國 立 交 通 大 學

統 計 學 研 究 所

碩 士 論 文

具有 AR(1)誤差的迴歸模型的線性修正平均值估計

Linear Trimmed Means for the Linear Regression with AR(1) Errors Model

研究生：彭豐洋

指導教授：陳鄰安博士

中華民國 九十四 年 六 月

具有 AR(1)誤差的迴歸模型的線性修正平均值估計

Linear Trimmed Means for the Linear Regression with AR(1) Errors Model

研究生：彭豐洋　　Student：Feng-Yang Peng

指導教授：陳鄰安博士　　Advisor：Dr. Lin-An Chen

國立交通大學理學院

統計研究所

碩士論文

A thesis

Submitted to Institute of Statistics

College of Science

National Chiao Tung University

In partial Fulfillment of the Requirements

For the Degree of

Master in Statistics

June 2005

Hsinchu Taiwan Republic of China

中華民國 九十四 年 六 月

具有 AR(1)誤差的迴歸模型的線性修正平均值估計

研究生：彭豐洋

指導教授：陳鄰安博士

國立交通大學統計研究所

中文摘要

延續 Lai (2003)其具有 AR(1)誤差的線性迴歸模型的穩健性估計基本架構，我們證明了在大樣本的情形下廣義修正平均值估計量能夠有類似 Gauss Markov Theorem 的性質。我們稱其為穩健型態的 Gauss Markov Theorem。

我們進而利用模擬的方法以及實例的分析，說明該估計量的特性與效率。

Linear Trimmed Means for the Linear Regression with AR(1) Errors Model

Student: Feng-Yang Peng Advisor: Dr. Lin-An Chen

Institute of Statistics, National Chiao Tung University,

Hsinchu, Taiwan.

Abstract

For the linear regression with AR(1) errors model, a robust type generalized and feasible generalized estimators of Lai et al. (2003) of regression parameters are shown to have the desired property of robust type Gauss Markov theorem. It is done by shown that these two estimators are, respectively, the best among classes of linear trimmed means. Monte Carlo and data analysis for this technique have been performed.

Keywords: Gauss Markov theorem; Generalized least squares estimator; linear trimmed mean; robust estimator.

誌謝

感謝指導老師 陳鄰安教授孜孜不倦地教導，本論文方能順利完成。

承蒙交通大學 彭南夫教授、清華大學統計學研究所 徐南蓉教授與 江永進教授於百忙之中撥冗審稿、校正，並給予寶貴的意見，僅此表達誠摯的謝忱。

感謝所上諸位老師的指導，郭姐的照顧，以及一起研究的同伴們，育仕、景嵐、穎劭、雅靜、玉均、辰昀、揚波、婉菁、瑩琪等，對我的鼓勵與包容。感謝媽媽、姊姊陪我一路走來，無怨無悔的付出所有愛與關懷，有你們的支持我才成克服一切的困難，沒有後顧之憂地完成我的學業。

僅將本論文獻給各位。

彭豐洋　　　僅誌于

國立交通大學統計研究所

中華民國九十四年六月

Contents

1. Introduction

Consider the linear regression model

$$y = X\beta + \epsilon \tag{1.1}$$

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. An interesting question in robust regression is if there is robust type Gauss-Markov theorem, i.e., if there is a robust estimator that is (asymptotically) more efficinet than a class of linear robust estimators? This has been done by Chen et al. (2001) that they considered a class of estimators based on Winsorized observations and show that the trimmed mean of Welsh (1987) is asymptotically the best among it.

Suppose that the error vector $\epsilon = (\epsilon_1, ..., \epsilon_n)'$ has the covariance matrix structure

$$\text{Cov}(\epsilon) = \sigma^2\Omega \tag{1.2}$$

where Ω is a positive definite matrix. From the regression theory of the estimation of β, it is known that any estimator having an (asymptotic) covariance matrix of the form

$$\delta(X'\Omega^{-1}X)^{-1} \tag{1.3}$$

is more efficient than the estimator having (asymptotic) covariance matrix of the form

$$\delta(X'X)^{-1}(X'\Omega X)(X'X)^{-1} \tag{1.4}$$

where δ is some positive constant. In the least squares estimation when the matrix Ω is known, Aitken (1935) introduced the generalized least squares estimator (LSE) and showed that it has a covariance matrix of the form (1.3) and the LSE has a covariance matrix of the form (1.4) with $\delta = \sigma^2$. It is also well known that, when Ω is unknown, the feasible generalized LSE has the asymptotic covariance matrix of the form (1.3). Then these two generalized type estimators are more efficient than the LSE.

Although the generalized and feasible generalized LSE's are asymptotically more efficient than the LSE in many regression problems, they are highly sensitive to even a very small departure from normality and to the presence of outliers. Therefore developing robust type generalized and feasible generalized estimators in each specific regression problem is interesting. Let's consider the linear regression with AR(1) errors model, a structure of (1.2), as follows

$$y_i = x_i'\beta + \epsilon_i, i = 1, ..., n$$
$$\epsilon_i = \rho\epsilon_{i-1} + e_i \qquad (1.5)$$

where $e_1, ..., e_n$ are independent and identically distributed (iid) random variables, is one of the most popular models. Suppose that $|\rho| < 1$ and e_i has a distribution function F.

