Analysis of least absolute deviation

By KANI CHEN

Department of Mathematics, HKUST, Kowloon, Hong Kong

makchen@ust.hk

ZHILIANG YING

Department of Statistics, Columbia University, NY, NY, 10027, U.S.A.

zying@stat.columbia.edu

HONG ZHANG

Department of Statistics and Finance, USTC, Heifei, Anhui, 230026, China

zhangh@ustc.edu.cn

AND LINCHENG ZHAO

Department of Statistics and Finance, USTC, Heifei, Anhui, 230026, China

lczhao@ustc.edu.cn

SUMMARY

The least absolute deviation or L_1 method is a widely known alternative to the classical least squares or L_2 method for statistical analysis of linear regression models. Instead of minimizing the sum of squared errors, it minimizes the sum of absolute values of errors. Despite its long history and many ground-breaking works (cf. Portnoy and Koenker (1997) and references therein), the former has not been explored in theory as well as in application to the extent as the latter. This is largely due to the lack of adequate general inference procedures under the L_1 approach. There is no counterpart to the simple and elegant analysis-of-variance approach, which is a standard tool in L_2 method for testing linear hypotheses. The asymptotic variance of the L_1 estimator involves the error density, or conditional densities in the case of heterogeneous errors, thereby making the usual standard error estimation difficult to obtain. This paper is an attempt to fill some of the gaps by developing a unified analysis-of-variance-type method for testing linear hypotheses. Like the classical L_2-based analysis of variance, the method is coordinate free in the sense that it is invariant under any linear transformation of the covariates or regression parameters. Moreover, it does not require the design matrix to be of full rank. It also allows the error terms to be heterogeneous. A simple yet intriguing approximation using stochastic perturbation is proposed to obtain cutoff values for the resulting test statistics. Both test statistics and distributional approximations can be computed using the standard linear programming. An asymptotic theory is derived to give theoretical justification of the proposed method. Special cases of one- and multi-way analysis of variance and analysis of covariance models are worked out in details. Extensive simulations are reported, showing that the proposed method works well in practical settings. The method is also applied to a data set from General Social Surveys.

Some Key words: Analysis of variance; Analysis of covariance; Asymptotic expansion; Distributional approximation; Factorial design; Linear constraints; Linear hypothesis; Linear programming; Linear regression; One-way layout; Random perturbation.

1. Introduction

The least squares (LS) method is one of the oldest and most widely used statistical tools for linear models. Its theoretical properties have been extensively studied and are fully understood. Despite its many superior properties, the LS estimate can be sensitive to outliers and, therefore, non-robust. Its performance in terms of accuracy and statistical inferences may be compromised when the errors are large and heterogeneous. The least absolute deviation (LAD) method, which is also known as the L_1 method and has an equally long history (Portnoy and Koenker, 1997), provides a useful and plausible alternative.

Unlike the LS method, the LAD method is not sensitive to outliers and produces robust estimates. Since Charnes, Cooper and Ferguson (1955) reduced the LAD method to a linear programming problem, the computational difficulty is now entirely overcome by the availability of computing power and the effectiveness of linear programming (Koenker and D'orey, 1987). A comprehensive summary of the subject can be found in Portnoy and Koenker (1997). Large sample properties of the LAD estimates are obtained in Bassett and Koenker (1978), Chen, Bai, Zhao and Wu (1990) and Pollard (1991). Due to these developments in theoretical and computational aspects, the LAD method has become increasingly popular. In particular, it has many applications in econometrics and biomedical studies; see Koenker and Bassett (1978), Powell (1984), Buchinsky (1998), Jin, Ying and Wei (2001), among many others.

Compared with the LS method, the LAD is handicapped by two main drawbacks: its lack of convenient inference procedure and lack of valid analysis-of-variance approach. The asymptotic distribution of the LAD estimate involves the density of the errors, making the plug-in inference procedure uncertain and unreliable. The LS method includes the elegant analysis of variance (ANOVA) that provides a unified approach for testing general nested linear hypotheses (Scheffé, 1959). The LAD, on the other hand, has no analogue of the ANOVA approach. In our view, these two drawbacks are the main obstacles for the LAD to be explored and applied to the same extent as the LS method.

The aim of this paper is to fill the gaps so as to make the method of LAD more effective and convenient. The paper develops a general analysis-of-variance-type procedure for testing nested linear hypotheses. Like the LS method, the resulting procedure is invariant to linear transformation of the covariates. It avoids density estimation by introducing a novel and intriguing resampling scheme, so that the distribution of the test statistic can be approximated. The theoretical justification and useful properties are proved via certain stochastic approximations that make use of the modern empirical process theory. Like the analysis of variance, the proposed analysis of absolute deviation is general and retains many nice properties. In particular, it remains invariant regardless of whether the design matrix is of full rank or degenerate. Indeed, the design matrices are often degenerate in the usual formulation of multi-way layouts, which are used to illustrate the proposed method in Section 4. Use of the linear programming greatly facilitates the computation and makes the method easy to implement. Simulation results show that the method works well for practical sample sizes.

The rest of the paper is organized as follows. In the next section, the usual linear model and accompanying linear hypotheses are specified and relevant notations introduced. In Section 3, a new least absolute deviation statistic is introduced for testing nested linear hypotheses. Justification of such a test is given by deriving the usual large sample properties. For the purpose of illustration, one-way and two-way layouts are examined as special cases of linear models in Section 4. The main results are extended to the case of heterogeneous errors in Section 5. In Section 6, simulations are carried out to assess the finite sample performance of the method. An illustration of the method to a real example is given in Section 7. Some concluding remarks are given in Section 8. All proofs are presented in the appendices.

2. Notation and model specification

Consider linear regression model

$$y_i = x_i'\beta + e_i, \quad i = 1, \cdots, n, \tag{1}$$

where the y_i and x_i are, respectively, the univariate response variable and p-variate explanatory variables which may include intercept, β is the p-vector of regression coefficients and e_i is the unobserved error. We assume $med(e_i|x_i) = 0$, i.e., the conditional median of y_i given x_i is $\beta' x_i$. Alternatively, model (1) can be expressed in a matrix form:

$$\mathbf{y} = \mathbf{x}\beta + \mathbf{e}, \tag{2}$$

where $\mathbf{y} = (y_1, \cdots, y_n)'$, $\mathbf{x} = (x_1, \cdots, x_n)'$ and $\mathbf{e} = (e_1, \cdots, e_n)'$. Unless otherwise stated, x_i are assumed to be nonrandom.

The design matrix $\mathbf{x}$ is not necessarily of full rank. This is important as, in many situations, over-parametrization is inherited in designs. Examples of such kind include the commonly used multi-way layouts, where identification of parameters are through additional linear constraints. More details on these examples are provided in Section 4. Let $S_n = \sum_{i=1}^n x_i x_i' = \mathbf{x}'\mathbf{x}$. Non-singularity or full rank of the design matrix $\mathbf{x}$ implies that S_n is positive definite or, equivalently, the rank of S_n is p.

Define linear null hypothesis

$$H_0 : \beta \in \Omega_0 \tag{3}$$

where Ω_0 is a $q(< p)$ dimensional hyperplane in R^p and can be expressed as

$$\Omega_0 = \{b \in R^p : G'b = c_0\}, \tag{4}$$

where $G = (g_1, ..., g_{p-q})$ and c_0 are specified $p \times (p-q)$ matrix and $(p-q)$-vector, respectively, with g_j, $j = 1, \ldots, p-q$, being linearly independent p-vectors.

When the design matrix is intrinsically singular, the observations do not contain sufficient information to identify all parameters without additional constraints. In other words, not all parameters are identifiable and, consequently, not all linear hypotheses are testable. In order for (4) to be testable, $g_j, j = 1, ..., p-q$, must be in Ω, the linear space spanned by $x_1, x_2, \cdots$. To see this, let $p' \leq p$ denote the dimension of Ω and let $u_1, ..., u_p$ denote a set of orthogonal bases of R^p that such that $u_1, ..., u_{p'}$ span Ω. Define $p' \times p$ matrix $U_1 = (u_1, ..., u_{p'})'$ and $(p-p') \times p$ matrix $U_2 = (u_{p'+1}, ..., u_p)'$. Then, $(U_1', U_2')'$ is a $p \times p$ orthonormal matrix. Because x_i is orthogonal to $u_{p'+1}, ..., u_p$, we know that $U_2 x_i = 0$ and that model (1) can be written as

$$y_i = x_i'\beta + e_i = x_i'(U_1', U_2')\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}\beta + e_i = v_i'\gamma + e_i \tag{5}$$

where $v_i = U_1 x_i = (u_1' x_i, ..., u_{p'}' x_i)'$ and $\gamma = U_1\beta = (u_1'\beta, ..., u_{p'}'\beta)'$ are p'-vectors. It follows that values of $U_2\beta$ have no effect on the response y_i and, therefore, observations $(y_i, x_i), i \geq 1$ contain no information about $U_2\beta$. Therefore, in order for the hypothesis (4) to be testable, $g_j, j = 1, ..., p-q$, must belong to the linear space spanned by $u_1, ..., u_{p'}$ or, equivalently, they must be linear combinations of $u_1, ..., u_{p'}$. As a consequence, $p - q \leq p'$. In general, Ω can be expressed as

$$\Omega = \{b \in R^p : \tilde{G}'b = 0\},$$

where $\tilde{G}$ is a $p \times (p - p')$ matrix of rank $p - p'$ such that $\tilde{G}'\mathbf{x} = 0$ for all large n.

