

ESTIMATORS FOR THE LINEAR REGRESSION MODEL BASED ON WINSORIZED OBSERVATIONS

L-A Chen, A. H. Welsh and W. Chan

*National Chiao Tung University, Australian National University
and University of Texas-Houston*

Abstract: We develop an asymptotic, robust version of the Gauss-Markov theorem for estimating the regression parameter vector β and a parametric function $c'\beta$ in the linear regression model. In a class of estimators for estimating β that are linear in a Winsorized observation vector introduced by Welsh (1987), we show that Welsh's trimmed mean has smallest asymptotic covariance matrix. Also, for estimating a parametric function $c'\beta$, the inner product of c and the trimmed mean has the smallest asymptotic variance among a class of estimators linear in the Winsorized observation vector. A generalization of the linear Winsorized mean to the multivariate context is also given. Examples analyzing American lobster data and the mineral content of bones are used to compare the robustness of some trimmed mean methods.

Key words and phrases: Linear regression, robust estimation, trimmed mean, Winsorized mean.

1. Introduction

Consider the linear regression model

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

where y is a vector of observations for the dependent variable, X is a known $n \times p$ design matrix with 1's in the first column, and ϵ is a vector of independent and identically distributed disturbance variables. We consider the problem of estimating the parameter vector β and the parametric function $c'\beta$ of β .

From the Gauss-Markov theorem, it is known that the least squares estimator has the smallest covariance matrix in the class of unbiased linear estimators My where M satisfies $MX = I_p$. Also, the inner product of c and the least squares estimator has smallest variance among all linear unbiased estimators of $c'\beta$. However, the least squares estimator is sensitive to departures from normality and to the presence of outliers so we need to consider robust estimators. One approach to robust estimation is to construct a weighted observation vector

y^* and then construct a consistent estimator which is linear in y^* ; see for example, Ruppert and Carroll (1980), Welsh (1987), Koenker and Portnoy (1987), Kim (1992), Chen and Chiang (1996) and Chen (1997). There are two types of weighted observation vectors in this literature. First, y^* can represent a trimmed observation vector Ay with A a trimming matrix constructed from regression quantiles (see Koenker and Bassett (1978)) or residuals based on an initial estimator (see Ruppert and Carroll (1980) and Chen (1997)). Second, y^* can be a Winsorized observation vector defined as in Welsh (1987). In this paper, we consider the Winsorized observation vector of Welsh (1987), study classes of linear functions based on y^* for estimation of β and $c'\beta$, and develop a robust version of the Gauss-Markov theorem.

In Section 2, we introduce various types of linear Winsorized means and derive their large sample properties in Section 3. We discuss instrumental variables and bounded-influence Winsorized means in Section 4 and generalize the results to the multivariate linear model in Section 5. Examples analyzing the American lobster data and a set of bone data are given in Section 6. Proofs of theorems are in Section 7.

2. Linear Estimation Based on Winsorized Responses

In the regression model (1.1), let y_i be the i th element of y and x'_i be the i th row of X for $i = 1, \dots, n$. Let $\hat{\beta}_0$ be an initial estimator of β . The regression residuals from $\hat{\beta}_0$ are $e_i = y_i - x'_i\hat{\beta}_0$. For $0 < \alpha_1 < 0.5 < \alpha_2 < 1$, let $\hat{\eta}(\alpha_1)$ and $\hat{\eta}(\alpha_2)$ represent, respectively, the α_1 th and α_2 th empirical quantiles of the regression residuals. The Winsorized observation defined by Welsh (1987) is

$$y_i^* = y_i I(\hat{\eta}(\alpha_1) \leq e_i \leq \hat{\eta}(\alpha_2)) + \hat{\eta}(\alpha_1) (I(e_i < \hat{\eta}(\alpha_1)) - \alpha_1) + \hat{\eta}(\alpha_2) (I(e_i > \hat{\eta}(\alpha_2)) - (1 - \alpha_2)). \quad (2.1)$$

This definition reduces the influence of observations with residuals lying outside the quantile-interval $(\hat{\eta}(\alpha_1), \hat{\eta}(\alpha_2))$ and bounds the influence in the error variable ϵ . Alternative definitions of Winsorized observations can be entertained: for example, we could replace $\hat{\eta}(\alpha_i)$ by $\hat{\eta}(\alpha_i) + x'_i\hat{\beta}_0$. It is more convenient to work on the scale of the independent and identically distributed errors ϵ than on the scale of the non-identically distributed observations y , so we retain Welsh's definition. Let $y^* = (y_1^*, \dots, y_n^*)'$ and denote the trimming matrix by $A = \text{diag}(a_1, \dots, a_n)$, where $a_i = I(\hat{\eta}(\alpha_1) \leq e_i \leq \hat{\eta}(\alpha_2))$.

Any linear unbiased estimator has the form My with M a $p \times n$ nonstochastic matrix satisfying $MX = I_p$. Since M is a full-rank matrix, there exist matrices H and H_0 such that $M = HH_0'$. Thus, an estimator is a linear unbiased estimator

if there exists a $p \times p$ nonsingular matrix H and a $n \times p$ full-rank matrix H_0 such that the estimator can be written as

$$HH_0'y. \quad (2.2)$$

We generalize linear unbiased estimators defined on the observation vector y to estimators defined on y^* by requiring them to be of the form My^* with $M = HH_0'$, where H and H_0 are chosen to ensure that the estimator is consistent.

Definition 2.1. A statistic $\hat{\beta}_{tw}$ is asymptotically linear in the Winsorized observations (ALWO) y^* if

$$\hat{\beta}_{tw} = My^*, \quad (2.3)$$

and M can be decomposed as $M = HH_0'$ with H a $p \times p$ stochastic or nonstochastic matrix and H_0 a $n \times p$ matrix which is independent of the error variables ϵ , satisfying the following two conditions:

- (a1) $nH \rightarrow \tilde{H}$ in probability, where \tilde{H} is a full rank $p \times p$ matrix.
- (a2) $HH_0'X = (\alpha_2 - \alpha_1)^{-1}I_p + o_p(n^{-1/2})$, where I_p is the $p \times p$ identity matrix.

This is similar to the usual requirements for unbiased estimation except that we have introduced a Winsorized observation vector to allow for robustness and considered asymptotic instead of exact unbiasedness.

For estimating the parametric function $c'\beta$, we define a class of estimators analogously.

Definition 2.2. A linear function $a'y^*$ is asymptotically linear in the Winsorized observations (ALWO) y^* for a parametric function $c'\beta$ if the vector a can be decomposed as $a' = h_0'H_0'$ with column p -vector h_0 stochastic or nonstochastic and H_0 a $n \times p$ matrix which is independent of the error variables ϵ , satisfying the following two conditions:

- (a1*) $nh_0 \rightarrow \tilde{h}$ in probability, where \tilde{h} is a nonzero $p \times 1$ vector.
- (a2*) $h_0'H_0'X = (\alpha_2 - \alpha_1)^{-1}c' + o_p(n^{-1/2})$.

Suppose that My^* is an ALWO estimator for the parameter vector β . Then clearly $a'y^*$ with $a' = c'M$ is an ALWO estimator for the parametric function $c'\beta$. This means that results on the optimal estimation of $c'\beta$ can be derived from those on estimation of β .

Two questions arise for the class of ALWO estimators. First, does this class of estimators contain interesting estimators? We can answer in the affirmative because the class of ALWO estimators defined in this paper contains Welsh's (1987) trimmed mean ($H = (X'AX)^{-1}$ and $H_0 = X$), the subclass of linear Winsorized instrumental variables means ($H = (S'AX)^{-1}$ and $H_0 = S$ with S a $n \times p$ matrix of instrumental variables; see Section 4) and the Mallows-type bounded influence trimmed means ($H = (X'WAX)^{-1}$ and $H_0 = X'W$ with W

a diagonal matrix of weights); see De Jongh, De Wet and Welsh (1988). Second, can one find a best estimator in this class? This question will be answered in the next section.

3. Large Sample Properties of ALWO Estimators

Let ϵ have distribution function F with probability density function f . Denote by h'_i the i th row of H_0 . Let z_i represent either the vector x_i or h_i , and z_{ij} be its j th element. 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):

- (a3) $n^{-1} \sum_{i=1}^n z_{ij}^A = O(1)$ for $z = x$ or h 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) $n^{-1} \sum_{i=1}^n z_i = \theta_z + o(1)$, for $z = x$ or h , 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 $\alpha \in (0, 1)$.
- (a7) $n^{1/2}(\hat{\beta}_0 - \beta) = O_p(1)$.

The following theorem gives a Bahadur representation for ALWO estimators. Note that the results for Welsh's trimmed mean discussed by Ren (1994) and Jureckova and Sen (1996, pp.173-175) apply only for the case $x_i = h_i$.

Theorem 3.1. *Under conditions (a1)-(a7), we have*

$$n^{1/2}(\hat{\beta}_{lw} - (\beta + \gamma_{lw})) = n^{-1/2} \tilde{H} \sum_{i=1}^n h_i \psi(\epsilon_i, F) + o_p(1)$$

with $\psi(\epsilon, F) = \epsilon I(F^{-1}(\alpha_1) \leq \epsilon \leq F^{-1}(\alpha_2)) - \lambda + F^{-1}(\alpha_1) I(\epsilon < F^{-1}(\alpha_1)) + F^{-1}(\alpha_2) I(\epsilon > F^{-1}(\alpha_2)) - ((1 - \alpha_2)F^{-1}(\alpha_2) + \alpha_1 F^{-1}(\alpha_1))$, and $\gamma_{lw} = \lambda \tilde{H} \theta_h$ and where $\lambda = \int_{F^{-1}(\alpha_1)}^{F^{-1}(\alpha_2)} \epsilon dF(\epsilon)$.

From the above theorem, it is seen that the asymptotic properties of ALWO estimators do not depend on the initial estimator. The limiting distribution of ALWO estimators follows from the Central Limit Theorem (see, e.g. Serfling (1980, p.30)).

Corollary 3.2. *Under the conditions of Theorem 3.1, the normalized ALWO estimator $n^{1/2}(\hat{\beta}_{lw} - (\beta + \gamma_{lw}))$ has an asymptotic normal distribution with zero mean vector and asymptotic covariance matrix $(\alpha_2 - \alpha_1)^2 \sigma^2(\alpha_1, \alpha_2) \tilde{H} Q_h \tilde{H}'$, where $\sigma^2(\alpha_1, \alpha_2) = (\alpha_2 - \alpha_1)^{-2} [\int_{F^{-1}(\alpha_1)}^{F^{-1}(\alpha_2)} (\epsilon - \lambda)^2 dF(\epsilon) + \alpha_1 (F^{-1}(\alpha_1) - \lambda)^2 + (1 - \alpha_2) (F^{-1}(\alpha_2) - \lambda)^2 - (\alpha_1 F^{-1}(\alpha_1) + (1 - \alpha_2) F^{-1}(\alpha_2))^2]$.*

If we further assume that F is symmetric at 0 and let $\alpha_1 = 1 - \alpha_2 = \alpha$, $0 < \alpha < 0.5$, then $\gamma_{lw} = 0$ and $\hat{\beta}_{lw}$ is a consistent estimator of β . In general, when F is asymmetric, $\hat{\beta}_{lw}$ is a biased estimator of β and the asymptotic bias is given by γ_{lw} . If we center the columns of H_0 so that θ_z has all but the first element equal to 0, then the asymptotic bias affects the intercept alone and not the slope.