Denote the transformed vector $u = \Omega^{-1/2'} y$. One approach to robust estimation is to construct a weighted observation vector u^* and then construct a consistent estimator which is linear in u^*, in case that ρ is unknown, all vectors are replaced by the ones with estimating ρ by estimator $\hat{\rho}$; see for example, Lai et al. (2003). There are two types of weighted observation vectors in this literature. First, u^* can represent a trimmed observation vector Au 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, u^* can be a Winsorized observation vector defined as in Welsh (1987). In this paper, we consider the trimmed observation vector of Koenker and Bassett (1978), study classes of linear functions based on u^* for estimation of β, and develop a robust version of the Gauss-Markov theorem. Based on regression quantiles, Lai et al. (2003) proposed generalized and feasible generalized trimmed means for estimating regression parameters β. Then a robust type generalized and feasible generalized estimation technique have been developed.

With the result that we have robust version of Gauss Markov theorem for linear regression with iid errors model, it is then interesting to see if there is any robust type generalized and feasible generalized estimators for the linear regression with AR(1) errors model that also play the same version of Gauss Markov theorem. Our aim in this paper is to show that the Lai et al. (2003) does have this desired property.

We introduce a class of linear trimmed means when ρ is known in Section 2 and establish their large sample theory in Section 3. We also establish the theory for a class of linear trimmed means when ρ is unknown in Section 4. In both cases, we show that the generalized and feasible generalized trimmed means are the best, respectively, in these two classes of linear trimmed means in terms of asymptotic covariance matrix. Finally, the proofs of the theorems are displayed in Appendix.

2. Linear Trimmed Mean When ρ is known

For the linear regression with AR(1) errors model (1.5), to obtain a linear trimmed mean we need to specify the quantile for determining the observation trimming and to make a transformation of the linear model to obtain generalized estimators. For given i-th dependent variable for model (1.5), assuming that $i \geq 2$, one way to derive a generalized estimator is to consider the transformation by Cochrane and Orcutt (C-O, 1949) as $y_i = \rho y_{i-1} + (x_i - \rho x_{i-1})'\beta + e_i$. For error variable e, we assume that it has distribution function F with probability density function f. With the transformation for generalized estimation, a quantile could be defined through variable e or a linear conditional quantile of y_{i-1} and y_i. By the fact that x_i is vector with first element 1, the following two events determined by two quantiles are equivalent:

$$e_i \leq F^{-1}(\alpha) \tag{2.1}$$

and

$$(-\rho, 1)\begin{pmatrix} y_{i-1} \\ y_i \end{pmatrix} \leq (-\rho, 1)\begin{pmatrix} x'_{i-1} \\ x'_i \end{pmatrix}\beta(\alpha) \tag{2.2}$$

with $\beta(\alpha) = \beta + \begin{pmatrix} \frac{1}{1-\rho}F^{-1}(\alpha) \\ 0_{p-1} \end{pmatrix}$. The event in inequality (2.1) specifies the quantile of the error variable e and it through inequality (2.2) specifies the conditional quantile of linear function $(-\rho, 1)\begin{pmatrix} y_{i-1} \\ y_i \end{pmatrix}$. Here $\beta(\alpha)$ is called the population regression quantile by Koenker and Bassett (1978). With the specification of quantiles and transformation, we may define the linear trimmed means.

For defining the linear trimmed means, we consider the C-O transformation on the matrix form of the linear regression with AR(1) error model of (1.5) which is

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

where it is seen that $\mathrm{Cov}(\epsilon) = \sigma^2\Omega$ with

$$\Omega = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}. \tag{2.3}$$

Define the half matrix of Ω^{-1} as

$$\Omega^{-1/2'} = \begin{pmatrix} (1-\rho^2)^{1/2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}.$$

With the above half matrix of Ω, we consider the model for the transformation $u = \Omega^{-1/2'}y$ as

$$u = Z\beta + ((1-\rho^2)^{1/2}\epsilon_1, e_2, e_3, ..., e_n)' \tag{2.4}$$

where $Z = \Omega^{-1/2'}X$. Note that the vector u and the matrix Z are both functions of parameter ρ. The usual descriptive statistics, robust or nonrobust, based on model (1.1) can be carried over straightforwardly to transformed model (2.4) when ρ is known. However, when ρ is unknown, u and Z need to be replaced by the ones that place its ρ by the estimator. Knowing the fact that generalized LSE is simply the LSE of β for model (2.4), we may consider the linear trimmed mean defining on this transformed model. To validate the terminology calling the linear trimmed means with ρ known and unknown, we will show that they are asymptotically equivalent in the sense of having the same asymptotic covariance matrix. This is what the generalized and feasible generalized LSE's performed.