The analysis of variance provides an elegant way to test nested linear hypotheses. Let Ω_1 be a q_1 dimensional hyperplane in R^p that contains Ω_0 as a subspace. Then, Ω_1 can be expressed as

$$\Omega_1 = \{b \in R^p : \ G_1'b = c_1\},$$

where $G_1 = (g_1, ..., g_{p-q_1})$ with g_j, $j = 1, ..., p - q_1$, being linearly independent p-vectors and c_1 a $(p - q_1)$-vector. It is common to consider linear alternative hypothesis

$$H_1 : \beta \in \Omega_1 \setminus \Omega_0.$$

Letting $\Omega_1 = R^p$ gives the important special case of $H_1 : \beta \notin \Omega_0$. In the next section, we shall develop a general approach to hypothesis testing under the least absolute deviation criterion. Although it is analogous to the analysis of variance, an easy-to-implement inference procedure is quite nontrivial and rigorous justification requires sophisticated theoretical tools.

3. Methods of analysis of absolute deviation

The classical least squares-based analysis of variance uses the difference between the "least squares" when β is constrained under the two nested hypotheses: $\min_{\beta \in \Omega_0} \sum_{i=1}^n (x_i'\beta - y_i)^2 - \min_{\beta \in \Omega_1} \sum_{i=1}^n (x_i'\beta - y_i)^2$. This difference is further scaled by the "residual sum of squares" to produce the F-ratio for testing H_0 against $H_1 : \beta \in \Omega_1 \setminus \Omega_0$. To develop an analogue with the least absolute deviation, it is natural to consider test statistic

$$\min_{\beta \in \Omega_0} \sum_{i=1}^n |x_i'\beta - y_i| - \min_{\beta \in \Omega_1} \sum_{i=1}^n |x_i'\beta - y_i|.$$

For clarity of presentation, we focus our attention on the important special case of $\Omega_1 = R^p$. Then, the test statistic becomes

$$M_n := \min_{\beta \in \Omega_0} \sum_{i=1}^n |x_i'\beta - y_i| - \min_{\beta \in R^p} \sum_{i=1}^n |x_i'\beta - y_i|. \tag{6}$$

Let $\hat{\beta}_c \in \Omega_0$ and $\hat{\beta} \in R^p$ be any two values of β achieving the first and second minimums in the right hand side of (6), respectively. Note that when the error distribution is double exponential,

M_n coincides with the log-likelihood ratio test statistic. Since the double-exponential distribution has a heavier tail than the normal distribution does, M_n should be less sensitive to outliers and thus more robust.

The minimums as well as $\hat{\beta}_c$ and $\hat{\beta}$ can be computed via linear programming algorithm. Specifically, the first minimum in the right hand side of (6) is the same as the minimum of $\sum_{i=1}^n a_i$ subjective to linear constraints: $G'\beta = c_0$, $a_i \geq x_i'\beta - y_i$ and $a_i \geq -x_i'\beta + y_i$, $1 \leq i \leq n$. Therefore, the standard linear programming applies and the computation becomes routine. Under suitable regularity conditions, the LAD-type statistic M_n converges in distribution as $n \to \infty$. In particular, assuming homogenous error terms, Theorem 1 shows that M_n converges in distribution to $\chi^2_{p-q}/\{4f(0)\}$, where $f(\cdot)$ is the common density function of e_i. Unlike the least squares method, the limiting distribution of M_n, the test statistic based on LAD, involves the density function of the error terms. To avoid density estimation, we propose the following distributional approximation based on random weighting by exogenously generated i.i.d. random variables and on suitable centering adjustments. The approach can again be implemented with the simple linear programming.

Let $w_1, \cdots, w_n$ be a sequence of independent and identically distributed (i.i.d.) nonnegative random variables, with mean and variance both equal to 1. The standard exponential distribution has mean and variance equal to 1. Define

$$M_n^* := \min_{\beta \in \Omega_0} \sum_{i=1}^n w_i |x_i'\beta - y_i| - \min_{\beta \in R^p} \sum_{i=1}^n w_i |x_i'\beta - y_i| - (\sum_{i=1}^n w_i |x_i'\hat{\beta}_c - y_i| - \sum_{i=1}^n w_i |x_i'\hat{\beta} - y_i|). \tag{7}$$

We intend to use the resampling distribution of M_n^* to approximate the distribution of M_n. This is justified if it can be shown that the conditional distribution of M_n^* given data converges to the same limiting distribution as that of M_n. Because the resampling distribution can be approximated arbitrarily close by repeatedly generating a large number of i.i.d. sequences $w_1, \cdots, w_n$, we can use the conditional empirical distribution of M_n^* to get critical regions for M_n. Clearly this approach avoids any density estimation. Moreover, M_n^* can be computed using linear programming algorithm.

REMARK 1. The bootstrap method cannot provide a correct approximation of the distribution of M_n. The bootstrap statistic is

$$\min_{\beta \in \Omega_0} \sum_{i=1}^n |\beta' x_i^* - y_i^*| - \min_{\beta \in R^p} \sum_{i=1}^n |\beta' x_i^* - y_i^*|,$$

where $(x_i^*, y_i^*), i = 1, ..., n$, are the bootstrap sample based on $(x_i, y_i), i = 1, ..., n$. It can be shown that the conditional distribution of the above bootstrap statistic given data does not converge to the limiting distribution of M_n.

REMARK 2. The use of M_n^* is motivated from ideas of Parzen et al. (1994) and Jin et al. (2001) but with a nontrivial modification. A direct mimic of Parzen et al. (1994) and Jin et al. (2001) would use

$$\min_{\beta\in\Omega_0}\sum_{i=1}^n w_i|x_i'\beta - y_i| - \min_{\beta\in R^p}\sum_{i=1}^n w_i|x_i'\beta - y_i|,$$

to approximate the distribution of M_n. However, it can be shown that the resampling distribution of the above statistic does not approximate the distribution of M_n. Therefore a modification is necessary.

REMARK 3. If S_n is of full rank p, i.e., $p' = p$, both $\hat\beta$ and $\hat\beta_c$ as minimizers are asymptotically unique in the sense that they converge to some fixed values under regularity conditions. When S_n is of rank $p' < p$, the sets of minimizers $\hat\beta$ and $\hat\beta_c$ contain unbounded linear spaces, but their projections onto Ω, denoted by $\Pi(\hat\beta|\Omega)$ and $\Pi(\hat\beta_c|\Omega)$ respectively, are asymptotically unique. In fact,

$$\min_{\beta\in\Omega_0}\sum_{i=1}^n w_i|x_i'\beta - y_i| = \min_{\beta\in\Omega_0\cap\Omega}\sum_{i=1}^n w_i|x_i'\beta - y_i| \quad \text{and} \quad \min_{\beta\in R^p}\sum_{i=1}^n w_i|x_i'\beta - y_i| = \min_{\beta\in\Omega}\sum_{i=1}^n w_i|x_i'\beta - y_i|.$$

The above equations still hold when w_i are replaced by 1. The sets of minimizers of the first and second minimums of (6) are, respectively, $\{\Pi(\hat\beta|\Omega)+b : b\in\Omega_\perp\}$ and $\{\Pi(\hat\beta_c|\Omega)+b : b\in\Omega_\perp\}$, where $\Omega_\perp$ is the collection of all p-vectors orthogonal to Ω. Despite the non-uniqueness of $\hat\beta$ and $\hat\beta_c$, M_n^* remains the same for any choices of $\hat\beta$ and $\hat\beta_c$. Denote respectively by $\hat\beta_c^*\in\Omega_0$ and $\hat\beta^*\in R^p$ any two values of β achieving the first and second minimums in the definition of M_n^*. Then, M_n^* can be expressed as

$$M_n^* = \sum_{i=1}^n w_i|x_i'\hat\beta_c^* - y_i| - \sum_{i=1}^n w_i|x_i'\hat\beta^* - y_i| - (\sum_{i=1}^n w_i|x_i'\hat\beta_c - y_i| - \sum_{i=1}^n w_i|x_i'\hat\beta - y_i|). \tag{8}$$

Similarly, $\hat\beta_c^*$ and $\hat\beta^*$ are not asymptotically unique if S_n is degenerate but $\Pi(\hat\beta^*|\Omega)$ and $\Pi(\hat\beta_c^*|\Omega)$ are asymptotically unique.