We briefly sketch a large-sample methodology for statistical inference for β based on an ALWO estimator. To do this, we first need to estimate the asymptotic covariance matrix of $\hat{\beta}_{lw}$. Let $\hat{Q}_h = n^{-1} \sum_{i=1}^n h_i h_i' = n^{-1} H_0' H_0$ and $V = (\alpha_2 - \alpha_1)^{-2} [n^{-1} \sum_{i=1}^n e_i^2 I(\hat{\eta}(\alpha_1) < e_i < \hat{\eta}(\alpha_2)) + \alpha_1 \hat{\eta}^2(\alpha_1) + (1 - \alpha_2) \hat{\eta}^2(\alpha_2) - (\alpha_1 \hat{\eta}(\alpha_1) + (1 - \alpha_2) \hat{\eta}(\alpha_2) + \lambda)^2] H \hat{Q}_h H'$, where $\lambda = n^{-1} \sum_{i=1}^n e_i I(\hat{\eta}(\alpha_1) < e_i < \hat{\eta}(\alpha_2))$.

Theorem 3.3. $V \rightarrow \sigma^2(\alpha_1, \alpha_2)$ 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 a $\ell \times p$ matrix of rank ℓ and we want to test $H_0 : K\beta = v$. Let m be the number of e_i removed by trimming. Then the rejection region will be $(K\hat{\beta}_s - v)'(KV^{-1}K')^{-1}(K\hat{\beta}_s - v) \geq d_u(\ell, n - m - p)$ with size approximately equal to 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 $1 - u$.

Next we consider the question of optimal ALWO estimation. For any two positive definite $p \times p$ matrices Q_1 and Q_2 , we say that Q_1 is smaller than or equal to Q_2 if $Q_2 - Q_1$ is positive semidefinite. An estimator is said to be the best in an estimator-class if it is in this class and its asymptotic covariance matrix is smaller than or equal to that of any estimator in this class. The following lemma implies that any ALWO estimator with asymptotic covariance matrix

$$\sigma^2(\alpha_1, \alpha_2) Q_x^{-1} \quad (3.1)$$

is a best estimator in this class.

Lemma 3.4. For any matrices \tilde{H} and Q_h induced from conditions (a1) and (a4), the difference $(\alpha_2 - \alpha_1)^2 \tilde{H} Q_h \tilde{H}' - Q_x^{-1}$ is positive semidefinite.

The trimmed mean proposed by Welsh (1987) is

$$\hat{\beta}_w = (X'AX)^{-1} X' y^* \quad (3.2)$$

so put $H = (X'AX)^{-1}$ and $H_0 = X$. From Welsh (1987) we have $n^{-1} X'AX \rightarrow (\alpha_2 - \alpha_1) Q_x$ so we can see that conditions (a1) and (a2) hold for $\hat{\beta}_w$, and Welsh's trimmed mean is an ALWO estimator. Moreover, Welsh (1987) proved that

$n^{1/2}(\hat{\beta}_w - (\beta + \gamma_w))$ has an asymptotic normal distribution with zero mean and covariance matrix of the form (3.1).

Theorem 3.5. *Under conditions (a1)-(a7), Welsh's trimmed mean $\hat{\beta}_w$ defined in (3.2) is a best ALWO estimator.*

For estimating the parametric function $c'\beta$, we have the following corollary to Theorem 3.1 and Corollary 3.2.

Corollary 3.6. *Under conditions (a1*)-(a2*) and (a3)-(a7),*

- (a) $n^{1/2}(a'y^* - (c'\beta + \gamma^*)) = n^{-1/2} \sum_{i=1}^n h'\tilde{h}_i\psi(\epsilon_i, F) + o_p(1)$, where $\gamma^* = \lambda h'\theta_h$.
- (b) *The normalized ALWO estimator $n^{1/2}(a'y^* - (c'\beta + \gamma^*))$ has an asymptotic normal distribution with zero mean and asymptotic variance $(\alpha_2 - \alpha_1)^2\sigma^2(\alpha_1, \alpha_2) h'Q_h h$.*

It follows from Theorem 3.5 that the inner product of c and Welsh's trimmed mean is also asymptotically best in the class of (asymptotically) linear functions of the Winsorized observation vector y^* .

Corollary 3.7. *Under the conditions of Corollary 3.6, a best ALWO estimator for estimating $c'\beta$ is $c'\hat{\beta}_w$, where $\hat{\beta}_w$ is Welsh's trimmed mean.*

In the class of linear estimators based on the Winsorized observation vector y^* , we have shown that for estimating the parameter vector β and the parametric function $c'\beta$, Welsh's trimmed mean and the inner product of c and Welsh's trimmed mean are both best ALWO estimators. This establishes the robust version of the Gauss-Markov theorem.

4. Particular Estimators

We noted in Section 2 that the class of ALWO estimators includes a subclass of instrumental variables estimators and the Mallows type bounded-influence trimmed means. In this section, we specialise the general results of Section 3 to these estimators and, where appropriate, discuss their implications.

The ALWO instrumental variables estimator is defined by $\hat{\beta}_s = (S'AX)^{-1}S'y^*$, where S is a matrix of instrumental variables. That is, S is a $n \times p$ matrix with i th row s'_i and i, j th element s_{ij} such that

- (b1) $n^{-1} \sum_{i=1}^n s_{ij}^4 = O(1)$ for all j ,
- (b2) $n^{-1}S'X = Q_{sx} + o(1)$, and $n^{-1}S'S = Q_s + o(1)$, where Q_s is a $p \times p$ positive definite matrix and Q_{sx} is a full rank matrix,
- (b3) $n^{-1} \sum_{i=1}^n s_i = \theta_s + o(1)$.

Our first result shows that the ALWO instrumental variables estimator is an ALWO estimator.

Lemma 4.1. *Under conditions (b1)-(b3), $n^{-1}S'AX$ converges in probability to the full rank matrix $(\alpha_2 - \alpha_1)^{-1}Q_{sx}$.*

This lemma implies that, with $H = (S'AX)^{-1}$ and $H_0 = S$ in (2.2), condition (a1) holds. One can also check that condition (a2) holds. Thus the ALWO instrumental variables estimator is an ALWO estimator.

The large sample properties of $\hat{\beta}_s$ follow immediately from Theorem 3.1 and Corollary 3.2. It can be shown that Welsh's trimmed mean is a best ALWO instrumental variables estimator. That is, it is optimal to use X rather than a matrix of instruments S .

For the class of Mallows-type bounded influence trimmed means $\hat{\beta}_{bi} = (X'WAX)^{-1} X'Wy^*$, we assume that the following additional assumption is valid.

(b4) $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n w_i x_i x_i' = Q_w$, $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n w_i^2 x_i x_i' = Q_{ww}$, where Q_w and Q_{ww} are $p \times p$ positive definite matrices.

De Jongh et al (1988) proved that $n^{1/2}(\hat{\beta}_{bi} - \beta)$ has an asymptotic normal distribution with zero mean vector and asymptotic covariance matrix $(\alpha_2 - \alpha_1)^2 \sigma^2 (\alpha_1, \alpha_2) Q_w^{-1} Q_{ww} Q_w^{-1}$. As Welsh's trimmed mean is a Mallows-type bounded influence trimmed mean ($W = I_n$), it follows that Welsh's trimmed mean is also the best Mallows-type bounded influence trimmed mean. This result is based solely on considerations of the asymptotic variance and ignores the fact that Welsh's trimmed mean does not have bounded influence in the space of independent variables. It confirms that bounded influence is achieved at the cost of efficiency.

5. Multivariate ALWO Estimators

Consider the classical multivariate regression model

$$Y = XB + V,$$

where Y is a $n \times m$ matrix of observations of m dependent variables, X is a known $n \times p$ design matrix with 1's in the first column, and V is a $n \times m$ matrix of independent and identically distributed disturbance random m -vectors. Let $\hat{B}_0 = (\hat{\beta}_1, \dots, \hat{\beta}_m)$ be an initial estimator of B with the property $n^{1/2}(\hat{\beta}_j - \beta_j) = O_p(1)$ for $j = 1, \dots, m$. The regression residuals are $e_{ij} = y_{ij} - x_i' \hat{\beta}_j$, $i = 1, \dots, n$ and $j = 1, \dots, m$, where y_{ij} is the (i, j) th element of matrix Y . For $0 < \alpha_{j1} < 0.5 < \alpha_{j2} < 1$, let $\hat{\eta}_j(\alpha_{j1})$ and $\hat{\eta}_j(\alpha_{j2})$ represent, respectively, the α_{j1} and α_{j2} th empirical quantiles of the regression residuals for the j th equation. Then the Winsorized observation vector for the j th equation is $y_j^* = (y_{1j}^*, \dots, y_{nj}^*)$ where

$$y_{ij}^* = y_{ij} I(\hat{\eta}_j(\alpha_{j1}) \leq e_{ij} \leq \hat{\eta}_j(\alpha_{j2})) + \hat{\eta}_j(\alpha_{j1}) (I(e_{ij} < \hat{\eta}_j(\alpha_{j1})) - \alpha_{j1}) \\ + \hat{\eta}_j(\alpha_{j2}) (I(e_{ij} > \hat{\eta}_j(\alpha_{j2})) - (1 - \alpha_{j2})).$$

Denote the j th trimming matrix by $A_j = \text{diag}(a_{j1}, \dots, a_{jn})$, where $a_{ji} = I(\hat{\eta}_j(\alpha_{j1}) \leq e_{ij} \leq \hat{\eta}_j(\alpha_{j2}))$. Estimation is defined for the parameter vector $\beta = (\beta'_1, \dots, \beta'_m)'$ with $B = (\beta_1, \dots, \beta_m)$.

Definition 5.1. A statistic $\hat{\beta}_{mlw}$ is called a multivariate ALWO estimator if there exists $p \times p$ matrices H_j , stochastic or nonstochastic, $j = 1, \dots, p$ and a $n \times p$ matrix H_0 which is independent of the error variables ϵ , such that

$$\hat{\beta}_{mlw} = \begin{bmatrix} H_1 H'_0 & 0 & \cdots & 0 \\ 0 & H_2 H'_0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & H_m H'_0 \end{bmatrix} \begin{bmatrix} y_1^* \\ \vdots \\ y_m^* \end{bmatrix},$$

where the matrices H_j and H_0 satisfy

- (c1) $nH_j \rightarrow (\alpha_{j2} - \alpha_{j1})^{-1} \tilde{H}_0$,
(c2) $H_j H'_0 X = (\alpha_{j2} - \alpha_{j1})^{-1} I_p + o_p(n^{-1/2})$.

Comparing the notation used in Definition 2.1, we replace \tilde{H} by $(\alpha_{j2} - \alpha_{j1})^{-1} \tilde{H}_0$ where \tilde{H}_0 is a constant matrix independent of j . Let \otimes represent the Kronecker product defined as $C \otimes B = (c_{ij} B)$ if matrix $C = (c_{ij})$. The following theorem follows from Theorem 3.1 and Corollary 3.2.