At this moment that we want to study generalized robust estimator, we assume that ρ is known. For $0 < \alpha < 1$, the α-th (sample) regression quantile of Koenker and Bassett (1978) for the linear regression with AR(1) errors model is defined as

$$\hat{\beta}(\alpha) = \arg_{b\in R^p}\min\sum_{i=1}^{n}(u_i - z_i'b)(\alpha - I(u_i \le z_i'b))$$

where u_i and z_i' are the i-th rows of u and Z respectively. Define the trimming matrix as $A = \mathrm{diag}\{a_i = I(z_i'\hat{\beta}(\alpha_1) \le u_i \le z_i'\hat{\beta}(\alpha_2)) : i = 1, ..., n\}$. After outliers are trimmed by regression quantiles $\hat{\beta}(\alpha)$ and $\hat{\beta}(1-\alpha)$, we have the following submodel

$$Au = AZ\beta + A \begin{pmatrix} (1-\rho^2)^{1/2}\epsilon_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}. \tag{2.5}$$

Since A is random, the error vector in the above transformed model is now not a set of independent variables. The Koenker and Bassett's type generalized trimmed mean (proposed by Lai et al. (2003)) is defined as

$$\hat{\beta}_{tm} = (Z'AZ)^{-1}Z'Au. \tag{2.6}$$

We now move to define the linear trimmed means. Any linear unbiased estimator defined in model of (2.4) has the form Mu with M being a $p \times n$ nonstochastic matrix satisfying $MZ = I_p$. 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 an $n \times p$ full-rank matrix H_0 such that the estimator can be written as

$$HH_0'u.$$

We generalize linear unbiased estimators defined on the observation vector u to estimators defined on Au by requiring them to be of the form MAu with $M = HH_0'$.

Definition 2.1. A statistic $\hat{\beta}_{ltm}$ is called a (α_1, α_2) linear trimmed mean 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:

$$\hat{\beta}_{ltm} = HH_0'Au, \tag{2.7}$$

where H and H_0 satisfy the following two conditions:
(a1) $nH \to \tilde{H}$ in probability, where $\tilde{H}$ is a full rank $p \times p$ matrix.
(a2) $HH_0'Z = (\alpha_2 - \alpha_1)^{-1}I_p + o_p(n^{-1/2})$ 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 trimmed observation vector to allow for robustness and considered asymptotic property instead of unbiasedness.

Two questions arise for the class of linear trimmed means. First, does this class of means contain estimators that have already appeared in the literature? The answer is affirmative because the class of linear trimmed means defined in this paper contains the generalized trimmed mean of Lai et al. (2003) ($H = (Z'AZ)^{-1}$ and $H_0 = Z$), and the set of Mallows-type bounded influence trimmed means ($H = (Z'WAZ)^{-1}$ and $H_0' = Z'W$ with W, a diagonal matrix of weights; see Section 3). Second, is there a best estimator in this class of linear trimmed means and can we find it if it exists? This question will be answered in the next section.

With the C-O transformation, the half matrix $\Omega^{-1/2'}$ has rows with only a finite number (not depending on n) of elements that depend on the unknown parameter ρ. This trick, traditionally used in econometrics literature for regression with AR(1) errors (see, for example, Fomby, Hill and Johnson (1984, p210-211)), makes the study of asymptotic theory for $\hat{\beta}_{ltm}(\alpha)$ similar to what we have for the classical trimmed mean for linear regression. Large sample representations of the linear trimmed mean and its role as generalized robust estimator will be introduced in the next section.

3. Asymptotic Properties of Linear Trimmed Mean

Denoting by h_i' the ith row of H_0, $\theta_h = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n h_i$, $Q_{hz} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^n h_i z_i'$ and $Q_z = \lim_{n\to\infty} n^{-1} Z'Z$, the following theorem gives a "Bahadur" representation of the (α_1, α_2) linear trimmed mean.

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

$$n^{1/2}(\hat{\beta}_{ltm} - (\beta + \gamma_{ltm})) = n^{-1/2}\tilde{H}\sum_{i=1}^n \{h_i(e_i I(F^{-1}(\alpha_1) \le e_i \le F^{-1}(\alpha_2)) - \lambda)$$
$$+ [F^{-1}(\alpha_1)I(e_i < F^{-1}(\alpha_1)) + F^{-1}(\alpha_2)I(e_i > F^{-1}(\alpha_2)) - ((1-\alpha_2)F^{-1}(\alpha_2)$$
$$+ \alpha_1 F^{-1}(\alpha_1))]Q_{hz}Q_z^{-1}z_i\} + o_p(1),$$

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

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

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

$$[\int_{F^{-1}(\alpha_1)}^{F^{-1}(\alpha_2)} e^2 dF(e) - \lambda^2]\tilde{H}Q_h\tilde{H}' + (\alpha_2 - \alpha_1)^{-2}[\alpha_1(F^{-1}(\alpha_1))^2 + (1-\alpha_2)(F^{-1}(\alpha_2))^2 \\ - (\alpha_1 F^{-1}(\alpha_1) + (1-\alpha_2)F^{-1}(\alpha_2))^2 - 2\lambda(\alpha_1 F^{-1}(\alpha_1) + (1-\alpha_2)F^{-1}(\alpha_2))]Q_z^{-1}. \tag{3.1}$$

