

ROBUST TESTS OF INEQUALITY CONSTRAINTS AND
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. 1879

May 1990

**ROBUST TESTS OF INEQUALITY CONSTRAINTS
AND ONE-SIDED HYPOTHESES IN
THE LINEAR MODEL**

MERVYN J. SILVAPULLE

University of North Carolina at Chapel Hill and La Trobe University

Abbreviated Title: Robust Tests in the Linear Model

ABSTRACT

For the linear regression model $y_i = x_i^t \beta + \epsilon_i$, $1 \leq i \leq n$, where $\epsilon_1, \dots, \epsilon_n$ are independent and identically distributed, robust tests of various hypotheses on β are developed; by 'robust', we mean robustness of size and power against long tailed error distributions which may not be symmetric. The proposed test statistic, denoted MR, resembles the likelihood ratio statistic since, ignoring scaling and normalizing factors, MR is essentially the same as $\{L(\hat{\beta}^0) - L(\hat{\beta}^1)\}$, where $L(b) = \sum \rho(y_i - x_i^t b)$ is the usual loss function used in M-estimation, ρ is convex, and $\hat{\beta}^0$ and $\hat{\beta}^1$ are the M-estimates under the null and alternative hypotheses respectively. Thus, $L(b)$ takes the place of log likelihood in these tests.

The two main results are (i) under the null hypothesis, the asymptotic distribution of MR and that of the normal theory likelihood ratio (ML) statistic are the same; and (ii) in terms of asymptotic efficiency, the behaviour of MR relative to ML is similar to that of the corresponding M-estimator relative to the least squares estimator. The first result is useful since it enables us to make use of an extensive literature on ML tests; in particular, the critical values and procedures for computing p-values for ML are applicable for MR as well. The second is useful since it tells us that MR is in fact power-robust against long tailed errors. The general results incorporate the important situations such as $H_0: R\beta = 0$ against $H_1: R\beta \geq 0$, $R\beta \neq 0$ and $H_2: R\beta \geq 0$ against $H_3: R\beta \not\geq 0$, where R is a matrix. An example and simulation results illustrate that MR has desirable robustness properties compared to ML, and that the implementation of the former is not any more difficult than that of the latter.

Key Words: asymmetric errors; composite hypotheses; likelihood ratio type tests; linear regression; M-estimator; Pitman efficiency.

1. INTRODUCTION

To motivate the type of situations that we are interested in, let us first consider a simple example. Let y = Blood Pressure, x_1 = level of Fat and x_2 = level of Cholesterol. Consider the linear model $y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \text{error}$. Suppose that β_2 and β_3 are known to be non-negative. A basic question of interest is “does an increase in the level of Fat and/or the level of Cholesterol lead to an increase in Blood Pressure ? This type of questions arise particularly at an early stage of the investigation. If it turns out that Fat and/or Cholesterol affect the Blood Pressure, then it may be of interest to investigate it further. The above question may be formulated as a test of $H_0: R\beta = 0$ against $H_1: R\beta \geq 0, R\beta \neq 0$ where R is a 2×3 matrix with rows $(0, 1, 0)$ and $(0, 0, 1)$; matrix equalities and inequalities are interpreted coordinatewise. We can test H_0 against H_1 using the normal theory likelihood ratio statistic (see Gourieroux *et al* (1980)). In this situation, it would not be desirable to simply test $H_0: R\beta = 0$ against the two-sided alternative $H_2: R\beta \neq 0$, and ignore the prior information available about the signs of β_2 and β_3 ; to do so would result in loss of power (see Robertson *et al* (1988)). In the above example, the null hypothesis involved only equality constraints. However, there are many situations, particularly in econometrics, where the null hypothesis involves inequalities of functions of β ; some of these fall into the framework considered here.

To formulate the problem in a general context, let us consider the linear regression model $y_i = x_i^t \beta + \epsilon_i, i = 1, \dots, n$ where $x_i^t = (1, x_{i2}, \dots, x_{ip})$ are non-stochastic, $\beta^t = (\beta_1, \dots, \beta_p)$ is a vector of unknown parameters and $\epsilon_1, \dots, \epsilon_n$ are independent and identically distributed. For testing a large class of one-sided hypotheses on β , different test statistics which are asymptotically equivalent to the normal theory likelihood ratio statistics are available (see Gourieroux *et al* (1980) and Wolak (1987)). Such normal theory statistics are likely to be sensitive to outliers in the y -variable. For testing $H_0: R\beta = 0$ against the two-sided alternative $H_2: R\beta \neq 0$ where R is a full-rank matrix, Schrader and Hettmansperger (1980) considered robust tests by adapting the theory of M-estimation and obtained quite attractive results. Here we extend their main results to more general one-sided hypotheses testing situations.

