

This article was downloaded by: [National Chiao Tung University 國立交通大學]

On: 28 April 2014, At: 05:15

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Communications in Statistics - Theory and Methods

Publication details, including instructions for authors and subscription information:
<http://www.tandfonline.com/loi/lsta20>

Trimmed least squares estimator as best trimmed linear conditional estimator for linear regression model

Lin-An Chen^a & Peter Thompson^b

^a Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan

^b Mathematics Department, Wabash College, Crawfordsville, IN, 47933, U.S.A

Published online: 27 Jun 2007.

To cite this article: Lin-An Chen & Peter Thompson (1998) Trimmed least squares estimator as best trimmed linear conditional estimator for linear regression model, *Communications in Statistics - Theory and Methods*, 27:7, 1835-1849

To link to this article: <http://dx.doi.org/10.1080/03610929808832193>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

TRIMMED LEAST SQUARES ESTIMATOR AS BEST
TRIMMED LINEAR CONDITIONAL ESTIMATOR FOR
LINEAR REGRESSION MODEL

Lin-An Chen¹ and Peter Thompson²

¹Institute of Statistics, National Chiao Tung University, Hsinchu,
Taiwan.

²Mathematics Department, Wabash College,
Crawfordsville, IN 47933, U.S.A.

Key words: Instrumental variables estimator; linear conditional estimator;
linear regression; regression quantile; trimmed least squares estimator.

ABSTRACT

A class of trimmed linear conditional estimators based on regression quantiles for the linear regression model is introduced. This class serves as a robust analogue of non-robust linear unbiased estimators. Asymptotic analysis then shows that the trimmed least squares estimator based on regression quantiles (Koenker and Bassett (1978)) is the best in this estimator class in terms of asymptotic covariance matrices. The class of trimmed linear conditional estimators contains the Mallows-type bounded influence trimmed means (see De Jongh et al (1988)) and trimmed instrumental variables estimators. A large sample methodology based on trimmed instrumental variables estimator for confidence ellipsoids and hypothesis testing is also provided.

1. INTRODUCTION

Consider the linear regression model

$$y = X\beta + \epsilon \quad (1.1)$$

where y is a vector of observations of dependent variable, X is a $n \times p$ matrix of observations of $p - 1$ independent variables with values 1's on the first column, and ϵ is a vector of i.i.d. disturbance variables. The interest is to estimate the parameter vector β . It is well known that the least squares estimator is the best in covariance matrix among the unbiased subclass of linear estimators. However, the least squares estimator is highly sensitive to quite small departures from normality and to the presence of outliers. Thus, there are already a great number of papers in the literature developing robust alternatives for analyzing the linear regression model. For example, see Ruppert and Carroll (1980), Welsh (1987), Koenker and Portnoy (1987), Kim (1992), Chen (1997) and Chen and Chiang (1996). We then consider the question: Is there a robust estimator which is best in asymptotic covariance matrix in some class of robust estimators. To be specific, suppose let y^t be the subvector of y after all suspected outliers are trimmed. The vector y^t has a corresponding trimmed model that we take as

$$y^t = X^t\beta + \epsilon^t. \quad (1.2)$$

It is then natural to ask in this way if there is an estimator which is best in terms of asymptotic covariance matrices in some subclass of linear estimators for this trimmed model. For large sample comparison of covariance matrices, we replace the condition of unbiasedness in linear unbiased estimation with a condition on the trimmed linear estimation. The purpose of this paper is to introduce a class of trimmed linear estimators specified by a trimming procedure and to derive the best estimator in this class.