Theorem 5.2. Under conditions (c1)-(c2) and (a3)-(a7), we have

$$(a) \quad n^{1/2}(\hat{\beta}_{mlw} - (\beta + \gamma_{mlw})) = I_m \otimes \tilde{H}_0 n^{-1/2} \sum_{i=1}^n \begin{bmatrix} (\alpha_{12} - \alpha_{11})^{-1} \psi(\epsilon_{1i}, F_1) \\ (\alpha_{22} - \alpha_{21})^{-1} \psi(\epsilon_{2i}, F_2) \\ \vdots \\ (\alpha_{m2} - \alpha_{m1})^{-1} \psi(\epsilon_{mi}, F_m) \end{bmatrix} \otimes h_i + o_p(1),$$

where ϵ_{ij} , the (ij) th element of V , has distribution function F_j , and, with $\lambda_j =$

$$\int_{F_j^{-1}(\alpha_{j1})}^{F_j^{-1}(\alpha_{j2})} \epsilon dF_j(\epsilon), \quad \gamma_{mlw} = \begin{bmatrix} (\alpha_{12} - \alpha_{11})^{-1} \lambda_1 \\ \vdots \\ (\alpha_{m2} - \alpha_{m1})^{-1} \lambda_m \end{bmatrix} \otimes \tilde{H}_0 \theta_h.$$

- (b) $n^{1/2}(\hat{\beta}_{mlw} - (\beta + \gamma_{mlw}))$ has an asymptotic normal distribution with zero mean vector and asymptotic covariance matrix $\Sigma \otimes \tilde{H} Q_h \tilde{H}'$ where Σ is defined as the following matrix

$$\begin{bmatrix} \sigma_1^2(\alpha_{11}, \alpha_{12}) & \sigma_{12}(\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}) & \cdots & \sigma_{1m}(\alpha_{11}, \alpha_{12}, \alpha_{m1}, \alpha_{m2}) \\ \sigma_{21}(\alpha_{21}, \alpha_{22}, \alpha_{11}, \alpha_{12}) & \sigma_2^2(\alpha_{21}, \alpha_{22}) & \cdots & \sigma_{2m}(\alpha_{21}, \alpha_{22}, \alpha_{m1}, \alpha_{m2}) \\ \vdots & \vdots & & \vdots \\ \sigma_{m1}(\alpha_{m1}, \alpha_{m2}, \alpha_{11}, \alpha_{12}) & \sigma_{m2}(\alpha_{m1}, \alpha_{m2}, \alpha_{21}, \alpha_{22}) & \cdots & \sigma_m^2(\alpha_{m1}, \alpha_{m2}) \end{bmatrix}$$

with

$$\begin{aligned}\sigma_j^2(\alpha_{j1}, \alpha_{j2}) &= (\alpha_{j2} - \alpha_{j1})^{-2} \left[\int_{F_j^{-1}(\alpha_{j1})}^{F_j^{-1}(\alpha_{j2})} \epsilon^2 dF_j(\epsilon) + \alpha_{j1} (F_j^{-1}(\alpha_{j1}))^2 \right. \\ &\quad \left. + (1 - \alpha_{j2}) (F_j^{-1}(\alpha_{j2}))^2 - (\alpha_{j1} F_j^{-1}(\alpha_{j1}) \right. \\ &\quad \left. + (1 - \alpha_{j2}) F_j^{-1}(\alpha_{j2}) + \lambda_j)^2 \right],\end{aligned}$$

$$\begin{aligned}\sigma_{jk}(\alpha_{j1}, \alpha_{j2}, \alpha_{k1}, \alpha_{k2}) &= (\alpha_{j2} - \alpha_{j1})^{-1} (\alpha_{k2} - \alpha_{k1})^{-1} \left[\int_{F_k^{-1}(\alpha_{k1})}^{F_k^{-1}(\alpha_{k2})} \int_{F_j^{-1}(\alpha_{j1})}^{F_j^{-1}(\alpha_{j2})} \epsilon_j \epsilon_k dF_{jk} \right. \\ &\quad \left. + F_k^{-1}(\alpha_{k1}) \int_{-\infty}^{F_k^{-1}(\alpha_{k1})} \int_{F_j^{-1}(\alpha_{j1})}^{F_j^{-1}(\alpha_{j2})} \epsilon_j dF_{jk} \right. \\ &\quad \left. + F_k^{-1}(\alpha_{k2}) \int_{F_k^{-1}(\alpha_{k2})}^{\infty} \int_{F_j^{-1}(\alpha_{j1})}^{F_j^{-1}(\alpha_{j2})} \epsilon_j dF_{jk} \right. \\ &\quad \left. + F_j^{-1}(\alpha_{j1}) \int_{F_k^{-1}(\alpha_{k1})}^{F_k^{-1}(\alpha_{k2})} \int_{-\infty}^{F_j^{-1}(\alpha_{j1})} \epsilon_k dF_{jk} \right. \\ &\quad \left. + F_j^{-1}(\alpha_{j2}) \int_{F_k^{-1}(\alpha_{k1})}^{F_k^{-1}(\alpha_{k2})} \int_{F_j^{-1}(\alpha_{j2})}^{\infty} \epsilon_k dF_{jk} \right. \\ &\quad \left. + F_j^{-1}(\alpha_{j1}) F_k^{-1}(\alpha_{k1}) P(\epsilon_j < F_j^{-1}(\alpha_{j1}), \epsilon_k < F_k^{-1}(\alpha_{k1})) \right. \\ &\quad \left. + F_j^{-1}(\alpha_{j1}) F_k^{-1}(\alpha_{k2}) P(\epsilon_j < F_j^{-1}(\alpha_{j1}), \epsilon_k > F_k^{-1}(\alpha_{k2})) \right. \\ &\quad \left. + F_j^{-1}(\alpha_{j2}) F_k^{-1}(\alpha_{k1}) P(\epsilon_j > F_j^{-1}(\alpha_{j2}), \epsilon_k < F_k^{-1}(\alpha_{k1})) \right. \\ &\quad \left. + F_j^{-1}(\alpha_{j2}) F_k^{-1}(\alpha_{k2}) P(\epsilon_j > F_j^{-1}(\alpha_{j2}), \epsilon_k > F_k^{-1}(\alpha_{k2})) - ((1 - \alpha_{j2}) F_j^{-1}(\alpha_{j2}) \right. \\ &\quad \left. + \alpha_{j1} F_j^{-1}(\alpha_{j1}) + \lambda_j)((1 - \alpha_{k2}) F_k^{-1}(\alpha_{k2}) + \alpha_{k1} F_k^{-1}(\alpha_{k1}) + \lambda_k) \right],\end{aligned}$$

where F_{jk} represents the joint p.d.f. of variables ϵ_j and ϵ_k .

The multivariate trimmed mean generalized from Welsh (1987) is

$$\hat{\beta}_{mw} = \begin{bmatrix} (X' A_1 X)^{-1} & 0 & \cdots & 0 \\ 0 & (X' A_2 X)^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (X' A_m X)^{-1} \end{bmatrix} (I_m \otimes X') \begin{bmatrix} y_1^* \\ \vdots \\ y_m^* \end{bmatrix}.$$

It is obvious that $\hat{\beta}_{mw}$ is a multivariate ALWO estimator and it has an asymptotic normal distribution with zero mean and covariance matrix $\Sigma \otimes Q_x^{-1}$. From Lemma 3.3, we have the following.

Theorem 5.3. *The Welsh type multivariate trimmed mean is the best multivariate ALWO estimator.*

For large sample inference, we need to estimate the asymptotic covariance matrix of the multivariate Welsh's trimmed mean. We now exhibit an estimator of the matrix Σ . Let

$$\begin{aligned} v_{jk} &= (\alpha_{j2} - \alpha_{j1})^{-1}(\alpha_{k2} - \alpha_{k1})^{-1}n^{-1} \sum_{i=1}^n \left\{ [e_{ij}I(\hat{\eta}_j(\alpha_{j1}) \leq e_{ij} \right. \\ &\quad \leq \hat{\eta}_j(\alpha_{j2})) + \hat{\eta}_j(\alpha_{j1})I(e_{ij} < \hat{\eta}_j(\alpha_{j1})) + \hat{\eta}_j(\alpha_{j2})I(e_{ij} > \hat{\eta}_j(\alpha_{j2}))][e_{ik}I(\hat{\eta}_k(\alpha_{k1}) \\ &\quad \leq e_{ik} \leq \hat{\eta}_k(\alpha_{k2})) + \hat{\eta}_k(\alpha_{k1})I(e_{ik} < \hat{\eta}_k(\alpha_{k1})) + \hat{\eta}_k(\alpha_{k2})I(e_{ik} > \hat{\eta}_k(\alpha_{k2}))] \\ &\quad \left. - [\alpha_{j1}\hat{\eta}_j(\alpha_{j1}) + (1 - \alpha_{j2})\hat{\eta}_j(\alpha_{j2}) + \hat{\lambda}_j][\alpha_{k1}\hat{\eta}_k(\alpha_{k1}) + (1 - \alpha_{k2})\hat{\eta}_k(\alpha_{k2}) + \hat{\lambda}_k] \right\}, \end{aligned}$$

where $\hat{\lambda}_m = n^{-1} \sum_{i=1}^n e_{im}I(\hat{\eta}_m(\alpha_{m1}) \leq e_{im} \leq \hat{\eta}_m(\alpha_{m2}))$ for $m = j$ and k . Then an estimator of Σ is

$$\hat{\Sigma} = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \vdots & & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mm} \end{bmatrix}.$$

The multivariate ALWO estimator is not equivariant. In fact, the component-wise trimming used in its construction means that it cannot be made equivariant. Equivariance is an attractive mathematical property but is arguably of limited relevance in practice. The absence of equivariance simply means that we need to be careful about choosing a meaningful coordinate system for the data so that the components make sense. The above results (Theorem 5.2-5.3) apply to any fixed coordinate system. However, we may sometimes want to use a coordinate system which is estimated from the data. We therefore introduce a weighted multivariate ALWO estimator in which the weights are estimated from the data.

We denote the independent and identically distributed disturbance random m -vectors of V by $\bar{v}_i, i = 1, \dots, n$, i.e., $\bar{v}_i = (\epsilon_{1i}, \dots, \epsilon_{mi})'$. Let G be an estimator of a $m \times m$ dispersion matrix Ξ with the property that $n^{1/2}(G - \Xi) = O_p(1)$. Then let $\hat{B}_0^g = (\hat{\beta}_1^g, \dots, \hat{\beta}_m^g) = \hat{B}_0 G^{-1/2}$, where \hat{B}_0 is an initial estimator of B satisfying $n^{1/2}(\hat{B}_0 - B) = O_p(1)$. The transformed multivariate regression model is

$$Y^g = XB^g + V^g \tag{5.1}$$

with $Y^g = YG^{-1/2}$, $B^g = BG^{-1/2}$ and $V^g = VG^{-1/2}$. To construct Winsorized observations, consider the residuals of the transformed observations Y^g from the initial estimator \hat{B}_0^g , namely $e_{ij}^g = y_{ij}^g - x_i' \hat{\beta}_j^g, i = 1, \dots, n$ and $j = 1, \dots, m$, where y_{ij}^g is the (i, j) th element of matrix Y^g . For $0 < \alpha_{j1} < 0.5 < \alpha_{j2} < 1$, let $\hat{\eta}_j^g(\alpha_{j1})$ and $\hat{\eta}_j^g(\alpha_{j2})$ represent, respectively, the α_{j1} and α_{j2} th empirical quantiles of the

regression residuals $e_{ij}^g, i = 1, \dots, n$. Then the Winsorized observation vector for the j th transformed equation model (5.1) is $y_j^{g*} = (y_{1j}^{g*}, \dots, y_{nj}^{g*})$, where

$$y_{ij}^{g*} = y_{ij}^g I(\hat{\eta}_j^g(\alpha_{j1}) \leq e_{ij}^g \leq \hat{\eta}_j^g(\alpha_{j2})) + \hat{\eta}_j^g(\alpha_{j1}) \{I(e_{ij}^g < \hat{\eta}_j^g(\alpha_{j1})) - \alpha_{j1}\} \\ + \hat{\eta}_j^g(\alpha_{j2}) \{I(e_{ij}^g > \hat{\eta}_j^g(\alpha_{j2})) - (1 - \alpha_{j2})\}.$$

Denote the j th trimming matrix by $A_j = \text{diag}(a_{j1}, \dots, a_{jn})$, where $a_{ji} = I(\hat{\eta}_j^g(\alpha_{j1}) \leq e_{ij}^g \leq \hat{\eta}_j^g(\alpha_{j2}))$. Estimation is defined for the parameter vector $\beta = (\beta_1', \dots, \beta_m')'$ with $B = (\beta_1, \dots, \beta_m)$.

