Robust Estimation of Regression Models with AR(1) Error

Student: Yi-Hsuan Lai　　Advisor: Lin-An Chen

Institute of Statistics

National Chiao Tung University

ABSTRACT

In this thesis we focus on the linear and non-linear regression with AR(1) error models. In the linear regression part, the generalized median estimator, feasible generalized median estimator, generalized trimmed mean (GTM) and the feasible generalized trimmed mean (FGTM) are proposed. Before defining the estimators, we use the covariance matrix of the error terms to do a Cochran-Orcutt transformation on the regression model such that the transformed one is a usual linear regression model with i.i.d. error variables. Then we discuss the robust estimators of this transformed model.

The generalized and feasible generalized median estimators are defined by the l_1 norm method. So they are robust to outliers. For the generalized and feasible generalized trimmed mean estimators, we apply sample regression quantile which is defined by Koenker and Bassett (1978) to trim data first and then define the estimators based on the rest of data. Due to trimming, these estimators are robust to outliers also. Besides the linear regression with AR(1) error model, we extended the idea of trimmed mean to introduce generalized and feasible generalized trimmed means for the nonlinear regression with AR(1) error model.

The corresponding Bahadur representations and asymptotic normality are proved in this thesis also. Besides these theoretical results, we also do simulations to discuss the effect of the estimation of correlation coefficients ρ to the model. And an application on a real data set is also given.

Key words and phrases : robust, AR(1), generalized median estimators, generalized trimmed mean estimators, Bahadur representation, regression quantile, the linear and non-linear regression.

Acknowledgements

I am very grateful to my advisor, Prof. Lin-An Chen, for his guidance and enthusiastic support. I also thank professors Hui-Nien Hung, Wen-Han Huang, Shin-Chen Huang, and Berlin Wu for serving as members of the committee and providing thoughtful suggestions. I would like to thank my family for kindly mental and financial support. In addition, I thank all my colleagues for their mental support and the pleasant learning environment.

Contents

List of Tables

List of Notations

y	observation
x_i	design point corresponding to y_i
ϵ_i	error variable
β	parameter vector
ρ	correlation coefficient
$e_i^\prime s$	i.i.d. random variables
Ω	covariance matrix
u	Cochran-Orcutt transformation of vector y
Z	Cochran-Orcutt transformation of vector X
$r\hat{h}o$	estimator of ρ
$\hat{\beta}_G$	the generalized median estimator
$\hat{\beta}_{FG}$	the feasible generalized median estimator
F	distribution function of e_i
f	density function of F
Q_ρ	the limit matrix of $n^{-1}X^\prime\Omega X$
$\hat{\beta}_G(\alpha)$	the α-th regression quantile
A_n	the trimming matrix
$L_G(\alpha_1, \alpha_2)$	the Koenker and Bassetts type generalized trimmed mean
$L_{MFG}(\alpha_1, \alpha_2)$	the Mallows-type bounded influence FGTM
$\eta_n(\alpha)$	the α-th residual quantile

Chapter 1

Introduction

1.1 Why Robustness?

Statistical inferences are based on observations and prior assumptions about the underlying situation. There are explicit or implicit assumptions about randomness and independence, about distributional models, prior distributions for some unknown parameters, and so on. In model selection one needs to look critically at the assumptions and examine the effects of plausible departures from model assumptions in various possible directions. Most classical statistical procedures are based on two assumptions, that the sample observations are independently and identically distributed, and that the underlying distribution is normal. While the first assumption may not be unrealistic in certain situations, it is the second assumption that is rather unrealistic from a practical point of view. Nonnormal distributions are more prevalent in practice, and to assume normality instead might lead to erroneous statistical inferences.

What is desired is an inference procedure which in some sense does almost as well as possible if the assumption is true, but does not perform much worse within a range of alternatives to the assumption. Such a procedure is **robust** to departures from the assumption on which it is based. For example, a robust to deviations from normality procedure is defined to be one that is nearly as efficient as the classical procedure for a normal distribution but is considerably more efficient overall for nonnormal distributions.

There are several definitions of robustness. First question we have is from what deviations from assumptions we are seeking robustness. The most usual answer is deviations due to a different form of p. d. f. $f(x;\theta)$.

Robustness with respect to outliers is known as being resistant to outliers. Another question in studies of robustness is the range of distributions over which a procedure should be robust. The general shape of a data set should allow one to narrow the range considerably.

Statisticians have been receptive to the basic notion of robustness and its effective implementations in statistical methodology. A significant contributor to the development of robust statistical procedure is Peter J. Huber, and his graduate level textbook (1981) provides a good coverage of the basic theory developed during the 1960s and 1970s. The best-known approach to robust estimation has its roots in Huber (1964). Parametric results began to be questioned in the 1950s, when asymptotic methods developed to the point where the inefficiency of the t-test to the Wilcoxon test could be demonstrated [Pitman(1949), Hodges and Lehmann (1956)]. The sensitivity of some classical estimators was documented by Tukey (1960).

Other factors also contributed to the growing interest in robustness. Hampel (1973) has given a comprehensive list of them and a provocative introduction to robust methods; the reader is encouraged to read his essay. More statisticians are now familiar with the rudiments of functional analysis and computing. Huber (1964) and Hampel (1968) both have employed ideas from functional analysis in their pioneering work; and the Princeton robustness study [Andrews et al. (1972)] is typical of the computational approach to gaining insight into robust proposals. The idea is widespread accepted that every new statistical study is incomplete without some serious attention to the effect of slight changes in the assumptions. One reason for the continued proliferation of such work is the almost universal applicability and intuitive appeal of Hampels influence function and breakdown point (1971, 1973).

Two monographs, by Serfling (1980) and Sen (1981a) contain some detailed discussions on the asymptotic properties of robust estimators and test statistics for the location as well as simple linear models. The book by Hampel, Ronchetti, Rousseeuw, and Stahel (1986) addresses some of the robustness issues from (mostly) a finite sample point of view. In this perspective robustness against plausible departures from the assumed form of the error distribution, heteroscadasticity of the errors and lack of independence of these error components (serial dependence, interclass correlations etc.), and presence of outliers (or error-contamination) have all been identified as principal issues. Two other monographs, by Shorack and Wellner (1986) and Koul (1992) on asymptotics of the (weighted) empirical process, have touched on robustness primarily from an asymptotic point of view. The book by Rieder (1994) is devoted solely to robust asymptotic statistics, but it emphasizes mostly the infinitesimal concept of robustness and nonparametric optimality based on the notion of least favorable local alternatives.

There are other perceptions and definition of robustness, which in application seem to be limited to long-tailed symmetric distributions, and the reader is referred to the authoritative works of Hoaglin et al. (1983) and Huber (1981).

1.2 Regression Quantile

Regression analysis is an important statistical tool that is routinely applied in most sciences. Out of many possible regression techniques, the least squares (LS) method has been generally adopted because of tradition and ease of computation. However, there is presently a widespread awareness of the dangers posed by the occurrence of outliers. Outliers occur very frequently in real data, and they often go unnoticed because nowadays computers process much data. Not only the response variable can be outlying, but also the explanatory part, leading to so-called leverage points. Both types of outliers may totally spoil an ordinary LS analysis.

To avoid this problem, new statistical techniques have been developed that are not so easily affected by outlieers. These are the robust methods, the results of which remain trustworthy even if a certain amount of data is contaminated. The outliers are far away from the robust fit and hence can be detected by their large residuals from it, whereas the standardized residuals from ordinary LS may not expose outliers at all.

Among other noteworthy developments over the past 15 years, regression quantiles, have steadily reshaped the domain and scope of robust statistics. For sample quantiles the celebrated Bahadur (1966) representation, provided a novel approach to the study of the asymptotic theory of order statistics, quantile functions, and a broad class of statistical functionals. Some of these developments are reported in Serfling (1980, ch. 2), Sen

(1981a, ch.7) and other contemporary advanced monographs. Bahadurs (1966) own results, as further extended by Kiefer (1967) and supplemented by Ghosh (1971) in a weaker and yet elegant form, let the way to various types of representations for statistics and estimators.

It started with a humble aim of regression L-estimators by Koenker and Bassett (1978) and traversed the court of robustness onto the domain of regression rank scores, see Gutenbrunner and Jureckova (1992). In this way it provides a natural link to various classes of robust estimators and strengthens their interrelationships as well. The asymptotic theory is further streamlined to match the needs of practical applications. In this respect it would be helpful for readers to have familiarity with the basic theory of robustness e.g., the introductory discussion in Huber (1981).

