Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Statistics & Probability Letters 67 (2004) 203–211
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Generalized and pseudo-generalized trimmed means for the linear regression with AR(1) error model

Yi-Hsuan Lai[a], Peter Thompson[b], Lin-An Chen[a,*]

[a] *Institute of Statistics, National Chiao Tung University, 1001 Ta Hsueh Rd, Hsinchu, Taiwan*
[b] *Mathematics Department, Wabash College, Crawfordsville, IN 47933, USA*

Received April 2003

Abstract

We propose a generalized and pseudo-generalized trimmed means for the linear regression with AR(1) errors model. These will play the role of robust-type generalized and pseudo-generalized estimators for this regression model. Their asymptotic distributions are developed.

Keywords: Generalized estimation; Linear regression; Trimmed mean

1. Introduction

For some regression models such as linear regression with AR(1) errors or the seemingly unrelated regression model, the generalized least-squares estimator (GLSE) and the pseudo-generalized least-squares estimator (PGLSE) have the advantage that their variances (or asymptotic variances) are smaller than that of the least-squares estimator (LSE). However, the GLSE and the PGLSE are sensitive to departures from normality and to the presence of outliers. Hence, extending these concepts to robust estimation is an interesting topic in regression analysis. The concept of developing robust-type generalized estimators in regression analysis is not new. Koenker and Portnoy (1990) introduced this interesting idea and developed the generalized M-estimators for the estimation of regression parameters of the multivariate regression model. Although considering only generalized estimation, their approach initiated interest in robust type generalized and pseudo-generalized estimators for estimation of regression parameters. Rather than multivariate regression, we consider

* Corresponding author.
E-mail address: lachen@stat.nctu.edu.tw (L.-A. Chen).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2003.08.003

the linear regression with AR(1) errors model

$$y_i = x_i'\beta + \varepsilon_i, \quad i = 1, \ldots, n,$$
$$\varepsilon_i = \rho\varepsilon_{i-1} + e_i, \tag{1.1}$$

where $|\rho| < 1$, $e_i, i = 1, \ldots, n$ are i.i.d. variables with mean zero and variance σ^2, and x_i is a known design p-vector with value 1 in its first element. From the regression theory on the estimation of β, it is known that, when ρ is known, the GLSE and, when ρ is unknown, the PGLSE have (or asymptotically have) the same covariance matrix, which is smaller than that of the LSE. To see the sensitivity of the GLSE and the PGLSE, let $X' = (x_1, \ldots, x_n)$ and $\Omega = Cov(\varepsilon)$ with $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$, both GLSE and the PGLSE have a (asymptotic) covariance matrix of the form

$$\sigma^2(X'\Omega^{-1}X)^{-1}. \tag{1.2}$$

The sensitivity is clear from the fact that σ^2 could be arbitrarily large when e_i has a heavy tailed distribution.

The fact that σ^2 is sensitive to the error distribution motivates us to consider robust estimators that have a (asymptotic) covariance matrix of the form

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

where robustness means that γ is insensitive to heavy tailed distributions. Based on the regression quantiles of Koenker and Bassett (1978), we will introduce the generalized trimmed mean (GTM) and the pseudo-generalized trimmed mean (PGTM) to play the role of robust-type generalized and pseudo-generalized estimators for the linear regression with AR(1) errors model.

We introduce the concepts of GTM and PGTM in Section 2 and establish their large sample theory in Section 3. Finally, the proofs of the theorems are displayed in Appendix.

2. Generalized and pseudo-generalized trimmed means

For the linear regression with AR(1) errors model (1.1), to obtain a GTM 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 the given ith-dependent variable for model (1.1), assuming that $i \geqslant 2$, 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 a vector with first element 1, the following two events determined by two quantiles are equivalent:

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

and

$$(-\rho, 1)\begin{pmatrix} y_{i-1} \\ y_i \end{pmatrix} \leqslant (-\rho, 1)\begin{pmatrix} x_{i-1}' \\ x_i' \end{pmatrix}\beta(\alpha), \tag{2.2}$$

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 the 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 generalized trimmed means.

