한국통계학회논문집 제2권 2호
1995년 12월 pp. 288-295

Test of Hypotheses based on LAD Estimators in Nonlinear Regression Models

Seung Hoe Choi, Hae Kyung Kim[1)]

Abstract

In this paper a hypotheses test procedure based on the least absolute deviation estimators for the unknown parameters in nonlinear regression models is investigated. The asymptotic distribution of the proposed likelihood ratio test statistic are established both under the null hypotheses and a sequence of local alternative hypotheses. The asymptotic relative efficiency of the proposed test with classical test based on the least squares estimator is also discussed.

1. Introduction

We consider the following nonlinear regression mode

$$y_t = f(x_t, \theta_o) + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{1.1}$$

where y_t is the tth response, x_t is the tth input m-vector, θ_o is p-vector of unknown parameter, $f: R^q \times R^p \rightarrow R^1$ is a continuously differentiable up to of order 2, and ε_t are independent and identically distributed unobservable random variables with finite variance.

Let Θ be the set of possible values of unknown parameter θ_o and be compact subset of R^p. The least absolute deviation (LAD) estimator $\widehat{\theta_n}$ of θ_o based on (x_t, y_t) is a vector which minimizes

$$D_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} |y_t - f(x_t, \theta)|. \tag{1.2}$$

A statistical problem is make inference about θ by estimating and hypothesis testing. The problem of testing hypotheses about the unknown parameter θ_o in linear model has been investigated based on the LAD estimators by Mckean and Hettmansperger (1976, R-analysis), Koenker and Bassett (1982, L_1-analysis) and some others. For nonlinear model, Gallant (1987,

1) Department of Mathematics, Yonsei University, Seoul, 120-749, KOREA

L_2-analysis) and Kim (1989, R-analysis) studied testing hypotheses about unknown parameter θ_o.

Most of theoretical results for nonlinear regression models are asymptotic due to non-linearity. Complete information on the asymptotic properties of a test is provided by its asymptotic power. However, the limiting distribution of the test statistic under the alternative hypothesis is difficult or even impossible. Therefore we consider the sequence of local alternative hypothesis. Such local alternative tends to the null hypothesis as the sample sizes increaseing. For examples, see Koenker and Bassett (1982) and Kim (1989).

In this paper, we investigate the asymptotic distribution of the likelihood ratio test statistics based on the LAD estimator in section 2 and examine the asymptotic relative efficiency (ARE) of the proposed test with respect to the classical test based on the least squares (LS) estimator in section 3.

2. Asymptotic Distribution Of Test Statistics

In this section we propose the likelihood ratio test statistics and investigate limiting distribution of the test statistics not only under the
null hypotheses but under alternative hypothesis in model (1.1).

Let $h : R^p \rightarrow R^q$ be a function such that the matrix function $H(\theta)$, $H(\theta) = \left[\frac{\partial}{\partial \theta_j} h(\theta) \right]_{(q \times p)}$, is continuous in θ and $H(\theta)$ has full rank q. Our interest is to test the hypothesis

$$H_0 : h(\theta) = 0 \quad \text{against} \quad H_n : h(\theta) = \frac{\gamma}{\sqrt{n}} \tag{2.1}$$

where $\gamma \in R^q$. If h is a linear function about θ, there exists a $q \times p$ matrix R such that $h(\theta) = R\theta$. Hence, test of linear hypothesis

$$H_0 : R\theta = \gamma \quad \text{against} \quad A : R\theta \neq \gamma$$

for linear model is a special case of (2.1). To simplify the notation, we denote,

$$f_t(\theta) = f(x_t, \theta), \quad \nabla f_t(\theta) = \left[\frac{\partial}{\partial \theta_j} f_t(\theta) \right]_{(p \times 1)}, \quad J_{nj}(\theta) = \frac{1}{n} \sum_{t=1}^{n} \psi(r_t(\theta)) \frac{\partial}{\partial \theta_j} f_t(\theta),$$

where $\psi(x) = 1, 0$, or -1 according as $x > 0$, $x = 0$, or, $x < 0$.

Let (R^m, A, P_X) denotes the probability space. Throughout this paper we make the following assumptions on model (1.1).

A_1 : $V_n(\theta_o) = \frac{1}{n} \sum_{t=1}^{n} \nabla f_t(\theta_o) \nabla^T f_t(\theta)$ converges to a positive definite matrix $V(\theta)$ $n \rightarrow \infty$.

A_2 : ε has continuous density function $g(x)$ such that $G'(x)=g(x)$ and $G(0)=\frac{1}{2}$ uniquely.