The (α_1, α_2) generalized trimmed mean proposed by Lai et al. (2003) is defined by

$$\hat{\beta}_{tm} = (Z'AZ)^{-1}Z'Au. \tag{3.2}$$

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

$$n^{-1}Z'AZ \to (\alpha_2 - \alpha_1)Q_z.$$

By letting $H = (Z'AZ)^{-1}$ and $H_0 = Z$, can see that condition (a2) also holds for $\hat{\beta}_{tm}$. So, the (α_1, α_2) generalized trimmed mean is in the class of (α_1, α_2) linear trimmed mean's. Moreover, Lai et al. (2003) provided the result that $n^{1/2}(\hat{\beta}_{tm} - (\beta + \gamma_{tm}))$, where $\gamma_{tm} = (\alpha_2 - \alpha_1)^{-1}\lambda Q_z^{-1}\theta_z$, has an asymptotic normal distribution with zero means and covariance matrix $\sigma^2(\alpha_1, \alpha_2)Q_z^{-1}$, where

$$\sigma^2(\alpha_1, \alpha_2) = (\alpha_2 - \alpha_1)^{-2}[\int_{F^{-1}(\alpha_1)}^{F^{-1}(\alpha_2)} (e - \lambda)^2 dF(e) + (\alpha_1(F^{-1}(\alpha_1))^2 + (1-\alpha_2) \\ (F^{-1}(\alpha_2))^2 - (\alpha_1 F^{-1}(\alpha_1) + (1-\alpha_2)F^{-1}(\alpha_2))^2 - 2\lambda(\alpha_1 F^{-1}(\alpha_1) + (1-\alpha_2)F^{-1}(\alpha_2)))]. \tag{3.3}$$

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

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

$$\tilde{H}Q_h\tilde{H}' - (\alpha_2 - \alpha_1)^{-2}Q_z^{-1} \tag{3.4}$$

is positive semidefinite.

The relation in (3.4) then implies the following main theorem.

Theorem 3.4. Under the conditions (a.3)-(a.6), the (α_1, α_2) generalized trimmed mean $\hat{\beta}_{tm}$ of (3.2) is the best (α_1, α_2) linear trimmed mean.

Since the (α_1, α_2) generalized trimmed mean always exists, then the best (α_1, α_2) linear trimmed mean always exists. A further question is that how big is the class of (α_1, α_2) linear trimmed mean's? We are not going to study the scope of the linear trimmed means.

In the literature, consideration has been given to the development of estimators of regression parameters β that limit the effects of the error variable and the independent variables. Among them, approaches which simultaneously bound the influence of the design points and the residuals for the linear regression model include Krasker and Welsch (1982) and Krasker (1985). On the other hand, the approach of the Mallow$'$s type bounded-influence trimmed mean is to bound the influence of the design points and the residuals separately as applied in the AR(1) regression model by De Jongh and De Wet (1985) and in the linear regression model by De Jongh et al.(1988). In a study by Giltinan et al.(1986), they found these two approaches are competitive in a way that neither is preferable to the other one. They also note that the Mallow$'$s type estimators should theoretically give more stable inference than the Krasker-Welsch approach.

Let $w_i, i = 1, ..., n$, be real numbers. For $0 < \alpha < 1$, the Mallow$'$s type bounded-inference regression quantile, denoted by $\hat{\beta}_w(\alpha)$, is defined as the solution for the minimization problem

$$\min_{b \in R^p} \sum_{i=1}^{n} w_i(u_i - z_i'b)(\alpha - I(u_i \leq z_i'b)).$$

With W the diagnal matrix of $\{w_i, i = 1, ..., n\}$, the bounded influence trimmed mean is defined as

$$\hat{\beta}_{BI} = (Z'WA_wZ)^{-1}Z'WA_wu$$

where $A_w = \text{diag}\{a_i : I(z_i'\hat{\beta}_w(\alpha_1) \leq u_i \leq z_i'\hat{\beta}_w(\alpha_2)), i = 1, ..., n\}$.

Let $H = (Z'WA_wZ)^{-1}$ and $H_0 = WZ$. This shows that the bounded influence trimmed means also form a subclass of linear trimmed means$'$s (see De Jongh et al (1988) for their large sample properties).

Theorem 3.5 If assumptions (a1)-(a5) hold, then

(a) $n^{1/2}(\hat{\beta}_{BI} - (\beta + \gamma_w)) = (\alpha_2 - \alpha_1)^{-1} Q_w^{-1} n^{-1/2} \sum_{i=1}^{n} w_i z_i [(e_i I(F^{-1}(\alpha_1) \leq e_i \leq F^{-1}(\alpha_2))$
$- \lambda) + (F^{-1}(\alpha_1) I(e_i < F^{-1}(\alpha_1)) + F^{-1}(\alpha_2) I(e_i > F^{-1}(\alpha_2)) - ((1-\alpha_2)F^{-1}(\alpha_2)$
$+ \alpha_1 F^{-1}(\alpha_1)))] + o_p(1),$

where $\gamma_w = (\alpha_2 - \alpha_1)^{-1} \lambda Q_w^{-1} \theta_w$, $Q_w = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n w_i z_i z_i'$ and $\theta_w = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n w_i z_i$ and
(b) $n^{1/2}(\hat{\beta}_{BI} - (\beta + \gamma_w)) \to N(0, \sigma^2(\alpha_1, \alpha_2) Q_w^{-1} Q_{ww} Q_w^{-1})$ where $Q_{ww} \lim_{n\to\infty} n^{-1} \sum_{i=1}^n w_i^2 z_i z_i'$.