For defining the generalized trimmed means, we consider the C–O transformation on the matrix form of the linear regression with AR(1) error model of (1.1) which is

$$y = X\beta + \varepsilon,$$

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

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

Define the half matrix of Ω^{-1} as

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

The C–O transformation is

$$u = Z\beta + ((1-\rho^2)^{1/2}\varepsilon_1, e_2, e_3, \ldots, e_n)', \tag{2.3}$$

where $u = \Omega^{-1/2'} y$ and $Z = (z_1, \ldots, z_n)' = \Omega^{-1/2'} X$. It is known that the GLSE is simply the LSE of β for model (2.3).

For $0 < \alpha < 1$, the αth (sample) regression quantile of Koenker and Bassett (1978) for the linear regression with AR(1) errors model is defined as

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

where u_i and z_i' are the ith rows of u and Z, respectively. We are now ready to define a generalized trimmed mean based on regression quantiles.

Definition 2.1. Define the trimming matrix as $A_n = \text{diag}\{a_i = I(z_i'\hat{\beta}_G(\alpha_1) \leqslant u_i \leqslant z_i'\hat{\beta}_G(\alpha_2)): i = 1, \ldots, n\}$. The Koenker and Bassett's-type GTM is defined as

$$L_G(\alpha_1, \alpha_2) = (Z'A_nZ)^{-1}Z'A_nu. \tag{2.4}$$

After the development of the GTM, the next interesting problem is whether when the parameter ρ is unknown, the trimmed mean of (2.4) with ρ replaced by a consistent estimator $\hat{\rho}$, will have the same asymptotic behavior as displayed by $L_G(\alpha_1, \alpha_2)$. 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 $\hat{e} = \hat{\Omega}^{-1/2'}\varepsilon$. Let the regression quantile when the parameter ρ is unknown be defined as

$$\hat{\beta}_{\text{PG}}(\alpha) = \arg_{b \in R^p} \min \sum_{i=1}^{n} (\hat{u}_i - \hat{z}_i'b)(\alpha - I(\hat{u}_i \leqslant \hat{z}_i'b)),$$

where $\hat{u}_i$ and $\hat{z}_i'$ are ith rows of $\hat{u}$ and $\hat{Z}$, respectively.

Definition 2.2. Define the trimming matrix as $\hat{A}_n = \text{diag}\{a_i = I(\hat{z}_i'\hat{\beta}_{\text{PG}}(\alpha_1) \leqslant \hat{u}_i \leqslant \hat{z}_i'\hat{\beta}_{\text{PG}}(\alpha_2)): i = 1, \ldots, n\}$. The Koenker and Bassett's-type PGTM is defined as

$$L_{\text{PG}}(\alpha_1, \alpha_2) = (\hat{Z}'\hat{A}_n\hat{Z})^{-1}\hat{Z}'A_n\hat{u}.$$

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 et al., 1984, p. 210–211), makes the study of asymptotic theory for $\hat{\beta}_{\text{PG}}(\alpha)$ and PGTM $L_{\text{PG}}(\alpha_1, \alpha_2)$ similar to what we have for the classical regression quantile and trimmed mean for linear regression. Large sample representations of the GTM and the PGTM and their role as generalized and pseudo-generalized robust estimators will be introduced in the next section.

3. Asymptotic theory of GTM and PGTM

We state a set of assumptions (a1–a5) related to the design matrix X and the distribution of the error variable e in the Appendix that are assumed to be true throughout the paper. Denote the distribution function of $(1-\rho)e$ by F_ρ. In the following, we give a Bahadur representation for the generalized regression quantile which follows in a straightforward way from Theorem 3 of Ruppert and Carroll (1980).

