not the same (Silvapulle (1985), Carroll and Welsh (1988)). Fortunately, in most practical situations, the hypotheses involve only the slope component β of θ ; let us assume that this is the case. So, the hypotheses in (2.1) reduce to $H_0 : \beta \in K^*$ and $H_1 : \beta \in C^*$ where K^* and C^* are respectively appropriate projections of K and C on R^{p-1} . Let $\hat{\beta}$ be a robust estimator of β such that $(\hat{\beta} - \beta)$ is asymptotically $N\{0, \tau^2(Z^t Z)^{-1}\}$. So $\hat{\beta}$ could be an M- or R- estimator (see, Silvapulle (1985) and Carrol and Welsh (1988)). Let $\hat{\beta}_L$ be the slope component of $\hat{\theta}_L$. Assume that Conditions A and B are satisfied with $\hat{\beta}$ and β in place of $\hat{\theta}$ and θ respectively. Now, in view of (2.2), (2.3) and the comments made at the end of Section 3, W and RW take the following forms :

$$W = S^{-2} \left[\inf\{(\hat{\beta}_L - b)^t Z^t Z (\hat{\beta}_L - b) : b \in C^*\} \right] \quad (4.1)$$

$$RW = \hat{\tau}^{-2} \left[\inf\{(\hat{\beta} - b)^t Z^t Z (\hat{\beta} - b) : b \in K^*\} - \inf\{(\hat{\beta} - b)^t Z^t Z (\hat{\beta} - b) : b \in C^*\} \right]. \quad (4.2)$$

So, provided that $\sigma^2 < \infty$, we may conclude that the asymptotic null distributions of RW and W are the same, and the Pitman ARE of RW with respect to W is (σ^2/τ^2) which is precisely the asymptotic efficiency of $\hat{\beta}$ relative to $\hat{\beta}_L$. So, again we observe that (i) the asymptotic critical values and procedures for computing p-values for W are also applicable to RW ; and (ii) the efficiency-robustness properties of $\hat{\beta}$ relative to $\hat{\beta}_L$ translate to robustness of RW relative to W .

4.3 Asymmetric Errors and Tests Based on Bounded Influence Estimators

In general, the regularity conditions for the asymptotic normality of Bounded Influence and High Breakdown Point estimators, require the errors to be symmetrically distributed. One exception is the so called Mallows-type estimator or a one-step version of it (see Mallows (1975), Simpson *et al* (1990) and

Giltinan *et al* (1986)); these are essentially weighted M-estimators with appropriately chosen weights which depend only on X . Let $\tilde{\beta}$ be one such estimator of the slope component β of θ , and assume that $(\tilde{\beta}-\beta)$ is asymptotically $N(0, \tau^2 Q_n^{-1})$, where Q_n depends on the design matrix X and $\tau > 0$. With $\hat{\tau}$ as in Simpson *et al* (1990), we may define, based on $\tilde{\beta}$, a robust Wald-type statistic RW_M similar to RW in (4.2) with $\hat{\tau}^{-2}(Z^t Z)$ replaced by $\hat{\tau}^{-2} Q_n$. In general, ignoring small order terms, Q_n is not proportional to $(Z^t Z)$, and hence RW_M and W do not have the same asymptotic null distributions, and Theorem 2 does not help to compute the asymptotic efficiency of RW_M with respect to W . However, if $\tilde{\beta}_M$ is a Mallows-type weighted *least squares* estimator with the weights being the same as for $\tilde{\beta}$, then $(\tilde{\beta}_M-\beta)$ is asymptotically $N(0, \sigma^2 Q_n^{-1})$. Let W_M denote the corresponding Wald-type statistic similar to W in (4.1) with $(Z^t Z)$ replaced by Q_n . Since the asymptotic covariances of $\tilde{\beta}$ and $\tilde{\beta}_M$ are proportional, RW_M and W_M have the same asymptotic null distributions and, the Pitman ARE of RW_M with respect to W_M is (σ^2/τ^2) .

5. SIMULATION STUDY

A simulation study was carried out to evaluate the small sample behaviour of RW relative to W . The main finding of this study is that the asymptotic superiority of RW with respect to W is realized, at least to some extent, in realistic situations.

4.1 Design of the Simulation Study

We considered a two-way analysis of variance model with two rows, three columns, no interaction, and iid errors: $y_{ijk} = \mu + \alpha_i + \gamma_j + \epsilon_{ijk}$, $i = 1, 2, 3$ and $k, j = 1, 2, 3$. To identify the parameters, we chose $\alpha_2 = \gamma_3 = 0$. Now, the 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 to be investigated are H_0 against H_1 , H_1 against H_2 and H_0 against H_2 .