1.3 Summary

In this paper we focus on the linear/non-linear regression with AR(1) error models. In the linear regression part, the generalized trimmed mean and the feasible generalized trimmed mean are proposed. They are based on the regression quantile of Koenker and Bassett (1978). The first step we do to this model is a transformation. We use the covariance matrix of the error term to do a Cochrane0Orcutt transformation such that the model with AR(1) error has the usual linear regression form. Then we discuss the regression quantile estimator of this transformed model.

Due to the AR(1) error model the autoregressive parameter ρ appears in the transformed design matrix. If ρ is known, the transformed design matrix is known. When the autoregressive parameter ρ is unkown, we replace it by a $\sqrt{n}-$consistent estimator $\hat{\rho}$.

Then from the transformed model, we get the $\alpha-$th sample regression quantile. We use this estimator the design matrix to trim some data. After trimming, the general trimmed mean (GTM) is defined. Due to trimming, this is a robust estimator of the regression coefficient vector which is robust to outlier. If ρ is replaced by the $\sqrt{n}-$consistent estimator $\hat{\rho}$, the feasible generalized trimmed mean (FGTM) is defined.

After the definitions of the robust estimators (GTM, FGTM), the large sample theory of the estimators are given. After some regular assumptions, we prove the Bahadur representations of the generalized regression quantile, GTM, feasible generalized regression quantile, and also the asymptotic normality of the GTM and FGTM.

Besides, the linear regression with AR(1) error model, we apply the idea of trimmed mean to introduce generalized and feasible generalized trimmed means for the nonlinear regression with AR(1) error model. We show that these estimators are asymptotically more efficient than the trimmed means. These results extend the concept of generalized and feasible generalized least squares estimators for linear regression with AR(1) error model to the robust estimators for nonlinear regression models.

The organization of this paper is as follows. In Chapter 2 criterions are introduced, such as maximum (asymptotic) bias, (asymptotic) breakdown point, influence function, asymptotic variance under mixed normal, minimax bias. A short introduction of the robust techniques is given in Chapter 3. We focus on R-, M-, L-estimators, minimum distance estimator, and also Kolmogorove MD estimator. For more advanced results on robust estimations, reader is referred to Jureckova and Sen (1996). Main results corresponding to generalized and feasible generalized trimmed median estimators for linear regression with AR(1) error models are given in Chapter 4. The corresponding results to generalized and feasible generalized trimmed means for linear regression with AR(1) error models are given in Chapter 5. And the extension to the nonlinear regression with AR(1) models are given in Chapter 6.

Chapter 2

Criterions

In this chapter we would like to describe quantitatively how greatly a small change in the underlying distribution F changes the distribution $\mathfrak{L}_F(T_n)$ of an estimate $T_n = T_n(x_1, \cdots, x_n)$. In the following sections, we will give short descriptions on maximum (asymptotic) bias, asymptotic breakdown point of T at F_0, influence function, minimax bias.

2.1 Maximum asymptotic bias and asymptotic breakdown point

We assume that $T_n = T(F_n)$ derives from a functional T, where F_n is the empirical distribution. In most cases of practical interest, T_n is then consistent:

$$T_n \to T(F) \text{ in probability,}$$

and asymptotically normal

$$\mathfrak{L}_F\{\sqrt{n}[T_n - T(F)]\} \to \mathfrak{N}(0, A(F,T)).$$

Then we will discuss the quantitative large sample robustness of T in terms of the behavior of its asymptotic bias $T(F)-T(F_0)$ and asymptotic variance $A(F,T)$ in some neighborhood $\mathcal{P}_\epsilon(F_0)$ of the model distribution F_0. For instance, $\mathcal{P}_\epsilon(F_0)$ might be a Levy neighborhood

$$\{F|\forall t, F_0(t-\epsilon) - \epsilon \leq F(t) \leq F_0(t+\epsilon) + \epsilon\}$$

or a contaminated set

$$\{F|F = (1-\epsilon)F_0 + \epsilon H, H \in \mathcal{M}\},$$

where $\mathcal{M}$ is the set of all probability measure on the sample space.

The two most important characteristics then are the maximum bias

$$b_1(\epsilon) = \sup_{F \in \mathcal{P}_\epsilon} |T(F) - T(F_0)|$$

and the maximum variance

$$v_1(\epsilon) = \sup_{F \in \mathcal{P}_\epsilon} A(F, T).$$

We should like to establish that, for sufficiently large n, the estimate T_n behaves well for all $F \in \mathcal{P}_\epsilon$. Let $M(F, T_n)$ be the median of $\mathfrak{L}_F[T_n - T(F_0)]$ and let $Q_t(F, T_n)$ be a normalized $t-$quantile range of $\mathfrak{L}_F(\sqrt{n}T_n)$, which is defined as

$$Q_t(F, T_n) = \frac{\mathfrak{L}_F^{-1}(\sqrt{n}T_n)(1-t) - \mathfrak{L}_F^{-1}(\sqrt{n}T_n)(t)}{\Phi^{-1}(1-t) - \Phi^{-1}(t)}, \tag{2.1}$$

Φ being the standard normal cumulative. The value t is arbitrary, but fixed, say $t = 0.25$ (interquartile range) or $t = 0.025$ (95% range, which is convenient in view of the traditional 95% confidence intervals). For a normal distribution, Q_t coincides with the standard deviation; therefore Q_t^2 is sometimes called pseudo-variance.

Then define the maximum asymptotic bias and variance, respectively, as

$$b(\epsilon) = \lim_n \sup_{F \in \mathcal{P}_\epsilon} |M(F, T_n)|,$$

$$v(\epsilon) = \lim_n \sup_{F \in \mathcal{P}_\epsilon} Q_t(F, T_n)^2.$$

Note that, $\mathcal{P}_1 = \mathcal{M}$ is the set of all probability measures on the sample space, so $b(1)$ is the worst possible value of b (usually ∞). We define the asymptotic breakdown point of T at F_0 as

$$\epsilon^* = \epsilon^*(F_0, T) = \sup\{\epsilon | b(\epsilon) < b(1)\}.$$

Roughly speaking, the breakdown point gives the limiting fraction of bad outliers the estimator can cope with.

Similarly we may also define an asymptotic variance breakdown point

$$\epsilon^{**} = \epsilon^{**}(F_0, T) = \sup\{\epsilon | v(\epsilon) < v(1)\},$$

but this is a much less useful notion.

2.2 Influence function

Consider the simple model where $X_1, X_2, \cdots, X_n$ are i.i.d. r.v.s with a d.f. F generally unknown. We have a parametric model $\{G_\theta : \theta \in \Theta\}$ formed by a dominated system of distributions and wish to estimate θ for which G_θ is as close to F as possible.

Let $T_n = T_n(X_1, X_2, \cdots, X_n)$ be an estimator of θ which we express as a functional $T(F_n)$ of the empirical d.f. F_n of $X_1, X_2, \cdots, X_n$. A suitable normed limiting influence on the value of an estimate $T(F_n)$ can be expressed as

$$IC(x, F_n, T) = \lim_{\epsilon \downarrow 0}\{\frac{T((1-\epsilon)F_n + \epsilon\delta_x) - T(F_n)}{\epsilon}\}, x \in \mathbb{R}^1,$$

where for every $t, x \in \mathbb{R}^1$, $\delta_x(t) = 0$ or 1 with respect to $x \neq t$ or $x = t$, respectively. The above quantity, considered as a function of x, has been introduced by Hampel (1968, 1974) under the name influence function.