Throughout the paper, the inverse of any possibly degenerate symmetric matrix A, denoted by A^{-1}, is its Moore-Penrose inverse such that $AA^{-1}A = A$, $A^{-1}AA^{-1} = A^{-1}$ and that AA^{-1} and $A^{-1}A$ are symmetric. More specifically, if $A = T'\Lambda T$ where T is an orthonormal matrix and Λ is a diagonal matrix with nonnegative diagonal elements $\lambda_i, i = 1, ..., p$, then $A^{-1} = T'\bar\Lambda T$ where $\bar\Lambda$ is a diagonal matrix with diagonal elements $1/\lambda_i I(\lambda_i > 0), i = 1, ..., p$. Let β_0 denote the true value of β.

In the following, we shall assume that the error terms are independent of the covariates. Namely, the errors are homogeneous. This is to provide intuitive understanding for situations with degenerate design matrices since the relevant limiting distributions have simple expressions under these assumptions. Extension to heterogeneous errors is given in Section 5.

(A1) The error terms e_i are i.i.d. with median 0 and common density function f such that $f(0) > 0$.

(A2) $d_n := \max_{1\le i\le n}\{x_i' S_n^{-1} x_i\} \to 0$ as $n \to \infty$.

(A3) The random weights $w_1, w_2, \cdots$ are i.i.d. nonnegative random variables such that $E(w_1) = \mathrm{var}(w_1) = 1$, and the sequences $\{w_i\}$ and $\{x_i, e_i\}$ are independent.

THEOREM 1. *Suppose that (A1)-(A3) hold. Then, under the null hypothesis (3),*

$$\mathcal{L}(M_n^*|y_1, \cdots, y_n) \to \chi^2_{p-q}/\{4f(0)\} \leftarrow \mathcal{L}(M_n), \tag{9}$$

as $n \to \infty$, where $\mathcal{L}$ denotes distribution or conditional distribution.

Condition (A1) is standard in median regression. Condition (A2) can be viewed as a Lindeberg-type condition. Condition (A3) is simply a restatement of how the random weights should be generated; see Jin et al. (2001). Theorem 1 establishes the validity of using the resampling distribution of M_n^* to approximate the distribution of M_n. The main advantage is that it avoids density estimation and is numerically easy to implement.

REMARK 4. For testing general nested linear hypotheses H_0 against $H_1 : \beta \in \Omega_1 \setminus \Omega_0$, the statistics M_n and M_n^* are defined the same except with minimization over R^p replaced by over Ω_1. Theorems 1 and 3 still hold with the degree of freedoms $p - q$ replaced by $q_1 - q$, where q_1 is the dimension of the hyperplane Ω_1.

The least absolute deviation method provides $\hat{\beta}$, though not necessarily unique, as a natural estimator of β. In the case of the LS-based linear regression, the LS estimator of any estimable (linear) function is unique by the Gauss-Markov theorem (Scheff'e, 1959). The same is true asymptotically for the LAD estimation. In fact, the asymptotic normality is presented in the following theorem.

THEOREM 2. *Under conditions (A1) and (A2), for any $b \in \Omega$,*

$$\frac{b'(\hat{\beta} - \beta_0)}{(b' S_n^{-1} b)^{1/2}} \to N\left(0, \frac{1}{\{2f(0)\}^2}\right), \tag{10}$$

in distribution as $n \to \infty$.

Theorem 2 essentially gives the asymptotic normality about the projector of $\hat{\beta} - \beta_0$ onto Ω. An equivalent presentation of (10) is

$$V_n^{-1/2}(\hat{\gamma} - \gamma) \to N\left(0, \frac{1}{\{2f(0)\}^2} I_{p'}\right),$$

in distribution as $n \to \infty$, where $I_{p'}$ is the $p' \times p'$ identity matrix, $\hat{\gamma} = U_1\hat{\beta}$, $\gamma = U_1\beta$ and $V_n = \sum_{i=1}^n U_1 x_i x_i' U_1'$, following the notation in Section 2. As discussed in Section 2, $(y_i,\ x_i), i \geq 1$, contain only parameter information about γ and no information about $U_2\beta$. Therefore $U_2\beta$ is not identifiable and $U_2\hat{\beta}$ can be any vector in $R^{p-p'}$.

4. Special cases

Of particular interest in the classical analysis of variance is the case of multi-way layout. To demonstrate parallel development in the analysis of least absolute deviation, we consider one-way and two-way layouts as special but typical cases. A key feature is that the design matrices are degenerate and the models in themselves are over-parametrized, leaving some parameters unidentifiable. We give explicit expressions of the test statistic M_n for some commonly considered hypotheses. Throughout this section, $\mathbf{1}_t$ and $\mathbf{0}_t$ respectively stand for t-vectors with all elements being 1 and 0.

4.1. One-way layout

The conventional one-way layout model assumes

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad j = 1, \cdots, n_i, \quad i = 1, \cdots, a, \tag{11}$$

where y_{ij} is the jth response on the ith level, μ is the overall median, α_i represents the effect of level i and e_{ij} is the random error with median 0. Let $n = \sum_{i=1}^a n_i$. This model can be written in the matrix form (2) with $\mathbf{y} = (y_{11}, \cdots, y_{1n_1}, \cdots, y_{a1}, \cdots, y_{an_a})'$, $\beta = (\mu, \alpha_1, \cdots, \alpha_a)'$, $\mathbf{e} = (e_{11}, \cdots, e_{1n_1}, \cdots, e_{a1}, \cdots, e_{an_a})'$ and

$$\mathbf{x} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{1}_{n_1} & \mathbf{0}_{n_1} & \cdots & \mathbf{0}_{n_1} \\ \mathbf{1}_{n_2} & \mathbf{0}_{n_2} & \mathbf{1}_{n_2} & \cdots & \mathbf{0}_{n_2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{n_a} & \mathbf{0}_{n_a} & \mathbf{0}_{n_a} & \cdots & \mathbf{1}_{n_a} \end{pmatrix}.$$

Following the notations in Section 2, $\mathbf{x}$ is an $n \times p$ matrix with reduced rank p', where $p' = a < p = a + 1$. Let $\tilde{H}_2 = (-1, 1, \cdots, 1)'$ be of dimension p. Then $\mathbf{x}\tilde{H}_2 = 0$ and $\text{rank}(\tilde{H}_2) = 1 = p - p'$.

A null hypothesis of common interest is

$$H_0 : \alpha_1 = \cdots = \alpha_a. \tag{12}$$

Other typical linear hypotheses include, for example,

$$H_0 : \alpha_k = \alpha_l, \tag{13}$$

for the purpose of comparing the effects of levels k and l. The test statistics for hypotheses (12) and (13), as special cases of M_n, are, respectively,

$$\min_{\mu} \sum_{i=1}^{a} \sum_{j=1}^{n_i} |y_{ij} - \mu| - \min_{\mu_1, \cdots, \mu_a} \sum_{i=1}^{a} \sum_{j=1}^{n_i} |y_{ij} - \mu_i|$$

and

$$\min_{\mu} [\sum_{j=1}^{n_k} |y_{kj} - \mu| + \sum_{j=1}^{n_l} |y_{lj} - \mu|] - \min_{\mu_k, \mu_l} [\sum_{j=1}^{n_k} |y_{kj} - \mu_k| + \sum_{j=1}^{n_l} |y_{lj} - \mu_l|].$$

Model (11) is saturated and can be written as

$$y_{ij} = \mu_i + e_{ij}, \quad j = 1, \cdots, n_i, \quad i = 1, \cdots, a.$$

The corresponding design matrix is of full rank. A necessary and sufficient condition for consistently estimating $\mu_1, ..., \mu_a$ or a consistent test of hypothesis (12) is

$$\min(n_i : 1 \leq i \leq a) \to \infty, \tag{14}$$

which is equivalent to condition (A2).

4.2. Two-way layout

Consider the following two-way layout without interactions:

$$y_{ijk} = \mu + \alpha_i + \gamma_j + e_{ijk}, \quad k = 1, ..., n_{ij}, \quad i = 1, \cdots, a, \quad j = 1, \cdots, b. \tag{15}$$

Let $n_{i\cdot} = \sum_{j=1}^{b} n_{ij}$, $n_{\cdot j} = \sum_{i=1}^{a} n_{ij}$ and $n = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}$. Model (15) can be presented in matrix form (2) with $\mathbf{y} = (y_{111}, \cdots, y_{11n_{11}}, \cdots, y_{abn_{ab}}, \cdots, y_{abn_{ab}})'$, $\beta = (\mu, \alpha_1, \cdots, \alpha_a, \gamma_1, \cdots, \gamma_b)'$, $\mathbf{e} = (e_{111}, \cdots, e_{11n_{11}}, \cdots, e_{abn_{ab}}, \cdots, e_{abn_{ab}})'$,

$$\mathbf{x} = \begin{pmatrix} \mathbf{1}_{n_{1\cdot}} & \mathbf{1}_{n_{1\cdot}} & & & A_1 \\ \vdots & & \ddots & & \vdots \\ \mathbf{1}_{n_{a\cdot}} & & & \mathbf{1}_{n_{a\cdot}} & A_a \end{pmatrix} \quad \text{and} \quad A_k = \begin{pmatrix} \mathbf{1}_{n_{k1}} & & \\ & \ddots & \\ & & \mathbf{1}_{n_{kb}} \end{pmatrix}, \quad k = 1, ..., a.$$