The trimmed linear regression model is determined by observations removed from model (1.1), which we take to be those lying outside the regres-

sion quantiles (see Koenker and Bassett (1978)). We then introduce a class of trimmed linear conditional estimators (LCE) (see (2.2) as an analogue of linear unbiased estimators to the trimmed regression model). The asymptotic properties of these estimators are then derived, and the trimmed least squares estimator (LSE) based on the regression quantiles, which was proposed by Koenker and Bassett (1978) and studied by Ruppert and Carroll (1980), is shown to be the best trimmed LCE. As a subclass of trimmed LCE, a class of trimmed instrumental variables estimators (IVE) is also introduced, where instrumental variables are variables independent of the disturbance variables and correlated with the independent variables (see Dhrymes (1970, p296-298)). It is also shown that the best trimmed IVE exists and is also a best trimmed LCE. We also note that the class of Mallows-type bounded influence trimmed means (see De Jongh et al (1988)) is also a subclass of trimmed LCE's. In Section 2, we introduce the class of trimmed LCE's and their large sample properties are derived in Section 3. In Section 4, we introduce a class of trimmed IVE's. We derive the best trimmed IVE in Section 5. A large sample methodology for confidence ellipsoids and hypothesis testing based on the trimmed IVE is introduced in Section 6. Section 7 gives the proofs of the theorems.

2. THE TRIMMED LINEAR CONDITIONAL ESTIMATORS

Recall that the regression model is

$$y = X\beta + \epsilon.$$

Let y_i be the i -th element of y and x'_i be the i -th row of X for $i = 1, \dots, n$. For $0 < \alpha < 1$, the α -th regression quantile under the model with intercept, $\hat{\beta}(\alpha)$, of β defined by Koenker and Bassett (1978) is any vector b that solves the following equation

$$\min_{b \in R^p} \sum_{i=1}^n \rho_{\alpha}(y_i - x'_i b) \text{ for } \alpha \in (0, 1)$$

where $\rho_\alpha(u) = u\psi_\alpha(u)$ with $\psi_\alpha(u) = \alpha - I(u < 0)$. Here $I(A)$ is the indicator function of the event A . As described in Koenker and Bassett (1978), the process, $\hat{\beta}(\alpha)$ is piecewise constant and uniquely defined between the breakpoints. It successfully generalizes almost all of the properties of one-sample quantiles, and may be computed very quickly using parametric linear programming (see Koenker and d'Orey (1987)).

For $0 < \alpha_1 < \alpha_2 < 1$, let $\hat{\beta}(\alpha_1)$ and $\hat{\beta}(\alpha_2)$ be the regression quantiles. We then define the trimming matrix $A = (a_{ij}, i, j = 1, \dots, p$ and $a_{ij} = I(i = j$ and $x'_i\hat{\beta}(\alpha_1) < y_i < x'_i\hat{\beta}(\alpha_2))$). After outliers are trimmed by the regression quantiles, the submodel (1.2) can be written as

$$Ay = AX\beta + A\epsilon. \quad (2.1)$$

Since A is random, the error vector $A\epsilon$ is now not a set of independent variables. We are now ready to define a subclass of linear trimmed estimators.

Definition 2.1. A statistic $\hat{\beta}_{uc}$ is called a (α_1, α_2) -trimmed LCE if there exists a stochastic $p \times p$ matrix H and a nonstochastic $n \times p$ matrix H_0 such that it has the following representation:

$$\hat{\beta}_{uc} = HH'_0Ay, \quad (2.2)$$

where matrices H and H_0 satisfy the following two conditions:

- (a1) $nH \rightarrow \tilde{H}$ in probability, where \tilde{H} is a full rank $p \times p$ matrix.
- (a2) $HH'_0AX = I_p + o_p(n^{-1/2})$, where I_p is the $p \times p$ identity matrix.

We note that "conditional" means "conditional on the sample being trimmed". Condition (a1) is similar to the usual condition that $n^{-1}X'X$ converges to a positive definite matrix. Analogously, Condition (a2) for this trimmed LCE plays an analogous role to that of unbiasedness for linear unbiased estimation. Suppose that B_y is a linear unbiased estimator of β . Then, with the fact that $BX = I$, nonstochastic matrices H and H_0 such that $HH'_0 = (\alpha_2 - \alpha_1)^{-1}B$ make $\hat{\beta}_{uc}$ an example of trimmed LCE. This implies

that the class of trimmed LCE's is at least as big with the size of the class of linear unbiased estimators.