The error distributions 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 $\Phi(t)$ is the standard normal distribution function. Note that the third error distribution is asymmetric; the shift in the contaminating distribution is large enough to make the error

distribution reasonably skewed. The above four error distributions represent a reasonably realistic neighborhood of $\Phi(t)$. The estimator $\hat{\theta}$ used was the M-estimator corresponding to the so called Huber's Proposal 2 (Huber (1977)) with the kinks in the ψ function being at ± 1.5 . Since the hypotheses H_0 , H_1 and H_2 do not involve the intercept term θ_1 , RW is applicable even though one of the error distributions is not symmetric. M-estimators were computed using the algorithm of Huber and Dutter (1974). The inequality constrained estimators were computed using the subroutine BCOAH in the IMSL; see also Robertson *et al* (1988, p 306).

4.2 Results of the Simulation Study

When the errors are normally distributed, the exact finite sample distributions of W based on $\hat{\theta}_L$ for H_0 against H_1 , H_0 against H_2 , and for H_1 against H_2 at the least favorable value $\theta = 0$ are known to be mixtures of F-distributions (see Wolak (1987)). In our simulation study, we found that these mixtures of F-distributions approximated the tails of the null distributions of W and RW better than their asymptotic distributions which, by Corollary 1, are mixtures of chi-squared distributions. Therefore, only the estimates of size and power corresponding to critical values from the mixture of F-distributions are given here (see Table 1). The main observations are the following:

(i) *Error distribution is normal:* The estimated sizes of W and RW are close to the nominal level 5%; in fact, we observed that the estimated sizes were close to nominal level when it was set at 20%, 10% and 1%. The powers of the two statistics are very close to each other. So, we find that there is no noticeable difference between W and RW in terms of size and power.

Table 1 about here

(ii) *Error distribution is contaminated normal:* The estimated sizes are still reasonably close to the nominal level, 5%. As expected, the power of RW is better than that of W ; as the tails of the error distribution become heavier, RW becomes more powerful than W . Even with moderate departures from normality, the power advantages of RW over W are substantial.

We repeated the above simulations with ten observations in each cell; thus $n=60$. The observations (i) and (ii) made above turned out to be applicable here as well, except that the performance of RW relative to W was much better when the error was contaminated normal.

5. A SIMPLE NUMERICAL EXAMPLE

In this section we consider a simple example with artificially generated data. Let us consider the regression model in the previous section with ten observations in each cell. Data for y_1, \dots, y_{60} were generated with $\theta_1=\theta_2=0$, $\theta_3=0.9$ and $\theta_4=1.2$; the error distribution is $0.75 \Phi(t) + 0.25 \Phi\{(t-3)/t\}$.

We are interested in testing $H_0 : \theta_3=\theta_4=0$ against $H_1 : \theta_3, \theta_4 \geq 0$.

Unrestricted least squares estimation gave $\hat{\theta}_L=(0.18, 0.28, 1.30, 1.51)$ and $s^2=8.0$. With the value of $\hat{\theta}_L$ given above, the first expression in the square braces of (2.2) is easy to compute, and the second expression is zero since the third and fourth components of $\hat{\theta}_L$ are positive. Unrestricted M-estimation using Huber's Proposal 2 (Huber (1977)) with the ψ function having kinks at ± 1.5 , gave $\hat{\theta}=(0.04, 0.27, 0.93, 1.48)$ and $\hat{\tau}^2=2.6$; here we used the algorithm in Huber and Dutter (1974). With the above values of $\hat{\theta}$ and $\hat{\tau}$, RW was computed by the same procedure as for W.

Having computed the test statistics, we need to compute approximate p-values. For testing $H_0 : \theta_3=\theta_4=0$ against $H_1 : \theta_3, \theta_4 \geq 0$, $\text{pr}(W \geq c)$ and $\text{pr}(RW \geq c)$ are approximately $\{w_0 + w_1 \text{pr}(\chi_1^2 \geq c) + w_2 \text{pr}(\chi_2^2 \geq c)\}$ where $w_i=w(2, i, \Pi)$ and $\Pi=R(X^t X)^{-1}R^t$. The weights w_i for $i = 0, 1, 2$ were computed using the explicit formulae in Gourieroux *et. al.* (1980); the computed values are $w_0=0.167$, $w_1=0.5$ and $w_2=0.333$. With these weights, approximate p-values were computed easily.