If T is sufficiently regular and G is near F, then the leading terms of a Taylor expansion are

$$T(G) = T(F) + \int IC(x, F, T)\{G(dx)F(dx)\} + \cdots$$

We have $\int IC(x, F, T)F(dx) = 0$. Thus if we substitute the empirical distribution F_n for G in the above expansion, we obtain

$$\begin{aligned}\sqrt{n}(T(F_n) - T(F)) &= \sqrt{n}\int IC(x, F, T)F_n(dx) + \cdots \\ &= \frac{1}{\sqrt{n}}\sum IC(x_i, F, T) + \cdots\end{aligned}$$

By central limit theorem, $\frac{1}{\sqrt{n}}\sum IC(x_i, F, T)$ is asymptotically normal with mean 0, if x_i are independent with common distribution F. Moreover, the remaining terms are asymptotically negligible. Thus, $\sqrt{n}(T(F_n) - T(F))$ is then asymptotically normal with mean 0 and variance

$$A(F, T) = \int IC(x, F, T)^2 F(dx). \tag{2.2}$$

The influence function has two main uses. First, it allows us to assess the relative influence of individual observations toward the value of an estimate or test statistic. Its maximum absolute value,

$$\gamma^* = \sup_x |IC(x, F, T)|,$$

has been called gross error sensitivity by Hample. It is related to the maximum bias:

$$T(F) - T(F_0) \cong \epsilon \int IC(x, F_0, T)F(dx).$$

Hence

$$b_1(\epsilon) = \sup |T(F) - T(F_0)| \approxeq \epsilon\gamma^*.$$

Second, the influence function allows an immediate and simple, heuristic assessment of the asymptotic properties of an estimate, since it allows us to guess an explicit formula (2.2) for asymptotic variance.

2.3 Asymptotic variance

Let $(F_\theta)_{\theta\in\Theta}$ be a parametric family of distributions, and let the functional T be a Fisher consistent estimate of θ, that is

$$T(F_\theta) = \theta, \text{ for all } \theta.$$

Assume that T is Frechit differentiable at F. We would like to show that the corresponding estimate is asymptotically efficient at F_θ if and only if its influence function satisfies

$$IC(x; F_\theta, T) = \frac{1}{I(F_\theta)} \frac{\partial}{\partial\theta}(\log f_\theta). \tag{2.3}$$

Here, f_θ is the density of F_θ, and

$$I(F_\theta) = \int (\frac{\partial}{\partial\theta} \log f_\theta)^2 dF_\theta$$

is the Fisher information.

Assume that $d_L(F_\theta, F_{\theta+\delta}) = O(\delta)$, that

$$\frac{f_{\theta+\delta} - f_\theta}{\delta f_\theta} \to \frac{\partial}{\partial\theta} \log f_\theta \tag{2.4}$$

converges in the $L_2(F_\theta)$-sense, and that

$$0 < I(F_\theta) < \infty.$$

Then by definition of the Frechet derivative,

$$\begin{aligned} T(F_{\theta+\delta}) - T(F_\theta) - \int IC(x; F_\theta, T)(f_{\theta+\delta} - f_\theta)dx &= o(d_L(F_\theta, F_{\theta+\delta})) \\ &= o(\delta). \end{aligned}$$

We devide above by δ and let $\delta \to 0$. In view of (2.3) and (2.4) we obtain

$$\int IC(x; F_\theta, T)\frac{\partial}{\partial\theta}(\log f_\theta)f_\theta dx = 1. \tag{2.5}$$

The Schwartz inequality applied to (2.5) gives that the asymptotic variance $A(F_\theta, T)$ of $\sqrt{n}(T(F_n) - T(F_\theta))$ satisfies

$$A(F_\theta, T) = \int IC(x; F_\theta, T)^2 dF_\theta \geq \frac{1}{I(F_\theta)}; \tag{2.6}$$

and that he equality in (2.6) holds (i.e., asymptotic efficiency) only if $IC(x; F_\theta, T)$ is proportional to $\frac{\partial}{\partial\theta}\log f_\theta$. The factor of proportionality gives the result announced in (2.3).

2.4 Minimax asymptotic variance/bias

Assume that the true underlying distribution F lies I some neighborhood $\mathcal{P}_\epsilon$, of the assumed model distribution F_0, that the observations are independent with common distribution $F(x-\theta)$, and that the location parameter θ is to be estimated. In this section we would like to optimize the robustness properties of such a location estimate by minimizing its maximum asymptotic bias $b(\epsilon)$ for distributions $F \in \mathcal{P}_\epsilon$.

We now construct two ϵ-contaminated normal distributions F_+ and F_-, which are sy Metric about x_0 and $-x_0$, respectively, and which are translates of each other. F_+ is given by its density

$$f_+(x) = \begin{cases} (1-\epsilon)\phi(x) & \text{for } x \leq x_0, \\ (1-\epsilon)\phi(x - 2x_0) & \text{for } x > x_0, \end{cases} \tag{2.7}$$

where $\phi = \Phi'$ is the standard normal density, and

$$F_-(x) = F_+(x + 2x_0).$$

Thus $T(F_+) - T(F_-) = 2x_0$ for any translation invariant functional, and it is evident that none can have an absolute bias smaller than x_0 at F_+ and F_- simultaneously. This means that the median achieves the smallest maximum bias among all translation invariant functionals. Moreover, for symmetric unimodal distributions, the solution in variably is the sample median.

Let F_0 be the distribution having the smallest Fisher information

$$I(f) = \int (\frac{f'}{f})^2 f dx$$

among the members of $\mathcal{P}$.

For any sequence (T_n) os estimates, the asymptotic variance of $\sqrt{n}T_n$ at F_0 is at best $1/I(F_0)$. Thus if we can find a sequence (T_n) such that its asymptotic variance does not exceed $1/I(F_0)$ for any $F \in \mathcal{P}$, we have clearly solved the minimax problem. In particular, this sequence (T_n) must be asymptotically efficient for F_0.

Chapter 3

Robust techniques

3.1 Introduction

In this chapter we place emphasis on the motivations of several important classes of robust estimators and on their basic properties. We will consider two general situations, Location Models and Regression Models.

Location models. Let $X_1, \cdots, X_n$ be $n(\geq 1)$ i.i.d. random variables with unknown distribution function G, defined on the real line $\mathbb{R}^1$. Let $\{F_\theta : \theta \in \Theta\}$ be a parametric family of d.f.s and we wish to estimate θ for which F_θ provides the closest approximation of G. For the location model, we assume that

$$F_\theta(x) = F_0(x - \theta),$$

where θ is real and F_0 belongs to a class $\mathbb{F}_0$.

Regression models. Suppose that $X_1, \cdots, X_n$ are independent r.v.s where X_i has d.f. $F(x-\theta_i)$, for $i = 1, \cdots, n$, and the vector $\theta = (\theta_1, \cdots, \theta_n)$ of parameters satisfies the condition that for some p $(1 \leq p \leq n)$,

$$\theta \in \Pi_p,$$

where Π_p a linear p-dimensional subspace of $\mathbb{R}^n$. We conceive of a class $\mathbb{F}$ of distributions, and assume that $F \in \mathbb{F}$. The choice of $\mathbb{F}$ has an important bearing on the choice of robust estimators for the corresponding models. Note that the location model is a special case of the regression model for which $\theta = \theta_1$, for $\theta \in \Theta \subset \mathbb{R}^1$.

Among various robust estimators, three broad classes, namely the M-, L- and R-estimators, have turned out to be the most interesting, and they have been studied extensively in the literature. In Section 3.2, we consider the basic formulation of the M-estimators. They are well defined for a variety of models for which maximum likelihood estimators (MLE) are also defined. M-estimators cover both the MLE and LSE as subclasses.

In Section 3.3, R-estimators are considered. These estimators are based on the ranks of observations (or signed ranks), and generally correspond to suitable rank tests for symmetry or randomness against shift or regression alternative. We will consider the R-estimators of location and regression parameters. Section 3.4 is devoted to the study of L-estimators. These L-estimators were originally conceived as linear combinations of functions of order statistics for efficient estimation of location or scale parameters. Generally, they are computationally appealing and possess various desirable properties. L-estimators have also been considered for linear models.

Besides these principal classes of robust estimators, some other notable classes have been studied a lot, such as Minimum distance estimators (MDE). We will give a short introduction in Section 3.5.

3.2 M-Estimator

Let $X_1, \cdots, X_n$ be i.i.d. random variables with a distribution function (d.f.) $F(x, \theta)$ where $\theta \in \Theta$, an open set in $\mathbb{R}^p$. The true value of θ is denoted by θ_0. Let $\rho(x, \theta) : \mathbb{R} \times \Theta \to \mathbb{R}$ be an absolutely continuous function in the elements of θ and such that the function

$$h(\theta) = \mathbb{E}_{\theta_0} \rho(X_1, \theta) \tag{3.1}$$

exists for all $\theta \in \Theta$ and has a unique minimum over Θ at θ_0. An M-estimator (maximum likelihood type estimator) M_n of θ_0 is defined as a statistic which is a solution (with respect to θ) of the minimization of $\sum_{i=1}^n \rho(X_i, \theta)$ with respect to $\theta \in \Theta$, i.e.,

$$M_n = \arg\min\{\sum_{i=1}^n \rho(X_i, \theta) : \theta \in \Theta\}. \tag{3.2}$$

In many cases, the minimization problem in (3.2) leads to the implicit equation

$$\sum_{i=1}^n \psi(X_i, \theta) = 0$$

where $\psi(x,\theta) = \kappa\{(\partial/\partial\theta)\rho(x,\theta)\}$, for all x, θ, and κ which is a nonzero real number.

