Aligarh Journal of Statistics
Vol. 32 (2012), 1-9

Optimum Regression Quantiles for the Estimation and Tests of Hypothesis of the Parameters of a Simple Regression Model with Logistic Errors

A. Adatia and A. K. MD. Ehsanes Saleh

ABSTRACT

Consider the simple linear regression model errors are i.i.d. with logistic distribution. This paper deals with the estimation and tests of hypothesis regarding the parameters, $\theta = (\beta_0, \beta_1, \sigma)'$ based on a few "regression quantiles" introduced by Koenker and Bassett (1978). The question of optimum regression quantiles is addressed for the problems. Further, estimation of the conditional regression function is also considered along with the related optimum regression quantiles. In every case the optimum spacings are independent of the design matrix.

1. INTRODUCTION

Consider the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 x_i + \sigma z_i, \qquad (i = 1, 2, ..., n)$$

where $z_1, z_2, ...z_n$ are $i.i.d.$ errors with logistic distribution,

$$e^{-\pi z/\sqrt{3}} / (\sqrt{3}(1 + e^{-\pi z/\sqrt{3}})^2), -\infty < z < \infty.$$

It is well known that least squares estimators $(L.S.E.)$ of regression parameters are unbiased with minimum variance and the quadratic estimator of σ^2 is optimal in general. For the maximum likelihood estimators $(M.L.E.)$ the Fisher information matrix for the parameter $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma)'$ is given by the 3×3 matrix:

$$\frac{n}{\sigma^2}\begin{pmatrix} \frac{\pi^2}{9} & \frac{\pi^2\bar{x}}{9} & 0 \\ \frac{\pi^2\bar{x}}{9} & \frac{\pi^2(s^2+\bar{x}^2)}{9} & 0 \\ 0 & 0 & \frac{(3+\pi^2)}{9} \end{pmatrix}$$

In this paper we consider the estimation of, $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma)'$ based on a few selected regression quantiles (RQ) which is an extension of the sample quantiles in the location-scale model (See, Balakrishnan (1992), David and Nagaraja (2003), Hassanein (1969), Saleh and Adatia (2009) and Saleh, Hassanein and Brown, (1994)).

The objective of this paper is to basically, obtain (i) asymptotically best linear unbiased estimator $(ABLUE)$ of $\boldsymbol{\theta} = (\beta_0, \beta_1, \sigma)'$ based on k $(3 \le k \le n)$ optimum regression quantiles, (ii) propose test-statistics for jointly testing $(\beta_0, \beta_1, \sigma)'$ under local alternatives and discuss the related optimum spacings and finally, (iii) propose $(ABLUE)$ of a conditional quantile-function, $y(\xi) = \beta_0 + \beta_1 x_0 + \sigma\frac{\sqrt{3}}{\pi}\ln(\xi(1-\xi)^{-1}), 0 < \xi < 1$ and related optimum spacings. Thus, as a first step, we assume that n is large and

$$\lim_{n\to\infty} \bar{x}_n = \bar{x} \text{ and } \lim_{n\to\infty} n^{-1}\begin{pmatrix} n & n\bar{x}_n \\ \bar{x}_n & \sum_{i=1}^{n} x_i^2 \end{pmatrix} = \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & s^2+\bar{x}^2 \end{pmatrix}$$

Let $u = \frac{\sqrt{3}}{\pi}\ln\frac{\lambda}{(1-\lambda)}$ be the quantile-function of the logistic distribution corresponding to the spacing $\lambda (0 < \lambda < 1)$ and let $q_0(\lambda) = -\pi\lambda(1-\lambda)/\sqrt{3}$ be the corresponding density quantile function. Further, let on k $(3 \le k \le n)$ be a fixed integer and consider the spacing vector $\lambda = (\lambda_1, ..., \lambda_k)'$ satisfying the $0 < \lambda_1 <, ..., \lambda_k < 1$.