Notice that $\mathbf{x}$ is an $n \times (a+b+1)$ matrix of rank $p' = a+b-1 < p = a+b+1$. Consider hypothesis (12). The test statistic M_n can be written as

$$M_n = \min_{\gamma_j} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} |y_{ijk} - \gamma_j| - \min_{\alpha_i, \gamma_j} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} |y_{ijk} - \alpha_i - \gamma_j|.$$

Define, for $j = 1, ..., b$, $B_j = \{i : n_{ij} \to \infty, i = 1, ..., a.\}$. Then, $\alpha_{i_1} - \alpha_{i_2}$ is identifiable if $i_1, i_2 \in B_j$. Similar to the concept of connectivity in the state space of a Markov chain, define an equivalence relation among the integers $\{1, ..., a\}$ based on 1), members of B_j are equivalent to each other for any given j, and 2), this equivalence relation possesses the property of transitivity, i.e., if i_1 and i_2 are equivalent and i_2 and i_3 are equivalent, then i_1 and i_3 are equivalent. Then, it can be shown that a necessary and sufficient condition for a consistent test of (12) is that any two integers of $\{1, ..., a\}$ are equivalent to each other, which is implied in (A2). Notice that (A2) is a general condition to ensure the validity of testing any linear null hypothesis within Ω using the resampling method.

Consider now two-way layout with interactions:

$$y_{ijk} = \mu + \alpha_i + \gamma_j + \eta_{ij} + e_{ijk}, \quad k = 1, \cdots, n_{ij}, \quad i = 1, \cdots, a, \quad j = 1, \cdots, b, \tag{16}$$

where η_{ij} is the interaction term. Linear constraints $\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \gamma_j = \sum_{i=1}^{a} \eta_{ij} = \sum_{j=1}^{b} \eta_{ij} = 0$ can be imposed for the identifiability of all parameters. The design matrix $\mathbf{x}$ is an $n \times (ab+a+b+1)$ matrix of rank $p' = ab < p = ab + a + b + 1$. For testing hypothesis (12), the test statistic is

$$M_n = \min_{s_{\cdot 1} = \cdots = s_{\cdot b}} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} |y_{ijk} - s_{ij}| - \min_{s_{ij}} \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n_{ij}} |y_{ijk} - s_{ij}|,$$

where $s_{\cdot j} = \sum_{i=1}^{a} s_{ij}$. Model (16) is a saturated model, like the one-way layout model (11). Under this model, a necessary and sufficient condition for a consistent test of hypothesis (12) is $\min\{n_{i\cdot} : i = 1, ..., a\} \to \infty$. Again this condition is ensured by (A2).

5. Extension to heterogeneous errors

With heterogeneous errors, i.e., the errors are dependent on the covariates, the limiting distributions of M_n and M_n^* become much more complicated than $\chi^2_{p-q}/\{4f(0)\}$ given in (9) for the case of homogeneous errors. However, the distribution of M_n and the resampling distribution of M_n^* still converge to the same limit. Therefore the proposed resampling method is valid. This is proved in Theorem 3. Some regularity conditions as given below are needed.

(A1)' $\{(x_i, e_i), i \geq 1\}$ are i.i.d. random variables; the conditional median of e_1 given x_1 is 0; the conditional density function $f(\cdot|x_1)$ of e_1 given x_1 is continuous and uniformly bounded; and $f(0|x_1) > 0$.

(A2)' $J_0 := E\{f(0|x_1)x_1x_1'\}$ and $V := E(x_1x_1')$ are both finite and nonsingular.

Theorem 3. *Suppose (A1)', (A2)' and (A3) hold. Then, under the null hypothesis in (3),*

$$\mathcal{L}(M_n^*|x_1, y_1, \cdots, x_n, y_n) \to \mathcal{L}(Z) \leftarrow \mathcal{L}(M_n), \tag{17}$$

as $n \to \infty$, where Z is sum of squares of $p - q$ normal random variables. Moreover,

$$n^{1/2}(\hat{\beta} - \beta_0) \to N(0, \frac{1}{4}J_0^{-1}VJ_0^{-1}). \tag{18}$$

In particular, if the covariates are independent of the errors, then $J_0 = f(0)V$, the distribution of Z is $\chi^2_{p-q}/\{4f(0)\}$ and

$$n^{1/2}(\hat{\beta} - \beta_0) \to N(0, \frac{1}{\{2f(0)\}^2}V^{-1}), \tag{19}$$

as $n \to \infty$, where $f(0) = f(0|x_1)$.

Theorem 3 holds in broader scenarios than restricted by conditions (A1)' and (A2)', which are assumed here mainly for clarity of presentation. In particular, when the matrices J_0 and V are singular, (17) still holds and asymptotic normality of components of $\hat{\beta}$ can be presented in the fashion of Theorem 2. Under a Lindeberg type condition and some further conditions on the limits of $(1/n)\sum_{i=1}^n x_ix_i'$ and $(1/n)\sum_{i=1}^n x_ix_i'f(0|x_i)$, Theorem 3 can be proved without requiring the covariates x_i follow the same distribution. Typical examples of such kind are multi-way layouts and the analysis of covariance model

$$y_{ij} = \beta' x_{ij} + \mu + \alpha_i + e_{ij}, \quad j = 1, \cdots, n_i, \quad i = 1, \cdots, a.$$

The details are technical and are omitted.

Because the errors can be dependent on covariates, the limiting distribution in (17) has a rather complex expression, unlike that given in (9). It involves matrix algebra and the conditional densities of the errors given the covariates; see Appendix C. Therefore the plug-in inference procedures become increasingly unreliable and difficult, especially when the covariates are multi-dimensional and the curse of dimensionality occurs in the estimation of conditional densities. This highlights yet another advantage of using the proposed resampling method to approximate the distribution of M_n.

6. Numerical studies

The aim of the numerical studies conducted extensively is to examine the asymptotic properties established in Theorems 1-3 for finite samples. The simulation results presented here are under model

$$y = \beta_1 x_1 + \beta_2 x_2 + e,$$

where x_1 and x_2 are two independent covariates and β_1 and β_2 are the regression parameters. The true values of β_1 and β_2 are respectively 1 and 0. We test the null hypothesis $H_0 : \beta_2 = 0$ under 16 models as combinations of different distributions of the covariates and errors. The covariate x_1 can be constant 1 or standard normal $N[0, 1]$, and x_2 follows the uniform distribution on $[-2, 2]$ or the standard normal distribution. For the purpose of presentation, we label A, B, C and D respectively for the four choices of distribution of (x_1, x_2): $(1, U[-2, 2]), (N[0, 1], U[-2, 2]), (1, N[0, 1])$ and $(N[0, 1], N[0, 1])$. The random error follows one of the four distributions: the standard normal $N[0, 1]$, the standard Cauchy $Cauchy[1]$, conditional normal $N[0,(1 + |x_2|)^2/4]$ and conditional Cauchy $Cauchy[(1 + |x_2|)/2]$ given (x_1, x_2), which are labelled as a, b, c and d, respectively. The former two are homogeneous errors and the latter two are heterogeneous errors. Therefore, the 16 models can be evidently represented as combinations of A, B, C and D with a, b, c and d. The sample size n is set as 200 and the resampling size N is 10000. The simulation results are based on 5000 replications. The following figures show quantile-quantile plots of M_n^* with respect to M_n.

INSERT FIGURE 1 HERE

The 16 plots are nearly all straight lines, implying that the distributions of M_n and M_n^* are indeed very close to each other in all 16 models. This indicates that the proposed resampling procedure for inference indeed works well.

Recall that, for tests at significance level α, the null hypothesis is rejected if M_n is larger than the $(1 - \alpha)$ quantile of the resampling distribution of M_n^*. Choosing $\alpha = 0.1, 0.05$ and 0.025, we compare the actual type I errors with the corresponding significance levels. For comparison we also present in the parentheses the actual type I errors for the F-test based on the LS method.

INSERT TABLE 1 HERE

It is seen that the type I errors are generally close to the corresponding significance levels. This demonstrates further numerical evidence in support of the validity of the resampling procedure for inference. In contrast, the F-tests based on the LS method are far less accurate when the errors are heterogeneous or do not follow normal distribution. In summary, the simulation results show that the proposed analysis of LAD is indeed an ideal alternative to the commonly used ANOVA based on the LS method.

7. Real example

The data set in analysis is a fragment of General Social Surveys (GSS) regularly conducted by the National Opinion Research Center in 1993, which is available in SPSS. The purpose of the study is to examine whether different groups defined by different Likert scales have different oldness represented by "age", which is the response variable. The factor "polviews" is a 7-point Likert scale including "Extremely Liberal", "Liberal", "Slightly Liberal", "Moderate", "Slightly Conservative", "Conservative", "Extremely Conservative". Figure 2 gives the histogram plot and quantile-quantile (Q-Q) plot of the residuals of the LS fit against the standard normal distribution.

