ON THE DISTRIBUTION OF QUANTILES OF
RESIDUALS IN A LINEAR MODEL

by
Raymond J. Carroll*

University of North Carolina at Chapel Hill

ABSTRACT

The representation for quantiles discussed by Bahadur (1966) and
Ghosh (1971) is extended to quantiles from the residuals in a linear
model. As an application, a robust test for symmetry is proposed.

AMS 1970 Subject Classifications: Primary 62G30; Secondary 62J05.
Key Words and Phrases: Quantiles, Order Statistics, Tests for Symmetry.

*This work was supported by the Air Force Office of Scientific Research
under contract AFOSR-75-2796.
I. Introduction

Let \( y_1, \ldots, y_n \) be independent and identically distributed with common distribution \( F \), and let \( F(\xi) = p \). Let \( k_n \) be a sequence of integers \( k_n = np + o\left(n^{1/2} \log n\right) \), and let \( V_n \) be the \( k_n \)th sample quantile. Bahadur (1966) and Ghosh (1971) have shown that if \( F_n \) is the empirical distribution function of \( y_1, \ldots, y_n \), then

\[
(1) \quad n^{1/2} \left( V_n - \xi - (k_n/n - F_n(\xi))/F'(\xi) \right) \overset{p}{\longrightarrow} 0.
\]

Actually, Bahadur's result is an almost sure result, but (1) suffices for many statistical applications. Our goal is to extend (1) to the quantiles of the residuals in a simple linear model, using this to conduct a quick test for symmetry.

We consider the simple linear model \( z_i = x_i'\beta + y_i \), where \( y_1, \ldots, y_n \) are i.i.d. with distribution \( F \) continuous and strictly increasing in a neighborhood of \( \xi \), with \( F(\xi) = p \).

If \( \hat{\beta}_n \) is an estimate of \( \beta \) (possibly the least squares estimate or one of the new robust estimates), the residuals are

\[
\epsilon_{n, i} = z_i - x_i'\hat{\beta}_n = y_i - x_i'(\hat{\beta}_n - \beta).
\]

Let \( E_n(F_n) \) be the empirical distribution function of the residuals (errors), and choose

\[
a_n = n^{-1/2} \log n, \quad b_n = (\log n)^2.
\]

Let \( k_n \) be as above, let

\[
I_n = (\xi - a_n, \xi + a_n),
\]

and define \( V_n \) to be the \( k_n \)th order statistic of the residuals. We make the following assumptions:

(A1) \( F \) has two bounded derivatives in a neighborhood of \( \xi \), and \( F'(\xi) = f(\xi) > 0 \).

(A2) \( a_n^{-1} (\hat{\beta}_n - \beta) \overset{a.s.}{\longrightarrow} 0 \)
(A3) For some $\delta > 0$, $a_n^{1/2} - \delta \max\{|x_i|: i=1, \ldots, n\} \to 0$.

(A4) There exists $x_0$ finite with $(\log n)n^{-1} \sum_1^n (x_i - x_0) \to 0$.

Theorem Under conditions (A1) - (A4),

$$n^{1/2}|(V_n - \xi) - (k_n/n - F_n(\xi))/f(\xi) + x_0'(\beta_n - \beta)| \to 0 \text{ (a.s.)}.$$  

Remarks Condition (A1) is the same as Bahadur's, while (A2) is quite reasonable. Condition (A3) guarantees that the design elements will not vary too wildly. The usual condition on the design matrix $X$ is that $n^{-1}X'X \to \Sigma$ (Drygas (1976)), and this implies that if there is an intercept term there is an $x_0$ with $n^{-1} \sum_1^n (x_i - x_0) \to 0$, so that (A4) is only a slight strengthening of a standard condition, which in fact is reduced to $n^{-1} \sum_1^n (x_i - x_0) \to 0$ at the end of Section 2. Note that (A3) and (A4) are guaranteed in the problem considered by Bahadur and Ghosh.

2. Proof of the Theorem

We first consider simple linear regression without an intercept $z_i = x_i\beta + y_i$. The problems of simple and multiple linear regression with an intercept are easy consequences of the proof presented here.

Define the following processes:

$$G_n(x) = n^{-1} \sum_{i=1}^n \{I(r_i \leq x) - I(y_i \leq \xi) - (F(x + x_i(\beta - \beta)) - F(x))\}$$

$$H_n = n^{1/2} \sup\{|G_n(x)|: x \in I_n\}$$

$$W_n(s_i) = n^{-1/2} \sum_{i=1}^n \left\{I(y_1 \leq \xi + a_{s+a_t x_i}) - I(y_1 \leq \xi)ight\} - F(\xi + a_{s+a_t x_i}) + F(\xi)$$

\[\]
where \( I(A) \) is the indicator of the event \( A \). We note that from

\[(A1) \rightarrow (A4), \]

\[
(3) \quad n^{-1} \sum_{i=1}^{n} \{F(\xi + sa_n) - F(\xi + sa_n + ta_n x_i) + ta_n x_i f(\xi)\} = O(n^{-1/2})
\]

uniformly for \( 0 \leq |s|, |t| \leq 1 \).

**Lemma 1** Under \((A1) \rightarrow (A4), \quad H_n \rightarrow O(a.s.) \).

**Proof of Lemma 1** Since \( r_{in} = y_i - x_i(\beta - \hat{\beta}) \), by \((A2)\) is suffices to show that \( \sup \{|W_n(s, t)| : 0 \leq s, t, \leq 1\} \rightarrow o \) \((a.s.)\).