For simplicity, let us temporarily assume that the errors are *symmetrically* distributed. Let $\hat{\beta}$ and $\hat{\beta}_L$ denote an M- and the least squares estimators respectively. Assume that $(\hat{\beta} - \beta)$ and $(\hat{\beta}_L - \beta)$ are

asymptotically $N\{0, \tau^2(X^tX)^{-1}\}$ and $N\{0, \nu^2(X^tX)^{-1}\}$ respectively; thus ν^2 is the error variance. Let us denote our robust test statistic that corresponds to $\hat{\beta}$ by MR, and the usual statistic based on the method of least squares by ML. The first main result that we establish is that for testing a large class of one-sided hypotheses, MR and ML have the same asymptotic distribution. Aesthetically, this is appealing since, under the null and hence under contiguous alternatives, the asymptotic distribution theory for MR is not any more complicated than it is for ML. Further, it is a useful result since the critical values and procedures for computing p-values for ML are also applicable to MR as well. However, the result that makes MR appealing is that the Pitman asymptotic efficiency of MR relative to ML turns out to be (ν^2/τ^2) which is precisely the asymptotic efficiency of $\hat{\beta}$ relative to $\hat{\beta}_L$. In other words, if m and n are large positive integers such that $(\nu^2/\tau^2) \simeq (m/n)$, then the local asymptotic power of MR with sample size n is approximately equal to that of ML with sample size m. We already have a vast literature which says that $\hat{\beta}$ is efficiency-robust compared to $\hat{\beta}_L$. The above result says that these efficiency-robustness properties translate to power-robustness of MR relative to ML.

The results discussed in the above paragraph hold for *asymmetrically* distributed errors with a minor modification. In fact, an important point of departure of our work from much other robust-regression literature as developed in Huber (1981) and Hampel *et al* (1986) is that we allow the errors in the regression model to be *asymmetrically* distributed.

2. ASSUMPTIONS, NOTATIONS AND TEST STATISTICS

Let $\hat{\beta}$ be an M-estimator of β obtained by minimizing $L(b, \hat{\sigma}) = \sum \rho\{(y_i - x_i^t b)/\hat{\sigma}\}$ where $\hat{\sigma}$ is a robust estimator of scale for the errors, and ρ is a convex function. Let $\psi(t) = \rho'(t)$ with the prime denoting differentiation. Since the regression model includes an intercept term, we may write it as $y_i = \beta_1 + z_i^t \zeta + \epsilon_i$ where $\zeta = (\beta_2, \dots, \beta_p)^t$ is the slope parameter, and the estimator $\hat{\beta}^t$ as $(\hat{\beta}_1, \hat{\zeta}^t)$. Assume that $\{z_i\}$ are centered by subtracting out their means and adjusting the intercept term accordingly. Let Z and X be the matrices formed by $\{z_1, \dots, z_n\}$ and $\{x_1, \dots, x_n\}$ respectively; so, $X = [a, Z]$ where a is a column of ones and $a^t Z = 0$.

In what follows, we shall assume that the following conditions are satisfied:

Condition C: (i) ψ is non-decreasing, continuous everywhere and twice continuously differentiable everywhere except at a finite number of points; (ii) $x^2\psi'(x)$ and $x^3\psi''(x)$ are bounded; (iii) $n^{-1}X^tX \rightarrow V$ as $n \rightarrow \infty$, where V is positive definite; (iv) For some $\sigma, \tau > 0$, either (a) the error distribution and $\rho(t)$ are symmetric about the origin, $n^{1/2}(\hat{\sigma} - \sigma) = O_p(1)$, $n^{1/2}(\hat{\beta} - \beta)$ is asymptotically $N\{0, \tau^2(X^tX)^{-1}\}$; or (b) $n^{1/2}(\hat{\sigma} - \sigma) = O_p(1)$, $E\{\psi(\epsilon/\sigma)\} = 0$ where E denotes expectation, $(\hat{\zeta} - \zeta)$ is asymptotically $N\{0, \tau^2(Z^tZ)^{-1}\}$ and $n^{1/2}(\hat{\beta}_1 - \beta_1) = O_p(1)$.

A main feature of the above regularity conditions is that they could be satisfied without symmetry conditions on the errors (see Silvapulle (1985)). In the robust regression literature (see Huber (1981), and Hampel et al (1986)), there has been some tendency to assume that the errors are either symmetric or they satisfy the symmetry-like condition, " $E\{\psi(\epsilon/s)\} = 0$ for s in an open neighbourhood of σ ". Such an assumption leads to the result that the asymptotic covariance matrix of an M-estimator of β is proportional to $(X^tX)^{-1}$; this result was used as a basic assumption in Schrader and Hettmansperger (1980). It is important to note that the asymptotic covariance matrix of an M-estimator is not proportional to $(X^tX)^{-1}$, in general (see Silvapulle (1985)). Under symmetry, $\hat{\sigma}$ and $\hat{\beta}$ are asymptotically uncorrelated, and this simplifies the proofs considerably.

2.1 Definition of the Hypotheses and the Robust Statistic

A subset P of the p -dimensional Euclidean space R^p is said to be positively homogeneous if $a \in P$ implies that $\lambda a \in P$ for every $\lambda \geq 0$. Let P_0 and P_1 be closed, convex and positively homogeneous in R^p . Assume that P_0 is a proper subset of P_1 and that the linear space generated by P_0 is contained in P_1 . Let the null and alternative hypotheses be defined as $H_0: \beta \in P_0$ and $H_1: \beta \in P_1, \beta \notin P_0$ respectively; in what follows, for brevity, we shall write H_1 as ' $\beta \in P_1$ ', without ' $\beta \notin P_0$ '. The above formulation is general enough to incorporate the blood pressure example discussed at the outset, and linear and some nonlinear inequalities in β in the null hypothesis. Let us define

$$MR = \hat{\lambda}^{-1}[\inf\{L(b, \hat{\sigma}) : b \in P_0\} - \inf\{L(b, \hat{\sigma}) : b \in P_1\}]$$

where $\hat{\lambda}$ is a consistent estimator of $\lambda = 2^{-1}\{E\psi^2(\epsilon/\sigma)\}\{E\psi'(\epsilon/\sigma)\}^{-1}$. For example, $\hat{\lambda} =$

$\hat{\sigma}^{-1}(n-p)^{-1}\{\sum \psi^2(e_i/\hat{\sigma})\}\{n^{-1}\sum \psi'(e_i/\hat{\sigma})\}^{-1}$, where $e_i = (y_i - x_i^t\hat{\beta})$; we used this estimator in the small sample studies and the example presented later in this paper.