A_3 : $P_X\{x\in R^m \mid f(x, \theta_o)\neq f(x, \theta)\}>0$ for each fixed $\theta_o\neq\theta$.

A_4 : There exists a fixed $\gamma\in R^p$ such that $h(\theta)=\frac{\gamma}{\sqrt{n}}$ for any sample of size n.

The likelihood ratio test statistics which is based on the difference between sum of absolute residuals, $r_t=y_t-f_t(\theta)$, in the restricted and full model is defined by

$$T=2n\tau\{D_n(\widehat{\theta_n^h})-D_n(\widehat{\theta_n})\},$$

where θ_n^h minimizes $D_n(\theta)$ subject to $h(\theta)=0$, and $\tau=2g(0)$. To derive the asymptotic distribution of the test statistics T, we need the following quadratic function

$$Q_n(\theta)=D_n(\theta_o)+(\theta-\theta_o)^TJ_n(\theta_o)+g(0)(\theta-\theta_o)^TV_n(\theta_o)(\theta-\theta_o), \tag{2.2}$$

where $J_n(\theta_o)=[J_{nj}(\theta_o)]_{(p\times 1)}$. The following lemma shows that $Q_n(\theta)$ provides a useful approximation to $D_n(\theta)$ and explains the relationship between $D_n(\theta)$ and $Q_n(\theta)$. The proofs of the following lemmas are given in Choi and Kim (1994).

Lemma 2.1 Let $S=\{\theta_n\in\Theta : \sqrt{n}|\theta_n-\theta_o|\leq M\}$ for any $M>0$. Suppose that the assumptions (A_1)-(A_3) are satisfied for model (1.1). Then we have

$\{i\}$ $\text{Sup}_S\,|Q_n(\theta_n)-D_n(\theta_n)|=o_p(n^{-1})$,

$\{ii\}$ $Q_n(\theta_n)-D_n(\theta_n)=o(1)$

for $\theta_n\in S$.

Let $\widetilde{\theta_n}$ be any vector value which minimizes $Q_n(\theta)$. The existence of $\widetilde{\theta_n}$ results from continuity of $Q_n(\theta)$ and compactness of parameter space Θ. The next result concerns with the asymptotic equivalent of $\widehat{\theta_n}$ and $\widetilde{\theta_n}$.

Lemma 2.2 Under the assumptions (A_1)-(A_3), $\sqrt{n}(\widehat{\theta_n}-\widetilde{\theta_n})$ converges in probability to zero.

In following lemmas we consider the strong consistency and the asymptotic normality of LAD estimator $\widehat{\theta_n}$. The proofs of the lemmas are given in Kim and Choi (1993).

Lemma 2.3 Suppose that the assumptions (A_1)-(A_3) are satisfied in model (1.1). Then the LAD estimator $\widehat{\theta_n}$ converges almost surely to θ_o.

Lemma 2.4 Under the same conditions of lemma 2.3, $\sqrt{n}(\hat{\theta}_n-\theta_o)$ converges in distribution to a p-variate normal random vector with mean zero and variance-covariance $V^{-1}(\theta_o)/\tau^2$, denoted by $N(0,\ V^{-1}(\theta_o)/\tau^2)$.

Let $\hat{\theta}_n^h$ and $\tilde{\theta}_n^h$ be the vector value which minimizes of $D_n(\theta)$ and $Q_n(\theta)$, respectively, subject to $h(\theta)=0$. In the next lemma we consider the asymptotic properties of $\hat{\theta}_n^h$ and $\tilde{\theta}_n^h$.

Lemma 2.5 Under the assumptions $(A_1)-(A_4)$, we have that $\tilde{\theta}_n^h$ converges almost surely to θ_o.

proof From Bolzano-Weierstrass Theorem, the sequence $\{\tilde{\theta}_n^h\}$ has at least one limit point θ^*. Let $\{\tilde{\theta}_{n_k}^h\}$ be a subsequence of $\{\tilde{\theta}_n^h\}$ which converges to θ^*. It is sufficient to show that $\theta^*=\theta_o$. Kolmogolove's SLLN implies that $Q_n(\theta)$ converges almost surely to $Q(\theta)=g(0)(\theta-\theta_0)^T V(\theta_o)(\theta-\theta_0)+Y$, where $Y=\frac{1}{n}\sum_{t=1}^{n}E(|\varepsilon_t|)+o(1)$. Since

$$|Q_{n_k}(\tilde{\theta}_{n_k}^h)-Q(\theta^*)|\le|Q_{n_k}(\tilde{\theta}_{n_k}^h)-Q(\tilde{\theta}_{n_k}^h)|+|Q(\tilde{\theta}_{n_k}^h)-Q(\theta^*)|,$$

we have

$$Q(\theta^*)=\lim_{k\to\infty}Q_{n_k}(\tilde{\theta}_{n_k}^h)\le\lim_{k\to\infty}Q_{n_k}(\theta_0)=Q(\theta_0).$$

On the other hand, $Q(\theta)$ is a strictly convex due to

$$(\theta_1-\theta_2)^T\{\nabla Q_n(\theta_1)-\nabla Q_n(\theta_2)\}=\tau(\theta_1-\theta_2)^T V(\theta_o)(\theta_1-\theta_2)>0.$$

Hence, the lemma follows from the assumption A_1.