Let ϵ has distribution function F with probability density function f . Denote by h'_i the i -th row of H_0 and z_{ij} the j -th element of vector z_i for $z = x$ and h . The following conditions are similar to the standard ones for linear regression models as given in Ruppert and Carroll (1980) and Koenker and Portnoy (1987), for example:

$$(a3) \quad n^{-1} \sum_{i=1}^n z_{ij}^4 = O(1) \text{ for } z = x \text{ and } h \text{ and all } j,$$

(a4) $n^{-1} X'X = Q_x + o(1)$, $n^{-1} H'_0 X = Q_{hx} + o(1)$ and $n^{-1} H'_0 H_0 = Q_h + o(1)$ where Q_x and Q_h are positive definite matrices and Q_{hx} is a full rank matrix.

(a5) $n^{-1} \sum_{i=1}^n z_i = \theta_x + o(1)$, for $z = x$ and h , and where θ_x is a finite vector with first element value 1.

(a6) The probability density function and its derivative are both bounded and bounded away from 0 in a neighborhood of $F^{-1}(\alpha)$ for $\alpha \in (0, 1)$.

For any two positive definite $p \times p$ matrices Q_1 and Q_2 , we say that Q_1 is smaller than Q_2 if $Q_2 - Q_1$ is positive semidefinite.

Definition 2.2. An estimator in the class of (α_1, α_2) -trimmed LCE's is called the best if it has asymptotic covariance matrix smaller than or equal to it of any estimator in this class.

In analogy with the case of the best linear unbiased estimator, we will show that the best (α_1, α_2) -trimmed LCE always exists. It can also be seen that the asymptotic covariance matrix of the best (α_1, α_2) -trimmed LCE will vary in the trimming percentage (α_1, α_2) . We then may further expect the existence of a uniformly best one.

Definition 2.3. A trimmed LCE for some trimming percentage is said to be a uniformly best trimming LCE if it has asymptotic covariance matrix

smaller than or equal to it of any best (α_1, α_2) -trimmed LCE, for all $0 < \alpha_1 < 0.5 < \alpha_2 < 1$.

We are not going to study when a uniformly best trimming LCE exists.

3. ASYMPTOTIC PROPERTIES OF THE TRIMMED LINEAR CONDITIONAL ESTIMATOR

The following theorem gives a "Bahadur" representation of the (α_1, α_2) -trimmed LCE.

Theorem 3.1. With assumptions (a1)-(a6), we have

$$\begin{aligned} n^{1/2}(\hat{\beta}_{blc} - (\beta + \gamma_{uc})) &= n^{-1/2} \tilde{H} \sum_{i=1}^n [h_i(\epsilon_i I(F^{-1}(\alpha_1) \leq \epsilon_i \leq F^{-1}(\alpha_2)) - \lambda) \\ &+ [F^{-1}(\alpha_1)I(\epsilon_i < F^{-1}(\alpha_1)) + F^{-1}(\alpha_2)I(\epsilon_i > F^{-1}(\alpha_2)) - ((1 - \alpha_2)F^{-1}(\alpha_2) \\ &+ \alpha_1 F^{-1}(\alpha_1))] Q_{hx} Q_x^{-1} x_i] + o_p(1), \end{aligned}$$

where $\gamma_{uc} = \lambda \tilde{H} \theta_h$, $\lambda = \int_{F^{-1}(\alpha_1)}^{F^{-1}(\alpha_2)} \epsilon dF(\epsilon)$ and θ_h is defined in assumption (a5).

The limiting distribution of the (α_1, α_2) -trimmed LCE follows from the central limit theorem (see, e.g. Serfling (1980, p. 30)).

Corollary 3.2. $n^{1/2}(\hat{\beta}_{blc} - (\beta + \gamma_{uc}))$ has an asymptotic normal distribution with zero mean vector and the following asymptotic covariance matrix:

$$\begin{aligned} &[\int_{F^{-1}(\alpha_1)}^{F^{-1}(\alpha_2)} \epsilon^2 dF(\epsilon) - \lambda^2] \tilde{H} Q_h \tilde{H}' + [\alpha_1 (F^{-1}(\alpha_1))^2 + (1 - \alpha_2) (F^{-1}(\alpha_2))^2 \\ &- (\alpha_1 F^{-1}(\alpha_1) + (1 - \alpha_2) F^{-1}(\alpha_2))^2 - 2\lambda(\alpha_1 F^{-1}(\alpha_1) + (1 - \alpha_2) F^{-1}(\alpha_2))] Q_x^{-1}. \end{aligned} \quad (3.1)$$