3.2.1 M-estimation of General Parameters

We consider the general vector parameter and impose the following regularity conditions on ρ and on F:

A1. *First-order derivatives.* The functions $\psi_j(x,t) = (\partial/\partial\theta_j)\rho(x,\theta)$ are assumed to be absolutely continuous in θ_k with the derivatives $\dot{\psi}_{jk}(x,\theta) = (\partial/\partial\theta_k)\psi_j(x,\theta)$, such that $\mathbb{E}[\dot{\psi}_{jk}(X_1,\theta_0)]^2 < \infty$, $(j,k = 1,\cdots,p)$. Further we assume that the matrices $\Gamma(\theta_0) = [\gamma_{jk}(\theta_0)]_{j,k=1}^p$ and $\mathbf{B}(\theta_0) = [b_{jk}(\theta_0)]_{j,k=1}^p$ are positive definite, where

$$\gamma_{jk}(\theta) = \mathbb{E}_\theta \dot{\psi}_{jk}(\theta),$$

and

$$b_{jk}(\theta) = Cov_\theta(\psi_j(X_1,\theta), \psi_k(X_1,\theta)),\ \ j,k = 1,\cdots,p.$$

A2. *Second- and third-order derivatives.* $\dot{\psi}_{jk}(x,\theta)$ are absolutely continuous in the components of θ and there exist functions $M_{jkl}(x,\theta_0)$ such that $m_{jkl} = \mathbb{E}M_{jkl}(X_1,\theta_0) < \infty$ and

$$|\ddot{\psi}_{jkl}(x,\theta_0+\theta)| \le M_{jkl}(x,\theta_0), x \in \mathbb{R}, ||\theta|| \le \delta, \delta > 0,$$

where

$$\ddot{\psi}_{jkl}(x,\theta) = \frac{\partial^2\psi_j(x,\theta)}{\partial\theta_k\partial\theta_l},\ \ j,k,l = 1,\cdots,p.$$

Under the conditions **A1** and **A2**, the minimum in (3.2) is one of the roots of the system of equations:

$$\sum_{i=1}^{n} \psi_j(X_i,\theta) = 0,\ \ j = 1,\cdots,p. \tag{3.3}$$

The following theorem states existence of a solution of (3.3) that is $\sqrt{n}$-consistent estimator of θ_0 and admits an asymptotic representation.
Theorem 3.2.1 *Let $X_1, X_2, \cdots$ be i.i.d. r.v.s with d.f. $F(x,\theta_0)$, $\theta_0 \in \Theta$, Θ*

being an open set of $\mathbb{R}^p$. *Let* $\rho(x,\theta) : \mathbb{R} \times \Theta \to \mathbb{R}$ *be a function absolutely continuous in the components of* θ *and such that the function* $h(\theta)$ *of (3.1) has a unique minimum at* $\theta = \theta_0$. *Then, under the conditions* **A1** *and* **A2**, *there exists a sequence* $\{M_n\}$ *of solutions of (3.3) such that*

$$n^{\frac{1}{2}}||M_n - \theta_0|| = O_p(1) \text{ as } n \to \infty,$$

and

$$M_n = \theta_0 - n^{-1}(\Gamma(\theta_0))^{-1}\sum_{i=1}^{n}\psi(X_i,\theta_0) + O_p(n^{-1}), \tag{3.4}$$

where $\psi(x,\theta) = (\psi_1(x,\theta), \cdots, \psi_p(x,\theta))'$.

The representation (3.4) implies the asymptotic normality of M_n.

Corollary 3.2.1 *Under the conditions of Theorem 3.2.1,* $n^{\frac{1}{2}}(M_n - \theta_0)$ *has asymptotically p-dimensional normal distribution* $N_p(0, A(\theta_0))$ *with* $A(\theta_0) = (\Gamma(\theta_0))^{-1}B(\theta_0)(\Gamma(\theta_0))^{-1}$.

3.2.2 M-estimators in Location Models

Let $X_1, X_2, \cdots$ be i.i.d. r.v.s with the d.f. $F(x-\theta)$. Let $\rho : \mathbb{R} \to \mathbb{R}$ be an absolutely continuous function with the derivative ψ and such that the function

$$h(\theta) = \int \rho(x-\theta)dF(x) \tag{3.5}$$

has a unique minimum at $\theta = 0$. Let us consider the case when ψ is an absolutely continuous function which could be decomposed as

$$\psi(t) = \psi_1(t) + \psi_2(t), t \in \mathbb{R} \tag{3.6}$$

where ψ_1 has an absolutely continuous derivative ψ_1' and ψ_1 is a piecewise linear continuous function, constant outside a bounded interval. We impose the following conditions on ψ_1, ψ_2 and F.

Bl. *Smooth component* ψ_1. ψ_1 is absolutely continuous with an absolutely continuous derivative ψ_1' such that

$$\int (\psi_1'(x+t))^2 dF(x) < K_1 \text{ for } |t| \leq \delta,$$

and ψ_1' is absolutely continuous and $\int |\psi_1''(x+t)|dF(x) < K_2$ for $|t| \le \delta$, where δ, K_1 and K_2 are positive constants.

B2. *Piecewise linear component* ψ_2. ψ_2 is absolutely continuous with the derivate

$$\psi_2'(x) = \alpha_\nu \text{ for } r_\nu < x \le r_{\nu+1}, \nu = 1, \cdots, k,$$

where $\alpha_0, \alpha_1, \cdots, \alpha_k$ are real numbers, $\alpha_0 = \alpha_k = 0$, and $-\infty = r_0 < r_1 < \cdots < r_k < r_{k+1} = \infty$.

B3. *F smooth around the difficult points.* F has two bounded derivatives f and f' and $f > 0$ in a neighborhood of $r_1, \cdots, r_k$.

B4. *Fishers consistency.* $\gamma = \gamma_1 + \gamma_2 > 0$, where $\gamma_i = \int \psi_i'(x)dF(x)$ for $i = 1, 2$, and $\int \psi^2(x)df(x) < \infty$.

The M-estimator M_n of θ is then defined as a solution of the minimization

$$\sum_{i=1}^{n} \rho(X_i - \theta) := \min \text{ with respect to } \theta \in \mathbb{R}.$$

Under the conditions **(B1)**-**(B4)**, M_n coincides with a root of the equation

$$\sum_{i=1}^{n} \psi(X_i - t) = 0. \tag{3.7}$$

If ρ is not convex, then the equation (3.7) may have more roots. The conditions **B1-B4** guarantee that there exists at least one root of (3.7) that is $\sqrt{n}$-consistent estimator of θ and that admits an asymptotic representation. This is formally expressed in the following theorem.

Theorem 3.2.2 *Let $X_1, X_2, \cdots$ be i.i.d. r.v.s with d.f. $F(x-\theta)$. Let $\rho : \mathbb{R} \to \mathbb{R}$ be an absolutely continuous function whose derivative ψ can be decomposed as (3.6) and such that the function $h(\theta)$ of (3.5) has a unique minimum at $\theta = 0$. Then under the conditions* **(B.1)**-**(B.4)**, *there exists a sequence $\{M_n\}$ of roots of the equation (3.7) such that*

$$n^{\frac{1}{2}}(M_n - \theta) = O_p(1)$$

and

$$M_n = \theta + (n\gamma)^{-1} \sum_{i=1}^{n} \psi(X_i - \theta) + R_n, \text{ where } R_n = O_p(n^{-1}). \tag{3.8}$$

The asymptotic representation (3.8) immediately implies that the sequence $n^{\frac{1}{2}}$ is asymptotically normally distributed as $n \to \infty$; hence we have the following corollary:

Corollary 3.2.2 *Under the conditions of Theorem 3.2.2, there exists a sequence $\{M_n\}$ of solutions of the equation (3.7) such that $n^{\frac{1}{2}}(M_n - \theta)$ has asymptotically normal distribution $N(0, \sigma^2(\psi, F))$ with $\sigma^2(\psi, F) = \gamma^{-2} \int \psi^2(x) dF(x)$.*

3.2.3 M-estimators in Regression Models

Consider the linear model

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E}, \tag{3.9}$$

where $\mathbf{Y} = (Y_1, \cdots, Y_n)'$ is the vector of observations, $\mathbf{X} = \mathbf{X_n}$ is a (known or observable) design matrix of order $(n \times p)$, $\beta = (\beta_1, \cdots, \beta_p)'$ is an unknown parameter, and $\mathbf{E} = (E_1, \cdots, E_n)'$ is a vector of i.i.d. errors with a distribution function F.