Now, following Koenker and Bassett (1978) we obtain the k regression quantiles $\hat{\boldsymbol{\beta}}_{jn} = (\hat{\boldsymbol{\beta}}_{jn}(\lambda_1), ..., \hat{\boldsymbol{\beta}}_{jn}(\lambda_k))'$, $j = 0,1$ by minimizing

$$\sum_{j=1}^{n} \xi_{\lambda_j}(y_j - \beta_0 - \beta_1 x_j)$$

where

$$\xi(z) = |z|\{\lambda I(Z > 0) + (1-\lambda) I(Z < 0)\}$$

with $I(A)$ as the indicator function of set A. Thus, using Theorem 4.2 of Koenker and Bassett (1978) we see that the $2k$-dimensional random variable

$$(\sqrt{n}(\beta_{on}(\lambda) - \beta_0 \mathbf{1}_k - \sigma \boldsymbol{u})', \sqrt{n}(\hat{\beta}_n(\lambda) - \beta_1 \mathbf{1}_k)')'$$

converges in law $(n \to \infty)$ to the $2k-$dimensional normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\sigma^2 \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & s^2 + \bar{x}^2 \end{pmatrix}^{-1} \otimes \mathbf{\Omega}$$

where

$$\mathbf{\Omega} = \left(\left(\frac{\min(\lambda_i, \lambda_j) - \lambda_i \lambda_j}{(1-\lambda_i)(1-\lambda_j)} \right) \right)$$ and $\mathbf{1}_k = (1,1,...,1)'$ a $k-$tuple of ones and

$\mathbf{u} = (u_1, u_2, \ldots, u_k)'$, and $u_j = \ln(1-\lambda_j)^{-1}$, $j = 1, \ldots, k$.

These results will be used in the subsequent sections.

2. JOINT ESTIMATION OF $(\beta_0, \beta_1, \sigma)'$

We obtain the $(ABLUE)$ of $(\beta_0, \beta_1, \sigma)'$ minimizing the quadratic form

$$\begin{aligned}
&[\hat{\boldsymbol{\beta}}_{0n}(\boldsymbol{\lambda}) - \beta_0 \mathbf{1}_k - \sigma \mathbf{u}]' \mathbf{\Omega}^{-1} [\hat{\boldsymbol{\beta}}_{0n}(\boldsymbol{\lambda}) - \beta_0 \mathbf{1}_k - \sigma \mathbf{u}] \\
&+ 2\bar{x} [\hat{\boldsymbol{\beta}}_{0n}(\boldsymbol{\lambda}) - \beta_0 \mathbf{1}_k - \sigma \mathbf{u}]' \mathbf{\Omega}^{-1} [\hat{\boldsymbol{\beta}}_{1n}(\boldsymbol{\lambda}) - \beta_1 \mathbf{1}_k] \\
&+ (s^2 + \bar{x}^2) [\hat{\boldsymbol{\beta}}_{1n}(\boldsymbol{\lambda}) - \beta_1 \mathbf{1}_k]' \mathbf{\Omega}^{-1} [\hat{\boldsymbol{\beta}}_{1n}(\boldsymbol{\lambda}) - \beta_1 \mathbf{1}_k]
\end{aligned}$$

with respect to β_0, β_1 and σ to obtain the normal equation

$$\mathbf{K}\boldsymbol{\theta}_n^* = \mathbf{V}$$

Where

$$\mathbf{K} = \begin{pmatrix} K_1 & \bar{x}K_1 & K_3 \\ \bar{x}K_1 & (s^2 + \bar{x}^2)K_1 & \bar{x}K_3 \\ K_3 & \bar{x}K_3 & K_2 \end{pmatrix}$$

$\boldsymbol{\theta}_{\mathrm{n}}^* = (\beta_{0n}^*, \beta_{1n}^*, \sigma_n^*)$ and $\mathbf{V} = (V_0, V_1, V_2)'$,

with

$$V_0 = Z_0 + \bar{x}Z_1,\ V_1 = \bar{x}V_0 + s^2 Z_1, \quad V_2 = Z_0^* + \bar{x}Z_1^*$$

$$Z_j = \mathbf{1}_k^{'}\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\beta}}_{jn},\ Z_j^* = \mathbf{u}_k^{'}\boldsymbol{\Omega}^{-1}\hat{\boldsymbol{\beta}}_{jn},\ j = 0,1$$

and $\Delta = K_1K_2 - K_3^2$. The explicit form of K_1, K_2 and K_3 are given by

$$K_1 = \sum_{i=1}^{k+1} -\pi^2(\lambda_i - \lambda_{i-1})(\lambda_i - \lambda_{i-1} - 1)^2$$

$$K_2 = \sum_{i=1}^{k+1} \frac{(\lambda_i(\lambda_i - 1)\ln(\lambda_i/(1-\lambda_i)) - \lambda_{i-1}(\lambda_{i-1} - 1)\ln(\lambda_{i-1}/(1-\lambda_{i-1})))}{(\lambda_i - \lambda_{i-1})}$$