The (α_1, α_2) -trimmed LSE proposed by Koenker and Bassett (1978) is defined by

$$\hat{\beta}_{tr} = (X'AX)^{-1}X'Ay.$$

From the result of this estimator studied by Ruppert and Carroll (1980), we have

$$n^{-1}X'AX \rightarrow (\alpha_2 - \alpha_1)Q_x.$$

By letting $H = (X'AX)^{-1}$ and $H_0 = X$, can see that condition (a2) also holds for $\hat{\beta}_{tr}$. So, the (α_1, α_2) -trimmed LSE is in the class of (α_1, α_2) -trimmed LCE's. Moreover, Ruppert and Carroll (1980) provided the result that $n^{1/2}(\hat{\beta}_{tr} - (\beta + \gamma_{tr}))$, where $\gamma_{tr} = (\alpha_2 - \alpha_1)^{-1} \lambda Q_x^{-1} \theta_x$, has an asymptotic normal distribution with zero means and covariance matrix $\sigma^2(\alpha_1, \alpha_2)Q_x^{-1}$, where

$$\begin{aligned} \sigma^2(\alpha_1, \alpha_2) = & (\alpha_2 - \alpha_1)^{-2} \left[\int_{F^{-1}(\alpha_1)}^{F^{-1}(\alpha_2)} (\epsilon - \lambda)^2 dF(\epsilon) + \alpha_1 (F^{-1}(\alpha_1) - \lambda)^2 \right. \\ & \left. + (1 - \alpha_2)(F^{-1}(\alpha_2) - \lambda)^2 - (\alpha_1 F^{-1}(\alpha_1) + (1 - \alpha_2)F^{-1}(\alpha_2))^2 \right]. \end{aligned} \quad (3.2)$$

The following lemma orders the matrices $\tilde{H}Q_h\tilde{H}'$ and Q_x .

Lemma 3.3. For any matrices \tilde{H} and Q_h induced from conditions (a1) and (a4), the difference

$$\tilde{H}Q_h\tilde{H}' - (\alpha_2 - \alpha_1)^{-2}Q_x^{-1} \quad (3.3)$$

is positive semidefinite.

The relation in (3.3) then implies the following main theorem.

Theorem 3.4. Under the conditions (a.3)-(a.6), the (α_1, α_2) -trimmed LSE $\hat{\beta}_{tr}$ is the best (α_1, α_2) -trimmed LCE.

Since the (α_1, α_2) -trimmed LSE always exists, then the best (α_1, α_2) -trimmed LCE always exists. However, the existence of uniformly best trimming LCE depends on the underlying distribution. Suppose that there is a (α_1^*, α_2^*) so that $\sigma^2(\alpha_1^*, \alpha_2^*) = \inf_{0 < \alpha_1 < \alpha_2 < 1} \sigma^2(\alpha_1, \alpha_2)$, then the best (α_1^*, α_2^*) -trimmed LCE is the uniformly best trimming LCE. Two questions induced from above discussion are then raised. First, how big is the class of (α_1, α_2) -trimmed LCE's? Secondly, are the best (α_1, α_2) -trimmed LCE and the uniformly best trimming LCE unique if they exist? We are not going to study the

scope of the trimmed LCE's. However, we will introduce a class of (α_1, α_2) -trimmed IVE's which is shown to be a subclass of the (α_1, α_2) -trimmed LCE's. We will also show that if there is a best (α_1, α_2) -trimmed LCE, then it is asymptotically equivalent to the best (α_1, α_2) -trimmed IVE. Let $H = (X'WAX)^{-1}$ and $H_0 = WX$ with W a diagonal matrix of weights. This shows that the Mallows-type bounded influence trimmed means also form a subclass of trimmed LCE's (see De Jongh et al (1988) for their large sample properties). In particular, $\hat{\beta}_{tr}$ is the one with W the identity matrix and then belongs to this subclass. A direct result from Theorem 3.4 is that $\hat{\beta}_{tr}$ is the best Mallows-type bounded influence trimmed mean. In the next section, we will introduce the trimmed IVE.