The M-estimator of location parameter extends to the model (3.9) in a straightforward way: Given an absolutely continuous $\rho : \mathbb{R} \to \mathbb{R}$ with derivative ψ, we define an M-estimator of β as a solution of the minimization

$$\sum_{i=1}^{n} \rho(Y_i, \mathbf{x}_i'\mathbf{t}) := \min$$

with respect to $\mathbf{t} \in \mathbb{R}^p$, where $\mathbf{x}_i'$ is the i-th row of $\mathbf{X}_n, i = 1, \cdots, n$. Such M-estimator $\mathbf{M}_n$ is regression equivariant:

$$\mathbf{M}_n(\mathbf{Y} + \mathbf{X}\mathbf{b}) = \mathbf{M}_n(\mathbf{Y}) + \mathbf{b} \text{ for } \mathbf{b} \in \mathbb{R}^p,$$

but $\mathbf{M_n}$ is generally not scale equivariant: It does not satisfy

$$\mathbf{M_n}(c\mathbf{Y}) = c\mathbf{M_n}(\mathbf{Y}) \text{ for } c > 0.$$

On the other hand, the studentization leads to estimators that are scale as well as regression equivariant. The studentized M-estimator is defined as a solution of the minimization

$$\sum_{i=1}^{n} \rho((Y_i - \mathbf{x}_i'\mathbf{t})/S_n) := min, \tag{3.10}$$

where $S_n = S_n(\mathbf{Y}) \geq 0$ is an appropriate scale statistic. For the best results S_n should be regression invariant and scale equivariant:

$$S_n(c(\mathbf{Y} + \mathbf{X}\mathbf{b})) = cS_n(\mathbf{Y}) \text{ for } \mathbf{b} \in \mathbb{R}^p \text{ and } c > 0.$$

The minimization (3.10) should be supplemented by a rule how to define $\mathbf{M_n}$ if $S_n(\mathbf{Y}) = 0$. However, in typical cases it appears with probability zero, and the specific rule does not affect the asymptotic properties of $\mathbf{M_n}$.

Before giving the basic results on studentized M-estimator of regression model, we imposed the following conditions on (3.10):

M1. $S_n(\mathbf{Y})$ is regression invariant and scale equivariant, $S_n > 0$ a.s. and

$$n^{\frac{1}{2}}(S_n - S) = O_p(1)$$

for some functional $S = S(F) > 0$.

M2. The function $h(t) = \int \rho((z-t)/S)dF(z)$ has the unique minimum at $t = 0$.

M3. For some $\delta > 0$ and $\eta > 1$,

$$\int_{-\infty}^{\infty} \{|z| \sup_{|u|\leq\delta} \sup_{|v|\leq\delta} |\psi_a''(e^{-v}(z+u)/S)|\}^{\eta} dF(z) < \infty$$

and

$$\int_{-\infty}^{\infty} \{|z|^2 \sup_{|u|\leq\delta} |\psi_a''((z+u)/S)|\}^{\eta} dF(z) < \infty,$$

where $\psi_a'(z) = (d/dz)\psi_a(z)$, and $\psi_a''(z) = (d^2/dz^2)\psi_a(z)$.

M4. ψ_c is a continuous, piecewise linear function with knots at $\mu_1, \cdots, \mu_k$, which are constants in a neighborhood of $\pm\infty$. Hence the derivative ψ_c' of ψ_c is a step function

$$\psi_c'(z) = \alpha_\nu \text{ for } \mu_\nu < z < \mu_{\nu+1}, \nu = 0, 1, \cdots, k,$$

where $\alpha_0, \alpha_1, \cdots, \alpha_k \in \mathbb{R}$, $\alpha_0 = \alpha_k = 0$ and $-\infty = \mu_0 < \mu_1 < \cdots < \mu_k < \mu_{k+1} = \infty$. We assume that $f(z) = \frac{dF(z)}{dz}$ is bounded in neighborhoods of $S_{\mu_1}, \cdots, S_{\mu_k}$.

M5. $\psi_s(z) = \lambda_\nu$ for $q_\nu < z \leq q_{\nu+1}, \nu = 1, \cdots, m$ where $-\infty = q_0 < q_1 < \cdots < q_m < q_{m+1} = \infty$, $-\infty < \lambda_0 < \lambda_1 < \cdots < \lambda_m < \infty$. We assume that $0 < f(z) = (d/dz)F(z)$ and $f'(z) = (d^2/dz^2)F(z)$ are bounded in neighborhoods of $S_{q1}, \cdots, S_{qm}$.

The asymptotic representation for $\mathbf{M}_n$ will involve the functionals

$$\gamma_1 = S^{-1} \int_{-\infty}^{\infty} (\psi_a'(z/S) + \psi_c'(z/S))dF(z), \tag{3.11}$$

$$\gamma_2 = S^{-1} \int_{-\infty}^{\infty} z(\psi_a'(z/S) + \psi_c'(z/S))dF(z),$$

$$\gamma_1^* = \sum_{\nu=1}^{m} (\lambda_\nu - \lambda_{\nu-1}) f(S_{q_\nu}),$$

$$\gamma_2^* = S \sum_{\nu=1}^{m} (\lambda_\nu - \lambda_{\nu-1}) q_\nu f(S_{q_\nu}),$$

Conditions **M4** and **M5** depict explicitly the trade-off between the smoothness of ψ and smoothness of F. The class of functions ψ_c covers the usual Hubers and Hampels proposals.

Moreover we impose the following conditions on the matrix $\mathbf{X}_n$:

X1. $x_{i1} = 1, i = 1, \cdots, n$.

X2. $n^{-1} \sum_{i=1}^{n} ||\mathbf{x}_i||^4 = O_p(1)$.

X3. $\lim_{n\to\infty} \mathbf{Q}_n = \mathbf{Q}$, where $\mathbf{Q}_n = n^{-1}\mathbf{X}_n'\mathbf{X}_n$ and $\mathbf{Q}$ is a positive definite $p \times p$ matrix.

Let $\mathbf{M}_n$ be a solution of the minimization (3.10). If $\psi = \rho'$ is continuous (i.e., $\psi_s \equiv 0$), then $\mathbf{M}_n$ is a solution of the system of equations

$$\sum_{i=1}^{n} \mathbf{x}_i \psi(\frac{Y_i - \mathbf{x}_i' t}{S_n}) = 0. \tag{3.12}$$

The basic results on studentized M-estimators of regression are summarized in the following three theorems.

Theorem 3.2.3 *Consider the model(3.9) and assume the conditions* **M1-M4**, **X1-X3**, *and that* γ_1 *defined in (3.11) is different from zero. Then, provided* $\psi_s \equiv 0$, *there exists a root* $\mathbf{M_n}$ *of the system (3.12) such that*

$$n^{\frac{1}{2}}||\mathbf{M}_n - \beta|| = O_p(1) \text{ as } n \to \infty. \tag{3.13}$$

Moreover any root $\mathbf{M}$ *of (3.12) satisfying (3.13) admits the representation*

$$\mathbf{M_n} - \beta = (n\gamma_1)^{-1}\mathbf{Q}_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\psi(E_i/S) - \frac{\gamma_1}{\gamma_2}(\frac{S_n}{S} - 1)e_1 + \mathbf{R}_n,$$

where $||\mathbf{R}_n|| = O_p(n^{-1})$ *and* $e_1 = (1, 0, \cdots, 0)\prime \in \mathbb{R}_p$.