Lemma 3.1. *The generalized regression quantile has the representation*,

$$n^{1/2}(\hat{\beta}_G(\alpha) - \beta(\alpha)) = Q_\rho^{-1} f^{-1}(F_\rho^{-1}(\alpha)) n^{-1/2} \sum_{i=1}^{n} z_i(\alpha - I(e_i \leqslant F_\rho^{-1}(\alpha))) + o_p(1),$$

where $Q_\rho = \lim_{n\to\infty} X'\Omega^{-1}X$ and $F_\rho^{-1}(\alpha) = (1-\rho)^{-1}F^{-1}(\alpha)$. *Furthermore*, $n^{1/2}(\hat\beta_G(\alpha) - \beta(\alpha))$ *has a normal asymptotic distribution with mean zero vector and covariance matrix*

$$\alpha(1-\alpha)f^{-2}(F_\rho^{-1}(\alpha))Q_\rho^{-1}.$$

In accordance with (1.3), the quantile estimator $\hat\beta_G(\alpha)$ has asymptotic covariance of the form $\gamma(X'\Omega X)^{-1}$ with $\gamma = \alpha(1-\alpha)f^{-2}(F_\rho^{-1}(\alpha))$ which is then asymptotically a generalized estimator of $\beta(\alpha)$, the population regression quantile for the linear regression with AR(1) error model. The representation of $L_G(\alpha_1, \alpha_2)$ is also a direct result of Theorem 4 of Ruppert and Carroll (1980).

Theorem 3.2. *The GTM has the following representation*:

$$n^{1/2}(L_G(\alpha_1,\alpha_2) - (\beta + \lambda Q_\rho^{-1}\theta_x)) = \frac{1}{\alpha_2-\alpha_1} Q_\rho^{-1} n^{-1/2}\sum_{i=1}^{n} z_i(\phi(e_i) - E(\phi(e))) + o_p(1),$$

where $\lambda = \frac{1-\rho}{\alpha_2-\alpha_1}\int_{F_\rho^{-1}(\alpha_1)}^{F_\rho^{-1}(\alpha_2)} ef(e)de$, $\theta_x = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} x_i$ *and*

$$\phi(e) = \begin{cases} F_\rho^{-1}(\alpha_1) & \text{if } e < F_\rho^{-1}(\alpha_1) \\ e & \text{if } F_\rho^{-1}(\alpha_1) \leqslant e \leqslant F_\rho^{-1}(\alpha_2). \\ F_\rho^{-1}(\alpha_2) & \text{if } e > F_\rho^{-1}(\alpha_2) \end{cases}$$

The above theorem shows that the GTM is a generalization of the trimmed mean from the linear regression model with i.i.d. errors to that with AR(1) errors.

Corollary 3.3. *The normalized GTM* $n^{1/2}(L_G(\alpha_1,\alpha_2) - (\beta + \lambda(1-\rho)\theta_x))$ *has an asymptotic normal distribution with zero mean vector and asymptotic covariance matrix*

$$\sigma^2(\alpha_1,\alpha_2)Q_\rho^{-1},$$

where $\sigma^2(\alpha_1,\alpha_2) = (\alpha_2-\alpha_1)^{-2}\left[\int_{F_\rho^{-1}(\alpha_1)}^{F_\rho^{-1}(\alpha_2)} (e-\lambda)^2\, dF(e) + \alpha_1(F_\rho^{-1}(\alpha_1)-\lambda)^2 + (1-\alpha_2)(F_\rho^{-1}(\alpha_2)-\lambda)^2 - (\alpha_1 F_\rho^{-1}(\alpha_1) + (1-\alpha_2)F_\rho^{-1}(\alpha_2))^2\right]$.

The asymptotic covariance matrix of $L_G(\alpha_1,\alpha_2)$ is also of the form $\gamma(X'\Omega X)^{-1}$ with $\gamma = \sigma^2(\alpha_1,\alpha_2)$ which is the asymptotic variance of the trimmed mean for the location model. If we center the columns of X so that θ_x has all but the first element equal to 0, then the asymptotic bias affects the intercept alone and not the slope.

In the case where F is symmetric at 0, the asymptotic distribution of the GTM can be simplified.