4. TRIMMED INSTRUMENTAL VARIABLES ESTIMATORS

Let S be the $n \times k$, $k \geq p$, observation matrix of instrumental variables. Each instrument is a variable independent of the disturbance variables and correlated with the independent variables. Denote by s'_i the i -th row of S and s_{ij} the j -th element of s_i . We add the following conditions:

$$(a7) \quad n^{-1} \sum_{i=1}^n s_{ij}^4 = O(1) \text{ for all } j,$$

(a8) $n^{-1} S'X = Q_{sx} + o(1)$, and $n^{-1} S'S = Q_s + o(1)$, where Q_s is a $k \times k$ positive definite matrix and Q_{sx} is a full rank matrix,

$$(a9) \quad n^{-1} \sum_{i=1}^n s_i = \theta_s + o(1).$$

We then define the trimmed IVE.

Definition 4.1. Let P_s be the idempotent matrix $S(S'S)^{-1}S'$. The trimmed IVE is defined by

$$\hat{\beta}_s = ((AX)'P_sAX)^{-1}(AX)'P_sAy. \quad (4.1)$$

It will be shown that the trimmed IVE is a (α_1, α_2) -trimmed LCE.

Even in this trimmed regression model, it is apparent that there may exist many sets of instruments that one might consider using. Thus, we shall be

concerned with finding the instruments for which the corresponding trimmed IVE has smallest covariance matrix.

Definition 4.2. (1) An estimator in the class of (α_1, α_2) -trimmed IVE's is called the best if it has asymptotic covariance matrix smaller than or equal to any estimator in this class.

(2) A trimmed IVE for some trimming percentage is called the uniformly best trimming IVE if it has asymptotic covariance matrix smaller than or equal to it of any best (α_1, α_2) -trimmed IVE, for $0 < \alpha_1 < 0.5 < \alpha_2$.

We first show the relation between the IVE and the LCE.

Lemma 4.3. $((AX)'P_sAX)^{-1}(AX)'S(S'S)^{-1}$ converges to the full rank matrix $(\alpha_2 - \alpha_1)^{-1}(Q'_{sx}Q_s^{-1}Q_{sx})^{-1}Q'_{sx}Q_s^{-1}$ in probability.

This lemma then implies that condition (a1) holds. One can check that condition (a2) also holds. Then the trimmed IVE's form a subclass of trimmed LCE's.

We state the asymptotic properties of the trimmed IVE. The following theorem gives a "Bahadur" representation of $\hat{\beta}_s$.

Theorem 4.4. With the assumptions, we have

$$n^{1/2}(\hat{\beta}_s - (\beta + \gamma_s)) = n^{-1/2}(\alpha_2 - \alpha_1)^{-1}(Q'_{sx}Q_s^{-1}Q_{sx})^{-1}Q'_{sx}Q_s^{-1} \sum_{i=1}^n \{s_i(\epsilon_i I(F^{-1}(\alpha_1) \leq \epsilon_i \leq F^{-1}(\alpha_2)) - \lambda) + [F^{-1}(\alpha_1)I(\epsilon_i < F^{-1}(\alpha_1)) + F^{-1}(\alpha_2) I(\epsilon_i > F^{-1}(\alpha_2)) - ((1 - \alpha_2)F^{-1}(\alpha_2) + \alpha_1 F^{-1}(\alpha_1))]Q_{sx}Q_x^{-1}x_i\} + o_p(1)$$

where $\gamma_s = (\alpha_2 - \alpha_1)^{-1}\lambda(Q'_{sx}Q_s^{-1}Q_{sx})^{-1}Q'_{sx}Q_s^{-1}\theta_s$ and θ_s has been defined as $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n s_i$.

The limiting distribution of the trimmed IVE is stated in the following corollary.