In particular, $\hat{\beta}_{tm}$ is the one of $\hat{\beta}_{BI}$ with W the identity matrix and then belongs to this subclass. We may also show that $Q_w^{-1} Q_{ww} Q_w^{-1} - Q_z^{-1}$ is positive semidefinite which shows that $\hat{\beta}_{tm}$ is the best bounded influence trimmed mean.

Theorem 3.6. The (α_1, α_2) generalized trimmed mean is the best bounded influence trimmed mean.

This result is based solely on considerations of the asymptotic variance and ignores the fact that generalized 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.

4. Linear Trimmed Means When ρ is Unknown

After the development of the theory of the linear trimmed means for that ρ is known, the next interesting problem is whether when the parameter ρ is unknown, the linear trimmed mean of (2.7) with ρ replaced by a consistent estimator $\hat{\rho}$, will have the same asymptotic behavior as displayed by $\hat{\beta}_{ltm}$. If yes, the theory of generalized least squares estimation is then carried over to the theory of robust estimation in this specific linear regression model. Let $\hat{\Omega}$ be the matrix of Ω with ρ replaced by its consistent estimator $\hat{\rho}$. Define matrices $\hat{u} = \hat{\Omega}^{-1/2'} y$, $\hat{Z} = \hat{\Omega}^{-1/2'} X$ and $\hat{e} = \hat{\Omega}^{-1/2'} \epsilon$. Let the regression quantile when the parameter ρ is unknown be defined as

$$\hat{\beta}^*(\alpha) = \arg_{b \in R^p} \min \sum_{i=1}^{n} (\hat{u}_i - \hat{z}_i' b)(\alpha - I(\hat{u}_i \leq \hat{z}_i' b))$$

where $\hat{u}_i$ and $\hat{z}_i'$ are i-th rows of $\hat{u}$ and $\hat{Z}$ respectively. Define the trimming matrix as $\hat{A} = \text{diag}\{a_i = I(\hat{z}_i' \hat{\beta}^*(\alpha_1) \leq \hat{u}_i \leq \hat{z}_i' \hat{\beta}^*(\alpha_2)) : i = 1, ..., n\}$.

Definition 4.1. A statistic $\hat{\beta}^*_{ltm}$ is called a (α_1, α_2) linear trimmed mean if there exists a stochastic $p \times p$ and nonstochastic $n \times p$ matrices, respectively, H and H_0 such that it has the following representation:

$$\hat{\beta}^*_{ltm} = HH_0'\hat{A}\hat{u},$$

where H and H_0 satisfy conditions (a1) and (a2) for these H and H_0.

The Koenker and Bassett's feasible generalized trimmed mean is defined as

$$\hat{\beta}^*_{tm} = (\hat{Z}'\hat{A}\hat{Z})^{-1}\hat{Z}'\hat{A}\hat{u}.$$

From Lai et al. (2003), we may see that $n^{-1}\hat{Z}'\hat{A}\hat{Z} \xrightarrow{p} (\alpha_2 - \alpha_1)Q_z$. By letting $H = (\hat{Z}'\hat{A}\hat{Z})^{-1}$ and $H_0 = \hat{Z}$, we see that $\hat{\beta}^*_{tm}$ is in the class of (α_1, α_2) linear trimmed means. Lai et al. (2003) also showed that $\hat{\beta}^*_{tm}$ and $\hat{\beta}_{tm}$ have the same Bahadur representation and then they have the same asymptotic distribution. The following theorem states that the linear trimmed means for that ρ is known and unknown have the same large sample properties.

Theorem 4.2. $\sqrt{n}(\hat{\beta}^*_{ltm} - \hat{\beta}_{ltm}) = o_p(1)$.

We then have the result that the feasible generalized trimmed mean is the best linear trimmed mean when ρ is uknown.

Theorem 4.3. The feasible generalized trimmed mean is the best linear trimmed mean.

5. Monte Carlo Study and Example

In this section, we first consider a simulation study to compare the feasible generalized LSE $\hat{\beta}_{FG}$ and the feasible generalized trimmed mean $\hat{\beta}^*_{tm}$. By letting $\hat{\epsilon}_i = y_i - x_i'\hat{\beta}_{ls}$ where $\hat{\beta}_{ls}$ is the LSE of β, we note that the C-O method defines $\hat{\rho}$ by $\frac{\sum_{i=2}^n \hat{\epsilon}_i\hat{\epsilon}_{i-1}}{\sum_{i=2}^n \hat{\epsilon}_i^2}$. With sample size $n = 30$, the simple linear regression model, $y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$ where ϵ_i follows the AR(1) error is considered. For this simulation, we let the true parameter values of β_0 and β_1 $1's$ and ρ be 0.3. This simulation is conducted with the same data generation system except that the error variable e_i is generated from the mixed normal distribution $(1-\delta)N(0,1) + \delta N(0, \sigma^2)$ with $\delta = 0, 0.1, 0.2, 0.3$ and $\sigma = 3, 5, 10$ and x_i are independent normal random variables with mean $i/2$ and variance 1. A total of 10000 replications were performed and we compute the mean squares errors for the feasible generalized LSE $\hat{\beta}_{FG}$ and feasible generalized trimmed mean $\hat{\beta}^*_{tm}$ for $\alpha_1 = 1 - \alpha_2 = \alpha = 0.1, 0.2, 0.3$ where the total mean squared error is the square of the Euclidean distance between the estimator and true regression parameter β. For covenience, we here after in this section re-denote the feasible generalized trimmed mean by $\hat{\beta}_{tm}(\alpha)$. The mean squares errors are listed in Tables 1 and 2.