Lemma 2.6 Under the same conditions of lemma 2.5, we obtain

$$\{i\}\quad Q_n(\tilde{\theta}_n)-Q_n(\hat{\theta}_n)=o_p(n^{-1}),$$

$$\{ii\}\quad Q_n(\tilde{\theta}_n^h)-Q_n(\hat{\theta}_n^h)=o_p(n^{-1}).$$

Proof In virtue of Linderberg Central Limit theorem, we have

$$\sqrt{n}J(\theta_o)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi(\varepsilon_t)\nabla f_t(\theta_o)\xrightarrow{d}N(0,\ V(\theta_o)).$$

From (2.2), we have

$$n\{Q_n(\tilde{\theta}_n)-Q_n(\hat{\theta}_n)\}=\sqrt{n}(\tilde{\theta}_n-\hat{\theta}_n)^T\sqrt{n}J_n(\theta_o)$$
$$+\sqrt{n}(\tilde{\theta}_n-\hat{\theta}_n)^T g(0)V_n(\theta_o)\{\sqrt{n}(\tilde{\theta}_n-\theta_o)-\sqrt{n}(\hat{\theta}_n-\theta_o)\}.$$

Hence, the proof follows from lemma 2.2, lemma 2.3 and Slutsky theorem. An analogous argument implies that $Q_n(\hat{\tilde{\theta}}_n)-Q_n(\tilde{\hat{\theta}}_n)=o_p(n^{-1})$. This part is omitted here.

The main result of this section is the asymptotic behavior of T. In the following theorem we derive the limiting distribution of T.

Theorem 2.7 Suppose that the assumptions $(A_1)-(A_4)$ are satisfied in model (1.1). Then the test statistics T has asymptotically a noncentral chi-square $\chi^2(q,\lambda)$ distribution with q degree of freedom and noncentrality λ, where $\lambda=\frac{1}{2}\tau^2\gamma^T[HV^{-1}H^T]^{-1}\gamma$. Under the null hypotheses, we have $\lambda=0$.

Proof Let $B_n=\{D_n(\hat{\tilde{\theta}}_n)-D_n(\hat{\theta}_n)\}$. Then we can rewrite that

$$B_n=\{D_n(\hat{\tilde{\theta}}_n)-Q_n(\hat{\tilde{\theta}}_n)\}+\{Q_n(\hat{\tilde{\theta}}_n)-Q_n(\tilde{\hat{\theta}}_n)\}+\{Q_n(\tilde{\hat{\theta}}_n)-Q_n(\tilde{\theta}_n)\}$$
$$+\{Q_n(\tilde{\theta}_n)-Q_n(\hat{\theta}_n)\}+\{Q_n(\hat{\theta}_n)-D_n(\hat{\theta}_n)\}.$$

By lemma 2.1 and lemma 2.6, we have

$$nB_n=n\{Q_n(\tilde{\hat{\theta}}_n)-Q_n(\hat{\theta}_n)\}+o_p(1).$$

From (2.2) and first order Taylor's theorem, we get

$$(\tilde{\hat{\theta}}_n-\tilde{\theta}_n)=\tau^{-1}V_n^{-1}(\theta_o)\nabla Q_n(\tilde{\hat{\theta}}_n). \qquad (2.3)$$

Due to second order Taylor's theorem and (2.3), we have

$$Q_n(\tilde{\hat{\theta}}_n)-Q_n(\tilde{\theta}_n)=\frac{1}{2}(\tilde{\hat{\theta}}_n-\tilde{\theta}_n)^T[\tau V_n(\theta_o)](\tilde{\hat{\theta}}_n-\tilde{\theta}_n)$$

$$=\frac{1}{2\tau}\nabla^T Q_n(\tilde{\hat{\theta}}_n)V_n^{-1}(\theta_o)\nabla Q_n(\tilde{\hat{\theta}}_n).$$

By similar method, we obtain

$$\nabla Q_n(\tilde{\hat{\theta}}_n)=\nabla Q_n(\theta_0)+\tau V_n(\theta_o)(\tilde{\hat{\theta}}_n-\theta_0).$$