The above definition of MR is a direct generalization of that considered by Schrader and Hettmansperger (1980). We used the abbreviation MR to suggest "M-Ratio" type tests. Hampel *et al* (1986, Chapter 6) also considered test statistics similar to MR which they called τ -tests; their treatment incorporates more general estimation procedures but is restricted to simple null and two-sided alternative hypotheses. Sen (1982) also adapted M-estimation procedures to test one-sided hypothesis in β , but his development is restricted to hypotheses involving a single parameter only.

In the next section, we will give some general results on MR. They will be stated in such a way to enable us to compare them easily with those for normal theory likelihood ratio tests. So, let us define the statistic ML as

$$ML = S^{-2} \left[\inf\{(\hat{\beta}_L - b)'X'X(\hat{\beta}_L - b) : b \in P_0\} - \inf\{(\hat{\beta}_L - b)'X'X(\hat{\beta}_L - b) : b \in P_1\} \right]$$

where $\hat{\beta}_L$ is the unrestricted least squares estimator of β and S^2 is the corresponding error mean square. Note that ML is the normal theory likelihood ratio statistic with S^2 replacing the error variance. Although ML is not the proper normal theory likelihood ratio statistic, they are asymptotically equivalent. An advantage of ML is that, when the errors are exactly normal, the exact null distribution of ML turns out to be a mixture of F-distributions.

If C(iv)(a) is not satisfied then we shall assume that $(b_1, 0, \dots, 0) \in P_0$ for every real b_1 which is the same as saying that the null hypothesis does not impose any restrictions on the intercept, θ_1 . This restriction is unlikely to cause difficulties in most practical applications. Under asymmetry, the intercept term is defined as a function of ψ by the condition $E\psi(\epsilon/\sigma) = 0$ (see, C(iv)(b)). So, it is unclear as to in what context one might be interested to test hypotheses about such an intercept. However, under symmetry, the intercept has a simple interpretation independent of ψ , and our results hold for hypotheses involving it. To avoid a possible confusion, let us note that a treatment effect represented by a difference between two intercepts as in analysis of covariance, is a slope parameter; hence our results are applicable for tests involving such differences in intercepts.

3. THE MAIN RESULTS

In this section we discuss the main results; the proofs are given in the Appendix. Essentially, the main results are (i) the statistics MR and ML have the same asymptotic distribution under the null hypothesis,

(see Theorem 1) and (ii) the efficiency-robustness properties of the M-estimator $\hat{\beta}$ relative to $\hat{\beta}_L$ translates to power-robustness of MR relative to ML (see Theorem 2).

THEOREM 1. (i) The origin " $\beta = 0$ " corresponds to least favourable null distributions for ML and MR; that is, $\sup\{pr_b(ML \geq c) : b \in P_0\} = pr_0(ML \geq c)$ and $\sup\{pr_b(MR \geq c) : b \in P_0\} = pr_0(MR \geq c)$, where pr_b is the probability evaluated at $\beta = b$. (ii) Given $c > 0$ and $b \in P_0$, there exists a function $F(c, b)$ such that $pr_b(MR \geq c) \rightarrow F(c, b)$ as $n \rightarrow \infty$; further, if $\text{var}(\text{error}) < \infty$, then $pr_b(ML \geq c) \rightarrow F(c, b)$ as $n \rightarrow \infty$.

If the null hypothesis is of the form $R\beta=0$ for some matrix R , then P_0 is a linear space and it may be shown that the null distributions of ML and MR do not depend on the unknown β . Suppose that P_0 is not a linear space. Then, the null distributions of ML and MR depend on the unknown β . In view of Theorem 1, we can obtain conservative critical and p-values by assuming that $\beta = 0$.

Let us assume, as is the case for many M-estimators, that condition C(iv) could be shown to be satisfied without any moment conditions on the errors (see Silvapulle (1985) and Carroll and Welsh (1988)). Note that the asymptotic results for ML are applicable only if $\text{var}(\text{error}) < \infty$, however those of MR are applicable without such moment conditions. The above theorem says that for a given hypothesis testing problem, we can use the same set of critical values for ML and MR. If $c(\alpha)$ is the α -level asymptotic critical value for ML computed under the assumption that $\text{var}(\text{error}) < \infty$ then, it is also the α -level asymptotic critical value for MR even if the errors have no moments. On the other hand, if we have a procedure for computing the asymptotic p-value for ML under the assumption $\text{var}(\text{error}) < \infty$, then we can use the same procedure for computing the asymptotic p-value for MR as well even if the errors have no finite moments. Therefore, once MR has been computed, the rest of the computations required for applying MR are the same as those for ML. From a practical point of view, this is an important result since, for some one-sided hypothesis testing problems, computation of the critical and p-values may be a non-trivial problem. The example and simulation in the following sections show that it is rather easy to compute ML and MR, and their corresponding p-values in some special cases of interest.