Table 1. MSE's for $\hat{\beta}_{FG}$ and $\hat{\beta}^*_{tm}$ under contaminated normal distribution ($n = 30$)

σ	$\hat{\beta}_{FG}$	$\hat{\beta}^*_{tm}(0.1)$	$\hat{\beta}^*_{tm}(0.2)$	$\hat{\beta}^*_{tm}(0.3)$
($\delta = 0$)				
	0.2096	0.2241	0.2179	0.2556
($\delta = 0.1$)				
3	0.3746	0.2874	0.2697	0.2698
5	0.7326	0.3644	0.3075	0.2964
10	2.3055	0.5463	0.4184	0.3714
($\delta = 0.2$)				
3	0.5543	0.3963	0.3600	0.3300
5	1.2306	0.5819	0.4530	0.4229
10	4.4579	1.4236	0.7820	0.6012
($\delta = 0.3$)				
3	0.7075	0.5380	0.4448	0.4101
5	1.7109	0.9723	0.6503	0.5749
10	6.5214	2.8921	1.5105	1.0893

Table 2. MSE's for $\hat{\beta}_{FG}$ and $\hat{\beta}^*_{tm}$ under contaminated normal distribution ($n = 100$)

σ	$\hat{\beta}_{FG}$	$\hat{\beta}^*_{tm}(0.1)$	$\hat{\beta}^*_{tm}(0.2)$	$\hat{\beta}^*_{tm}(0.3)$
($\delta = 0$)				
	0.0751	0.0791	0.0804	0.0805
($\delta = 0.1$)				
3	0.1308	0.0964	0.0960	0.0904
5	0.2539	0.1067	0.1059	0.0988
10	0.8494	0.1253	0.1154	0.1111
($\delta = 0.2$)				
3	0.1962	0.1257	0.1182	0.1125
5	0.4249	0.1679	0.1343	0.1270
10	1.5763	0.2736	0.1574	0.1543
($\delta = 0.3$)				
3	0.2522	0.1670	0.1466	0.1362
5	0.6266	0.2744	0.1844	0.1643
10	2.1937	0.7071	0.2683	0.2147

We have several conclusions drawn from Tables 1 and 2:
(a) The MSE's of these two estimators both increase when the contaminated percentage δ increases or contaminated variance σ^2 increases. This verifies the performance of the usual estimators, robust or non-robust.
(b) The feasible generalized trimmed mean is relatively more efficient than the feasible generalized LSE in all cases of contaminated errors. This result shows that the feasible generalized trimmed mean is indeed, among the class of linear trimmed means, a robust one.

In the next, we consider real data regression analysis. Many firms use past sales to forecast future sales. Suppose a wholesale distributor of sporting goods is interested in forecasting its sales revenue for each of the next 5 years. Since an inaccurate forecast may have dire consequences to the distributor, efficiency of the estimation of regression parameters is an important indicator in accuracy of forecasting. A data of a firm's yearly sales revenue (thousands of dallars) with sample size $n = 35$ has been analyzed by Mendenhall and Sincich (1993). Since the scatter plot of the data revealed a linearly increasing trend, so a simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, ..., 35$$

seems to be reasonable to describe the trend. They first analyzed it with the least squares method that yields $R^2 = 0.98$ which indicates that it is appropriate to be formulated as a linear regression model. They further displayed a plot of the residuals that revealed the existence of $AR(1)$ errors and then the Durbin and Watson test has been performed that reject the hypothesis of null hypothesis $\rho = 0$. He also computed the prediction 95% confidence intervals for yearly revenues for years, 36-40, however, the interval estimates are wide that makes us less certain for the prediction of future observations (see this point in Mendenhall and Sincich (1993, p481)). We expect to have better analysis, based on the feasible generalized trimmed mean, in some sense.

We follow their idea in evaluating the predition of the yearly revenues for years 36-40. Since the observations of these are available, we may compute the following mean square errors (MSE),

$$\text{MSE} = \frac{1}{3} \sum_{i=33}^{35} (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

where $\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{pmatrix}$ is the estimate of $\begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix}$ corresponding the estimator. For this example, estimators considered include LSE $\hat{\beta}_{ls}$, feasible generalized LSE $\hat{\beta}_{FG}$, ℓ_1-norm estimator $\hat{\beta}_{\ell_1}$ and feasible generalized trimmed mean $\hat{\beta}_{tm}(\alpha)$ and their evaluated MSE's are listed in Table 3.