Corollary 3.4. *If* F *is symmetric at zero and we let* $\alpha = \alpha_1 = 1-\alpha_2$ *then* $n^{1/2}(L_G(\alpha, 1-\alpha) - \beta)$ *has an asymptotic normal distribution with zero mean vector and asymptotic covariance matrix* $\sigma^2(\alpha, 1-\alpha)Q_\rho^{-1}$, *where* $\sigma^2(\alpha, 1-\alpha) = (1-2\alpha)^{-2}\left[\int_{F_\rho^{-1}(\alpha)}^{F_\rho^{-1}(1-\alpha)} e^2\, \mathrm{d}F(e) + 2\alpha(F_\rho^{-1}(\alpha))^2\right]$.

How efficient is the GTM compared with the GLSE? Ruppert and Carroll (1980) computed the values of the term $\sigma^2(\alpha, 1-\alpha)$ for e following several contaminated normal distributions. In comparisons of it with σ^2, the variance of e, the GTM is strongly more efficient than the GLSE when the contaminated variance is large. Along with the results in Huber (1981) and Welsh (1987), Huber's M-estimator and Welsh's trimmed mean defined on model (2.3) are expected to have the same asymptotic distribution as in Corollary 3.3. These then serve as other types of generalized robust estimators. In general, the parameter ρ is unknown. An interesting question is then whether the PGTM has the same representation as that of the GTM. Before we state this result, we need to give a representation of the regression quantile $\hat{\beta}_{\mathrm{PG}}(\alpha)$.

Lemma 3.5. *The regression quantile $\hat{\beta}_{\mathrm{PG}}(\alpha)$ has the representation,*

$$n^{1/2}(\hat{\beta}_{\mathrm{PG}}(\alpha) - \beta(\alpha)) = Q_\rho^{-1} f^{-1}(F_\rho^{-1}(\alpha)) \left[n^{-1/2} \sum_{i=1}^{n} z_i(\alpha - I(e_i \leqslant F_\rho^{-1}(\alpha))) + f(F_\rho^{-1}(\alpha))\theta_z n^{1/2}(\hat{\rho} - \rho) F_\rho^{-1}(\alpha) \right] + o_p(1),$$

where $\theta_z = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} z_i$.

The asymptotic representation of $\hat{\beta}_{\mathrm{PG}}(\alpha)$ is not the same as that of $\hat{\beta}_G(\alpha)$. In fact, it relies on the asymptotic representation of $\hat{\rho}$. In the large sample expansion for the PGTM, we see that the representation for the part $\hat{Z}' A_n \hat{u}$ involves $n^{1/2}(\hat{\rho} - \rho)$ and $n^{1/2}(\hat{\beta}_{\mathrm{PG}}(\alpha) - \beta(\alpha))$ with $\alpha = \alpha_1$ and α_2. Since the representation of $\hat{\beta}_{\mathrm{PG}}(\alpha)$ also involves $n^{1/2}(\hat{\rho} - \rho)$, the terms with $n^{1/2}(\hat{\rho} - \rho)$ will automatically cancel out so the PGTM has a representation free of $\hat{\rho}$ in its formulation.

Theorem 3.6. *The PGTM has the same representation as that expressed for the GTM in Theorem* 3.2.

From Theorem 3.6, the PGTM indeed plays the role of a Pseudo generalized estimator for estimating the regression parameter β.

Acknowledgements

We are grateful to the editor and one referee for comments which improved the presentation of this paper. This research work was partially supported by the National Science Council of the Taiwan Grant No. NSC 91-2118-M-009-005.