Theorem 3.2.4 *Consider the linear model (3.9) and assume the conditions* **M2**, **M5**, *and* **X1-X3**. *Let* $\mathbf{M_n}$ *be the point of global minimum of (3.10). Then, provided that* $\psi_a = \psi_c \equiv 0$,

$$n^{\frac{1}{2}}||\mathbf{M}_n - \beta|| = O_p(1) \text{ as } n \to \infty,$$

and $\mathbf{M_n}$ *admits the representation*

$$\mathbf{M_n} - \beta = (n\gamma_1^*)^{-1}\mathbf{Q}_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\psi(E_i/S)\frac{\gamma_1^*}{\gamma_2^*}(\frac{S_n}{S} - 1)e_1 + \mathbf{R}_n,$$

where $||\mathbf{R}_n|| = O_p(n^{-3/4})$ *and* $e_1 = (1, 0, \cdots, 0)' \in \mathbb{R}^p$.

Theorem 3.2.5 *Consider the model (3.9) and assume the conditions* **M1-M4**, *and* **X1-X3**. *Let* ψ *be either continuous or monotone, and let* $\gamma_1 + \gamma_1^* \neq 0$. *Then, for any M-estimator* $\mathbf{M}_n$ *satisfying* $n^{1/2}(\mathbf{M}_n - \beta) = O_p(1)$,

$$\mathbf{M}_n - \beta = (n(\gamma_1+\gamma_1^*))^{-1}\mathbf{Q}_n^{-1}\sum_{i=1}^{n}\mathbf{x}_i\psi(E_i/S) - (\gamma_1+\gamma_1^*)^{-1}(\gamma_2+\gamma_2^*)(\frac{S_n}{S} - 1)e_1 + \mathbf{R}_n$$

where

$$||\mathbf{R}_n|| = \begin{cases} O_p(n^{-1}), & \text{if } \psi_x \equiv 0, \\ O_p(n^{-3/4}), & \text{otherwise.} \end{cases}$$

Theorem 3.2.3-3.2.5 have several interesting corollaries, parallel to those in the location model.

Corollary 3.2.3 *Under the conditions of Theorem 3.2.3, let* $\mathbf{M}_n^{(1)}$ *and* $\mathbf{M}_n^{(2)}$ *be any pair of roots of the system of equations (3.12), both satisfying (3.13). Then*

$$||\mathbf{M}_n^{(1)} - \mathbf{M}_n^{(2)}|| = O_P(n^{-1}).$$

Corollary 3.2.4 *Assume that*

$$\sigma^2 = \int_{-\infty}^{\infty} \psi^2(z/S)dF(z) < \infty.$$

Then, under the conditions of Theorem 3.2.3-3.2.5, respectively, the sequence

$$n^{\frac{1}{2}}\{\hat{\gamma}_1(\mathbf{M}_n - \beta) + \hat{\gamma}_2(\frac{S_n}{S} - 1)e_1\}$$

has the asymptotic p-dimensional normal distribution $N_p(\mathbf{0}, \sigma^2\mathbf{Q}^{-1})$*; here* $\hat{\gamma}_i$ *stands for* γ_i, γ_i^* *or* $\gamma_i + \gamma_i^*$*,* $i = 1, 2$*, respectively.*

3.3 R-Estimator

In this section we give a simple introduction of estimates derived from rank tests, which is called R-estimates. Consider a two-sample rank test for shift: let $x_1, \cdots, x_m$ and $y_1, \cdots, y_n$ be two independent samples from the distributions $F(x)$ and $G(x) = F(x - \Delta)$, respectively. Merge the two samples into one of size $m+n$ and let R_i be the rank of x_i in the combined sample. Let $a_i = a(i)$, $1 \leq i \leq m+n$, be some given scores; then base a test of $\Delta = 0$ against $\Delta > 0$ on the test statistic

$$S_{m,n} = \frac{1}{m}\sum_{i=1}^{m} a(R_i). \tag{3.14}$$

We assume that the scores a_i are generated by some function J as follows:

$$a_i = (m+n)\int_{(i-1)/(m+n)}^{i/(m+n)} J(s)ds. \tag{3.15}$$

To simplify the presentation, we assume that $m = n$ and $0 < J < 1$. In terms of functionals (3.14) can be written as

$$S(F,G) = \int J\big[\frac{1}{2}F(x) + \frac{1}{2}G(x)\big]F(dx), \tag{3.16}$$

or, if we substitute $F(x) = s$,

$$S(F,G) = \int J\big[\frac{1}{2}s + \frac{1}{2}G(F^{-1}(s))\big]ds. \tag{3.17}$$

We also assume that

$$\int J(s)ds = 0, \tag{3.18}$$

corresponding to

$$\sum a_i = 0. \tag{3.19}$$

Then the expected value of (3.14) under the null hypothesis is 0. We can derive estimates of shift Δ_n and of location T_n from such rank tests:

1. In the two sample case, adjust Δ_n such that $S_{n,n} \approx 0$ when computed from $(x_1, \cdots, x_n)$ and $(y_1 - \Delta_n, \cdots, y_n - \Delta_n)$.
2. In the one sample case, adjust T_n such that $S_{n,n} \approx 0$ when computed from $(x_1, \cdots, x_n)$ and $(2T_n x_1, \cdots, 2T_n x_n)$. In this case a mirror image of the first sample serves as a stand-in for the missing second sample.

Thus our location estimate T_n derives from a functional $T(F)$, defined by the implicit equation

$$\int J\{\frac{1}{2}\big[s + 1 - F(2T(G) - F^{-1}(s))\big]\}ds = 0. \tag{3.20}$$

3.3.1 Influence Function of R-Estimates

We now derive the influence function of $T(F)$. To shorten the notation we introduce the distribution function of the pooled population:

$$K(x) = \frac{1}{2}\big[F(x) + 1 - F(2T(F) - x)\big]. \tag{3.21}$$

Assume that F has a strictly positive density f. We insert $F_t = (1-t)F + tG$ for F in (3.20) and take the derivative $\partial/\partial t$ (denoted by a dot $\dot{}$) at $t = 0$. This gives

$$\int J'(K(F^{-1}(s)))\big[\dot F(2T - F^{-1}(s)) + \frac{f(2T - F^{-1}(s))}{f(F^{-1}(s))}\dot F(F^{-1}(s)) + 2f(2T - F^{-1}(s))\dot T\big]\,ds \tag{3.22}$$

To separate this expression in a sum of three integrals and substitute $x = 2T - F^{-1}(s)$ in the first (thus $s = F(2T - x)$), but $x = F^{-1}(s)$ in the second and third integrals, we obtain

$$\dot T \int J'(K(x))f(2T - x)f(x)dx + \int \frac{1}{2}[J'(K(x)) + J'(1 - K(x))]f(2T - x)\dot F(x)dx = 0. \tag{3.23}$$

Now we assume that the scores-generating function is symmetric

$$J(1-t) = -J(t), 0 < t < 1 \tag{3.24}$$

then we simplify (3.23) by introduction the function $U(x)$, being an indefinite integral of

$$U'(x) = J'\{\frac{1}{2}[F(x) + 1 - F(2T(F) - x)]\}f(2T(F) - x). \tag{3.25}$$

Then (3.23) turns into

$$\int U'(x)f(x)dx + \int U'(x)\dot F(x)dx = 0. \tag{3.26}$$

Integration by parts of the second integral yields

$$\int U'(x)\dot F(x)dx = -\int U(x)\dot F(dx).$$

With $G = \delta_x$ we obtain the influence function from (3.26) by solving for $\dot T$:

$$IC(x; F, T) = \frac{U(x) - \int U(x)f(x)dx}{\int U'(x)f(x)dx}. \tag{3.27}$$

for symmetric F this can be simplified considerably, since then $U(x) = J(F(x))$:

$$IC(x; F, T) = \frac{J(F(x))}{\int J'(F(x))f(x)^2 dx}.$$

Example The influence function of the Hodges-Lehmann estimate ($J(t) = t - \frac{1}{2}$) is

$$IC(x; F, T) = \frac{\frac{1}{2}F(2T(F) - x)}{\int f(2T(F) - x)f(x)dx}, \tag{3.28}$$

with $T(F)$ defined by

$$\int F(2T(F) - x)F(dx) = \frac{1}{2} \tag{3.29}$$

For symmetric F this simplifies to

$$IC(x; F, T) = \frac{F(x) - \frac{1}{2}}{\int f(x)^2 dx}, \tag{3.30}$$

and the asymptotic variance of $\sqrt{n}[T(F_n) - T(F)]$ is indeed known to be

$$A(F, T) = \int IC^2 dF = \frac{1}{12[\int f(x)^2 dx]^2}. \tag{3.31}$$

3.3.2 Robustness of R-Estimates

We assume that the scores function J is monotone increasing and $J(1-t) = -J(t)$. In order that (3.17) by well defined, we require

$$\int |J(s)| ds < \infty. \tag{3.32}$$

The function

$$\lambda(t : F) = \int J\{\frac{1}{2}[s + 1 - F(2t - F^{-1}(s))]\} ds \tag{3.33}$$

is monotone decreasing in t, and it increases if F is made stochastically larger.