Definition 5.4. An estimator \hat{B}_{mlw} is called a weighted multivariate ALWO estimator if it satisfies $\hat{B}_{mlw} = \hat{B}_{mlw}^g G^{1/2'}$, where $\hat{B}_{mlw}^g = (\hat{\beta}_1^g, \hat{\beta}_2^g, \dots, \hat{\beta}_m^g)$, there are $p \times p$ stochastic or nonstochastic matrices $H_j, j = 1, \dots, p$, and a nonstochastic $n \times p$ matrix H_0 , such that $\hat{\beta}_{mlw}^g = (\hat{\beta}_1^{g'}, \hat{\beta}_2^{g'}, \dots, \hat{\beta}_m^{g'})'$ has the representation:

$$\hat{\beta}_{mlw}^g = \begin{bmatrix} H_1 H_0' & 0 & \cdots & 0 \\ 0 & H_2 H_0' & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & H_m H_0' \end{bmatrix} \begin{bmatrix} y_1^{g*} \\ \vdots \\ y_m^{g*} \end{bmatrix},$$

where the matrices H_j and H_0 satisfy

(c1) $nH_j \rightarrow (\alpha_{j2} - \alpha_{j1})^{-1} \tilde{H}_0$,

(c2) $H_j H_0' X = (\alpha_{j2} - \alpha_{j1})^{-1} I_p + o_p(n^{-1/2})$.

Denote by F^{ξ_j} the distribution function of $\bar{v}' \xi_j$ and

$$\psi_j(\bar{v}) = (\alpha_{j2} - \alpha_{j1})^{-1} [\bar{v}' \xi_j I(\eta_j^\xi(\alpha_{j1}) \leq \bar{v}' \xi_j \leq \eta_j^\xi(\alpha_{j2})) + \eta_j^\xi(\alpha_{j1}) I(\bar{v}' \xi_j \leq \eta_j^\xi(\alpha_{j1})) \\ + \eta_j^\xi(\alpha_{j2}) I(\bar{v}' \xi_j \geq \eta_j^\xi(\alpha_{j2})) - \{\lambda_j + \alpha_{j1} \eta_j^\xi(\alpha_{j1}) + (1 - \alpha_{j2}) \eta_j^\xi(\alpha_{j2})\}],$$

where ξ_j is the j th column of $\Xi^{-1/2}$, $\eta_j^\xi(\alpha)$ is the α th quantile of distribution F^{ξ_j} and $\lambda_j = \int_{\eta_j^\xi(\alpha_{j1})}^{\eta_j^\xi(\alpha_{j2})} \epsilon dF^{\xi_j}(\epsilon)$. For large sample analysis, we make the following assumptions.

(c3) There exists $\epsilon > 0$ such that p.d.f of $\bar{v}'(\xi_j + \delta)$ is uniformly bounded in a neighborhood of $\eta_j^\xi(\alpha)$ for $\|\delta\| \leq \epsilon$ and the p.d.f of $\bar{v}'(\xi_j + \delta) \bar{v}' u(\bar{v}' \xi_j)$ is uniformly away from zero for $\|u\| = 1$ and $\|\delta\| \leq \epsilon$.

(c4) $E((\bar{v}' \xi_j)^2 \|\bar{v}\|) < \infty$.

Our main result is the following theorem.

Theorem 5.5. Under conditions (c1)-(c4) and (a3)-(a7), we have

(a) $n^{1/2}(\hat{\beta}_{mlw} - (\beta + \gamma_{mlw})) = I_m \otimes \tilde{H}_0 n^{-1/2} \sum_{i=1}^n \begin{bmatrix} (\psi_1(\bar{v}_i), \psi_2(\bar{v}_i) \cdots \psi_m(\bar{v}_i)) \xi_1^* \\ \vdots \\ (\psi_1(\bar{v}_i), \psi_2(\bar{v}_i) \psi_m(\bar{v}_i)) \xi_m^* \end{bmatrix} \otimes$

$$h_i + o_p(1), \text{ where } \xi_j^* \text{ is the } j\text{th column of } \Xi^{1/2'} \text{ and } \gamma_{mlw} = \begin{bmatrix} (\gamma_1 \cdots \gamma_m) \xi_1^* \\ \vdots \\ (\gamma_1 \cdots \gamma_m) \xi_m^* \end{bmatrix}$$

$$\text{with } \gamma_j = \lambda_j \tilde{H}_0 \theta_h \text{ and } \lambda_j = \int_{\eta_j^\xi(\alpha_{j1})}^{\eta_j^\xi(\alpha_{j2})} \epsilon dF^{\xi_j}(\epsilon).$$

- (b) $n^{1/2}(\hat{\beta}_{mlw} - (\beta + \gamma_{mlw}))$ has an asymptotic normal distribution with zero mean vector and asymptotic covariance matrix $\Sigma \otimes \tilde{H}_0 Q_h \tilde{H}_0'$ where

$$\Sigma = \text{cov} \left(\begin{bmatrix} (\psi_1(\bar{v}), \psi_2(\bar{v}) \cdots \psi_m(\bar{v})) \xi_1^* \\ \vdots \\ (\psi_1(\bar{v}), \psi_2(\bar{v}) \cdots \psi_m(\bar{v})) \xi_m^* \end{bmatrix} \right).$$

The weighted multivariate trimmed mean generalized from Welsh (1987) is

$$\hat{\beta}_{mw} = \begin{bmatrix} (X' A_1 X)^{-1} & 0 & \cdots & 0 \\ 0 & (X' A_2 X)^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (X' A_m X)^{-1} \end{bmatrix} (I_m \otimes X') \begin{bmatrix} y_1^{q*} \\ \vdots \\ y_m^{q*} \end{bmatrix}.$$

It is obvious that $\hat{\beta}_{mw}$ is a weighted multivariate ALWO estimator and it has an asymptotic normal distribution with zero mean and covariance matrix $\Sigma \otimes Q_x^{-1}$. From Lemma 3.3, we have the following.

Theorem 5.6. *The Welsh type weighted multivariate trimmed mean is the best multivariate ALWO estimator.*

Consider the special design $\alpha_{j1} = 0$ and $\alpha_{j2} = 1$ for $j = 1, \dots, m$ and Ξ is the covariance matrix $\text{cov}(\bar{v})$. Then the asymptotic covariance matrix of $\hat{\beta}_{mlw}$ is $\Sigma = \Xi \otimes \tilde{H} Q_h \tilde{H}'$ while the asymptotic covariance matrix of the least squares estimator is $\Xi \otimes Q_x^{-1}$.

6. Examples

Before we can use the ALWO estimators such as Welsh's trimmed mean and the multivariate generalization proposed in Section 5, we need to specify the initial estimator and the trimming proportions. The simplest initial estimator is the least squares estimator. To improve robustness in small samples, it may be better to use a robust initial estimator such as the ℓ_1 estimator (see Koenker's discussion of Welsh (1987)). Other robust estimators can also be considered. Similarly, in the multivariate case, if we choose to use a data determined coordinate system, the simplest dispersion estimator G is the sample variance of the residuals but robustness considerations may lead us to consider using a robust

dispersion estimator. The simplest way to choose the trimming proportions is to specify them in advance. The use of 10% trimming in both tails is widely recommended (see for example Ruppert and Carroll (1980)). On the other hand, the trimming proportions can be determined adaptively from the data (see for example Welsh (1987), Jureckova, Koenker and Welsh, (1994) and references therein). It appears to be largely a philosophical question as to which approach individual users prefer.

The choice of initial estimator and the method of choosing the trimming proportions impact on the computation of the estimators. Given an initial estimator and given trimming proportions, the calculation is straightforward. First, the componentwise residuals are sorted, then the Winsorized observation vectors y_j^* for $j = 1, \dots, m$ are constructed and, finally, the estimator is computed from its explicit definition by elementary matrix operations. Thus, the extent of the computational burden depends on the burden involved in calculating the initial estimator and the trimming proportions. The least squares and ℓ_1 estimators are readily computed but other robust estimators may be computationally more burdensome. Similarly, some choices of G may increase the computational burden. Adaptive methods for choosing the trimming proportions require the estimator to be computed over a number of trimming proportions. In practice, this is usually done by fixing a grid of possible trimming proportions. While this does increase the computational burden, it is generally by only a small amount.

Lobster Catch Data

In this section, the trimmed mean methods proposed by Koenker and Bassett (1978) and Welsh (1987) are applied to analyze a data set which consists of $n = 44$ observations on the American lobster resource displayed in Morrison (1983). In this data set, the response is the annual catch (in metric tons) of lobsters (y) and the independent variables (predictors) expected to affect the response include: the number of registered lobster traps (X_1), the number of fishermen (X_2), the mean annual sea temperature (X_3) and the year (T). From economic theory, we anticipate the mean regression function of y to be nondecreasing in the variables X_1 and X_2 . We also expect the mean regression function to be nondecreasing in X_3 .

Morrison (1983) studied the relationship between y and the above predictors by using a linear regression model including X_1, X_2, X_3 and polynomials in T with degree up to 4. Unfortunately, the estimate of the coefficient for the variable X_1 (lobster trap) was negative, which violates economic theory.

To perform the analyses using trimmed mean methods, we first identify the appropriate regression function. To achieve this goal, we fit the naive multiple regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon. \quad (6.1)$$

From the experience of Morrison (1983), we would not expect this model to fit the data. However, the residual plot provides insight into the data set and is useful for building a realistic model. The residual plot for the ℓ_1 -norm fit for model (6.1) is displayed in Figure 1.

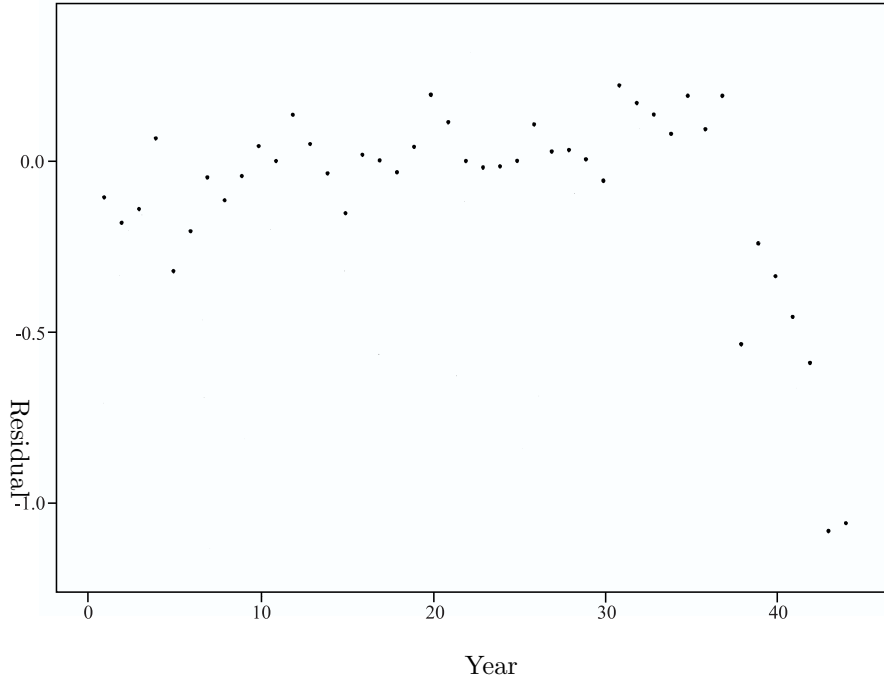


Figure 1. Residual plot based on ℓ_1 -norm for model (6.1).