Appendix

The following conditions concerning the design matrices X and H_0 and the distribution of error variable e are similar to the standard ones for linear regression models as given in

Ruppert and Carroll (1980) and Koenker and Portnoy (1990):

(a1) $n^{-1}\sum_{i=1}^{n} x_{ij}^4 = O(1)$ for all j,
(a2) $n^{-1}X'\Omega X = Q_\rho + o(1)$, where Q_ρ is positive definite matrix.
(a3) $n^{-1}\sum_{i=1}^{n} x_i = \theta_x + \mathrm{o}(1)$, where θ_x is a finite vector with first element value 1.
(a4) The probability density function and its derivative are both bounded and bounded away from 0 in a neighborhood of $F_\rho^{-1}(\alpha)$ for $\alpha \in (0,1)$.
(a5) $n^{1/2}(\hat{\rho} - \rho) = \mathrm{O_p}(1)$.

Proof of Lemma 3.5. Let

$$M(t_1,t_2) = n^{-1/2}\sum_{i=1}^{n} z_i\{\alpha - I(e_i - n^{-1/2}t_1\varepsilon_{i-1} \leqslant (z_i - n^{-1/2}t_1x_{i_1})'(n^{-1/2}t_2 + F_\delta^{-1}(\alpha)))\}.$$

We want to show that

$$\sup_{\|(t_1,t_2)\|\leqslant k} |M(t_1,t_2) - M(0,0) - F_\delta^{-1}(\alpha)f(F_\delta^{-1}(\alpha))n^{-1/2}\sum_{i=1}^{n} z_i(z_i't_2 - t_1F_\delta^{-1}(\alpha))| = \mathrm{o_p}(1). \quad (4.1)$$

By letting, for $k > 0$, $S_n(t_1,t_2) = M(t_1,t_2) - M(0,0)$, we will prove (4.1) in two steps. In the first step, we will show that

$$\sup_{\|(t_1,t_2)\|\leqslant k} |S_n(t_1,t_2) - ES_n(t_1,t_2)| = \mathrm{o_p}(1) \quad (4.2)$$

based on Lemma 3.2 in Bai and He (1999).

Now we prove (4.2) by checking the three conditions L_1, L_2 and L_3 in the hypothesis of Lemma 3.2 in Bai and He (1999). First, we prove

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} z_i'z_iE|&I(e_i - n^{-1/2}t_1\varepsilon_{i-1} \leqslant (z_i - n^{-1/2}t_1x_{i-1})'(n^{-1/2}t_2 + F_\delta^{-1}(\alpha)) \\ &- I(e_i - n^{-1/2}t_1^*\varepsilon_{i-1} \leqslant (z_i - n^{-1/2}t_1^*x_{i-1})'(n^{-1/2}t_2^* + F^{*-1}(\alpha))| \\ &\leqslant M(\|t_1 - t_1^*\| + \|t_2 - t_2^*\|), \quad \text{for some } M > 0. \end{aligned} \quad (4.3)$$

Define

$$\begin{aligned} A = n^{-1}\sum_{i=1}^{n} z_i'z_iE|&I(e_i - n^{-1/2}t_1\varepsilon_{i-1} \leqslant (z_i - n^{-1/2}t_1x_{i-1})'(n^{-1/2}t_2 + F_\delta^{-1}(\alpha)) \\ &- I(e_i - n^{-1/2}t_1^*\varepsilon_{i-1} \leqslant (z_i - n^{-1/2}t_1^*x_{i-1})'(n^{-1/2}t_2 + F_\delta^{-1}(\alpha)))| \end{aligned}$$

and

$$\begin{aligned} B = n^{-1}\sum_{i=1}^{n} z_i'z_iE|&I(e_i - n^{-1/2}t_1^*\varepsilon_{i-1} \leqslant (z_i - n^{-1/2}t_1^*x_{i-1})'(n^{-1/2}t_2 + F_\delta^{-1}(\alpha)) \\ &-I(e_i - n^{-1/2}t_1^*\varepsilon_{i-1} \leqslant (z_i - n^{-1/2}t_1^*x_{i-1})'(n^{-1/2}t_2^* + F_\delta^{-1}(\alpha)))|. \end{aligned}$$