$$K_3 = \sum_{i=1}^{k+1} \frac{\pi}{\sqrt{3}}(\lambda_i + \lambda_{i-1} - 1)(\lambda_i(\lambda_i - 1)\ln(\lambda_i/(1-\lambda_i)) - \lambda_{i-1}(\lambda_{i-1} - 1)\ln(\lambda_{i-1}/(1-\lambda_i)))$$

with $\lambda_0 = 0$ and $\lambda_{k+1} = 1$

Now, as $n \to \infty$ the asymptotic distribution of

$$[\sqrt{n}(\beta_{0n}^* - \beta_0), \sqrt{n}(\beta_{1n}^* - \beta_1), \sqrt{n}(\sigma_n^* - \sigma_n)]$$

follows the $3-$dimensional normal distribution with mean $\mathbf{0}$ and dispersion matrix $\sigma^2\mathbf{K}^{-1}$ where $|\mathbf{K}| = s^2 K_1\Delta$. Hence, the joint asymptotic relative efficiency ($JARE$) of θ_n^* relative to the MLE, say $\bar{\theta}_n$ is given by

$$JARE(\theta_n^* : \bar{\theta}_n) = \frac{729K_1\Delta}{\pi^4(3+\pi^2)},$$

Thus, in order to determine the optimum spacings for $ABLUE$ of $(\beta_0, \beta_1, \sigma)'$, we maximize $K_1\Delta$ with respect to $\lambda_1, \lambda_2, ..., \lambda_k$ The optimum spacings and $JARE(\theta_n^* : \bar{\theta}_n)$ are given in Table 1 for $k = 3(1)10$.

Table 1: Values of $JARE(\theta_n^* : \bar{\theta}_n)$ and Optimum Spacings for Selected k = 3(1)10

$JARE$	λ_1	λ_2	λ_3	λ_4	λ_5
0.5384	0.1381	0.5000	0.8619		
0.6649	0.0895	0.3331	0.6669	0.9105	
0.7471	0.0611	0.2348	0.5000	0.7652	0.9389
0.8030	0.0433	0.1718	0.3787	0.6213	0.8282
0.8425	0.0317	0.1294	0.2925	0.5000	0.7075
0.8713	0.0239	0.0998	0.2302	0.4041	0.5959
0.8929	0.0184	0.0784	0.1841	0.3297	0.5000
0.9095	0.0144	0.0626	0.1495	0.2719	0.4206

JARE	λ_6	λ_7	λ_8	λ_9	λ_{10}
0.5384					
0.6649					
0.7471					
0.8030	0.9567				
0.8425	0.8706	0.9683			
0.8713	0.7698	0.9002	0.9761		
0.8929	0.6703	0.8159	0.9216	0.9816	
0.9095	0.5794	0.7281	0.8505	0.9374	0.9856

3. TEST OF HYPOTHESIS ON $(\beta_0, \beta_1, \sigma)'$

In this section, we consider the joint test of hypothesis:

$$H_0 : (\beta_0, \beta_1, \sigma)' = (\beta_0^0, \beta_1^0, \sigma^0)'$$

against

$$H_A : (\beta_0, \beta_1, \sigma)' \neq (\beta_0^0, \beta_1^0, \sigma^0)'$$

based on $\hat{\boldsymbol{\beta}}_{jn} = (\hat{\boldsymbol{\beta}}_{jn}(\lambda_1), ..., \hat{\boldsymbol{\beta}}_{jn}(\lambda_1))', (j = 0,1)$ where $(\beta_0^0, \beta_1^0, \sigma^0)'$ is a specified vector. In this context, our objective is to assess the asymptotoic relative efficiency (ARE) of a test based on $(\beta_{0n}^*, \beta_{1n}^*, \sigma_n^{**})'$ relative to a test based on $(\overline{\beta}_{0n}, \overline{\beta}_{1n}, \overline{\sigma}_n)'$. It is shown that the optimum spacings for this problem remains the same as in the estimation problem.

We now define the test statistics Q_n^* for testing H_0 against H_A as follows:

$$Q_n^* = n(\sigma^0)^{-2}(K_1(\beta_{0n}^* - \beta_0^0)^2 + K_2(\sigma_n^* - \sigma^0)^2 + (s^2 + \bar{x}^2)K_1(\beta_{1n}^* - \beta_1^0)^2$$
$$+ 2\bar{x}K_1(\beta_{0n}^* - \beta_0^0)^2(\beta_{1n}^* - \beta_1^0) + 2\bar{x}K_3(\beta_{1n}^* - \beta_1^0)^2(\sigma_n^* - \sigma^0)$$
$$+ 2K_3(\beta_{0n}^* - \beta_0^0)(\sigma_n^* - \sigma^0)$$