Table 3. MSE's for predictors based on some estimators

Estimator	estimate	observation	predition	MSE
$\hat{\beta}_{ls}$	$\begin{pmatrix}1.053\\ 4.239\end{pmatrix}$	$\begin{pmatrix}146.10\\ 151.40\\ 150.90\end{pmatrix}$	$\begin{pmatrix}140.94\\ 145.17\\ 149.41\end{pmatrix}$	67.503
$\hat{\beta}_{FG}$	$\begin{pmatrix}0.142\\ 4.319\end{pmatrix}$		$\begin{pmatrix}142.67\\ 146.99\\ 151.31\end{pmatrix}$	31.336
$\hat{\beta}_{\ell_1}$	$\begin{pmatrix}0.531\\ 4.268\end{pmatrix}$		$\begin{pmatrix}141.38\\ 145.64\\ 149.91\end{pmatrix}$	56.304
$\hat{\beta}^*_{tm}(0.1)$	$\begin{pmatrix}-0.859\\ 4.386\end{pmatrix}$		$\begin{pmatrix}143.88\\ 148.27\\ 152.65\end{pmatrix}$	17.786
$\hat{\beta}^*_{tm}(0.2)$	$\begin{pmatrix}0.072\\ 4.364\end{pmatrix}$		$\begin{pmatrix}144.106\\ 148.47\\ 152.83\end{pmatrix}$	16.302
$\hat{\beta}^*_{tm}(0.3)$	$\begin{pmatrix}0.051\\ 4.336\end{pmatrix}$		$\begin{pmatrix}143.15\\ 147.49\\ 151.83\end{pmatrix}$	24.775

Surprisingly the feasible generalized trimmed means for several symmetric trimming proportions are with MSE's all smaller than those of the other three estimators. The feasible generalized trimmed mean not only has asymptotic optimal properties in the class of linear trimmed means but also shows an interesting fact in prediction of future observations.

6. Appendix

Let ϵ have distribution function F with probability density function f. Let z_{ij} represents the jth element of vector z_i. 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) $n^{-1}\sum_{i=1}^n z_{ij}^4 = O(1)$, (a4) $n^{-1}Z'Z = Q_z + o(1)$, $n^{-1}H_0'Z = Q_{hz} + o(1)$ and $n^{-1}H_0'H_0 = Q_h + o(1)$ where Q_z and Q_h are positive definite matrices and Q_{hz} is a full rank matrix.

(a5) $n^{-1}\sum_{i=1}^n h_i = \theta_h + o(1)$, where θ_h is a finite vector.

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

Proof of Theorem 3.1. From condition (a2) and (A.10) of Ruppert and Carroll (1980), $HH_0'A_nZ\beta = \beta + o_p(n^{-1/2})$. Inserting (2.4) in equation (2.7), we have

$$n^{1/2}(\hat{\beta}_{lt} - \beta) = n^{1/2}HH_0'Ae \tag{6.1}$$

where we replace $(1-\rho^2)^{1/2}\epsilon_1$ by e_1 that have the same asymptotic representation. Now we develop a representation of $n^{-1/2}H_0'Ae$. Let $U_j(\alpha, T_n) = n^{-1/2}\sum_{i=1}^n h_{ij}e_iI(e_i < F^{-1}(\alpha) + n^{-1/2}z_i'T_n)$ and $U(\alpha, T_n) = (U_1(\alpha, T_n), ..., U_p(\alpha, T_n))$. Also, let $T_n^*(\alpha) = n^{1/2}[\hat{\beta}(\alpha) - \beta(\alpha)]$. Then $n^{-1/2}H_0'A_ne = U(\alpha_2, T_n^*(\alpha_2)) - U(\alpha_1, T_n^*(\alpha_1))$. From Jureckova and Sen's (1987) extension of Billingsley's Theorem (see also Koul (1992)), we have

$$|U_j(\alpha, T_n) - U_j(\alpha, 0) - n^{-1}F^{-1}(\alpha)f(F^{-1}(\alpha))\sum_{i=1}^n h_{ij}z_i'T_n| = o_p(1) \tag{6.2}$$

for $j = 1, ..., p$ and $T_n = O_p(1)$. We know that, from Lai et al. (2004),

$$n^{1/2}(\hat{\beta}(\alpha) - \beta(\alpha)) = Q_z^{-1}f^{-1}(F^{-1}(\alpha))n^{-1/2}\sum_{i=1}^n z_i(\alpha - I(e_i \le F^{-1}(\alpha))) + o_p(1). \tag{6.3}$$

From (6.2) and (6.3)

$$\begin{aligned} n^{-1/2}H_0'A_ne &= n^{-1/2}\sum_{i=1}^n h_ie_iI(F^{-1}(\alpha_1) \le e_i \le F^{-1}(\alpha_2)) \\ &+ F^{-1}(\alpha_2)Q_{hz}Q_z^{-1}n^{-1/2}\sum_{i=1}^n z_i(\alpha_2 - I(e_i \le F^{-1}(\alpha_2)) \\ &+ F^{-1}(\alpha_1)Q_{hz}Q_z^{-1}n^{-1/2}\sum_{i=1}^n z_i(\alpha_1 - I(e_i \le F^{-1}(\alpha_1)). \end{aligned} \tag{6.4}$$

Then the theorem is follwoed from (6.1) and (6.4).