Represent $A = A_1 + A_2$ as follows:

$$
\begin{aligned}
A &= n^{-1}\sum_{i=1}^{n} z_i' z_i EI(e_i - n^{-1/2}t_1\varepsilon_{i-1} \leqslant (z_i - n^{-1/2}t_1 x_{i-1})'(n^{-1/2}t_2 + F_\delta^{-1}(\alpha)),\\
&\quad e_i - n^{-1/2}t_1^*\varepsilon_{i-1} > (z_i - n^{-1/2}t_1^* x_{i-1})'(n^{-1/2}t_2 + F_\delta^{-1}(\alpha)))\\
&\quad + n^{-1}\sum_{i=1}^{n} z_i' z_i EI(e_i - n^{-1/2}t_1\varepsilon_{i-1} > (z_i - n^{-1/2}t_1 x_{i-1})'(n^{-1/2}t_2 + F_\delta^{-1}(\alpha)),\\
&\quad e_i - n^{-1/2}t_1^*\varepsilon_{i-1} \leqslant (z_i - n^{-1/2}t_1^* x_{i-1})'(n^{-1/2}t_2 + F_\delta^{-1}(\alpha)))\\
&= A_1 + A_2.
\end{aligned}
$$

Let $\xi_n = n^{1/2}t_2 + F^{*-1}(\alpha)$ and $U_{i-1} = \varepsilon_{i-1} - x_{i-1}'\xi_n$. Then,

$$
\begin{aligned}
A_1 &= n^{-1}\sum_{i=1}^{n} z_i' z_i EI(e_i \leqslant z_i'\xi_n - n^{-1/2}t_1 U_{i-1}, e_i > z_i'\xi_n - n^{-1/2}t_1^* U_{i-1})\\
&= n^{-1}\sum_{i=1}^{n} z_i' z_i E\{f(z_i'\xi_n) n^{-1/2}\|t_1 - t_1^*\| U_{i-1}\}\\
&\leqslant M n^{-1/2}\|t_1 - t_1^*\|.
\end{aligned}
$$

Similarly, $A_2 \leqslant Mn^{-1/2}\|t_1 - t_1^*\|$ and $B \leqslant Mn^{-1/2}\|t_2 - t_2^*\|$. Hence (4.3) holds and so does condition (L1) in the hypothesis of Lemma 3.2 in Bai and He (1999). Condition (L2) is satisfied automatically since the indicator function is bounded.

Next, similar arguments to those used to prove (4.3) can be used to prove that

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n} z_i' z_i E\left\{\sup_{\|t_1 - t_1^*\| + \|t_2 - t_2^*\| \leqslant d} |I(e_i - n^{-1/2}t_1\varepsilon_{i-1} \leqslant (z_i - n^{-1/2}t_1 x_{i-1})'(n^{-1/2}t_2\right.\\
&\quad \left. + F_\delta^{-1}(\alpha))) - I(e_i - n^{1/2}t_1^*\varepsilon_{i-1} \leqslant (z_i - n^{-1/2}t_1^* x_{i-1})'(n^{-1/2}t_2^* + F_\delta^{-1}(\alpha)))\right\}
\end{aligned}
$$

is bounded by $Mn^{-1/2}d$, which implies that condition (L3) holds. Therefore, from Lemma 3.2 in Bai and He (1999), we obtain

$$\sup_{\|(t_1,t_1)\| \leqslant K} |S_n(t_1,t_2) - ES_n(t_1,t_2)| = o_p(1). \tag{4.4}$$

On the other hand, following the technique of Chen et al. (2001), we get that

$$\sup_{\|(t_1,t_2)\| \leqslant k} |E(S_n(t_1,t_2)) - F_\delta^{-1}(\alpha) f(F_\delta^{-1}(\alpha)) n^{-1/2}\sum_{i=1}^{n} z_i(z_i' t_2 - t_1 F_\delta^{-1}(\alpha))| = o_p(1). \tag{4.5}$$

Combining (4.2) and (4.5), statement (4.1) holds. Using the method of Jurečková (1977, Lemmas (4.2) and (4.1)) again, $n^{1/2}(\hat{\beta}(\alpha) - \beta(\alpha)) = O_p(1)$ is obtained. Thus, Lemma 3.5 is proved. □