In practice, hypotheses involving linear functions of β are encountered frequently. Therefore, it would be helpful to apply the above theorem to some such special cases for which the results take somewhat simpler forms. To state the results, let $w(k, i, A)$ denote the probability that exactly i components of a k -

dimensional $N(0, A)$ random vector are positive, where A is a $k \times k$ positive definite matrix; for more discussions on these quantities, see Wolak (1987). The next corollary, which covers a very large proportion of hypothesis testing problems of practical interest, follows from the above theorem and the results in Wolak (1987).

COROLLARY 1. Let R be a $q \times p$ matrix, $\text{rank}(R) = q \leq p$, and R be partitioned as $R^t = [R_1^t, R_2^t]$ where R_1 is $r \times p$.

(i) Let $H_0 : R\beta = 0$ and $H_1 : R_1\beta \geq 0$; further, let $\Pi = R_1(X^tX)^{-1}R_1^t$. Then, under H_0 ,

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

(ii) Let $H_0 : R_1\beta \geq 0, R_2\beta = 0$ and $H_1 : \text{No restriction on } \beta$. Define

$\Pi = R_1(X^tX)^{-1}R_1^t - R_1(X^tX)^{-1}R_2^t\{R_2(X^tX)^{-1}R_2^t\}^{-1}R_2(X^tX)^{-1}R_1^t$. Then $\beta = 0$ corresponds to the least favourable null distribution for MR . Further, for $c > 0$,

$$\text{pr}_0(MR \geq c) - \sum_{i=0}^r w(r, r-i, \Pi) \text{pr}\{\chi^2(q-r+i) \geq c\} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

The above corollary holds with ML in place of MR provided that the errors have finite variance (see, Wolak (1987)); this corollary also includes all the different hypotheses testing situations discussed in Wolak (1987). The main results discussed so far are centered around the null distributions of ML and MR . While these are important for obtaining critical values and p -values, the main reason for introducing MR was that we anticipated that it would have better power-robustness properties against long tail error distributions than does ML . We consider this issue in the next subsection.

3.1 Behaviour of the Test Statistic Under the Alternative Hypothesis

Let us consider a sequence of local hypotheses as in Hajek and Sidak (1967) and Hannan (1956) to illustrate the asymptotic power of MR relative to ML . Some care is needed in defining local alternatives when the null parameter space P_0 is not a linear space. Let β^* be a point on the boundary of P_0 , and Δ be a fixed point such that $(\beta^* + N^{-1/2}\Delta)$ lies in P_1 for any $N > 1$. Now, we define a sequence of local hypotheses as $H_{0N} : \beta = \beta^*$ and $H_{1N} : \beta = \beta^* + N^{-1/2}\Delta$. Suppose that we can choose sample sizes $l(N)$ and $m(N)$ such that as $N \rightarrow \infty$, $\{l(N)/m(N)\} \rightarrow K > 0$, $l(N) \rightarrow \infty$, $m(N) \rightarrow \infty$ and $0 < \lim \text{pr}\{ML \geq c | H_{1N}, n = l(N)\} = \lim \text{pr}\{MR \geq c | H_{1N}, n = m(N)\} < 1$ for some K and any

$c > 0$. Then, we may say that K is the *Pitman Asymptotic Relative Efficiency (ARE) of MR with respect to ML*. Now, we have the following:

THEOREM 2. *The Pitman ARE of MR with respect to ML is $\{\text{var}(\text{error})|\tau^2\}$ where τ is as in Condition C(iv).*

Recall that the asymptotic efficiency of the M-estimator $\hat{\zeta}$ which is the slope component of $\hat{\beta}$, relative to the least squares estimator $\hat{\zeta}_L$ is $\{\text{var}(\text{error})|\tau^2\}$. Thus, the above theorem says that the efficiency-robustness properties of $\hat{\zeta}_L$ relative to $\hat{\zeta}$ carry over to power-robustness of MR relative to ML. To put it somewhat differently, we can say that MR behaves relative to ML in the same way as $\hat{\zeta}$ behaves relative to $\hat{\zeta}_L$. This is an important result since we already have a vast literature which shows that, in general, M-estimators are more efficiency-robust against long tailed errors than is $\hat{\zeta}_L$. Theorem 2 enables us to restate these results in terms of power-robustness.

4. SIMULATION STUDY

The main results given above are asymptotic. So, we carried out a simulation study to compare the performance of MR with that of ML in small samples.

4.1 Design of the Simulation Study

We considered a 2×3 randomized complete block design without interaction : $y_{ijk} = \mu + \alpha_i + \gamma_j + \epsilon_{ijk}$, $i = 1, 2$, $j, k = 1, 2, 3$ and $\alpha_2 = \gamma_3 = 0$. Let us rewrite this model as $Y = X\beta + E$, where $\beta^t = (\mu, \alpha_1, \gamma_1, \gamma_2) = (\beta_1, \beta_2, \beta_3, \beta_4)$. The hypotheses under consideration are : $H_0 : \beta_3 = \beta_4 = 0$, $H_1 : \beta_3, \beta_4 \geq 0$ and $H_2 : \beta \in \mathbb{R}^4$. We will consider all three possible combinations of the above three hypotheses. 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 $\Phi(t)$ is the standard normal distribution function. The second and fourth distributions are symmetrically contaminated normals while the third is asymmetrically contaminated normal with heavy tails to the right. The degree of deviation away from the normal distribution, represented by these distributions is not unrealistic but large enough to affect least squares procedures, as we will see.