INSERT FIGURE 2 HERE

It is seen that the residuals are evidently skewed to the right. It indicates that using the traditional ANOVA method based on the LS method and the associated F-test or likelihood ratio test may not be appropriate, as the errors deviate from normality quite significantly. On the other hand, the analysis of LAD proposed in this paper does not require the normality or homogeneity of the errors.

We conduct a test to examine whether the effects of the 7 levels are the same. $10,000$ independent random weights from mean 1 exponential distribution are generated to calculate the p-value. The resulting p-value is 0.000, showing high significance of the differences in effects of the 7 levels. We then carry out pairwise comparison tests. The p-values are listed in the upper triangle of Table 2. The estimated differences of the medians are given in the parentheses, and the significant ones at 0.05 level are marked with asterisks. The traditional multiple comparisons based on LS method are also shown in the lower triangle of the table.

INSERT TABLE 2 HERE

It is interesting to see that the p-values by the proposed method are generally smaller. For example, using the proposed test, we conclude that the difference between "Slightly Liberal" and "Liberal" is significant at 0.05 level (with p-value 0.006), while it is not significant by the traditional method (with p-value 0.575). In summary, we believe the LAD method might be more trustworthy than the LS method as the former is more robust, especially when the errors appear to be quite skewed, which is the case in this example.

8. Discussion

The aim of this paper is to provide convenient inference for analysis-of-variance type of tests of linear hypothesis based on the LAD criterion. The proposed inference procedure via resampling avoids the difficulty of density estimation and is convenient to implement with the availability of the standard linear programming and computing power. As the LAD method is more robust than the LS method, the result obtained in this paper may broaden the applications of the former. Some of the conditions assumed for the main results may be dropped or relaxed and, in particular, the errors may not have to follow the same distribution. In contrast, the LS method, especially the associated F-test, generally requires the errors follow the same normal distribution. In addition, we believe the method can be extended to cases with presence of censorship, which are common in survival analysis and econometrics; cf. Powell (1984) and Ying, Jung and Wei (1995).

Appendix A: *Proof of Theorem 1*

The null hypothesis H_0 are assumed throughout the proof, i.e., $\beta_0 \in \Omega_0$. Following the notation in Section 2 and without loss of generality, let $h_1 = u_1, ..., h_{p-q} = u_{p-q}$ and $K = (u_{p-q+1}, ..., u_{p'})$. Then, K is of rank $p' - p + q$ and $H'K = 0$.

We first assume S_n is of full rank p, i.e., $p' = p$. Let $\hat{\eta}^* = K'(\hat{\beta}_c^* - \beta_0)$ and $\hat{\eta} = K'(\hat{\beta}_c - \beta_0)$ be vectors in R^q. Under conditions (A1)-(A3), Theorem 1.1 of Rao and Zhao (1992) implies that

$$2f(0)S_n^{1/2}(\hat{\beta}^* - \beta_0) = \sum_{i=1}^{n} w_i \mathrm{sgn}(e_i) S_n^{-1/2} x_i + o_p(1), \tag{20}$$

$$2f(0)(K'S_nK)^{1/2}\hat{\eta}^* = \sum_{i=1}^{n} w_i \mathrm{sgn}(e_i)(K'S_nK)^{-1/2}K'x_i + o_p(1), \tag{21}$$

$$2f(0)S_n^{1/2}(\hat{\beta} - \beta_0) = \sum_{i=1}^{n} \mathrm{sgn}(e_i) S_n^{-1/2} x_i + o_p(1), \tag{22}$$

$$2f(0)(K'S_nK)^{1/2}\hat{\eta} = \sum_{i=1}^{n} \mathrm{sgn}(e_i)(K'S_nK)^{-1/2}K'x_i + o_p(1), \tag{23}$$

where "sgn(·)" is the sign function. In fact, (20) and (22) hold without the null hypothesis. Lemma 2.2 of Rao and Zhao (1992) implies that

$$\sum_{i=1}^{n} w_i(|e_i - x_i'S_n^{-1/2}\gamma_1| - |e_i - x_i'S_n^{-1/2}\gamma_2|) + \sum_{i=1}^{n} w_i\mathrm{sgn}(e_i)x_i'S_n^{-1/2}(\gamma_1 - \gamma_2) - f(0)(\gamma_1'\gamma_1 - \gamma_2'\gamma_2) \to 0 \tag{24}$$

in probability, uniformly for γ_1 and γ_2 in any given bounded subset of R^p. Let $\gamma_1 = S_n^{1/2}(\hat{\beta}^* - \beta_0)$, $\gamma_2 = S_n^{1/2}(\hat{\beta} - \beta_0)$ and apply (20). We have

$$\sum_{i=1}^{n} w_i(|y_i - x_i'\hat{\beta}^*| - |y_i - x_i'\hat{\beta}|) = f(0)(\hat{\beta}^* - \hat{\beta})'S_n(\hat{\beta}^* - \hat{\beta}) + o_p(1). \tag{25}$$

It can be similarly shown that

$$\sum_{i=1}^{n} w_i(|y_i - x_i'\hat{\beta}_c^*| - |y_i - x_i'\hat{\beta}_c|) = f(0)(\hat{\eta}^* - \hat{\eta})'(K'S_nK)(\hat{\eta}^* - \hat{\eta}) + o_p(1). \tag{26}$$

Then, by the definition of M_n^* in (7),

$$\begin{aligned}
M_n^* &= -f(0)(\hat{\eta}^* - \hat{\eta})'(K'S_nK)(\hat{\eta}^* - \hat{\eta}) + f(0)(\hat{\beta}^* - \hat{\beta})'S_n(\hat{\beta}^* - \hat{\beta}) + o_p(1) \\
&= -\frac{1}{4f(0)}\|\sum_{i=1}^{n}(w_i - 1)\mathrm{sgn}(e_i)(K'S_nK)^{-1/2}K'x_i\|^2 + \frac{1}{4f(0)}\|\sum_{i=1}^{n}(w_i - 1)\mathrm{sgn}(e_i)S_n^{-1/2}x_i\|^2 \\
&\quad + o_p(1) \\
&= \frac{1}{4f(0)}(-\|\xi_n\|^2 + \|\zeta_n\|^2) + o_p(1), \qquad \text{say.}
\end{aligned}$$

Here and in the sequel, $\|\cdot\|$ is the Euclidean norm. By the checking the Lindeberg condition, we know that the conditional distributions of ξ_n and ζ_n given $y_1, ..., y_n$ converge to standard normal distribution of q and p dimensions, respectively. Moreover, ξ_n is a linear transformation of ζ_n. Therefore, the conditional distribution of M_n^* given $y_1, ..., y_n$ converges to $\chi^2_{p-q}/\{4f(0)\}$. One can show in an analogous fashion that

$$\begin{aligned}
M_n &= \left(\sum_{i=1}^{n}(|y_i - x_i'\hat{\beta}_c| - |y_i - x_i'\beta_0|)\right) - \left(\sum_{i=1}^{n}(|y_i - x_i'\hat{\beta}| - |y_i - x_i\beta_0|)\right) \\
&= -f(0)\hat{\eta}'K'S_nK\hat{\eta} + f(0)(\hat{\beta} - \beta_0)'S_n(\hat{\beta} - \beta_0) + o_p(1) \\
&= -\frac{1}{4f(0)}\|\sum_{i=1}^{n}\mathrm{sgn}(e_i)(K'S_nK)^{-1/2}K'x_i\|^2 + \frac{1}{4f(0)}\|\sum_{i=1}^{n}\mathrm{sgn}(e_i)S_n^{-1/2}x_i\|^2 + o_p(1).
\end{aligned}$$

Similarly, M_n converges to $\chi^2_{p-q}/\{4f(0)\}$ in distribution. Then (9) follows.

Now suppose S_n is degenerate for all large n, i.e., $p' < p$. Recall that, in this case, model (1) is equivalent to $y_i = v_i'\gamma + e_i, i = 1, ..., n$, where $\gamma = U_1\beta = (u_1'\beta, ..., u_{p'}'\beta)'$ and $v_i = U_1 x_i$ are p'-vectors. The hypothesis in (3) can be expressed as

$$H_0 : \gamma \in \Omega_0',$$

where $\Omega_0' = \{\gamma \in R^{p'} : (\gamma_1, ..., \gamma_{p-q})' = (c_1, ..., c_{p-q})'\}$ and γ_j and c_j are the j-th component of the p'-vector γ and the $p-q$-vector c_0 respectively. Ω_0' is a $p'-p+q$ dimension subspace in $R^{p'}$. Notice that M_n can be expressed as $\min_{\gamma\in\Omega_1}\sum_{i=1}^n |v_i'\gamma - y_i| - \min_{\gamma\in R^{p'}} |v_i'\gamma - y_i|$ and the expression of M_n^* can be similarly modified. Moreover, by using the definition of the Moore-Penrose inverse, condition (A.2) is equivalent to

$$\max_{1\le i\le n} v_i' \left(\sum_{i=1}^n v_i v_i'\right)^{-1} v_i = \max_{1\le i\le n} x_i' S_n^{-1} x_i \to 0, \quad \text{as } n \to \infty. \tag{27}$$

Since $\sum_{i=1}^n v_i v_i'$ is of full rank p' for all large n, the preceding proof with the positivity of S_n for all large n can be carried over to show that (9) still holds. Notice that the dimension of Ω_0' is $p'-p+q$ and therefore the limiting χ^2-distribution has degree of freedom $p' - \dim(\Omega_0') = p - q$. The proof is complete.