Figure 1 suggests evidence of a structural change that invalidates using a single regression equation to represent the data. An alternative model for fitting data with structural changes is obtained by adding dummy functions in time to model (6.1). By inspection, we select knots at $t = 10, 26$ and 38 because the residuals falling in regions $\{1, \dots, 9\}$, $\{26, \dots, 37\}$ and $\{38, \dots, 44\}$ are all, or almost all, of the same sign. We then consider the following regression model including dummy functions in t

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + (\beta_4 + \beta_7 t)I(t \geq 10) + (\beta_5 + \beta_8 t)I(t \geq 26) + (\beta_6 + \beta_9 t)I(t \geq 38) + \epsilon. \quad (6.2)$$

Two types of trimmed mean will be used to estimate the regression parameters of (6.2). Let $z'_i = (1, x_{1i}, x_{2i}, x_{3i}, I(t_i \geq 10), I(t_i \geq 26), I(t_i \geq 38), t_i I(t_i \geq 10), t_i I(t_i \geq 26), t_i I(t_i \geq 38))$.

The first approach proposed by Koenker and Bassett (1978) is based on regression quantiles. The regression quantile process $\hat{\beta}(\alpha)$, $0 < \alpha < 1$, is defined to be a solution to $\min_{b \in R^p} \sum_{i=1}^n \rho_\alpha(y_i - z_i' b)$ where $\rho_\alpha(u) = u(\alpha - I(u < 0))$. The trimmed mean based on regression quantiles is $\hat{\beta}_{KB} = (Z'W_\alpha Z)^{-1} Z'W_\alpha y$ where $W_\alpha = \text{diag}(w_1, \dots, w_n)$, $w_i = I(z_i' \hat{\beta}(\alpha) < y_i < z_i' \hat{\beta}(1 - \alpha))$ and matrix Z is the $n \times 10$ matrix with rows z_i' .

We list the estimates associated with some trimming proportions α in the following table.

Table 1. Koenker and Bassett's trimmed mean $\hat{\beta}_{KB}$.

α	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_9
.05	-1.36	.05	.92	.75	.08	.23	4.94	.01	-.13
.10	-1.40	.14	.88	.69	.14	.08	5.04	.01	-.13
.15	-2.82	.03	1.11	.82	.13	.05	5.59	.00	-.15
.20	-2.28	.18	0.94	.80	.25	-.13	5.70	.00	-.15
.25	-.80	.10	.82	.80	.17	.28	4.98	.00	-.13
.30	-.47	.10	0.78	.73	.28	.22	2.57	.00	-.07
.35	-3.32	.09	1.14	.77	.23	-.22	.00	.00	.00

* Estimates of β_8 are all zeros.

Basically, the estimates of β_1, β_2 and β_3 all have the right signs.

We use the ℓ_1 -estimate $\hat{\beta}_{\ell_1} \equiv \hat{\beta}(0.5)$ of (6.2) as the initial estimate for Welsh's trimmed mean. The residuals based on $\hat{\beta}_{\ell_1}$ are $e_i = y_i - z_i' \hat{\beta}_{\ell_1}$, $i = 1, \dots, n$. Let A be the trimming matrix defined in Section 2 based on residuals e_i .

Since trimmed means based on initial estimates are able to trim an arbitrary number of observations, here we select trimming proportions α so that the numbers of trimmed observations are $1, \dots, 10$. Table 2 gives the estimates $\hat{\beta}_w$.

As the true parameters are unknown, we are not able to compare the efficiencies of these estimates. However, comparison of these two tables gives the following conclusions.

- The estimates of the parameters β_1, β_2 and β_3 for the least squares, ℓ_1 -norm and both trimmed mean methods have the right signs. This means that the model at (6.2) improves on the model adopted by Morrison (1983).
- The estimates $\hat{\beta}_{KB}$ show fluctuation in trimming percentage and number of observations, respectively, without forming a convergent sequence. This makes it difficult to determine the trimming percentage or number. On the other hand, Table 2 shows that the trimmed mean $\hat{\beta}_w$ is relatively stable when the number of trimming observations increases. Welsh's trimmed mean performed quite robustly for this data set. Given that only a small number of outliers showed in Figure 1, Welsh's trimmed mean (in Table 2) with three

observations removed seems to be an appropriate estimate for the regression parameters of model (6.2).

Table 2. Welsh's trimmed mean $\hat{\beta}_w$.

Trim. no.	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_8	β_9
1	-1.30	.13	.85	.80	.17	.22	5.13	-.00	-.13
2	-1.30	.13	.86	.79	.16	.23	5.13	-.00	-.13
3	-1.64	.13	.90	.77	.13	.22	5.33	-.00	-.14
4	-1.47	.12	.89	.76	.14	.23	5.20	-.00	-.14
5	-1.49	.12	.89	.79	.15	.35	5.03	-.01	-.13
6	-2.08	.11	.97	.78	.09	.29	5.33	-.01	-.14
7	-2.00	.10	.96	.78	.11	.26	5.28	-.00	-.14
8	-1.54	.09	.91	.77	.13	.28	5.01	-.01	-.13
9	-1.64	.10	.92	.77	.15	.25	5.11	-.00	-.14
10	-1.95	.09	.97	.73	.21	.01	5.45	-.00	-.14

* Trim. no. is the trimming number of observations; estimates of β_7 are 0.1 for Trim. no. 1 - 9 and 0.0 for Trim. no. 10.

Mineral Content in Bones

Johnson and Wichern (1982, p.34) give data on the mineral content of the arm bones of 25 subjects and suggest the use of multivariate regression modelling to analyse the relationship between mineral content in the dominant radius (y_1) and the remaining radius (y_2), and the mineral content of the other four bones: the dominant humerus (x_1), the remaining humerus (x_2), the dominant ulna (x_3), and the remaining ulna (x_4).

Since the data consist of measurements of the same quantity (mineral content), it makes sense to keep them on the same scale. The coordinate system in which the data are presented is natural and meaningful so we will work with it. The scatterplot matrix of the data shows subjects 1 and 23 as slightly unusual by virtue of having a high mineral content in the humerus given the mineral content in the dominant humerus, but otherwise provides no evidence that a transformation is required. We therefore consider the bivariate regression model

$$(y_1 \ y_2) = (1 \ x_1 \ x_2 \ x_3 \ x_4) \begin{pmatrix} \beta_0 & \beta_0^* \\ \beta_{11} & \beta_{21} \\ \beta_{12} & \beta_{22} \\ \beta_{13} & \beta_{23} \\ \beta_{14} & \beta_{24} \end{pmatrix} + (\epsilon_1, \epsilon_2),$$

which has all the variables on the raw scale. The residual plot for the residuals from the ℓ_1 fit to the data for the dominant radius (Figure 2) shows some mild curvature and several potential outliers. There seems to be less curvature in the residual plot for the radius data (Figure 3) and more homogeneous variation, making it more difficult to determine whether outliers are present or not. The suggestion of curvature is not reduced by transforming all variables to the log scale so we retain the raw scale for simplicity. Normal quantile plots of the residuals show that the marginal distributions of the residuals have long tails. The marginal distribution of the residuals from the fit to the data for the dominant radius has a long lower tail consisting of subjects 23, 17, 25, and 14, and two mild outliers in the upper tail from subjects 1 and 19.

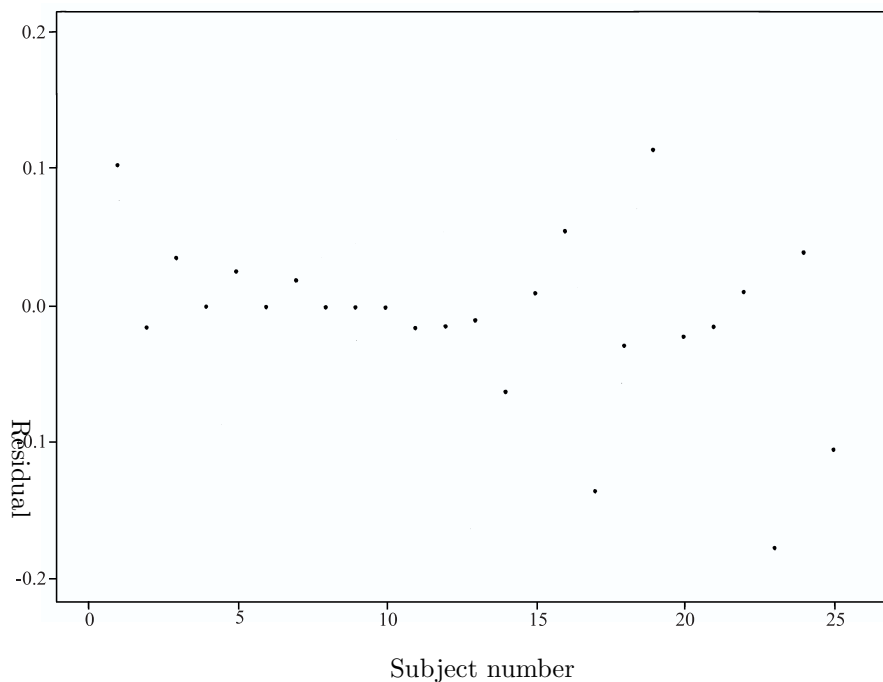


Figure 2. Residual plot based on ℓ_1 -norm for equation model.

The marginal distribution of the residuals from the fit to the data for the radius has two long tails rather than distinct outliers.

In Table 3, we give estimates of the β 's obtained using least squares (LS), some of Welsh's trimmed means with different numbers of observations Winsorized, and the ℓ_1 -norm (ℓ_1).

Notice that, apart from a small increase at $k = 2, 3$, and 5 with $j = 2$, the variance decreases as k , the number of observations trimmed, increases. This suggests that the distributions are long-tailed and that relatively severe trimming is required.