Corollary 4.5. $n^{1/2}(\hat{\beta}_s - (\beta + \gamma_s))$ has an asymptotic normal distribution with zero mean vector and the following asymptotic covariance matrix

not unique. If there is a (α_1^*, α_2^*) -trimmed LSE which is a uniformly best LCE, then the best (α_1^*, α_2^*) -trimmed IVE is also a uniformly best LCE. This says that if there is a trimmed LSE which is a uniformly best LCE then the uniformly best trimmed LCE is not numerically unique.

For a large sample inference methodology, we here give the limiting distribution of the trimmed IVE when the distribution F is symmetric.

Corollary 5.2. When the distribution F is symmetric and we let $\alpha_1 = 1 - \alpha_2 = \alpha, 0 < \alpha < 0.5$ then $n^{1/2}(\hat{\beta}_s - \beta)$ has an asymptotic normal distribution with zero mean vector and the following asymptotic covariance matrix

$$(1 - 2\alpha)^{-2} \left[\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \epsilon^2 dF(\epsilon) (Q'_{sx} Q_s^{-1} Q_{sx})^{-1} + 2\alpha (F^{-1}(\alpha))^2 Q_x^{-1} \right].$$

6. LARGE SAMPLE INFERENCE

Here we sketch a large-sample methodology for confidence ellipsoids and hypothesis testing based on the trimmed IVE for the case of symmetric distribution. To do this, we need first to estimate the asymptotic covariance matrix of $\hat{\beta}_s$. Let $\hat{Q}_x = n^{-1} \sum_{i=1}^n x_i x_i'$, $\hat{Q}_{sx} = n^{-1} \sum_{i=1}^n s_i x_i'$ and $\hat{Q}_s = n^{-1} \sum_{i=1}^n s_i s_i'$ and also $\hat{F}^{-1}(1 - \alpha) = \delta'(\hat{\beta}(1 - \alpha) - \hat{\beta}_s)$ where δ is a p -vector with first element value 1 and else zeros. Furthermore, let

$$V = (1 - 2\alpha)^{-2} \left[n^{-1} \sum_{i=1}^n e_i^2 I(x_i' \hat{\beta}(\alpha) < y_i < x_i' \hat{\beta}(1 - \alpha)) (\hat{Q}'_{sx} \hat{Q}_s^{-1} \hat{Q}_{sx})^{-1} + 2\alpha (\hat{F}^{-1}(1 - \alpha))^2 \hat{Q}_x^{-1} \right],$$

where $e_i = y_i - x_i' \hat{\beta}_s, i = 1, \dots, n$.

Theorem 6.1. $V \rightarrow (1 - 2\alpha)^{-2} \left[\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \epsilon^2 dF(\epsilon) (Q'_{sx} Q_s^{-1} Q_{sx})^{-1} + 2\alpha (F^{-1}(\alpha))^2 Q_x^{-1} \right]$ in probability.

For $0 < u < 1$, let $F_u(r_1, r_2)$ denote the $(1-u)$ quantile of the F distribution, with r_1 and r_2 degrees of freedom, and let

$$d_u(r_1, r_2) = (1 - 2\alpha)^{-1} r_1 F_u(r_1, r_2).$$

Suppose for some integer ℓ , K is matrix of size $\ell \times p$, and K has rank ℓ . Let m be the number of observations y_i lying outside the interval $(x_i' \hat{\beta}(\alpha), x_i' \hat{\beta}(1 - \alpha))$. Then the region of β

$$(\hat{\beta}_s - \beta)' K' (K V K')^{-1} K (\hat{\beta}_s - \beta) \geq d_u(\ell, n - m - p)$$

has probability of approximately u . If $K = I_p$, the confidence ellipsoid

$$(\hat{\beta}_s - \beta)' V^{-1} (\hat{\beta}_s - \beta) \leq d_u(\ell, n - m - p)$$

for β has an asymptotic confidence coefficient of approximately $1 - u$. Moreover, if we test $H_0 : K\beta = v$ by rejecting H_0 whenever

$$(K \hat{\beta}_s - v)' (K V K')^{-1} (K \hat{\beta}_s - v) \geq d_u(\ell, n - m - p)$$

has an asymptotic size of u .