Appendix B: *Proof of Theorem 2*

Observe that (22) holds under conditions (A1)-(A2) and the positivity of S_n. Since $V_n = \sum_{i=1}^n v_i v_i'$ is of full rank p' for all large n, an analogous version of (22) under model (5) is

$$2f(0)V_n^{1/2}(\hat\gamma - \gamma_0) = \sum_{i=1}^n \operatorname{sgn}(e_i) V_n^{-1/2} v_i + o_p(1).$$

It follows that

$$2f(0)(U_1 S_n U_1')^{1/2} U_1(\hat\beta - \beta_0) = 2f(0)V_n^{1/2}(\hat\gamma - \gamma_0) \to N(0, I_{p'})$$

in distribution as $n \to \infty$, where $I_{p'}$ is the $p' \times p'$ identity matrix. For $b \in \Omega$, let $b = U_1' a$, where a is a p'-vector. Since

$$\frac{2f(0)a'U_1(\hat\beta - \beta_0)}{\{a'(U_1 S_n U_1')^{-1} a\}^{1/2}} \to N(0,1)$$

in distribution as $n \to \infty$. The Moore-Penrose inverse implies $b'S_n^{-1}b = a'(U_1 S_n U_1')^{-1}a$ and therefore (10) follows. The proof is complete.

Appendix C: *Proof of Theorem 3*

Under the assumptions in Theorem 3, we have

$$\hat{\beta}^* - \beta_0 = \frac{1}{2n} J_0^{-1} \sum_{i=1}^{n} w_i \mathrm{sgn}(e_i) x_i + o_p(1) \tag{28}$$

and

$$\hat{\beta} - \beta_0 = \frac{1}{2n} J_0^{-1} \sum_{i=1}^{n} \mathrm{sgn}(e_i) x_i + o_p(1). \tag{29}$$

Therefore,

$$\hat{\beta}^* - \hat{\beta} = \frac{1}{2n} J_0^{-1} \sum_{i=1}^{n} (w_i - 1) \mathrm{sgn}(e_i) x_i + o_p(1). \tag{30}$$

Analogously,

$$\hat{\eta}^* - \hat{\eta} = \frac{1}{2n} (K' J_0 K)^{-1} K' \sum_{i=1}^{n} (w_i - 1) \mathrm{sgn}(e_i) x_i + o_p(1). \tag{31}$$

Let

$$\psi_n(\gamma) = \frac{1}{n} \sum_{i=1}^{n} w_i \left(|x_i'\gamma + e_i| - |e_i| + \mathrm{sgn}(e_i) x_i'\gamma \right).$$

Similar to (24), as $n \to 0$,

$$\psi_n(\gamma) = \gamma' J_0 \gamma + o_p(||\gamma||^2) \tag{32}$$

uniformly for $||\gamma|| \le C n^{-1/2}$, where C is any given constant. Expressions (28) and (29) imply that $\hat{\beta}^* - \beta_0 = O_p(n^{-1/2})$ and $\hat{\beta} - \beta_0 = O_p(n^{-1/2})$. Combining with (32) yields

$$\begin{aligned}
& \sum_{i=1}^{n} w_i |y_i - x_i'\hat{\beta}^*| - \sum_{i=1}^{n} w_i |y_i - x_i'\hat{\beta}| \\
= \;& n(\psi_n(\hat{\beta}^* - \beta_0) - \psi_n(\hat{\beta} - \beta_0)) - \Big(\sum_{i=1}^{n} w_i \mathrm{sgn}(e_i) x_i'(\hat{\beta}^* - \beta_0) - \sum_{i=1}^{n} w_i \mathrm{sgn}(e_i) x_i'(\hat{\beta} - \beta_0) \Big) \\
= \;& n(\hat{\beta}^* - \beta_0)' J_0 (\hat{\beta}^* - \beta_0) - n(\hat{\beta} - \beta_0)' J_0 (\hat{\beta} - \beta_0) \\
& - \Big(2n(\hat{\beta}^* - \beta_0)' J_0 (\hat{\beta}^* - \beta_0) - 2n(\hat{\beta}^* - \beta_0)' J_0 (\hat{\beta} - \beta_0) \Big) + o_p(1) \\
= \;& -n(\hat{\beta}^* - \hat{\beta})' J_0 (\hat{\beta}^* - \hat{\beta}) + o_p(1).
\end{aligned} \tag{33}$$

Similarly, we have

$$\sum_{i=1}^{n} w_i |y_i - x_i'\hat{\beta}_c^*| - \sum_{i=1}^{n} w_i |y_i - x_i'\hat{\beta}_c| = -n(\hat{\eta}^* - \hat{\eta})' (K' J_0 K) (\hat{\eta}^* - \hat{\eta}) + o_p(1). \tag{34}$$

Then, it follows from (30), (31), (33) and (34) that

$$M_n^* = \frac{1}{4n} \{ \sum_{i=1}^{n} (w_i - 1) \mathrm{sgn}(e_i) x_i \}' (J_0^{-1} - K(K' J_0 K)^{-1} K') \{ \sum_{i=1}^{n} (w_i - 1) \mathrm{sgn}(e_i) x_i \} + o_p(1). \tag{35}$$

Then, conditioning on $(y_1, x_1, ..., y_n, x_n, ...)$, the conditional distribution of M_n^* converges to that of $\xi' V^{1/2}(J_0^{-1} - K(K'J_0K)^{-1}K')V^{1/2}\xi$, where ξ is a standard normal random vector of p dimension. This distribution is that of sum of squares of $p-q$ normal random variables since $J_0^{-1} - K(K'J_0K)^{-1}K')$ has rank $p-q$. The proof of M_n having the same limiting distribution is analogous and is omitted. Then (17) holds. (18) follows from (29). If the errors are homogeneous, $f(0|x_1) = f(0)$, $J_0 = f(0)V$, the limiting distribution is $\chi^2_{p-q}/\{4f(0)\}$ and (19) follows. The proof is complete.

References

Bassett, G. and Koenker, R. (1978). Asymptotic Theory of Least Absolute Error Regression. *Journal of American Statistical Association* **73**, 618–622.

Buchinsky, M. (1998). Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research. *Journal of Human Resources* **33**(1), 88-126.

Charnes, A., Cooper, W. W. and Ferguson, R. O. (1955). Optimal Estimation of Executive Compensation by Linear Programming. *Management Sciences* **1**, 138–151.

Chen, X. R., Bai, Z. D., Zhao L. C. and Wu, Y. H. (1990). Asymptotic Normality of Minimum L_1-norm Estimates in Linear Models. *Science in China* **33**, 1311-1328.

Jin, Z., Ying, Z. and Wei, L. J. (2001). A Simple Resampling Method by Perturbing the Minimand. *Biometrika* **88**, 381–390.

Koenker, R. and Bassett, G. S. (1978). Regression Quantiles. *Econometrica* **46**, 33–50.

Koenker, R., and D'orey, V. (1987). Computing Regression Quantiles. *Journal of the Royal Statistical Society, Applied Statistics* **36**, 383–393.

Matern, B. (1949). Independence of Non-negative Quadratic Forms in Normally Correlated Variables. *Annals of Mathematical Statistics* **20**, 119-120.

Parzen, M. I., Wei, L. J., and Ying, Z. (1994). A Resampling Method Based on Pivotal Estimating Functions. *Biometrika* **81**, 341–350.

Portnoy, S. and Koenker, R. (1997). The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-error vs. Absolute-error Estimators, with Discussion. *Statistical Science* **12**, 279-300.

Pollard, D. (1991). Asymptotics for least absolute deviations regression estimators. *Econometric Theory* **7**, 186-199.

Powell, J. L. (1984). Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics* **25**, 303-325.

Rao, C. R. and Zhao, L. C. (1992). Approximation to the Distribution of M-estimates in Linear Models by Randomly Weighted Bootstrap. *Sankhya* **54**, 323–331.

Scheffé, H. (1959). *The Analysis of Variance.* John Wiley & Sons, Inc., New York.

Ying, Z., Jung, S.H. and Wei, L.J. (1995). Survival analysis with median regression model. *Journal of American Statistical Association* **90**, 178–184.

Table 1. Type I errors of the proposed tests and the F-tests.