The proof of Theorem 3.5 is analogous as it for the above and then is skipped.

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

$$C = HH_0' - (Z'A_nZ)^{-1}Z'.$$

With this, $\text{plim}(CZ) = \text{plim}(HH_0'Z) - \text{plim}(Z'A_nZ)^{-1}Z'Z = 0$.
Then

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

Proof of Theorem 4.2. We here sketch only briefly a proof of the theorem. For detail references, see Chen et al. (2001) and Lai et al. (2003). With the fact that $n^{1/2}(\hat{\rho} - \rho) = O_p(1)$ and condition (a1), we may see that

$$n^{1/2}(\hat{\beta}^*_{ltm} - \beta) = n^{1/2}HH_0'\hat{A}e + o_p(1). \tag{6.5}$$

By letting $M(t_1, t_2, \alpha) = n^{-1/2}\sum_{i=1}^n h_i e_i I(e_i - n^{-1/2}t_1\epsilon_{i-1} \leq F^{-1}(\alpha) + n^{-1/2}(z_i + n^{-1/2}t_1x_{i-1})'t_2 + n^{-1/2}t_1F^{-1}(\alpha))$, we see that

$$n^{-1/2}\hat{Z}'A_ne = M(T_1^*(\alpha_2), T_2^*, \alpha_2) - M(T_1^*(\alpha_1), T_2^*, \alpha_1) \tag{6.5}$$

with $T_1^*(\alpha) = n^{1/2}(\hat{\beta}^*(\alpha) - \beta(\alpha))$ and $T_2^* = n^{1/2}(\hat{\rho} - \rho)$. However, using the same methods in the proof of Lemma 3.5, we can see that

$$M(T_1, T_2, \alpha) - M(0, 0, \alpha) = F^{-1}(\alpha)f(F^{-1}(\alpha))n^{-1/2}\sum_{i=1}^n h_i(z_i'T_2 - T_1F^{-1}(\alpha)) + o_p(1) \tag{7.6}$$

for any sequences $T_1 = O_p(1)$ and $T_2 = O_p(1)$. Then, from (6.5) and (6.6), we see that $n^{-1/2}H_0'\hat{A}_ne$ has the same representation of (6.4). Then (a1) and (6.5) further implies the theorem. □

References

Aitken, A. C. (1935). On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburg.* 55, 42-48.

Bai, Z.-D. and He, X. (1999). Asymptotic distributions of the maximal depth estimators for regression and multivariate location. *The Annals of Statistics.* **27**, 1616-1637.

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

Chen, L-A, Welsh, A. H. and Chan, W. (2001) Linear winsorized means for the linear regression model. *Statistica Sinica.* **11**, 147-172.

Cochrane, D. and Orcutt, G. H. (1949). Application of least squares regressions to relationships containing autocorrelated error terms. *Journal of the American Statistical Association*, **44**, 32-61.

De Jongh, P. J. and De Wet, T. (1985), Trimmed Mean and Bounded Influence Estimators for the Parameters of the AR(1) Process, *Communications in Statistics - Theory and Methods*, 14, 1361-1357.

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.

Fomby, T. B., Hill, R. C. and Johnson, S. R. (1984). *Advanced Econometric Methods.* New York: Springer-Verlag.

Giltinan, D. M., Carroll, R. J. and Ruppert, D. (1986), Some New Estimation Methods for Weighted Regression When There Are Possible Outliers, *Technometrics*, 28, 219-230.

Huber, P. J. (1981). *Robust Statistics.* New York: Wiley.

Jurečková, J. (1977). Asymptotic relations of M-estimates and R-estimates in linear regression model. *Annals of Statistics* 5, 464-472.

Jureckova, J. and Sen, P. K. (1987). An extension of Billingsley's theorem to higher dimension M-processes. *Kybernetica* **23**, 382-387.

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

Koenker, R. and Portnoy, S. (1990). M estimation of multivariate regression. *Journal of the American Statistical Association*, **85**, 1060-1068.

Koul, H.L. (1992). *Weighted Empiricals and Linear Models.* IMS Lecture Notes 21.

Krasker, W. S. (1985), Two Stage Bounded-Influence Estimators for Simultaneous Equations Models, *Journal of Business and Economic Statistics* 4, 432-444.

Krasker, W. S. and Welsch, R. E. (1982), Efficient Bounded Influence Regression Estimation, *Journal of the American Statistical Association* 77, 595-604.

Lai, Y.-H., Thompson, P. and Chen, L.-A. (2003). Generalized and Pseudo Generalized Trimmed Means for the Linear Regression with AR(1) Error Model. *Statistics and Probability Letter*, 67, 203-211.

Mendenhall, W. and Sincich, T. (1993). *A Second Course in Business Statistics: Regression Analysis.* New York: Macmillan Publishing Company.

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.