The normal probability plot for the unrestricted least squares residuals indicated that there was some departure from normality; the largest positive and the largest negative residuals appeared to be somewhat large. So we repeated the above computations without these two "outliers". The test statistics and approximate p-values are given below:

	All the data (n=60)		Without the two outliers (n=58)	
W (p-value)	3.4	(0.10)	6.3	(0.025)
RW (p-value)	8.63	(0.02)	10.0	(0.009)

Note that when the outliers are present, RW detected the departure from H_0 but not W. However, when the two outliers are deleted, W detects the departure from H_0 . The change in the p-value of W is substantial compared to that of RW. These observations reflect the robustness of RW compared to W. It should however be noted that a p-value computed after deleting some outliers, do not have the usual interpretation of a p-value. Nevertheless, the above table of p-values are instructive.

REFERENCES

- Carroll, R. J. and Welsh, A. H. (1988). A note on the asymmetry and robustness in linear regression. *The American Statistician*, 42, 285-287.
- 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.
- Gourieroux, C., Holly, A. and Montfort, A. (1982). Likelihood ratio, Wald test and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica*, 50, 63-80.
- Hajek, J. and Sidak, Z. (1967). *Theory of rank tests*. Academic Press.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A. (1986). *Robust Statistics: the approach based on influence functions*. Wiley, New York.
- Hannan, E. J. (1956). The asymptotic powers of certain tests based on multiple correlations. *Jour. Roy. Statist. Soc. B*, 227-233.
- Heathcote, C. R. and Silvapulle, M. J. (1981). Minimum mean squared estimation of location and scale parameters under misspecification of the model. *Biometrika*, 68, 501-514.
- Huber, P. J. (1977). Robust methods of estimating regression coefficients. *Math. Operationsforsch. Statist., Series Statistics*, Vol. 8, No. 1, 41-53.
- Huber, P. J. (1981). *Robust Statistics*. New York, John Wiley and Sons.
- Huber, P. J. and Dutter, R. (1974). Numerical solutions of robust regression problems. In COMPSTAT 1974 (g. Bruckmann, ed.) Physika Verlag, Wien. 165-172.
- Krasker, W. S. and Welsch, R. E. (1983). Efficient bounded-influence regression estimation. *Jour. Amer. Statist. Assoc.* 77, 595-604
- Mallows, C. L. (1975). On some topics in robustness. Unpublished memorandum, Bell Telephone Laboratories, Murray Hill, N. J.
- Perlman, M. D. (1969). One-sided problems in multivariate analysis. *Ann. Math. Statist.*, 40, 549-567.

- Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order restricted statistical inference*. J. Wiley, N. Y.
- 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*. J. Wiley, N. Y.
- Sen, P. K. (1982). On M-tests in linear models. *Biometrika*, 69, 245-8.
- Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *Int. Statist. Rev.*, 56, 49-62.
- Silvapulle, M. J. (1985). Asymptotic behavior of robust estimators of regression and scale parameters with fixed carriers. *Ann. Statist.*, 13, 1490-1497.
- Simpson, D. G., Ruppert, D. and Carroll, R. J. (1990). One-step GM-estimates for regression with bounded influence and high breakdown point. Unpublished.
- Wolak, F. A. (1987). An exact test for multiple inequality and equality constraints in the linear model. *Jour. Amer. Statist. Assoc.*, 82, 782-793.
- Yohai, V. J. and Zamar, R. H. (1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Jour. Amer. Statist. Assoc.* 83, 406-413.

APPENDIX

Proof of Theorem 1. Assume that $\theta \in K$ is fixed. Let $T_{0n} = n^{1/2}\tau^{-1}(\hat{\theta}-\theta)$. Then $T_{0n} \rightarrow_d T_0$ as $n \rightarrow \infty$, where $T_0 \sim N(0, B^{-1})$. Using Conditions B (i) and (ii), it may be verified that

$$RW = n\tau^{-2} \inf\{(\hat{\theta}-b)^t B(\hat{\theta}-b) : b \in K\} - n\tau^{-2} \inf\{(\hat{\theta}-b)^t B(\hat{\theta}-b) : b \in C\} + o_p(1). \quad (\text{A.1})$$

Part (A): This follows since the first two terms on the right hand side of (A.1) are respectively

$$\inf\{(T_{0n}-b)^t B(T_{0n}-b) : b \in K\} \text{ and } \inf\{(T_{0n}-b)^t B(T_{0n}-b) : b \in C\}.$$