The M-estimator used was the so called "Huber's Proposal 2" (Huber (1977)), with the kinks in the ψ function being at ± 1.5 . For testing H_1 against H_2 , the origin corresponds to the least favourable null distributions for ML and MR (see, Theorem 1); hence the critical values were computed for the origin only.

4.2 Results of the Simulation Study

Under normality, the exact null distribution of ML for finite n is a weighted sum of F-distributions (Wolak (1987)). For each case considered in the simulation, we found that the distribution of MR in the right-tail was approximated by this mixture of F-distributions better than by the asymptotic null distribution of MR (see Corollary 1). This is consistent with the observation of Schrader and Hettmansperger (1980, page 95). Therefore, only the estimates corresponding to the critical values computed from a mixture of F-distributions are given here.

The main conclusion that may be drawn from Table 1 are the following:

- (i) *Error distribution is normal:* The estimated sizes of ML and MR are close to the nominal level, 5%; in fact, we observed that the estimated sizes were close to the nominal level when it was set at 20%, 10% and 1% as well. There is no noticeable difference between ML and MR in terms of size and power under every condition that we investigated.
- (ii) *Error distribution is $0.8 \Phi(t) + 0.2 \Phi(t/3)$, $0.8 \Phi(t) + 0.2 \Phi\{(t-2)/3\}$ or $0.8 \Phi\{(t)/3 + 0.2 \Phi\{(t)/5\}$:* The estimated sizes are still reasonably close to the nominal level, 0.05. However, MR has better power properties than has ML. As expected, the power advantages of MR over ML become substantial as the tails of the error distribution become heavier.
- (iii) Note that the same values of β were used for $0.8 \Phi(t) + 0.2 \Phi(t/3)$ and $0.8 \Phi(t) + 0.2 \Phi\{(t-2)/3\}$. For a given value of β , when the error distribution changes from the former to the latter, the powers of ML and MR decrease by almost the same amount. Thus, it appears that asymmetry does not affect MR any more than it affects ML.
- (iv) As expected, for testing H_0 against H_1 , the one sided tests are more powerful than their corresponding two sided tests. One would expect these differences to become larger as the number of one-sided restrictions in the alternative hypothesis increase.

We repeated the above simulation with ten observations per cell. The findings are essentially the same as in (i) - (iv) above except that the performance of MR relative to ML was even better. This was not

unexpected since we know from other studies that when the error distribution has long tails, an M-estimator does not realize its full asymptotic superiority over the least squares estimator in small samples (see Holland and Welsch (1977)).

Table 1 about here

5. A SIMPLE NUMERICAL EXAMPLE

The main objective of this numerical example is to illustrate the application of MR in a simple situation. The regression model, the M-estimator and the hypotheses H_i , $i = 0, 1, 2$, considered are the same as those in Section 4 with ten observations in each cell. Thus, the discussion in this section also provides some more details about the simulation.

Data for y_1, \dots, y_{60} were generated with $\beta_1 = \beta_2 = 0$ and $\beta_3 = \beta_4 = 1.1$, and with the asymmetric error distribution $0.7 \Phi(t) + 0.3 \Phi\{(t-1)/4\}$. Let $\hat{\beta}_L^i$ and $\hat{\beta}^i$ be the least squares and M-estimates of β under H_i , $i = 0, 1, 2$. We are interested to test H_0 against H_1 . So, to compute MR and ML, we need to compute $\hat{\beta}_L^i$, and $\hat{\beta}^i$ for $i = 0, 1, 2$, and the robust scale estimate, $\hat{\sigma}$.

The quantities $\hat{\beta}_L^0$, $\hat{\beta}_L^2$, and s^2 were computed using the standard methods. It may be shown that (see, Wolak (1987), (3.8)) $\hat{\beta}_L^1$ is the value of b which minimizes $f(b)$ subject to $Rb \geq 0$, where $f(b) = (\hat{\beta}_L^2 - b)^t X^t X (\hat{\beta}_L^2 - b)$; R is defined as $H_0 : R\beta = 0$. Since $f(b)$ is convex in b , this optimization problem is easy to solve using, for example, the subroutine BCOAH in IMSL.

The quantities $\hat{\beta}^0$, $\hat{\beta}^2$ and $\hat{\sigma}$ were computed using the algorithm of Huber and Dutter (1974). By definition, $\hat{\beta}^1$ is the value of b which solves "minimize $L(b, \hat{\sigma})$ subject to $Rb \geq 0$ ". Again, since $L(b, \hat{\sigma})$ is convex in b this optimization problem is easy to solve as above.