Covariates	Significance level	Homogeneous Errors		Heterogeneous Errors	
		a	b	c	d
A	0.1	0.103(0.100)	0.088(0.091)	0.108(0.163)	0.097(0.122)
	0.05	0.056(0.049)	0.046(0.034)	0.060(0.095)	0.051(0.051)
	0.025	0.030(0.025)	0.022(0.013)	0.033(0.058)	0.028(0.022)
B	0.1	0.101(0.104)	0.097(0.103)	0.109(0.201)	0.108(0.150)
	0.05	0.050(0.054)	0.045(0.054)	0.060(0.130)	0.055(0.085)
	0.025	0.027(0.029)	0.023(0.027)	0.032(0.081)	0.028(0.048)
C	0.1	0.098(0.103)	0.097(0.089)	0.106(0.157)	0.101(0.122)
	0.05	0.049(0.055)	0.050(0.035)	0.059(0.092)	0.055(0.054)
	0.025	0.027(0.026)	0.026(0.015)	0.032(0.055)	0.028(0.023)
D	0.1	0.104(0.101)	0.091(0.102)	0.106(0.205)	0.099(0.149)
	0.05	0.055(0.051)	0.049(0.051)	0.062(0.132)	0.046(0.084)
	0.025	0.028(0.024)	0.024(0.027)	0.033(0.085)	0.025(0.047)

Table 2. P-value for paired comparison test with difference between I and J medians/means in the parenthesis of General Social Surveys data (1993)

J\ I	EL	L	SL	M	SC	C	EC
EL	——	0.174	0.894	0.319	0.523	0.001	0.002
	——	(5.07)	(0.873)	(5.07)	(4.07)	(13.07*)	(17.80*)
L	0.529	——	0.006	0.771	0.511	0.003	0.010
	(-6.20)	——	(-4.29*)	(0.00)	(-0.00)	(8.00*)	(12.73*)
SL	0.974	0.575	——	0.051	0.155	0.000	0.001
	(-3.00)	(3.21)	——	(4.29)	(3.29)	(12.29*)	(17.02*)
M	0.404	1.000	0.194	——	0.699	0.002	0.016
	(-6.47)	(-0.26)	(3.47)	——	(-1.00)	(8.00*)	(12.73*)
SC	0.666	0.999	0.779	0.981	——	0.000	0.008
	(-5.37)	(0.84)	(-2.37)	(1.10)	——	(9.00*)	(13.73*)
C	0.007	0.026	0.000	0.002	0.001	——	0.448
	(-11.70*)	(-5.49*)	(-8.70*)	(-5.23*)	(-6.33*)	——	(4.73)
EC	0.003	0.034	0.001	0.022	0.009	0.853	——
	(-15.48*)	(-9.28*)	(-12.49*)	(-9.02*)	(-10.12*)	(-3.79)	——

NOTE: Upper triangle matrix is for our tests, lower triangle matrix is for . "EL": "Extremely Liberal", "L": "Liberal", "SL": "Slightly Liberal", "M": "Moderate", "SC": "Slightly Conservative", "C": "Conservative", 'EC": "Extremely Conservative". Random weights follow mean 1 exponential distribution.

* The median/mean difference is significant at the .05 level.

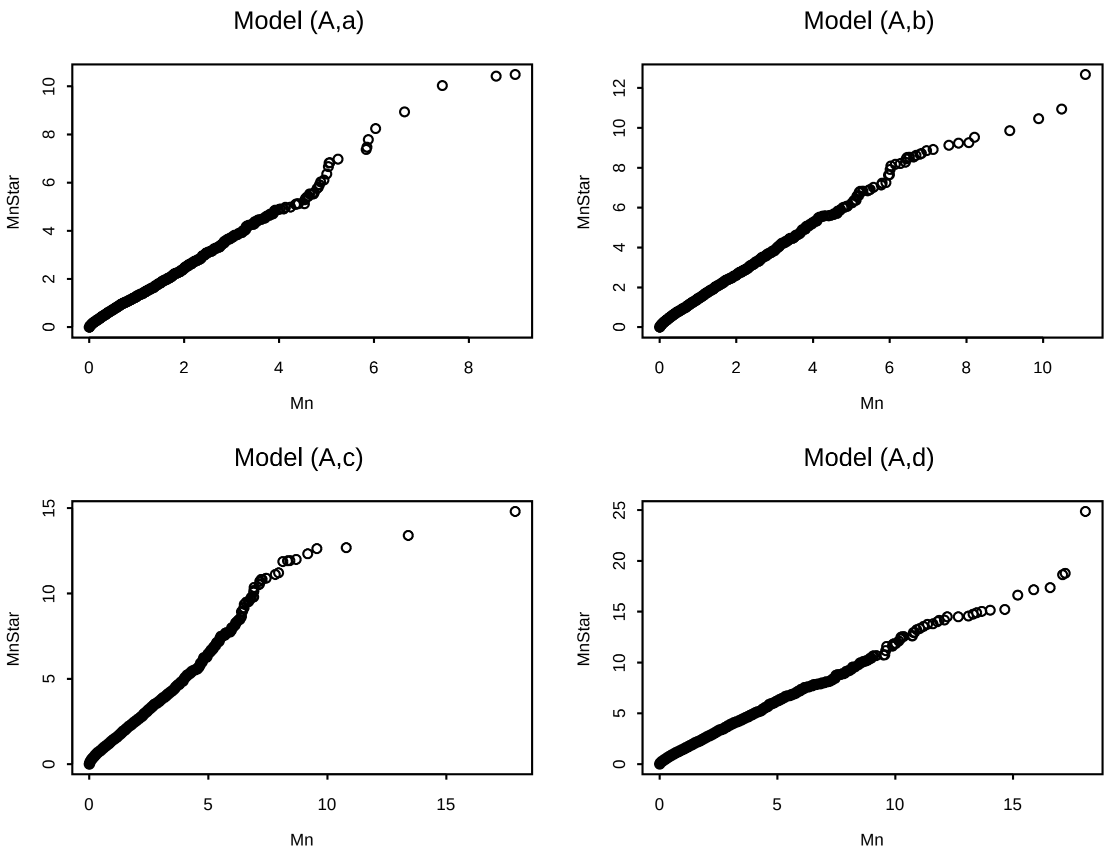

Figure 1. Q-Q plot of M_n^* v.s. M_n.

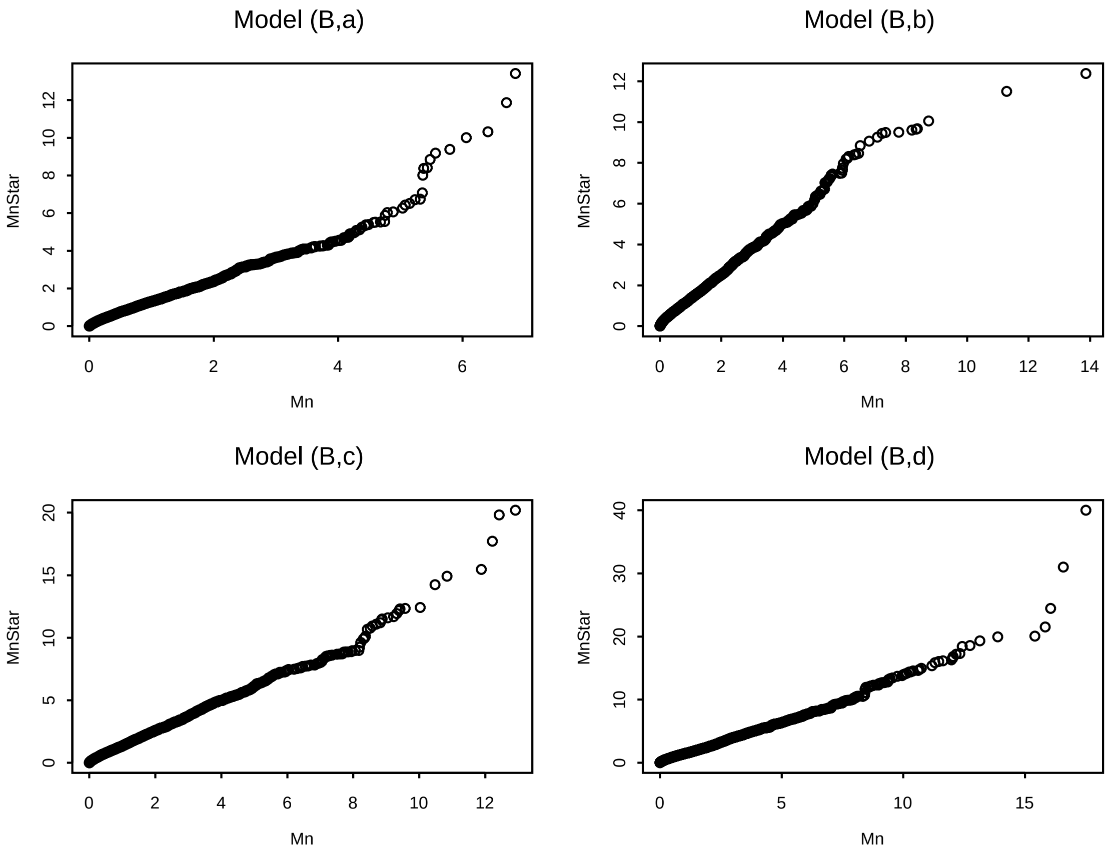

Figure 1 (continued).