7. APPENDIX

Proof of Theorem 3.1. Inserting (1.1) in the equation (2.2), we have

$$n^{1/2}(\hat{\beta}_{uc} - \beta) = n^{1/2} H H_0' A \epsilon + o_p(1).$$

Now, we consider a representation of $n^{-1/2} H_0' A \epsilon$. Let

$$U_j(\alpha, T_n) = n^{-1/2} \sum_{i=1}^n h_{ij} \epsilon_i I(\epsilon_i < F^{-1}(\alpha) + n^{-1/2} x_i' T_n)$$

and

$$U(\alpha, T_n) = (U_1(\alpha, T_n), \dots, U_p(\alpha, T_n)).$$

So

$$n^{-1/2} H_0' A \epsilon = U(\alpha_2, n^{1/2}(\hat{\beta}(\alpha_2) - \beta(\alpha_2))) - U(\alpha_1, n^{1/2}(\hat{\beta}(\alpha_1) - \beta(\alpha_1))).$$

The following result, which also uses the Jureckova and Sen (1987) extension of Billingsly's Theorem, will provides the representation for $n^{-1/2} H_0' A \epsilon$.

$$|U_j(\alpha, T_n) - U_j(\alpha, 0) - n^{-1}F^{-1}(\alpha)f(F^{-1}(\alpha)) \sum_{i=1}^n h_{ij}x_i'T_n| = o_p(1) \quad (7.1)$$

for $p = 1, \dots, p$ and $T_n = O_p(1)$.

To complete the proof of Theorem 3.1, from (7.1) and the representation of $\hat{\beta}(\alpha)$ (see Ruppert and Carroll (1980)) we have

$$\begin{aligned} n^{-1/2}H_0'Ac &= (U(\alpha_2, n^{1/2}(\hat{\beta}(\alpha_2) - \beta(\alpha_2))) - U(\alpha_2, 0)) \quad (7.2) \\ &\quad - U(\alpha_1, n^{1/2}(\hat{\beta}(\alpha_1) - \beta(\alpha_1))) - U(\alpha_1, 0) + (U(\alpha_2, 0) - U(\alpha_1, 0)) \\ &= n^{-1/2} \sum_{i=1}^n [h_i \epsilon_i I(F^{-1}(\alpha_1) \leq \epsilon_i \leq F^{-1}(\alpha_2)) + (F^{-1}(\alpha_2) I(\epsilon_i > F^{-1}(\alpha_2)) \\ &\quad + F^{-1}(\alpha_1) I(\epsilon_i < F^{-1}(\alpha_1))) Q_{hx} Q_x^{-1} x_i - ((1 - \alpha_2) F^{-1}(\alpha_2) + \alpha_1 F^{-1}(\alpha_1)) \\ &\quad Q_{hx} Q_x^{-1} x_i] + o_p(1). \end{aligned}$$

The theorem is then followed from (7.2) and Condition (a1).

Proof of Lemma 3.3. Denote by $\text{plim}(B_n) = B$ if B_n converges to B in probability. Let

$$C = HH_0 - (X'AX)^{-1}X'.$$

With this, $\text{plim}(CX) = \text{plim}(HH_0X) - \text{plim}(X'AX)^{-1}X'X = 0$.

Then

$$\begin{aligned} \tilde{H}Q_h\tilde{H}' &= \text{plim}(HH_0(HH_0)') \\ &= \text{plim}((C + (X'AX)^{-1}X')(C + (X'AX)^{-1}X'))' \\ &= \text{plim}(CC') + \text{plim}((X'AX)^{-1}X'X(X'AX)^{-1}) \\ &= \text{plim}(CC') + (\alpha_2 - \alpha_1)^{-2} \text{plim}(X'X)^{-1} \\ &\geq (\alpha_2 - \alpha_1)^{-2} Q_x^{-1}. \end{aligned}$$

Proof of Lemma 4.3. From Condition (a8), we need only to show that

$$n^{-1}S'AX = (\alpha_2 - \alpha_1)Q_{sx} + o_p(1). \quad (7.3)$$

The following result, which uses the Jureckova and Sen (1987) extension of Billingsley's Theorem, will give an expansion of the matrix $n^{-1}(S'AX)$.