By Corollary 1, for testing $H_0 : \beta_3 = \beta_4 = 0$ against $H_1 : \beta_3, \beta_4 \geq 0$, $\text{pr}(\text{MR} \geq c)$ and $\text{pr}(\text{ML} \geq c)$, under H_0 , are both approximately $\{w(2, 0, \Pi) + w(2, 1, \Pi) \text{pr}(\chi_1^2 \geq c) + w(2, 2, \Pi) \text{pr}(\chi_2^2 \geq c)\}$, where $\Pi = R(X^t X)^{-1} R^t$ and $c > 0$. The weights $w(2, i, \Pi)$, $i = 0, 1, 2$, were computed using the explicit formulae in Gourieroux *et. al.* (1982, p. 71); the computed values are 0.167, 0.5 and 0.333 respectively. For a general discussion on the computation of these weights, see Wolak (1987). In view of the remarks made in Section 4.2, the p-values for the test statistics, MR and ML, were computed with respect to the exact distribution of ML under normality, namely a weighted sum of F-distributions; the weights are the same as above.

For $i = 0, 1, 2$, let RSS_i and L_i be $\sum (y_j - x_j^t \hat{\beta}_L^i)^2$ and $\sum \rho\{(y_j - x_j^t \hat{\beta}^i)/\hat{\sigma}\}$, respectively. Let W and KT be the Wald and Kuhn-Tucker statistics for testing H_0 against H_1 (Gourieroux, *et. al.* (1982, Section 5)). Now,

$$MR = \hat{\lambda}^{-1}(L_0 - L_1) = 0.476^{-1} (48.7 - 45.7) = 6.3 \text{ (p} \approx 0.025\text{);}$$

$$ML = s^{-2}(RSS_0 - RSS_1) = 2.23^{-2} (288.5 - 273.9) = 3.0 \text{ (p} \approx 0.13\text{);}$$

$$W = n\{(RSS_0/RSS_1) - 1\} = 60\{(288.5/273.9) - 1\} = 3.2 \text{ (p} \approx 0.13\text{);}$$

$$KT = n\{1 - (RSS_1/RSS_0)\} = 60\{(273.9/288.5) - 1\} = 3.1 \text{ (p} \approx 0.13\text{).}$$

The normal probability plot (see Fig 1) of the unrestricted least squares residuals indicates that there is some departure from normality, and it points to the presence of possibly 2 or 3 outliers, but these outliers are not very large. The effects of the outliers are reflected in the p-values of the statistics above; only MR is significant. If the observations corresponding to the largest three outliers are deleted, then we have $MR = 0.46^{-1} (38.5 - 33.9) = 10.0$ and $ML = 1.62^{-2} (163.7 - 140.1) = 9.0$; the p-values for both these are just under 0.01. Note that the change in the p-value of MR is not large, but that of ML is quite large. This reflects the sensitivity of ML, compared to MR, to outliers in the response variable.

Acknowledgements : I wish to thank Professor P.K. Sen for some helpful discussions.

REFERENCES

- Carroll, R. J. and Welsh, A. H. (1988). A note on the asymmetry and robustness in linear regression. *The Amer. Statistician*, 42, 285-287.
- 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.
- Hannan, E. J. (1956). The asymptotic powers of certain tests based on multiple correlations. *Jour. Roy. Stat. Soc. B.*, 227-233.
- Holland, P. W. and Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Commun. Statist. - Theor. Meth.*, A6(9), 813-27.
- Huber, P. J. (1973), Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Statist.*, 1, 799-821.
- Huber, P. J. (1977). Robust methods of estimating regression coefficients. *Math., Operationsforsch. Statist., Ser. Statistics*, 8, 41-53.
- Huber, P. J. (1981). *Robust Statistics*. J. Wiley, New York.
- Huber, P. J. and Dutter, R. (1974). Numerical solutions of robust regression problems. In COMPSTAT 1974 (G. Bruckmann ed.), Physika Verlag, Wien, 165-172.
- 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, 73-101.
- Sen, P. K. (1981). *Sequential Nonparametrics*. J. Wiley, N. Y.
- Sen, P. K. (1982). On M-tests in linear models. *Biometrika*, 69, 245-248.
- Silvapulle, M. J. (1985). Asymptotic behaviour of robust estimators of regression and scale parameters with fixed carriers. *Ann. Statist.*, 13, 1490-1497.
- Wolak, F. A. (1987). An exact test for multiple inequality and equality constraints in the linear model. *Jour. Amer. Statist. Assoc.*, 82, 782-793.

APPENDIX

Let $B = \sigma^{-2} E\{\psi'(\epsilon/\sigma_0)\}V$, $\tau^2 = \sigma^2 E\psi^2(\epsilon/\sigma)\{E\psi'(\epsilon/\sigma)\}^2$, and $\sigma_1^2 = \text{var}(\text{error})$. By arguments similar to those in Silvapulle (1985, proofs of Theorem 1 and 2), we can show that

$$L(b, \hat{\sigma}) = L(\hat{\beta}, \hat{\sigma}) + 2^{-1}n(\hat{\beta} - b)^t B(\hat{\beta} - b) + E_n(b),$$

where $\sup\{|E_n(b)| : n^{1/2}\|b - \beta\| < K\} = o_p(1)$ for any $K > 0$. Now, the convexity of $L(b, \hat{\sigma})$ ensures the following: for $i = 0, 1$, if $\beta \in P_i$, then we have

$$\inf\{L(b, \hat{\sigma}) : b \in P_i\} = L(\hat{\beta}, \hat{\sigma}) + 2^{-1}n \inf\{(\hat{\beta} - b)^t B(\hat{\beta} - b) : b \in P_i\} + o_p(1) \quad (\text{A.1})$$

More detailed proofs of the above may be obtained from the author directly.