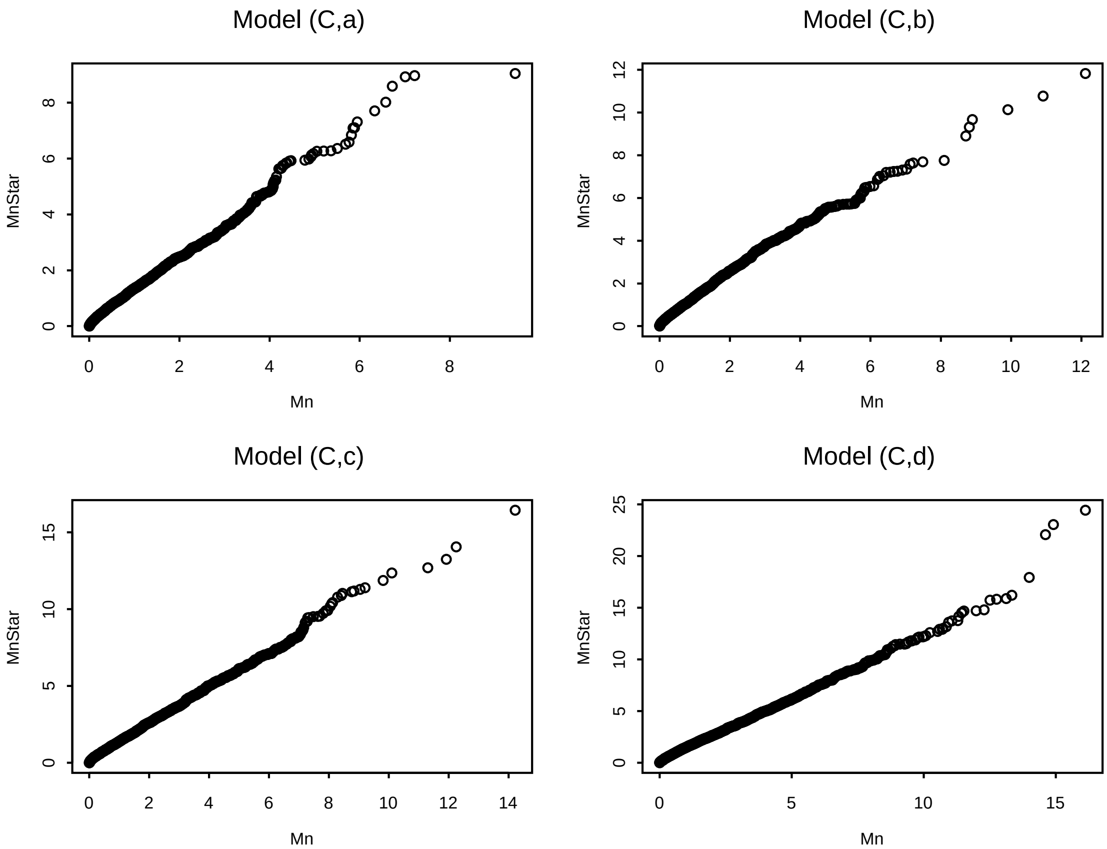

Figure 1 (continued).

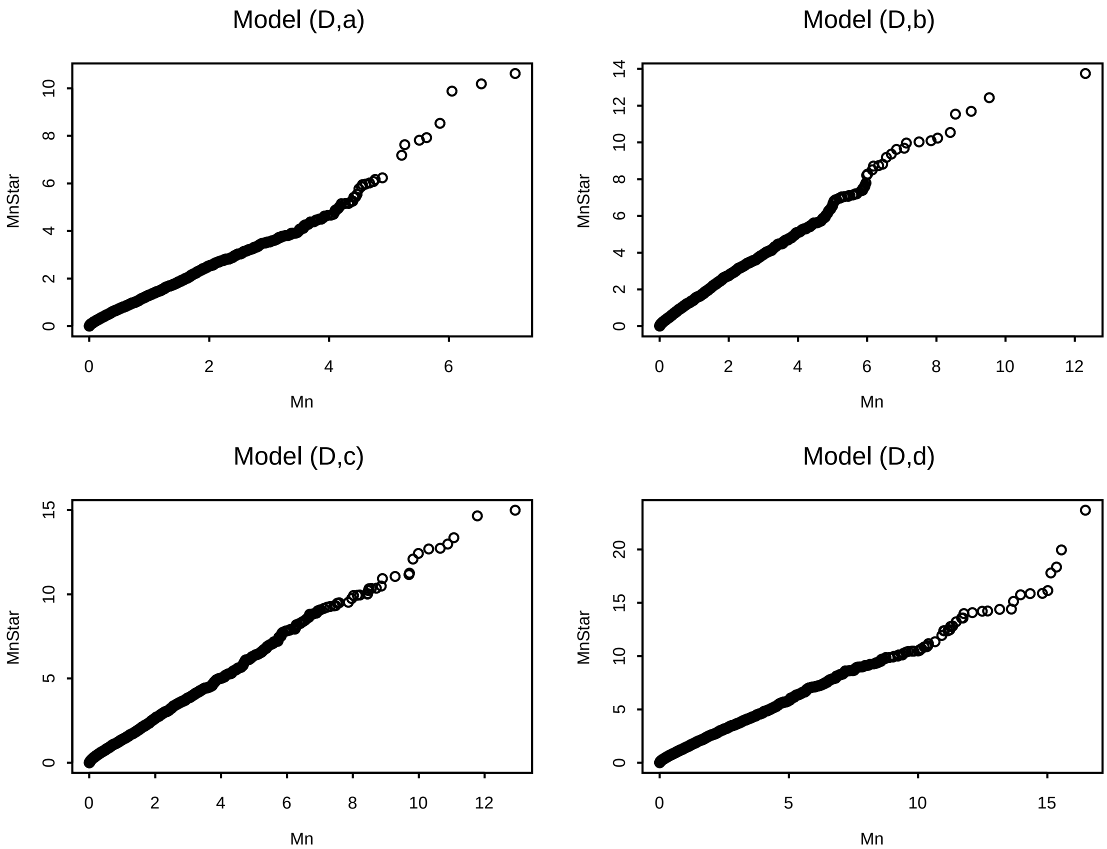

Figure 1 (continued).

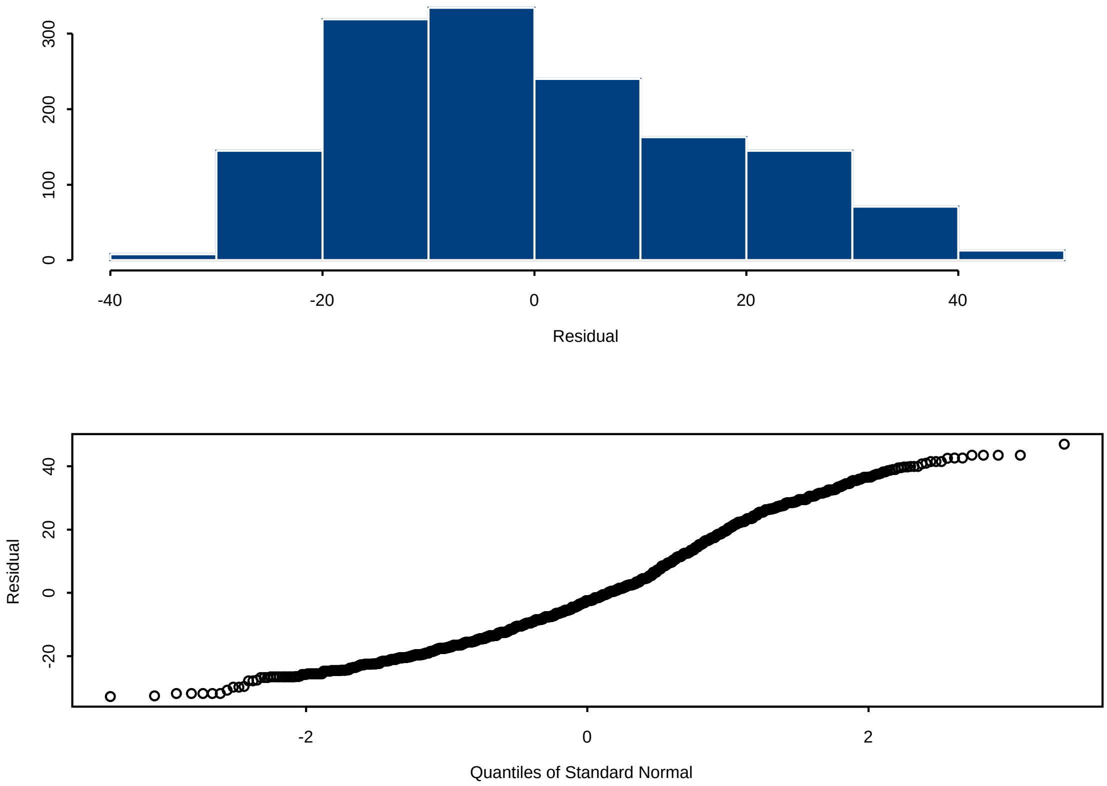

Figure 2. Histogram plot and Q-Q plot of residuals for General Social Surveys data (1993) after LS fitting.