$$n^{-1} \sum_{i=1}^n s_{ij} x_{ik} I(x'_i \hat{\beta}(\alpha_1) < y_i < x'_i \hat{\beta}(\alpha_2)) = (\alpha_2 - \alpha_1) q_{jk} + o_p(1)$$

where q_{jk} is the jk -th term of the matrix Q_{sx} , and s_{ij} , x_{ik} are the ij -th and ik -th terms of S and X , respectively. We then have the statement (7.3).

Theorems 4.4 and 4.6 are followed from the arguments for Theorem 3.1 and then their proofs are omitted.

Proof of Theorem 6.1. From the representation of regression quantile in Ruppert and Carroll (1980) and the trimmed IVE $\hat{\beta}_s$ we have $\hat{F}(1 - \alpha) \rightarrow F^{-1}(1 - \alpha)$ in probability. Now,

$$n^{-1} \sum_{i=1}^n e_i^2 I(x'_i \hat{\beta}(\alpha) < y_i < x'_i \hat{\beta}(1 - \alpha)) = n^{-1} (\hat{\beta}_s - \beta)' \sum_{i=1}^n x_i x_i' (\hat{\beta}_s - \beta) \\ I(x'_i \hat{\beta}(\alpha) < y_i < x'_i \hat{\beta}(1 - \alpha)) + n^{-1} \sum_{i=1}^n \epsilon_i^2 I(x'_i \hat{\beta}(\alpha) < y_i < x'_i \hat{\beta}(1 - \alpha)).$$

From the fact that $n^{1/2}(\hat{\beta}_s - \beta) = O_p(1)$ and the Condition (a8), the theorem follows from the result that

$$n^{-1} \sum_{i=1}^n \epsilon_i^2 I(x'_i \hat{\beta}(\alpha) < y_i < x'_i \hat{\beta}(1 - \alpha)) = n^{-1} \sum_{i=1}^n \epsilon_i^2 I(F^{-1}(\alpha) < \epsilon_i < \\ F^{-1}(1 - \alpha)) + o_p(1)$$

which follows from Lemma A.4 of Ruppert and Carroll (1980).

ACKNOWLEDGEMENT

We thank an associate editor and three referees for their insightful comments and suggestions, which led to a much improved version of this article.

REFERENCES

Chen, L.-A. (1997). An efficient class of weighted trimmed means for linear regression models, *Statistica Sinica*, 7, 669-686.

- Chen, L.-A. and Chiang, Y. C. (1996), Symmetric type quantile and trimmed means for location and linear regression model. *Journal of Nonparametric Statistics*, 7, 171-185.
- De Jongh, P. J., De Wet, T. and Welsch, A. H. (1988), Mallows-Type Bounded-Influence-Regression Trimmed Means, *Journal of the American Statistical Association*, 83, 805-810.
- Dhrymes, P. J. (1970). *Econometrics - Statistical Foundations and Applications*. Harper & Row, New York.
- Jureckova, J. and Sen, P. K. (1987). An extension of Billingsley's theorem to higher dimension M-processes. *Kybernetika* 23, 382-387.
- Kim, S. J. (1992). The metrically trimmed means as a robust estimator of location. *Annals of Statistics*. 20, 1534-1547.
- Koenker, R. and Bassett, G. J. (1978). Regression quantiles. *Econometrica*. 46, 33-50.
- Koenker, R. and Portnoy, S. (1987). L-estimation for linear model. *Journal of the American Statistical Association*, 82, 851-857.
- Koenker, R. and d'Orey, V. (1987). Computing regression quantiles. *Applied Statistics*. 36, 383-393.
- Maddala, G. S. (1988), *Introduction to econometrics*. Macmillan Publishing Company, New York.
- Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association*. 75, 828-838.
- Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. Wiley, New York.
- Welsh, A. H. (1987). The trimmed mean in the linear model. *Annals of Statistics* 15, 20-36.

Received Decmter, 1996; Revised January, 1998.