Proof of Theorem 1 :

(i) Let $\theta \in P_0$, and let $\hat{\beta}^1$ be defined by $L(\hat{\beta}^1, \hat{\sigma}) = \inf\{L(b, \hat{\sigma}) : b \in P_1\}$. Since the linear space generated by P_0 is contained in P_1 , it may be verified that $P_1 = \theta + P_1$. Now, by arguments similar to the proof of Lemma 8.2 in Perlman (1969), we have

$$\begin{aligned} \text{pr}_\theta(\text{MR} \geq c) &\leq \text{pr}_\theta[\inf\{L(b + \theta, \hat{\sigma}) - L(\hat{\beta}^1, \hat{\sigma}) : b \in P_0\} \geq c] \\ &= \text{pr}_\theta[\inf\{L(b, \hat{\sigma}) - L(\hat{\beta}^1, \hat{\sigma}) : b \in P_0\} \geq c]. \end{aligned}$$

The second step follows from the invariance properties of $\hat{\beta}^1$ and $\hat{\sigma}$ under translation. The proof for ML follows by similar arguments.

(ii) In the proofs below, we shall assume that the error distribution and/or $\rho(t)$ may be asymmetric and that the null hypothesis does not impose restrictions on the intercept. If the error distribution and $\rho(t)$ are symmetric then the null hypothesis may involve the intercept, and our proofs require only minor modifications. Since the X matrix is centered, V is block diagonal with the lower block being $\lim n^{-1}(Z^tZ)$ which we denote by W. Let $T_n = n^{1/2}\tau^{-1}(\hat{\zeta} - \zeta)$ and $U_n = n^{1/2}\sigma_1^{-1}(\hat{\zeta}_L - \zeta)$. Then, $T_n \rightarrow N(0, W^{-1})$ and $U_n \rightarrow N(0, W^{-1})$ in distribution as $n \rightarrow \infty$. Now, by (A.1) for $i = 0$ and 1, we have

$$\begin{aligned} \lambda^{-1}\inf\{L(b, \hat{\sigma}) : b \in P_i\} &= \lambda^{-1}2^{-1}n \inf\{(\hat{\beta} - b)^t B(\hat{\beta} - b) : b \in P_i\} + \lambda^{-1}L(\hat{\beta}, \hat{\sigma}) + o_p(1) \\ &= \inf\{(T_n - c)^t W(T_n - c) : c \in Q_{in}\} + \lambda^{-1}L(\hat{\beta}, \hat{\sigma}) + o_p(1), \quad (\text{A.2}) \end{aligned}$$

where $Q_{i_n} = Q_i - n^{1/2}\tau^{-1}\zeta$ and $Q_i = \{(b_2, \dots, b_p) : (b_1, b_2, \dots, b_p) \in P_i \text{ for all } b_1\}$. For $i = 0$ and 1 , let $Q_{i_\infty} = \lim Q_{i_n}$ as $n \rightarrow \infty$. Then (A.2) holds with Q_{i_n} replaced by Q_{i_∞} . Since $\hat{\lambda} = \lambda + o_p(1)$, we have

$$MR = \inf\{(T_n - c)^t W(T_n - c) : c \in Q_{0_\infty}\} - \inf\{(T_n - c)^t W(T_n - c) : c \in Q_{1_\infty}\} + o_p(1) \quad (A.3)$$

Similarly, if $\text{var}(\text{error}) < \infty$, then

$$ML = \inf\{(U_n - c)^t W(U_n - c) : c \in Q_{0_\infty}\} - \inf\{(U_n - c)^t W(U_n - c) : c \in Q_{1_\infty}\} + o_p(1). \quad (A.4)$$

The rest follows from (A.3) and (A.4) since T_n and U_n have the same limiting distributions.

Proof of Theorem 2. Let β^* be a point on the boundary of P_0 , and Δ be a fixed point such that $(\beta^* + n^{-1/2}\Delta)$ lies in P_1 for any $n > 1$. Now, we define a sequence of local hypotheses as $H_{0n} : \beta = \beta^*$ and $H_{1n} : \beta = \beta^* + n^{-1/2}\Delta$. By arguments similar to those in Hajek and Sidak (1967) (see also Sen (1981)), $\{H_{1n}\}$ is contiguous to H_{0n} . So (A.1) holds for $i = 0$ and 1 with $\beta = \beta^* + n^{-1/2}\Delta$. Let Δ_2 be the slope component of Δ , W be as in (A.2), Q_{0_∞} and Q_{1_∞} be as in (A.3) with ζ^* in place of ζ , and $m = m(n)$ be the integer closest to $\{n(\sigma_1/\tau)^2\}$. Let $(\hat{\zeta}, MR)$ and $(\hat{\zeta}_L, ML)$ be based on n and m observations respectively. Let us define $T_n = n^{1/2}\tau^{-1}(\hat{\zeta} - \zeta^* - n^{-1/2}\Delta_2)$ and $U_m = m^{1/2}\sigma_1^{-1}(\hat{\zeta}_L - \zeta^* - n^{-1/2}\Delta_2)$. Now, the following arguments hold under $\{H_{1N}\}$:

As $n \rightarrow \infty$, T_n and U_m converge to $N(0, W^{-1})$ in distribution. Further,

$$n\tau^{-2} \inf\{(\hat{\beta} - b)^t V(\hat{\beta} - b) : b \in P_i\} = \inf\{(T_n + \tau^{-1}\Delta_2 - c)^t W(T_n + \tau^{-1}\Delta_2 - c) : c \in Q_{i_\infty}\} + o_p(1). \quad (A.5)$$

Let the first term in (A.5) be A_i . Then $MR = (A_0 - A_1) + o_p(1)$. Let B_i be the first term in (A.5) with U_m in place of T_n . Then, by similar arguments, $ML = (B_0 - B_1) + o_p(1)$. The proof follows since $(A_0 - A_1)$ and $(B_0 - B_1)$ have the same asymptotic distribution as $n \rightarrow \infty$.