For this, we may consider only \( x_i \geq 0 \) by breaking the sum defining \( W_n \) into two parts, one with \( x_i \geq 0 \) and the other over the terms \( x_i < 0 \). By \((A1)\) as in Bahadur's proof, it also suffices to consider points \( s, t = n_{r,n} = r/b_n \) \((r = 0, \ldots, b_n)\). To see this, note that if

\[|s - n_{r,n}| \leq b_n, \quad |t - n_{p,n}| \leq b_n,\]

then

\[
|n^{-1} \sum_{i=1}^{n} \{F(\xi + sa_n + ta_n x_i) - F(\xi + n_{r,n} a_n + n_{p,n} a_n x_i)\}| = O(a_n b_n^{-1}) = o(n^{-1/2}).
\]

Now, by Bernstein's inequality (Hoeffding, eq. 2.13),

\[
P_r\{W_n(s, t) > \epsilon\} \leq c_0 \exp(-c_\epsilon n^{1/4})
\]

for some constants \( c_0, \xi \) depending on \( \epsilon \). Thus,
$$P_r\{ |W_n(s,t)| > \epsilon \text{ for some } n \geq n_o, (s,t) = (\eta_{tn}, \eta_{pn}) \}$$
$$\leq \sum_{n=n_o}^{\infty} c_o b_n^2 \exp(-c_1 n^{1/4}) \rightarrow o \text{ as } n_o \rightarrow \infty .$$

This completes the proof of the Lemma. \qed

Lemma 2 Almost surely as \( n \rightarrow \infty, V_n \in I_n \).

Proof of Lemma 2 We have

$$\Pr(V_n \geq \xi + a_n) = \Pr\{ \sum_1^n I(y_i \leq \xi + a_n + x_i(b_n-\beta)) \leq k_n \} .$$

It suffices to show the existence of an \( n > 0 \) for which

$$Q_{n-o}(n) + 0, \text{ where}$$

$$Q_{n-o}(n) = \Pr\{ \sum_1^n I(y_i \leq \xi + a_n + n x_i) \leq k_n \text{ for some } 0 \leq t \leq n, n \geq n_o \}$$

$$= \Pr(F_n(\xi+a_n) \leq k_n/n + n^{-1} \sum_1^n (F(\xi+a_n)-F(\xi+a_n+t x_i))$$

$$- n^{-1/2} W_n(1,t) \text{ for some } 0 \leq t \leq n, n \geq n_o \} .$$

By Lemma 1 and using equation (3),

$$Q_n(n) = \Pr(F_n(\xi+a_n)-F(\xi+a_n) \leq -a_n f(\xi)(1+tx) + o(a_n) \} + o(1).$$

Choosing \( n > 0 \) small, \( Q_{n-o}(n) + 0 \) as in the proof of Bahadur's

Lemma 2. \qed

Proof of the Theorem Since \( E_n(k_n) = k_n/n \), Lemmas 1 and 2 may
be applied to show that

$$n^{1/2} |F(V_n) - F(\xi) - (k_n/n - E_n(\xi))| \rightarrow 0 \text{ (a.s.)} .$$

Further, \( F(V_n) - F(\xi) = (V_n - \xi) f(\xi) + o(n^{-1/2}) \text{ (a.s.)} \) and

$$E_n(\xi) - F(\xi) = F_n(\xi) - F(\xi) + W_n(0,a_n^{-1}(b_n-\beta)) + n^{-1} \sum_{i=1}^n \{ F(\xi + x_i(b_n-\beta)) - F(\xi) \} .$$
Since $A_n^{-1}(\beta_n - \beta) + o(a.s.)$, the proof is complete by Lemma 1 and (A4).

3. An Application

Hinkley (1975) suggested a method based on order statistics for estimating the parameter of a power transformation; his method is most applicable in the location parameter case. The Theorem can be used to construct a quick test for symmetry in more complicated models. Let $p \neq 1/2$ and define $V_n^{(p)}$ to be the $\lfloor np \rfloor$th order residual. For $F(\xi_p) = p$,

$$n^{1/2}(V_n^{(p)} + V_n^{(1-p)} - 2V_n^{(1/2)} - (\xi_p + \xi_{1-p} - 2\xi_{1/2}))$$

has a limiting normal distribution (with mean zero and variance $\sigma^2$) independent of the limit distribution of $\beta_n$. Under the hypothesis of symmetry, $\xi_p + \xi_{1-p} = 2\xi_{1/2}$, so one rejects the hypothesis if

$$n^{1/2}|V_n^{(p)} + V_n^{(1-p)} - 2V_n^{(1/2)}| / \sigma$$

is too large. To estimate $\sigma$ one needs estimates of the density at the quantiles $\xi_p$, $\xi_{1-p}$, $\xi_{1/2}$, which can be accomplished as in Bloch and Gastwirth (1968) by using the Theorem.
References


On the Distribution of Quantiles of Residuals in a Linear Model

Raymond J. Carroll

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina 27514

Air Force Office of Scientific Research
Bolling AFB, Washington, D.C.

Approved for Public Release: Distribution Unlimited

Quantiles, Order Statistics, Tests for Symmetry

The representation for quantiles discussed by Bahadur (1966) and Ghosh (1971) is extended to quantiles from the residuals in a linear model. As an application, a robust test for symmetry is proposed.