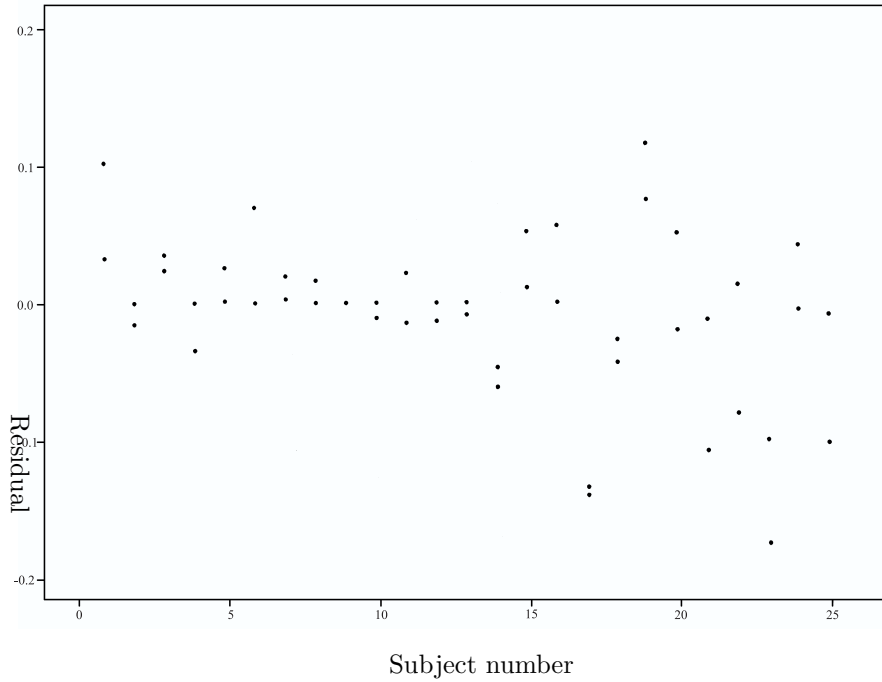
Figure 3. Residual plot based on ℓ_1 -norm for second equation model.

Table 3. Estimates by least squares, Welsh's trimmed mean.

Estimate	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}
LS	.0995	.2208	-.0877	.3605	.3564
$\hat{\beta}_{mlw}(1)$.1177	.2091	-.0832	.3547	.3568
$\hat{\beta}_{mlw}(2)$.1386	.2269	-.1112	.3384	.3660
$\hat{\beta}_{mlw}(3)$.1882	.1818	-.0819	.2918	.3905
$\hat{\beta}_{mlw}(4)$.1865	.1434	-.0670	.2728	.4789
$\hat{\beta}_{mlw}(5)$.1571	.1572	-.0912	.2829	.5356
ℓ_1	.1287	.1806	-.1328	.3244	.5742

Estimate	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}
LS	.1263	-.0154	.1561	.1940	.4486
$\hat{\beta}_{mlw}(1)$.1162	-.0149	.1617	.1976	.4454
$\hat{\beta}_{mlw}(2)$.1176	-.0156	.1654	.2121	.4250
$\hat{\beta}_{mlw}(3)$.1255	-.0012	.1564	.1788	.4351
$\hat{\beta}_{mlw}(4)$.1572	-.0371	.2026	.0521	.4959
$\hat{\beta}_{mlw}(5)$.1634	-.0427	.2102	.0276	.5077
ℓ_1	.1476	-.0368	.2103	-.0314	.5815

ps: $\hat{\beta}_{mlw}(k)$ represents the Welsh's trimmed mean with number k of Winsorized observations.

Table 4. Estimates of $\hat{\sigma}_j^2(k/n, 1 - k/n)$ and $\hat{\sigma}_{12}(k/n, 1 - k/n)$.

j	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
$j=1$.00612	.00609	.00444	.00296	.00179	.00137
$j=2$.00416	.00439	.00471	.00466	.00253	.00298
$\hat{\sigma}_{12}$.00299	.00289	.00243	.00177	.00079	.00055

7. Appendix

Proof of Theorem 3.1. From condition (a2) and (A.10) of Ruppert and Carroll (1980), $HH'_0AX\beta = \beta + o_p(n^{-1/2})$. Inserting (2.1) in equation (2.3), we have

$$\begin{aligned} n^{1/2}(\hat{\beta}_{lw} - \beta) &= n^{1/2}H[H'_0A\epsilon - \hat{\eta}(\alpha_1) \sum_{i=1}^n h_i(\alpha_1 - I(e_i \leq \hat{\eta}(\alpha_1))) \\ &\quad + \hat{\eta}(\alpha_2) \sum_{i=1}^n h_i(\alpha_2 - I(e_i \leq \hat{\eta}(\alpha_2)))] + o_p(1). \end{aligned}$$

Now we develop a representation of $n^{-1/2}H'_0A\epsilon$. 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 $U(\alpha, T_n) = (U_1(\alpha, T_n), \dots, U_p(\alpha, T_n))$. Also, let $T_n^*(\alpha) = n^{1/2} \left[\hat{\beta}_0 + \begin{bmatrix} \hat{\eta}(\alpha) \\ 0_{p-1} \end{bmatrix} - \left(\beta + \begin{bmatrix} F^{-1}(\alpha) \\ 0_{p-1} \end{bmatrix} \right) \right]$. Then $n^{-1/2}H'_0A\epsilon = U(\alpha_2, T_n^*(\alpha_2)) - U(\alpha_1, T_n^*(\alpha_1))$. From Jureckova and Sen's (1987) extension of Billingsley's Theorem (see also Koul (1992)), we have

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

for $j = 1, \dots, p$ and $T_n = O_p(1)$. From (7.1),

$$\begin{aligned} &n^{-1/2}H'_0A\epsilon \\ &= (U(\alpha_2, T_n^*(\alpha_2)) - U(\alpha_2, 0)) - (U(\alpha_1, T_n^*(\alpha_1)) - U(\alpha_1, 0)) + (U(\alpha_2, 0) - U(\alpha_1, 0)) \\ &= n^{-1/2} \sum_{i=1}^n h_i \epsilon_i I(F^{-1}(\alpha_1) \leq \epsilon_i \leq F^{-1}(\alpha_2)) + F^{-1}(\alpha_2)f(F^{-1}(\alpha_2))Q_{hx}T_n^*(\alpha_2) \\ &\quad + F^{-1}(\alpha_1)f(F^{-1}(\alpha_1))Q_{hx}T_n^*(\alpha_1) + o_p(1). \end{aligned} \quad (7.2)$$

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

$$\begin{aligned} n^{-1/2}H'_0A\epsilon &= n^{-1/2} \sum_{i=1}^n h_i \epsilon_i I(F^{-1}(\alpha_1) \leq \epsilon_i \leq F^{-1}(\alpha_2)) + (F^{-1}(\alpha_2)f(F^{-1}(\alpha_2)) \\ &\quad - F^{-1}(\alpha_1)f(F^{-1}(\alpha_1)))Q_{hx}n^{1/2}(-I_p + I^*)(\hat{\beta}_0 - \beta) \end{aligned}$$

$$\begin{aligned}
& + [F^{-1}(\alpha_2)n^{-1/2} \sum_{i=1}^n (\alpha_2 - I(\epsilon_i \leq F^{-1}(\alpha_2))) \\
& - F^{-1}(\alpha_1)n^{-1/2} \sum_{i=1}^n (\alpha_1 - I(\epsilon_i \leq F^{-1}(\alpha_1)))] Q_{hx} + o_p(1), \quad (7.3)
\end{aligned}$$

where I^* is a $p \times p$ diagonal matrix with the first diagonal element equal to 1. Similarly, we also have, for $0 < \alpha < 1$,

$$\begin{aligned}
& \hat{\eta}(\alpha)n^{-1/2} \sum_{i=1}^n h_i(\alpha - I(e_i \leq \hat{\eta}(\alpha))) \\
& = F^{-1}(\alpha)[f(F^{-1}(\alpha))Q_{hx}n^{1/2}(-I_p + I^*)(\hat{\beta}_0 - \beta) \\
& - Q_{hx}n^{-1/2} \sum_{i=1}^n (\alpha - I(\epsilon_i \leq F^{-1}(\alpha)))\delta_0 \\
& + n^{-1/2} \sum_{i=1}^n h_i(\alpha - I(\epsilon_i \leq F^{-1}(\alpha)))] + o_p(1), \quad (7.4)
\end{aligned}$$

where δ_0 is p -vector with first element equal to 1 and the remaining elements equal to 0. Combining (7.3) and (7.4) for $\alpha = \alpha_1$ and α_2 ,

$$\begin{aligned}
& n^{-1/2}[H'_0 A \epsilon - \hat{\eta}(\alpha_1)n^{-1/2} \sum_{i=1}^n h_i(\alpha_1 - I(e_i \leq \hat{\eta}(\alpha_1))) \\
& + \hat{\eta}(\alpha_2)n^{-1/2} \sum_{i=1}^n h_i(\alpha_2 - I(e_i \leq \hat{\eta}(\alpha_1)))] \\
& = n^{-1/2} \sum_{i=1}^n h_i[\epsilon_i I(F^{-1}(\alpha_1) \leq \epsilon_i \leq F^{-1}(\alpha_2)) + F^{-1}(\alpha_1)I(\epsilon_i \leq F^{-1}(\alpha_1)) \\
& + F^{-1}(\alpha_2)I(\epsilon_i \geq F^{-1}(\alpha_2)) - (\alpha_1 F^{-1}(\alpha_1) + (1 - \alpha_2)F^{-1}(\alpha_2))] + o_p(1). \quad (7.5)
\end{aligned}$$

The theorem is then obtained from (7.5) and Condition (a1).

Proof of Theorem 3.3. From the representation of sample quantiles in Ruppert and Carroll (1980) and the linear Winsorized instrumental variables mean $\hat{\beta}_s$, $\hat{\eta}(a) \rightarrow F^{-1}(a)$ in probability for $a = \alpha_1$ and α_2 . Now,

$$\begin{aligned}
& n^{-1} \sum_{i=1}^n e_i^2 I(\hat{\eta}(\alpha_1) < e_i < \hat{\eta}(\alpha_2)) \\
& = n^{-1}(\hat{\beta}_0 - \beta)' \sum_{i=1}^n x_i x_i' (\hat{\beta}_0 - \beta) I(\hat{\eta}(\alpha_1) < e_i < \hat{\eta}(\alpha_2)) \\
& + n^{-1} \sum_{i=1}^n \epsilon_i^2 I(\hat{\eta}(\alpha_1) < e_i < \hat{\eta}(\alpha_2)) + n^{-1}(\hat{\beta}_0 - \beta)' \sum_{i=1}^n x_i \epsilon_i I(\hat{\eta}(\alpha_1) < e_i < \hat{\eta}(\alpha_2)).
\end{aligned}$$

From the fact that $n^{1/2}(\hat{\beta}_s - \beta) = O_p(1)$, $n^{-1} \sum_{i=1}^n x_i \epsilon_i I(\hat{\eta}(\alpha_1) < e_i < \hat{\eta}(\alpha_2)) = o_p(1)$ and $n^{-1} \sum_{i=1}^n \epsilon_i^2 I(\hat{\eta}(\alpha_1) < e_i < \hat{\eta}(\alpha_2)) = n^{-1} \sum_{i=1}^n \epsilon_i^2 I(F^{-1}(\alpha_1) < \epsilon_i < F^{-1}(\alpha_2)) + o_p(1)$, where the last equation follows from Lemma A.4 of Ruppert and Carroll (1980). Analogous discussion shows that $\hat{\lambda}$ is consistent for λ . Then these results imply the theorem.