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

True β ⁽³⁾		Test Statistic ⁽²⁾					
β_3	β_4	ML02	MR02	ML01	MR01	ML12	MR12
Error distribution: $\Phi(t)$							
0.0	0.0	4	4	4	4	5	5
1.5	1.5	67	66	79	78	70	67
2.0	2.0	91	89	95	95	92	91
Error distribution: $0.8 \Phi(t) + 0.2 \Phi(t/3)$							
0.0	0.0	4	5	4	5	5	5
1.5	1.5	37	41	50	56	41	46
2.5	2.5	73	81	82	91	76	83
Error distribution: $0.8 \Phi(t) + 0.2 \Phi\{(t-2)/3\}$							
0.0	0.0	3	4	4	5	4	4
1.5	1.5	33	39	44	50	33	41
2.5	2.5	67	75	76	85	53	62
Error distribution: $0.8 \Phi(t) + 0.2 \Phi(t/5)$							
0.0	0.0	3	4	3	4	5	4
2.0	2.0	38	51	50	65	39	56
3.0	3.0	62	80	71	87	62	80

(1) All estimates are based on 1000 samples.

(2) W02 means the statistic W for H_0 against H_2 ; similarly, RW12 means the statistic RW for H_1 against H_2 ; similar interpretations are applicable for the other abbreviations.

(3) $\beta_1 = \beta_2 = 0$ throughout. The true values of β_3 and β_4 are for the first four tests statistics only (namely, ML02, ..., MR01). To obtain the corresponding values for the last two statistics (namely, ML12 and MR12,), simply change the signs of the values of β_3 .

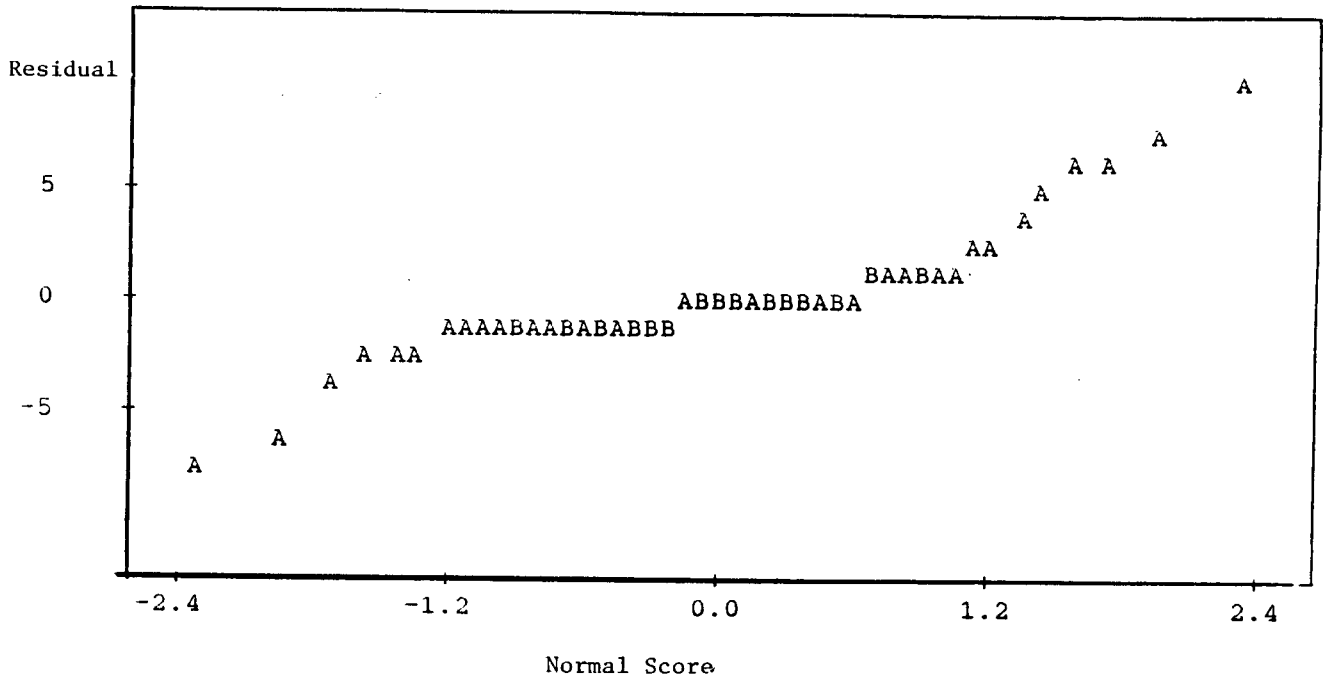


Figure 1. Normal probability plot for the residuals;
 A: one observation, B: two observations etc.