Part (B): Since the linear space generated by K is contained in C , we have $C + a\theta = C$ for any real number a , and $K + \theta \subset K$. Now, by arguments similar to those in the proof of Theorem 8.2 of Perlman (1969), we have, for $c > 0$,

$$\begin{aligned} \text{pr}_\theta(RW \geq c) &\leq \\ &\text{pr}_\theta\left[\tilde{\tau}^{-2} \inf\{(\hat{\theta}-b)^t B_n(\hat{\theta}-b) : b \in K+\theta\} - \tilde{\tau}^{-2} \inf\{(\hat{\theta}-b)^t B_n(\hat{\theta}-b) : b \in C\} \geq c\right] \\ &= \text{pr}_0(RW \geq c). \end{aligned}$$

The rest of the proof follows from part (A) and the fact that the first two terms on the right hand side of (A.1) are $\inf\{(T_{0n}-b - n^{1/2}\tau^{-1}\theta)^t B(T_{0n}-b - n^{1/2}\tau^{-1}\theta) : b \in K\}$ and $\inf\{(T_{0n}-b)^t B(T_{0n}-b) : b \in C\}$ respectively.

Proof of Theorem 2: $\{H_{1n}\}$ is contiguous to $\{H_{0n}\}$ (see Hajek and Sidak (1967), and Sen (1980)). Let $K(\theta^*) = \lim(K - n\theta^*)$ as $n \rightarrow \infty$, and $m = m(n)$ be the integer part of $n(\tau_2/\tau_1)^2$. Let $T_{1n} = n^{1/2}\tau_1^{-1}(\hat{\theta}_1 - \theta^* - n^{-1/2}\Delta)$ and $T_{2m} = m^{1/2}\tau_2^{-1}(\hat{\theta}_2 - \theta^* - n^{-1/2}\Delta)$, where $\hat{\theta}_1$ and $\hat{\theta}_2$ are based on samples of sizes n and m respectively. The following arguments hold under $\{H_{1n}\}$: T_{1n} and $T_{2m} \rightarrow_d N(0, B^{-1})$ as $n \rightarrow \infty$;

$$\begin{aligned} n\tau_1^{-2} \inf\{(\hat{\theta}_1-b)^t B_n(\hat{\theta}_1-b) : b \in K\} \\ = \inf\{(T_{1n} + \tau_1^{-1}\Delta - b)^t B(T_{1n} + \tau_1^{-1}\Delta - b) : b \in K(\theta^*)\} + o_p(1); \end{aligned} \quad (\text{A.2})$$

$$\begin{aligned} m\tau_2^{-2} \inf\{(\hat{\theta}_2-b)^t B_m(\hat{\theta}_2-b) : b \in K\} \\ = \inf\{(T_{2m} + \tau_1^{-1}\Delta - b)^t B(T_{2m} + \tau_1^{-1}\Delta - b) : b \in K(\theta^*)\} + o_p(1). \end{aligned} \quad (\text{A.3})$$

Now, the proof is complete since the approximations in (A.2) and (A.3) hold with K and $K(\theta^*)$ replaced by C .

TABLE 1: Estimated⁽¹⁾ Size and Power (%) Corresponding to 5% Nominal Level for n=18.

$\theta^{(3)}$		Test Statistic ⁽²⁾					
θ_3	θ_4	W02	RW02	W01	RW01	W12	RW12
Error distribution: $\Phi(t)$							
0.0	0.0	4	5	4	5	5	5
1.0	1.0	35	37	50	50	38	38
2.0	2.0	91	90	95	96	92	91
Error distribution: $0.8 \Phi(t) + 0.2 \Phi(t/3)$							
0.0	0.0	4	5	4	5	5	5
1.0	1.0	19	23	28	34	20	25
2.5	2.5	73	82	82	91	76	84
Error distribution: $0.8 \Phi(t) + 0.2 \Phi\{(t-2)/3\}$							
0.0	0.0	3	5	4	4	4	4
1.0	1.0	18	22	26	33	16	20
2.5	2.5	67	78	76	86	53	62
Error distribution: $0.8 \Phi(t) + 0.2 \Phi(t/5)$							
0.0	0.0	3	4	3	5	5	5
2.0	2.0	38	54	50	66	39	58
3.0	3.0	62	81	71	88	62	81

(1) All estimates are based on 1000 samples.

(2) W02 means the W statistic for $H_0: \theta_3 = \theta_4 = 0$ against $H_2: \theta_3$ and θ_4 are unrestricted; W01 means the W statistic for $H_0: \theta_3 = \theta_4 = 0$ against $H_1: \theta_3, \theta_4 \geq 0$. Similar interpretations are applicable to RW02, ... etc. The critical values are based on the exact distribution of the W statistic when the error is normally distributed.

(3) These true values of θ_3 and θ_4 are for W02, RW02, W01 and RW01 only. To obtain the corresponding true values for W12 and RW12, change the sign of θ_3 ; $\theta_1 = \theta_2 = 0.0$ throughout.