Proof of Lemma 3.4. Write $\text{plim}(B_n) = B$ if B_n converges to B in probability. Let $C = HH_0 - (X'AX)^{-1}X'$. Now $\text{plim}(CAX) = \text{plim}(HH_0'AX) - \text{plim}(X'AX)^{-1}X'AX = 0$. Hence

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

Proof of Corollary 3.7. It is obvious that $(\alpha_2 - \alpha_1)^2 h'Q_h h = \text{plim}(\alpha_2 - \alpha_1)^{-2} na'a$. The best linear Winsorized mean for $c'\beta$ satisfies $\min \text{plim}n(\alpha_2 - \alpha_1)^{-2} a'a$ subject to $c = \text{plim}(\alpha_2 - \alpha_1)X'a$. Equivalently, solve $\min \text{plim}L(a, \lambda) = (\alpha_2 - \alpha_1)^{-2} na'a + \lambda(c - (\alpha_2 - \alpha_1)X'a)$. Taking the partial derivative of $L(a, \lambda)$ with respect to a and λ , we have $a = (2n)^{-1}(\alpha_2 - \alpha_1)^3 X\lambda$ subject to $C = (\alpha_2 - \alpha_1)X'a$. Thus $a = (\alpha_2 - \alpha_1)^{-1} X(X'X)^{-1}c$. This we can estimate by $a = X(X'AX)^{-1}c$, and the best linear Winsorized mean for $c'\beta$ is $c'(X'AX)^{-1}X'y^* \equiv c'\hat{\beta}_{lw}$.

Proof of Lemma 4.1. Using the Jureckova and Sen (1987) extension of Billingsley's Theorem, we have $n^{-1} \sum_{i=1}^n s_{ij}x_{ik}I(\hat{\eta}(\alpha_1) < \epsilon_i < \hat{\eta}(\alpha_2)) = (\alpha_2 - \alpha_1)q_{jk} + o_p(1)$ where q_{jk} is the jk th term of the matrix Q_{sx} , and x_{ij} , s_{ik} are the ij th and ik th terms of X and S , respectively. We then have $n^{-1}S'AX = (\alpha_2 - \alpha_1)Q_{sx} + o_p(1)$.

Proof of Theorem 5.5. The vector $\hat{\beta}_{mlw}^g$ is a vertical stacking of $\hat{\beta}_j^g = H_j H_0' y_j^{g*}$. We have

$$\begin{aligned} \hat{\beta}_j^g &= H_j H_0' A_j X B g_j + H_j H_0' A_j V g_j - H_j \hat{\eta}_j^g(\alpha_1) \sum_{i=1}^n h_i \{\alpha_1 - I(e_{ij}^g \leq \hat{\eta}_j^g(\alpha_1))\} \\ &\quad + H_j \hat{\eta}_j^g(\alpha_2) \sum_{i=1}^n h_i \{\alpha_2 - I(e_{ij}^g \leq \hat{\eta}_j^g(\alpha_2))\}, \end{aligned} \quad (7.6)$$

where g_j is the j th column of $G^{-1/2}$. The proof is the same for each j so we drop j to simplify the notation.

We first derive the representation of $n^{-1/2}H'_0AVg$. Since $n^{1/2}(g-\xi) = O_p(1)$, we need to consider only the term $n^{-1/2}H'_0AV\xi$. For $\ell = 1, \dots, p$, let

$$S_\ell(b, g) = n^{-1/2} \sum_{i=1}^n h_{i\ell} \bar{v}'_i \xi [I\{\bar{v}'_i(\xi + n^{-1/2}g) \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b\} - I\{\bar{v}'_i \xi \leq \eta^\xi(\alpha)\}].$$

We want to show that

$$\begin{aligned} & \sup_{\|b\| \leq k, \|g\| \leq k'} |S_\ell(b, g) - \eta^\xi(\alpha) f_\xi(\eta^\xi(\alpha)) n^{-1} \sum_{i=1}^n h_{i\ell} \{d'_i b + g' E(\bar{v} | \bar{v}' \xi = \eta^\xi(\alpha))\}| \\ &= o_p(1), \end{aligned} \quad (7.7)$$

where f_ξ is the density of $\bar{v}' \xi$, to obtain a representation for H'_0AVg in (7.6).

To establish (7.7), we first show that

$$\begin{aligned} & n^{-1} \sum_{i=1}^n h_{i\ell}^2 E[(\bar{v}'_i \xi)^2 | I\{\bar{v}'_i(\xi + n^{-1/2}g_1) \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b_1\} - I\{\bar{v}'_i(\xi + n^{-1/2}g_2) \\ & \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b_2\}] \leq M(\|b_2 - b_1\| + \|g_2 - g_1\|) \end{aligned} \quad (7.8)$$

for some $M > 0$. Let $A = n^{-1} \sum_{i=1}^n h_{i\ell}^2 E[(\bar{v}'_i \xi)^2 | I\{\bar{v}'_i(\xi + n^{-1/2}g_1) \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b_1\} - I\{\bar{v}'_i(\xi + n^{-1/2}g_2) \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b_1\}]$ and $B = n^{-1} \sum_{i=1}^n h_{i\ell}^2 E[(\bar{v}'_i \xi)^2 | I\{\bar{v}'_i(\xi + n^{-1/2}g_2) \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b_1\} - I\{\bar{v}'_i(\xi + n^{-1/2}g_2) \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b_2\}]$. Note that (7.8) is bounded by $A+B$. We can decompose A as

$$\begin{aligned} A &= n^{-1} \sum_{i=1}^n h_{i\ell}^2 E[(\bar{v}'_i \xi)^2 | I\{\bar{v}'_i(\xi + n^{-1/2}g_1) \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b_1, \bar{v}'_i(\xi + n^{-1/2}g_2) \\ & > \eta^\xi(\alpha) + n^{-1/2}d'_i b_1\}] + n^{-1} \sum_{i=1}^n h_{i\ell}^2 E[(\bar{v}'_i \xi)^2 | I\{\bar{v}'_i(\xi + n^{-1/2}g_1) > \eta^\xi(\alpha) + n^{-1/2}d'_i b_1, \\ & \bar{v}'_i(\xi + n^{-1/2}g_2) \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b_1\}] = A_1 + A_2. \end{aligned}$$

Consider A_1 . Suppose that $g_1 \neq g_2$ and let $U_1 = \bar{v}'(\xi + n^{-1/2}g_1)$, $U_2 = \bar{v}' \frac{(g_2 - g_1)}{\|g_2 - g_1\|}$, and $U_3 = \bar{v}' \xi$. Then, using the conditional expectation $E(H(U_1, U_2, U_3)) = E(E(H(U_1, U_2, U_3) | U_2, U_3))$, $A_1 = n^{-1} \sum_{i=1}^n h_{i\ell}^2 E\{U_3^2 f_{U_1|U_2, U_3}(\eta^\xi(\alpha) + n^{-1/2}d'_i b_1) U_2\} n^{-1/2} \|g_2 - g_1\| \leq Mn^{-1/2} \|g_2 - g_1\|$. Similarly, we have $A_2 \leq Mn^{-1/2} \|g_2 - g_1\|$ and $B \leq Mn^{-1/2} \|b_2 - b_1\|$ so (7.8) holds.

Next, we consider

$$\begin{aligned} & n^{-1} \sum_{i=1}^n h_{i\ell}^2 E[(\bar{v}'_i \xi)^2 \sup_{\|g_1 - g\| + \|b_1 - b\| \leq k} |I\{\bar{v}'_i(\xi + n^{-1/2}g_1) \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b_1\} \\ & - I\{\bar{v}'_i(\xi + n^{-1/2}g) \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b\}|]. \end{aligned} \quad (7.9)$$

The expression (7.9) is bounded by $C_1 + C_2 + D$ with

$$\begin{aligned}
C_1 &= n^{-1} \sum_{i=1}^n h_{i\ell}^2 E[(\bar{v}'_i \xi)^2 \sup_{\|g_1 - g\| + \|b_1 - b\| \leq k} I\{\bar{v}'_i(\xi + n^{-1/2} g_1) \leq \eta^\xi(\alpha) + n^{-1/2} d'_i b_1, \\
&\quad \bar{v}'_i(\xi + n^{-1/2} g) > \eta^\xi(\alpha) + n^{-1/2} d'_i b_1\}]; \\
C_2 &= n^{-1} \sum_{i=1}^n h_{i\ell}^2 E[(\bar{v}'_i \xi)^2 \sup_{\|g_1 - g\| + \|b_1 - b\| \leq k} I\{\bar{v}'_i(\xi + n^{-1/2} g_1) > \eta^\xi(\alpha) + n^{-1/2} d'_i b_1, \\
&\quad \bar{v}'_i(\xi + n^{-1/2} g) \leq \eta^\xi(\alpha) + n^{-1/2} d'_i b_1\}]; \\
D &= n^{-1} \sum_{i=1}^n h_{i\ell}^2 E[(\bar{v}'_i \xi)^2 \sup_{\|g_1 - g\| + \|b_1 - b\| \leq k} |I\{\bar{v}'_i(\xi + n^{-1/2} g) \leq \eta^\xi(\alpha) + n^{-1/2} d'_i b_1\} \\
&\quad - I\{\bar{v}'_i(\xi + n^{-1/2} g) \leq \eta^\xi(\alpha) + n^{-1/2} d'_i b\}|].
\end{aligned}$$

Similar arguments to those used to prove (7.8) can be used to show that (7.9) is bounded by $n^{-1/2} M k$. For example, letting $U_1 = \bar{v}'(\xi + n^{-1/2} g)$, $U_2 = \sup_{g_1} \bar{v}' \frac{(g_1 - g)}{\|g_1 - g\|}$, and $U_3 = (\bar{v}' \xi)^2$, we see that from Assumption c4,

$$\begin{aligned}
C_1 &\leq n^{-1} \sum_{i=1}^n h_{i\ell}^2 E U_3^2 I\{U_1 \leq \eta^\xi(\alpha) + U_2 n^{-1/2} \|g_1 - g\| + n^{-1/2} \sup_{\|b_1 - b\| \leq k} |d'_i b_1|, \\
&\quad U_1 > \eta^\xi(\alpha) - U_2 n^{-1/2} \|g_1 - g\| - n^{-1/2} \sup_{\|b_1 - b\| \leq k} |d'_i b_1|\} \\
&= n^{-1} \sum_{i=1}^n h_{i\ell}^2 E U_3^2 \int_{\eta^\xi(\alpha) - u_2 n^{-1/2} \|g_1 - g\| - n^{-1/2} \sup_{\|b_1 - b\| \leq k} |d'_i b_1|}^{\eta^\xi(\alpha) + u_2 n^{-1/2} \|g_1 - g\| + n^{-1/2} \sup_{\|b_1 - b\| \leq k} |d'_i b_1|} f_{U_1|U_2, U_3}(u_1) du_1 \\
&\leq M n^{-1} \sum_{i=1}^n h_{i\ell}^2 E(U_2 U_3) n^{-1/2} \|g_1 - g\| \leq n^{-1/2} M k.
\end{aligned}$$

It follows that (7.9) is bounded by $n^{-1/2} M K$, so from lemma 3.2 of Bai and He (1998) and (7.8), we have

$$\sup_{\|b\| \leq k, \|g\| \leq k'} |S_\ell(b, g) - ES_\ell(b, g)| = o_p(1). \quad (7.10)$$

To establish (7.7), we still need to show that

$$\begin{aligned}
&\sup_{\|b\| \leq k, \|g\| \leq k'} |ES_\ell(b, g) - \eta^\xi(\alpha) f_\xi(\eta^\xi(\alpha)) n^{-1} \sum_{i=1}^n h_{i\ell} \{d'_i b + g' E(\bar{v} | \bar{v}' \xi \\
&= \eta^\xi(\alpha))\}| = o_p(1). \quad (7.11)
\end{aligned}$$