Proof of Theorem 3.6. The PGTM can be formulated as

$$n^{1/2}(L_{\mathrm{PG}}(\alpha_1,\alpha_2)-\beta)=(n^{-1}\hat{Z}'A_n\hat{Z})^{-1}n^{-1/2}\hat{Z}'A_n\hat{e}.$$

Since $n^{1/2}(\hat{\rho}-\rho)=\mathrm{O_p}(1)$, we have $n^{-1/2}\hat{Z}'A_n\hat{e}=n^{-1/2}\hat{Z}'A_n e+\mathrm{o_p}(1)$. By letting $M(t_1,t_2,\alpha)=n^{-1/2}\sum_{i=1}^n z_i e_i I(e_i-n^{-1/2}t_1\varepsilon_{i-1}\leqslant F_\delta^{-1}(\alpha)+n^{-1/2}(z_i+n^{-1/2}t_1x_{i-1})'t_2+n^{-1/2}t_1F_\delta^{-1}(\alpha))$, we see that

$$n^{-1/2}\hat{Z}'A_n e=M(T_1^*(\alpha_2),T_2^*,\alpha_2)-M(T_1^*(\alpha_1),T_2^*,\alpha_1) \tag{4.6}$$

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

$$M(T_1,T_2,\alpha)-M(0,0,,\alpha)=F^{*-1}(\alpha)f(F_\delta^{-1}(\alpha))n^{-1/2}\sum_{i=1}^n z_i(z_i'T_2-T_1F_\delta^{-1}(\alpha))+\mathrm{o_p}(1) \tag{4.7}$$

for any sequences $T_1=\mathrm{O_p}(1)$ and $T_2=\mathrm{O_p}(1)$. Then, from Lemmas 3.1, 4.6 and 4.7, we have

$$\begin{aligned}n^{-1/2}\hat{Z}'A_n e&=n^{-1/2}\sum_{i=1}^n z_i[e_iI(F_\delta^{-1}(\alpha_1)\leqslant e_i\leqslant F_\delta^{-1}(\alpha_2))+F_\delta^{-1}(\alpha_2)(\alpha_2-I(e_i\\&\qquad\leqslant F_\delta^{-1}(\alpha_2)))-F_\delta^{-1}(\alpha_1)(\alpha_1-I(e_i\leqslant F_\delta^{-1}(\alpha_1)))]+\mathrm{o_p}(1).\end{aligned} \tag{4.8}$$

Also, a similar discussion of the proof for Lemma 3.5 provides the result

$$n^{-1}\hat{Z}'A_n\hat{Z}=Q_\rho+\mathrm{o_p}(1). \tag{4.9}$$

Then (4.8) and (4.9) imply the theorem. □

References

Bai, Z.-D., He, X., 1999. Asymptotic distributions of the maximal depth estimators for regression and multivariate location. Ann. Statist. 27, 1616–1637.

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

Cochrane, D., Orcutt, G.H., 1949. Application of least squares regressions to relationships containing autocorrelated error terms. J. Amer. Statist. Assoc. 44, 32–61.

Fomby, T.B., Hill, R.C., Johnson, S.R., 1984. Advanced Econometric Methods. Springer, New York.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

Jurečková, J., 1977. Asymptotic relations of *M*-estimates and *R*-estimates in linear regression model. Ann. Statist. 5, 464–472.

Koenker, R.W., Bassett, G.W., 1978. Regression Quantiles. Econometrica 46, 33–50.

Koenker, R., Portnoy, S., 1990. *M* estimation of multivariate regression. J. Amer. Statist. Assoc. 85, 1060–1068.

Ruppert, D., Carroll, R.J., 1980. Trimmed least squares estimation in the linear model. J. Amer. Statist. Assoc. 75, 828–838.

Welsh, A.H., 1987. The trimmed mean in the linear model. Ann. Statist. 15, 20–36.