Then, the test function is defined by

$$\phi(Q_n^*) = \begin{cases} 1 & \text{if } Q_n^* \geq Q_{n,\alpha}^* \\ 0 & \text{otherwise} \end{cases}$$

Now under H_0, Q_n^* follows a central chi-squared distribution with three degrees of freedom (DF) , and we take $Q_{n,\alpha}^* = \chi_{3,\alpha}^2$ which is the upper $\alpha\% - tile$ of the chi-squared distribution. Similarly, we consider test-statistics based on $\overline{\boldsymbol{\theta}} = (\overline{\beta}_{0n}, \overline{\beta}_{1n}, \overline{\sigma}_n)'$ is given by

$$\overline{Q}_n = n(\sigma^0)^{-2}\left(\frac{\pi^2}{9}(\overline{\beta}_{0n} - \overline{\beta}_0)^2\right) + \left(\frac{(3+\pi)^2}{9}(\overline{\sigma}_n - \overline{\sigma}_0)^2\right) + \frac{\pi^2}{9}(s^2 + \overline{x}^2)(\overline{\beta}_{1n} - \overline{\beta}_1)^2$$
$$+ \frac{2\pi^2}{9}\overline{x}(\overline{\beta}_{0n} - \overline{\beta}_0^0)(\overline{\beta}_{1n} - \overline{\beta}_1^0)$$

giving the test-function

$$\phi(\overline{Q}_n) = \begin{Bmatrix} 1 & if\ \overline{Q}_n \geq \overline{Q}_{n,\alpha} \\ 0 & otherwise \end{Bmatrix}$$

As in the case $Q_n^*, \overline{Q}_n$ follows a central chi-squared distribution with three DF under H_0 and $\overline{Q}_{n,\alpha} = \chi_{3,\alpha}^2$ as before.

To find the asymptotic distribution of $Q_n^*(\overline{Q}_n)$ under H_A, we consider a sequence of local alternatives $\{A_n\}$, where

$$A_n = \beta_{0(n)} + \beta_0^0 + n^{-1/2}\delta_0,\ B_{1(n)} = \beta_1^0 + n^{-1/2}\delta_1,\ \overline{Q}_{(n)} = \sigma^0 + n^{-1/2}\delta_2$$

where $\boldsymbol{\delta} = (\delta_0, \delta_1, \delta_2) \neq (0,0,0)$ is some fixed real vector in $R^2 \times R^+$. Using the asymptotic distribution of $(\beta_{0n}^*, \beta_{1n}^*, \sigma_n^*)'$ and $(\overline{\beta}_{0n}, \overline{\beta}_{1n}, \overline{\sigma}_n)'$ under $\{A_n\}$, we find the asymptotic distribution of Q_n^* and $\overline{Q}_n$ follows the non-central chi-squared distribution with three degrees of freedom and the non-central parameters.

$$\Delta^* = \boldsymbol{\delta}'\mathbf{K}\boldsymbol{\delta} / (\sigma^0)^2 \text{ and } \overline{\Delta} = \boldsymbol{\delta}'\mathbf{I}\boldsymbol{\delta} /(\sigma^0)^2$$

respectively. To compare Q_n^* and $\overline{Q}_n$ we note that the classical Pitman ARE result is applicable since the tests have same size α and similar non-central chi-squared distribution. So using Puri and Sen (1971) we obtain

$$ARE[Q_n^* : \overline{Q}_n] = \frac{\boldsymbol{\delta}'\mathbf{K}\boldsymbol{\delta}}{\boldsymbol{\delta}'\mathbf{I}\boldsymbol{\delta}}$$

By Courant-Fisher theorem (Rao (1973) the extremes of the ratio of two quadratic forms in δ are given by

$$Ch_{\min}\,(\mathbf{KI}^{-1}) \le \frac{\boldsymbol{\delta}'\mathbf{K}\boldsymbol{\delta}}{\boldsymbol{\delta}'\mathbf{I}\boldsymbol{\delta}} \le Ch_{\max}\,(\mathbf{KI}^{-1})$$

where $Ch_{\min}(A)$ and $Ch_{\max}(A)$ are minimum and maximum characteristic roots of A. In this case,

$$(\mathbf{KI}^{-1}) = \begin{pmatrix} \frac{9\boldsymbol{K}_1}{\pi^2} & 0 & \frac{9\boldsymbol{K}_3}{(3+\pi^2)} \\ 0 & \frac{9\boldsymbol{K}_1}{\pi^2} & \frac{9\bar{\boldsymbol{x}}\boldsymbol{K}_3}{(3+\pi^2)} \\ \frac{9\boldsymbol{K}_3}{(3+\pi^2)} & 0 & \frac{9\boldsymbol{K}_2}{(3+\pi^2)} \end{pmatrix}$$

Further,

$$\left|\mathbf{KI}^{-1}\right| = \frac{729K_1\Delta}{\pi^4(3+\pi^2)}$$

which is same as the *JARE* expression. Thus the optimum spacings are the same as the estimation problem.