Consider the decomposition

$$ES_\ell(b, g) = n^{-1/2} \sum_{i=1}^n h_{i\ell} E \bar{v}'_i \xi [I\{\bar{v}'_i(\xi + n^{-1/2} g) \leq \eta^\xi(\alpha) + n^{-1/2} d'_i b\}$$

$$\begin{aligned}
& -I\{\bar{v}'_i\xi \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b\} \\
& + n^{-1/2} \sum_{i=1}^n h_{i\ell} E\bar{v}'_i\xi [I\{\bar{v}'_i\xi \leq \eta^\xi(\alpha) + n^{-1/2}d'_i b\} - I\{\bar{v}'_i\xi \leq \eta^\xi(\alpha)\}] \\
& = E_1 + E_2.
\end{aligned}$$

Let $U = \bar{v}'\xi$, $Z = \bar{v}'g$ and $\delta = n^{-1/2}d'_i b$. Then

$$\begin{aligned}
& |E_1 - n^{-1} \sum_{i=1}^n h_{i\ell} \eta^\xi(\alpha) f_\xi(\eta^\xi(\alpha)) g' E(\bar{v}|\bar{v}'\xi = \eta^\xi(\alpha))| \\
& \leq n^{-1/2} \sum_{i=1}^n h_{i\ell} \left| \int_{-\infty}^{\infty} \int_{\eta^\xi(\alpha) + \delta - n^{-1/2}z}^{\eta^\xi(\alpha) + \delta} u f(u, z) du dz \right. \\
& \quad \left. - n^{-1/2} \eta^\xi(\alpha) \int_{-\infty}^{\infty} z f(\eta^\xi(\alpha), z) dz \right| \\
& = n^{-1/2} \sum_{i=1}^n h_{i\ell} \int_{-\infty}^{\infty} \left| \int_{\eta^\xi(\alpha) + \delta - n^{-1/2}z}^{\eta^\xi(\alpha) + \delta} u f_{U|Z}(u|z) \right. \\
& \quad \left. - \eta^\xi(\alpha) f_{U|Z}(\eta^\xi(\alpha)|z) \right\} du \Big| f_Z(z) dz \\
& \leq n^{-1/2} \sum_{i=1}^n h_{i\ell} \int_{-\infty}^{\infty} \int_{\eta^\xi(\alpha) + \delta - n^{-1/2}z}^{\eta^\xi(\alpha) + \delta} \int_{\eta^\xi(\alpha)}^u \left| f_{U|Z}(t|z) + t f'_{U|Z}(t|z) \right| dt du f_Z(z) dz \\
& \leq M_1 n^{-1/2} \sum_{i=1}^n h_{i\ell} \int_{-\infty}^{\infty} \int_{\eta^\xi(\alpha) + \delta - n^{-1/2}z}^{\eta^\xi(\alpha) + \delta} (u - \eta^\xi(\alpha)) du f_Z(z) dz \\
& \leq M_2 n^{-1} \sum_{i=1}^n h_{i\ell} (\|d_i\| \|b\| \|g\| E\|\bar{v}\| + n^{-1/2} E\|\bar{v}\|^2 \|g\|^2) \leq M_3 \|b\|. \tag{7.12}
\end{aligned}$$

Similarly, $|E_2 - \eta^\xi(\alpha) f_\xi(\eta^\xi(\alpha)) n^{-1} \sum_{i=1}^n h_{i\ell} d'_i b| \leq M \|b\|$ so we have proved (7.11) and hence (7.7).

Provided

$$n^{1/2}(\hat{\eta}^g(\alpha) - \eta^\xi(\alpha)) = O_p(1), \tag{7.13}$$

similar arguments to those leading to (7.7) establish that

$$n^{-1} H'_0 A^g X = (\alpha_2 - \alpha_1) Q_{sx} + o_p(1). \tag{7.14}$$

Then from (7.7) and (7.13), we have

$$\begin{aligned}
n^{-1/2} H'_0 A^g V g & = n^{-1/2} \sum_{i=1}^n h_i \bar{v}'_i \xi I(\eta^\xi(\alpha_1) \leq \bar{v}'_i \xi \leq \eta^\xi(\alpha_2)) - \eta^\xi(\alpha_2) f_{\bar{v}'\xi}(\eta^\xi(\alpha_2)) \\
& \quad \cdot n^{-1/2} \sum_{i=1}^n h_i \{d'_i T_{n2} + T'_n E(\bar{v}|\bar{v}'\xi = \eta^\xi(\alpha_2))\} \\
& \quad + \eta^\xi(\alpha_1) f_{\bar{v}'\xi}(\eta^\xi(\alpha_1)) n^{-1/2} \sum_{i=1}^n h_i \{d'_i T_{n1} \\
& \quad + t'_n E(\bar{v}|\bar{v}'\xi = \eta^\xi(\alpha_1))\} + o_p(1), \tag{7.15}
\end{aligned}$$

where $T_{nk} = n^{1/2}(\hat{\beta}^g + \begin{pmatrix} \hat{\eta}^g(\alpha_k) \\ 0 \end{pmatrix}) - (\beta^g + \begin{pmatrix} \eta^g(\alpha_k) \\ 0 \end{pmatrix})$, $k = 1, 2$ and $T_n = n^{1/2}(g - \xi)$.

We still need to establish (7.13) and derive representations for the last two terms in (7.6). Let $\tilde{S}(g, b) = n^{-1/2} \sum_{i=1}^n [-I(\bar{v}'_i(\xi + n^{-1/2}g) \leq \eta^\xi(\alpha) + n^{-1/2}x'_i b) + I(\bar{v}'_i(\xi + n^{-1/2}g) \leq \eta^\xi(\alpha))]$. Then we need to prove that

$$\sup_{\|b\| \leq k, \|g\| \leq k'} |\tilde{S}(g, b) - f_\xi(\eta^\xi(\alpha))n^{-1} \sum_{i=1}^n h_{i\ell} \{d'_i b + g' E(\bar{v} | \bar{v}' \xi = \eta^\xi(\alpha))\}| = o_p(1). \quad (7.16)$$

Similar arguments to those leading to (7.10) show that $\sup_{\|b\| \leq k, \|g\| \leq k'} |S_n(g, b) - E(S_n(g, b))| = o_p(1)$ and similar arguments to those leading to (7.12) establish (7.16). Following Ruppert and Carroll (1980), we also have

$$n^{-1/2} \sum_{i=1}^n (\alpha - I(y_i^g - x'_i \hat{\beta}^g \leq \hat{\eta}^g(\alpha))) = o_p(1). \quad (7.17)$$

Moreover, as in the proof of Lemma 5.1 of Jureckova (1977), we obtain from (7.17) and (7.18) that, for every $\epsilon > 0$, there exists $K > 0, \ell > 0$ and N such that

$$P[\inf_{\|b\| \geq k} n^{-1/2} |\sum_{i=1}^n \{\alpha - I(\bar{v}'_i g \leq \eta^\xi(\alpha) + n^{-1/2} d'_i b)\}| < \ell] < \epsilon \quad (7.18)$$

for $n \geq N$. Then (7.13) follows from (7.17) and (7.18).

Combining (7.16) and (7.17), we have

$$\begin{aligned} & -\hat{\eta}^g(\alpha_1)n^{-1/2} \sum_{i=1}^n h_i \{\alpha_1 - I(e_i^g \leq \hat{\eta}^g(\alpha_1))\} + \hat{\eta}^g(\alpha_2)n^{-1/2} \sum_{i=1}^n h_i \{\alpha_2 - I(e_i^g \leq \hat{\eta}^g(\alpha_2))\} \\ & = \eta^\xi(\alpha_1)n^{-1/2} \sum_{i=1}^n h_i \{\alpha_1 - I(\bar{v}'_i \xi \leq \eta^\xi(\alpha_1))\} \\ & \quad + \eta^\xi(\alpha_2)n^{-1/2} \sum_{i=1}^n h_i \{\alpha_1 - I(\bar{v}'_i \xi \leq \eta^\xi(\alpha_2))\} \\ & \quad - \eta^\xi(\alpha_1) f_{\bar{v}' \xi}(\eta^\xi(\alpha_1))n^{-1} \sum_{i=1}^n h_i (d'_i T_{n1} + t'_n E(\bar{v} | \bar{v}' \xi = \eta^\xi(\alpha_1))) \\ & \quad + \eta^\xi(\alpha_2) f_{\bar{v}' \xi}(\eta^\xi(\alpha_2))n^{-1} \sum_{i=1}^n h_i \{d'_i T_{n2} + t'_n E(\bar{v} | \bar{v}' \xi = \eta^\xi(\alpha_2))\} + o_p(1). \quad (7.19) \end{aligned}$$

Combining (7.15) and (7.19), we have $n^{1/2}(\hat{\beta}^g - (B\xi - \gamma)) = \tilde{H}n^{-1/2} \sum_{i=1}^n h_i \psi(\bar{v}_i) + o_p(1)$, which implies that

$$n^{1/2}(\hat{B}_{mlw}^g - (B + (\gamma_1 \cdots \gamma_m) \Xi^{1/2'})) = \tilde{H}_0 n^{-1/2} \sum_{i=1}^n h_i (\psi_1(\bar{v}_i), \dots, \psi_m(\bar{v}_i)) \Xi^{1/2'} + o_p(1).$$

The theorem then follows.

Acknowledgement

We are grateful to an associate editor and two referees for their comments which improved the presentation of this paper.

References

- Bai, Z.-D. and He, X. (1999). Asymptotic distributions of the maximal depth estimators for regression and multivariate location. *Ann. Statist.* **27**, 1616-1637.
- Chen, L.-A. (1997). An efficient class of weighted trimmed means for linear regression models. *Statist. Sinica* **7**, 669-686.
- Chen, L. A. and Chiang, Y. C. (1996). Symmetric type quantile and trimmed means for location and linear regression model. *J. Nonparametr. Statist.* **7**, 171-185.
- De Jongh, P. J., De Wet, T. and Welsh, A. H. (1988). Mallows-type bounded-influence-regression trimmed means. *J. Amer. Statist. Assoc.* **83**, 805-810.
- Johnson, R. A. and Wichern, D. W. (1982). *Applied Multivariate Statistical Analysis*. Prentice-Hall, Inc., New Jersey.
- Jureckova, J., Koenker, R. and Welsh, A. H. (1994). Adaptive choice of trimming proportions. *Ann. Inst. Statist. Math.* **26**, 737-755.
- Jureckova, J. and Sen, P. K. (1987). An extension of Billingsley's theorem to higher dimension M-processes. *Kybernetika* **23**, 382-387.
- Jureckova, J. and Sen, P. K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations*. Wiley, New York.
- Kim, S. J. (1992). The metrically trimmed means as a robust estimator of location. *Ann. Statist.* **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. *J. Amer. Statist. Assoc.* **82**, 851-857.
- Koul, H. L. (1992). *Weighted Empiricals and Linear Models*. IMS Lecture Notes 21.
- Morrison, D. F. (1983). *Applied Linear Statistical Methods*. Prentice-Hall, Inc., New Jersey.
- Ren, J. J. (1994). Hadamard differentiability and its applications to R-estimation in linear models. *Statist. Dec.* **12**, 1-22.
- Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *J. Amer. Statist. Assoc.* **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. *Ann. Statist.* **15**, 20-36.

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

E-mail: lachen@stat.nctu.edu.tw

Centre for Mathematics and Its Applications, Australian National University, Canberra, Australia.

E-mail: Alan.Welsh@ann.edu.au

School of Public Health, University of Texas-Houston, Houston, Texas.

(Received February 1999; accepted April 2000)