We consider that

$$F_1^{-1}(s) = \begin{cases} F_0^{-1}(s + \epsilon) + \epsilon, & \text{for } 0 \le s \le 1 - \epsilon \\ \infty, & \text{for } s > 1 - \epsilon. \end{cases}$$

Thus provided the two side conditions

$$0 \le s \le 1 - \epsilon$$

and

$$2t - F_1^{-1}(s) \ge x_0 + \epsilon,$$

where $F_0(x_0) = \epsilon$ is satisfied, we have

$$F_1[2t - F_1^{-1}(s)] = F_0[2t - 2\epsilon - F_0^{-1}(s + \epsilon)] - \epsilon.$$

The second condition can be written as

$$s \le F_0(2t - 2\epsilon - x_0) - \epsilon.$$

Putting things together we obtain

$$\lambda(t; F_1) = \int_0^{s_0} J(\frac{1}{2}[s+\epsilon+1-F_0(2(t-\epsilon)-F_0^{-1}(s+\epsilon))])ds + \int_{s_0}^1 J[\frac{1}{2}(s+1)]ds, \tag{3.34}$$

with

$$s_0 = [F_0(2(t - \epsilon) - x_0) - \epsilon]^+.$$

We then obtain

$$b_+(\epsilon) = inf\{t|\lambda(t; F_1) < 0\},$$

and, symmetrically we also obtain $b_-(\epsilon)$; if F_0 is symmetric, we have

$$b_1(\epsilon) = b_+(\epsilon) = -b_-(\epsilon).$$

With regard to breakdown we note that $b_+(\epsilon) < \infty$ if and only if

$$\lim_{t\to\infty} \lambda(t; F_1) < 0.$$

Since

$$\begin{aligned}\lim_{t\to\infty} \lambda(t; F_1) &= \int_0^{1-\epsilon} J[\frac{1}{2}(s + \epsilon)]ds + \int_\epsilon^1 J[\frac{1}{2}(s + 1)]ds \\ &= 2\int_{\epsilon/2}^{1/2} J(s)ds + \int_{1-\epsilon/2}^1 J(s)ds \\ &= 2[\int_{1-\epsilon/2}^1 J(s)ds - \int_{1/2}^{1-\epsilon/2} J(s)ds],\end{aligned}$$

the breakdown point ϵ^* is that value ϵ for which

$$\int_{1/2}^{1-\epsilon/2} J(s)ds = \int_{1-\epsilon/2}^1 J(s)ds. \tag{3.35}$$

Example For the Hodges-Lehmann estimates, $J(t) = t - \frac{1}{2}$, we obtain as breakdown point

$$\epsilon^* = 1 - \frac{1}{\sqrt{2}} \approx 0.293.$$

When $\epsilon \downarrow 0$ the integrand in (3.34) decreases and converges to the integrand corresponding to F_0 for almost all s and t. It follows from the monotone convergence theorem that $\lambda(t; F_1) \downarrow \lambda(t; F_0)$ at the continuity points of $\lambda(\cdot; F_0)$. Hence if $\lambda(t; F_0)$ has a unique zero, that is, if $T(F_0)$ is uniquely defined, T is continuous at F_0. If $T(F_0)$ is not unique, then T of course cannot be continuous at F_0. A sufficient condition for uniqueness is that the derivative of $\lambda(t; F_0)$ with regard to t exists and is not equal to 0 At $T = T(F_0)$; this derivative occurred already as the denominator of (3.27) and (3.28).

We summarize the results in the following theorem.

Theorem 3.3.1 *Assume that the scores generating function J is monotone increasing, integrable, and $J(1-t) = -J(t)$. If the R-estimate $T(F_0)$ is uniquely defined by (3.20), then T is weakly continuous at F_0. The breakdown point of T is given by (3.35).*

3.4 L-Estimator

Consider a statistic that is a linear combination of order statistics, or more generally, of some function h of them:

$$T_n = \sum_{i=1}^{n} a_{ni} h(x_{(i)}).$$

We assume that the weights are generated by a measure M on $(0, 1)$:

$$a_{ni} = \frac{1}{2} M\{(\frac{i-1}{n}, \frac{i}{n})\} + \frac{1}{2} M\{[\frac{i-1}{n}, \frac{i}{n}]\}.$$

This choice preserves the total mass, $\sum_i a_{ni} = M\{(o, 1)\}$, and symmetry of the coefficients, if M is symmetric about $t = \frac{1}{2}$.

Then $T_n = T(F_n)$ derives from the functional

$$T(F) = \int h(F^{-1}(s))M(ds). \tag{3.36}$$

We have exact equality $T_n = T(F_n)$ if we regularize the integrand at its discontinuity points and replace it by

$$\frac{1}{2}h(F_n^{-1}(s-0)) + \frac{1}{2}h(F_n^{-1}(s+0)),$$

Here, the inverse of any distribution function F is defined in the usual way as

$$F^{-1}(s) = \inf\{x|F(x) \geq s\}, \;\; 0 < s < 1.$$

3.4.1 Influence Function of L-Estimates

To find the influence function $IC(x; F, T)$ of T: insert $F_t = (1-t)F + tG$ into (3.36), and take the derivative with respect to t at $t = 0$, for $G = \delta_x$. We begin with the derivative of $T_s = F_t^{-1}(s)$, that is, of the s-quantile. If we differentiate the identity

$$F_t(F_t^{-1}(s)) = s$$

with respect to t at $t = 0$, we obtain

$$G(F^{-1}(s)) - F(F^{-1}(s)) + f(F^{-1}(s))\dot{T}_s = 0,$$

or

$$\dot{T}_s = \frac{s - G(F^{-1}(s))}{f(F^{-1}(s))}.$$

If $G = \delta_x$ is the pointmass 1 at x, this gives the value of the influence function of T_s:

$$\begin{aligned} IC(x; F, T_s) &= \frac{s-1}{f(F^{-1}(s))}, \;\text{for } x < F^{-1}(s) \\ &= \frac{s}{f(F^{-1}(s))}, \;\text{for } x > F^{-1}(s). \end{aligned}$$

Quite clearly, these calculations make sense only if F has a nonzero finite derivative f at $F^{-1}(s)$, but then they are legitimate.

By the chain rule for differentiation, the influence function of $h(T_s)$ is

$$IC(x; F, h(T_s)) = IC(x; F, T_s)h'(T_s),$$

and that of T itself then is

$$IC(x; F, T) = \int IC(x; F, h(T_s))M(ds) \tag{3.37}$$

$$= \int \frac{sh'(F^{-1}(s))}{f(F^{-1}(s))}M(ds) - \int_{F(x)}^{1} \frac{h'(F^{-1}(s))}{f(F^{-1}(s))}M(ds). \tag{3.38}$$

If M has a density m, it may be more convenient to write (3.38) as

$$IC(x; F, T) = \int_{-\infty}^{x} h'(y)m(F(y))dy - \int_{-\infty}^{\infty} (1 - F(y))h'(y)m(F(y))dy. \tag{3.39}$$

This can be remembered through its derivative:

$$\frac{d}{dx}IC(x; F, T) = h'(x)m(F(x)).$$

Then we have the following alternative version of (3.36):

$$\begin{aligned} T(F) &= \int h(F^{-1}(s))m(s)ds \\ &= \int h(y)m(F(y))F(dy) \\ &= -\int h'(y)M(f(y))dy. \end{aligned}$$

Example 1. For the median ($s = \frac{1}{2}$) we have

$$\begin{aligned} IC(x; F, T_{1/2}) &= \frac{-1}{2f(F^{-1}(\frac{1}{2}))} \quad \text{for } x < F^{-1}(\frac{1}{2}), \\ &= \frac{1}{2f(F^{-1}(\frac{1}{2}))} \quad \text{for } x > F^{-1}(\frac{1}{2}). \end{aligned}$$

Example 2. The α-trimmed mean corresponds to $h(x) = x$ and

$$\begin{aligned} m(s) &= \frac{1}{1-2\alpha}, \quad \text{for } \alpha < s < 1 - \alpha \\ &= 0, \quad \text{otherwise;} \end{aligned}$$

thus

$$T(F) = \frac{1}{1-2\alpha} \int_{\alpha}^{1-\alpha} F^{-1}(s)ds. \tag{3.40}$$

Note that the α-trimmed mean $T(F_n)$, as defined by (3.40), has the following property: if αn is an integer, then αn observations are removed from each end of the sample and the mean of the rest is taken. If it is not an integer, say $\alpha n = \lfloor \alpha n \rfloor + p$, then $\lfloor \alpha n \rfloor$ observations are removed from each end, and the next observations $x_{(s[\alpha n]+1)}$ And $x_{(n-[\alpha n])}$ are given the reduced weight $1-p$.