4. ESTIMATION OF CONDITIONAL REGRESSION QUANTILES

Consider the conditional regression quantiles

$$Q(\xi) = \beta_0 + \beta_1 x_0 + \sigma\frac{\sqrt{3}}{\pi}\ln(\xi(1-\xi)^{-1}), 0<\xi<1$$

where x_0 and ξ are specified. We can estimate $Q(\xi)$ using the two estimators of $(\beta_0, \beta_1, \sigma)'$ namely $(\beta_{0n}^*, \beta_{1n}^*, \sigma_n^*)'$ and $(\bar{\beta}_{0n}, \bar{\beta}_{1n}, \bar{\sigma}_n)'$ yielding

$$Q_n^*(\xi) = \beta_{0n}^* + \beta_{1n}^* x_0 + \sigma_n^*\frac{\sqrt{3}}{\pi}\ln(\xi(1-\xi)^{-1})$$

and

$$\bar{Q}_n\,(\xi) = \bar{\beta}_{0n} + \bar{\beta}_{1n} x_0 + \bar{\sigma}_n\,\frac{\sqrt{3}}{\pi}\ln(\xi(1-\xi)^{-1})$$

with the respective asymptotic variance given by

$$Var[Q_n^*(\xi)] = \frac{\sigma^2}{n}\mathbf{1}'\mathbf{K}^{-1}\mathbf{1},\ \mathbf{l} = (1, x_{0,} \frac{\sqrt{3}}{\pi}\ln(\xi(1-\xi)^{-1}))'$$

and

$$Var[\overline{Q}_n(\xi)] = \frac{\sigma^2}{n}\mathbf{l}'\mathbf{I}^{-1}\mathbf{l},$$

The ARE of $Q_n^*(\xi)$ relative to $\overline{Q}_n(\xi)$ is then given by

$$\text{ARE}[Q_n^*(\xi) : \overline{Q}_n(\xi)] = \frac{\mathbf{l}'\mathbf{I}^{-1}\mathbf{l}}{\mathbf{l}'\mathbf{K}^{-1}\mathbf{l}}$$

where

$$Ch_{\min}(\mathbf{KI}^{-1}) \leq ARE[Q_n^*(\xi) : \overline{Q}_n(\xi)] \leq Ch_{\max}(\mathbf{KI}^{-1})$$

Thus, the optimum spacings of the k regression quantiles are the same as the spacings for the estimation and testing problems.

REFERENCES

Balakrishnan, N. (1992): *Handbook of the Logistic Distribution,* Marcel Dekker, New York

David, H. A. and Nagaraja, H. N.(2003): *Order Statistics*, John Wiley, USA

Hassanein, K. M. (1969): Estimation of the parameters of the logistic distribution by sample quantiles, *Biometrika*, **56**, 684-687.

Koenker, R. and Bassett, G. (1978). Regression uantiles. *Econometrica*, **46**, 33-50.

Puri, M. L. and Sen, P. K. (1971*). Nonparametric Methods In Multivariate Analysis*, John Wiley & Sons.

Rao, C. R.(1973). *Linear Statistical Inference and Its Application*, John Wiley, USA.

Saleh, A. K. Md. E. and AdatiA, A. (2009). Optimum regression quantiles for the estimation and tests of hypothesis of the parameters of a simple regression model with exponential errors, *J. of Statist. Res.*, **43**, 67-73.

Saleh, A. K. Md. E., Hassanein, K. M. and Brown, E. F.(1994). Joint estimation and tests of hypothesis of regression and scale parameters in a simple linear model with cauchy errors based on a few regression quantiles, *Nonparametrics Statistics*, **3**, 309-315.

Received: 21.09.2010

A. Adatia

Department of Mathematics and
Computer Science
University Of Lethbridge,
Lethbridge, Canada
Email:aminmohamed.adatia@uleth.ca

A. K. MD. Ehsanes Saleh

Carleton University
Ottawa, Canada
Email esaleh@math.carleton.ca