Thus,

$$\tilde{H}_n V_n^{-1}(\theta_o)\nabla Q_n(\tilde{\hat{\theta}}_n)=\tilde{H}_n V_n^{-1}(\theta_o)J_n(\theta_o)+\tau\tilde{H}_n(\tilde{\hat{\theta}}_n-\theta_o) \qquad (2.4)$$
$$=\tilde{H}_n V_n^{-1}(\theta_o)J_n(\theta_o)+\tau[h(\tilde{\hat{\theta}}_n)-h(\theta_o)],$$

where $\widetilde{H_n} = \nabla^T h(\widetilde{\theta_n})_{(q \times p)}$. From (2.4), we have

$$\sqrt{n}\,\nabla Q_n(\theta_n^{\widetilde{h}}) = \widetilde{H_n^T}\,[\widetilde{H_n} V_n^{-1}(\theta_0)\widetilde{H_n^T}\,]^{-1}\widetilde{H_n} V_n^{-1}(\theta_0)\sqrt{n}\,\nabla Q_n(\theta_n^{\widetilde{h}})$$

$$= \widetilde{H_n^T}\,[\widetilde{H_n} V_n^{-1}(\theta_0)\widetilde{H_n^T}\,]^{-1}\widetilde{H_n} V_n^{-1}(\theta_0)\sqrt{n} J_n(\theta_0)$$

$$-\tau \widetilde{H_n^T}\,[\widetilde{H_n} V_n^{-1}(\theta_0)\widetilde{H_n^T}\,]^{-1}\sqrt{n} h(\theta_0).$$

Moreover, by Linderverg Central Limit Theorem we have

$$\sqrt{n} J_n(\theta_0) \xrightarrow{d} N(0,\ V(\theta_o)).$$

Hence, from (2.5) we get

$$\sqrt{n}\,\nabla Q_n(\theta_n^{\widetilde{h}}) \xrightarrow{d} N\,[-\tau H^T(HV^{-1}H^T)^{-1}\gamma,\ H^T(HV^{-1}H^T)^{-1}H\,].$$

The proof follows the fact that $H^T(HV^{-1}H^T)^{-1}HV^{-1}$ is idempotent matrix.

If the density function $g(x)$ of ε_t is known, we reject or accept

$$H_0 : h(\theta)=0 \quad \text{according as} \quad T \geq \ \text{or} \ \leq \chi^2_{1-\alpha}(q)$$

with the α level of significance.

In the case where there is a nuisance parameter in the test statistics T, we need to have a consistent estimate of τ, i.e. $\tau = 2g(0)$. Now we introduce consistent estimator of τ. Suppose we have a sample of observation $X_1,\ldots,X_n$ from a population with density $g(x)$. The Kernel estimator is defined by

$$g_n(x) = \frac{1}{n\eta_n}\sum_{t=1}^{n} K\left\{\frac{X_t - x}{\eta_n}\right\},$$

where η_n is a sequence of positive numbers, and K is a Borel measurable function satisfying

$$K \geq 0,\ \int K(x)dx = 1 \ \text{(See [2])}.$$

Specially, we can construct the following estimators of $g(0) = \frac{\tau}{2}$;

$$\widehat{\tau_n} = \widehat{g_n}(0) = \frac{1}{n}\sum_{t=1}^{n}\frac{1}{\eta_n} K\left\{\frac{r_t(\widehat{\theta_n})}{\eta_n}\right\}.$$

The following theorem shows that $\widehat{\tau_n}$ is strong consistent estimators of $g(0)$.

Theorem 2.8 Suppose that the assumptions $(A_1)-(A_3)$ hold for model (1.1). Let the Kernel zK(x) satisfy the α-Lipschitz condition with some $\alpha > 0$ ($i.e., |K(x)-K(y)| \leq C|x-y|^{\alpha}$) and let η_n satisfy $\lim_{n\to\infty}\eta_n = 0$ and $\lim_{n\to\infty} n\eta_n = \infty$. Then $\widehat{\tau_n}$ converges almost surely to $g(0)$.

Proof Let $c>0$ and $M_n = \max_{1\le k\le p}\{\frac{\partial}{\partial\theta_k} f_t(\overline{\theta_n})\}$, where

$\overline{\theta_n} = (1-\lambda)\theta_0 + \lambda\widehat{\theta_n}$, $0<\lambda<1$. Using h $\ddot{o}$der inequality, we have

$$|\widehat{g_n}(0) - g_n(0)| \le \frac{c}{n\eta_n^{1+\alpha}} \sum_{t=1}^{n} |f_t(\widehat{\theta_n}) - f_t(\theta_o)|^{\alpha}$$

$$\le \frac{c}{n\eta_n^{1+\alpha}} |\widehat{\theta_n} - \theta_0|_l^{\alpha} \sum_{t=1}^{n} \{\sum_{k=1}^{p} |\frac{\partial}{\partial\theta_k} f_t(\overline{\theta_n})|^m\}^{\frac{\alpha}{m}}$$

$$\le \frac{c}{n\eta_n^{1+\alpha}} |\widehat{\theta_n} - \theta_0|_l^{\alpha} p^{\frac{\alpha}{m}} M_n^{\alpha},$$

for $l>0$, and $m>0$ such that $\frac{1}{l}+\frac{1}{m}=1$. Let $U_n = (c^{\alpha} p^{\frac{1}{m}} M_n)^{-1}$. For sufficiently large N, we have

$$P\{\mathrm{Sup}_{n\ge N} \mid \widehat{g_n}(0) - g_n(0)| > \varepsilon\} \le P\{\mathrm{Sup}_{n\ge N} \mid \widehat{\theta_n} - \theta_0| > \varepsilon^{\frac{1}{\alpha}} h_n^{1+\frac{1}{\alpha}} U_n\}.$$

The proof follows lemma 2.4 and the condition on η_n.

3. Asymptotic Relative Efficiency

In this section we compare the efficiciency of the proposed tests with classical tests based on least squares estimator and show that the efficiency is the ratio of the asymptotic variance of the LS estimator and the LAD estimator.

It is well known that $\breve{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n}(y_t - f_t(\breve{\theta_n}))^2$ is strongly consistent estimator and

$$\sqrt{n}(\breve{\theta_n} - \theta_o) \xrightarrow{d} N(0, \sigma^2 V^{-1}(\theta_o)),$$

where $\breve{\theta_n}$ is least square estimator in model (1.1). For the ARE of proposed tests, let us denote

$$S_n(\breve{\theta_n}) = \min_{\theta} \frac{1}{n}\sum_{t=1}^{n}(y_t - f_t(\theta))^2$$

and

$$S_n(\breve{\theta}_n^{\lambda}) = \min_{\theta_0} \frac{1}{n}\sum_{t=1}^{n}(y_t - f_t(\theta))^2.$$

We can derive that $S = n\sigma^{-2}[S_n(\theta_n) - S_n(\breve{\theta}_n^{\lambda})]$ converges to $\chi^2(q, \lambda)$ where

$\lambda = \frac{1}{2}\sigma^{-2}\gamma^{T}[HV^{-1}H^{T}]^{-1}\gamma$, where σ is the variance of the error ε_t (See [3]). The ARE of the proposed test statistics T to the classical test statistics S is ratio of the noncentrality parameter of the limiting χ^2 distribution. This is the ratio of the asymptotic variances of the LAD and LS estimator (See [4]). So we have the following theorem.

Theorem 3.1 Under the same conditions of theorem 2.7, the asymptotic relative efficiency of the T with respect to the S is τ^{-2}/σ^2 which coincides with the ratio of the varience of sample median and mean from the error distribution $G(x)$.

Theorem 3.1 implies that the test statistics T based on LAD estimator is more efficient than the test statistics S based on LS estimator whenever the error distribution is heavy-tailed distributions and have peaked density at the median, such as Cauchy, double-exponential, logistic distribution etc.

Referrences

[1] Choi, S.H. and Kim, H.K. (1994). Asymptotic Normality of LAD estimator in Nonlinear Regression Model, *Bulletin of the Honam Mathematical Society*, Vol. 11, 177-185.

[2] Devroy, L. and Gr örf, L. (1985). *Nonparametric Density Estimation*, John Wiley & Sons, New York.

[3] Gallant, A.R. (1987). *Nonlinear Regression Models*, John Wiley & Sons, New York.

[4] Kim, H.K. (1989). A Nonparametric Method for Nonlinear Regression Parameters, *Journal of the Korean Statistical Society*, Vol. 18, 46-61.

[5] Kim. H.K. and Choi. S.H. (1993). Asymptotic Properties of Nonlinear Least Absolute Deviation, *Journal of Korean Statistical Society*, Vol. 24, 127-140.

[6] Koenker. R. and Bassett. G. (1982). Tests of Linear Hypotheses and L_1 *Estimation, Econometrica*, Vol. 50, 1577-1583.

[7] Mckean and Hettmansperger. T.P. (1976). Tests of the Hypotheses Based on the General Linear Model, *Communication in Statistics, Theory and Method*, Vol. 5, 693-708.