The influence function of the α-trimmed mean is, according to (3.39),

$$\begin{aligned} IC(x) &= \frac{1}{1-2\alpha}[F^{-1}(\alpha)W(F)], \text{ for } x < F^{-1}(\alpha) \\ &= \frac{1}{1-2\alpha}[xW(F)], \text{ for } F^{-1}(\alpha) \le x \le F^{-1}(1-\alpha) \\ &= \frac{1}{1-2\alpha}[F^{-1}(1-\alpha) - W(F)], \text{ for } x > F^{-1}(1-\alpha). \end{aligned}$$

Here W is the functional corresponding to the so-called α-Winsorized mean:

$$\begin{aligned} W(F) &= \int_{\alpha}^{1-\alpha} F^{-1}(s)ds + \alpha F^{-1}(\alpha) + \alpha F^{-1}(1-\alpha) \\ &= (1-2\alpha)T(F) + \alpha F^{-1}(\alpha) + \alpha F^{-1}(1-\alpha). \end{aligned}$$

3.4.2 Robustness of L-Estimates

We now calculate the maximum bias b_1 for L-estimates. Assume that $h(x) = x$ and that M is a positive measure on $(0,1)$ with total mass 1. The resulting functional corresponds to a location estimate; if F_{aX+b} denotes the distribution of the random variable $aX+b$, We have

$$T(F_{aX+b}) = aT(F_X) + b, \quad \text{for } a \ge 0.$$

Let α be the largest real number such that $[\alpha, 1-\alpha]$ contains the support of M; then the breakdown point satisfies $\epsilon^* \le \alpha$. We now show that $\epsilon^* = \alpha$.

Assume that the target value is $T(F_0) = 0$, let $0 < \epsilon < \alpha$, and define b_+, b_- as follows:

$$\begin{aligned} b_+(\epsilon) &= sup\{T(F) | d_L(F_0, F) \le \epsilon\} \\ b_-(\epsilon) &= inf\{T(F) | d_L(F_0, F) \le \epsilon\}. \end{aligned}$$

Then, with $F_1(x) = (F_0(x-\epsilon) - \epsilon)^+$, we have

$$b_+(\epsilon) = \int F_1^{-1}(s)M(ds) = \epsilon + \int_{\alpha}^{1-\alpha} F_0^{-1}(s+\epsilon)M(ds),$$

and

$$b_-(\epsilon) = -\epsilon + \int_\alpha^{1-\alpha} F_0^{-1}(s-\epsilon)M(ds).$$

Let

$$b_1(\epsilon) = max\{b_+(\epsilon), -b_-(\epsilon)\}$$

As $F_0^{-1}(s+\epsilon) - F_0^{-1}(s-\epsilon) \downarrow 0$ for $\epsilon \downarrow 0$, except at the discontinuity points of F_0^{-1}, we conclude that $b_1(\epsilon) \leq b_+(\epsilon) - b_-(\epsilon) \downarrow 0$ iff the distribution function of M and F_0^{-1} do not have common discontinuity points, and then $T$is continuous at F_0. Since $b_1(\epsilon)$ is finite for $\epsilon < \alpha$, we must have $\epsilon^* \geq \alpha$.

In particular, the α-trimmed mean with $0 < \alpha < \frac{1}{2}$ is everywhere continuous. The α-Winsorized mean is continuous at F_0 if $F_0^{-1}(\alpha)$ and $F_0^{-1}(1-\alpha)$ are uniquely determined.

The generalization to signed measures is immediate, the results is in the following theorem.

Theorem3.4.1 *Let $M = M^+ - M^-$ be a finite signed measure on $(0,1)$ and let $T(F) = \int F^{-1}(s)M(ds)$. Let α be the largest real number such that $[\alpha, 1-\alpha]$ contains the support of M^+ and M^-. If $\alpha > 0$, then T is weakly continuous at F_0, provided m does not put any point mass on a discontinuity point of F_0^{-1}. The breakdown point satisfies $\epsilon^* \geq \alpha$. If M is positive, we have $\epsilon^* = \alpha$, and $\alpha = 0$ implies that T is discontinuous.*

For the asymptotic properties of L-estimates the following theorem is a useful version which is from Huber(1969) and Stigler (1969).

Theorem3.4.2 *Let M be an absolutely continuous signed measure with density m, whose support is contained in $[\alpha, 1-\alpha]$, $\alpha > 0$. Let $T(F) = \int F^{-1}(s)m(s)ds$. The $\sqrt{n}(T(F_n) - T(F))$ is asymptotically normal with mean 0 and variance $\int IC(x; F, T)^2 F(dx)$, provided both (1) and (2) hold:*

(1) *m is of bounded total variation (so all its discontinuities are jumps).*

(2) *No discontinuity of m coincides with a discontinuity of F^{-1}.*

3.5 Minimum Distance Estimates

Assuming the general parametric model

$$\mathcal{P} = \{P_\theta | \theta \in \Theta\} \subset \mathcal{M}_1(\mathcal{A})$$

on a general sample space $(\Omega, \mathcal{A})$, the minimum distance (MD) idea is to determine the value θ so that $\mathcal{P}_\theta$ fits best a given probability, respectively the empirical measure. The set of probabilities $\mathcal{M}_1(\mathcal{A})$ has to be mapped, and the parametric model $\mathcal{P}$ embedded, into some metric space (Ξ, d). The following conditions are imposed on the parametrization $\theta \to P_\theta$:

(1) $\xi \neq \theta \Longrightarrow d(P_\xi, P_\theta) > 0$

(2) $\xi \to \theta \Longrightarrow d(P_\xi, P_\theta) \to 0$

(3) For every $\theta \in \Theta$ there exist numbers $\eta_\theta, K_\theta \in (0, \infty)$ such that $|\xi - \theta| \leq \eta_\theta \Longrightarrow d(P_\xi, P_\theta) \geq K_\theta |\xi - \theta|$.

The open parameter space $\Theta \subset \mathbb{R}^n$ being locally compact separable, it has a representation

$$\Theta = \bigcup_{\nu=1}^{\infty} \Theta_\nu$$

with Θ_ν open, the closure $\bar{\Theta}_\nu$ compact, and $\bar{\Theta}_\nu \subset \Theta_{\nu+1}$ for all $\nu \geq 1$.

Due to (2), the parametrization $\theta \to P_\theta$ is uniformly continuous on each compact $\bar{\Theta}_\nu$: For every $\delta \in (0, \infty)$ there exists some $\epsilon_\nu(\delta) \in (0, \infty)$ such that for all $\zeta, \theta \in \bar{\Theta}_\nu$,

$$|\zeta - \theta| < \epsilon_\nu(\delta) \Longrightarrow d(P_\zeta, P_\theta) < \delta \tag{3.41}$$

The same is true for the inverse $P_\theta \to \theta$ restricted to the image set $\{P_\theta | \theta \in \bar{\Theta}_\nu\}$: For every $\epsilon \in (0, \infty)$ there is some $\delta_\nu(\epsilon) \in (0, \infty)$ such that for all $\zeta, \theta \in \bar{\Theta}_\nu$,

$$d(P_\zeta, P_\theta) < \delta_\nu(\epsilon) \Longrightarrow |\zeta - \theta| < \epsilon.$$

Choose three sequences $r_\nu, z_\nu, \rho_n \in (0, \infty)$ such that

$$\lim_{\nu \to \infty} r_\nu = \infty, \quad \lim_{\nu \to \infty} z_\nu = 0, \quad \lim_{n \to \infty} \sqrt{n} \rho_n = 0.$$

For $\delta = \rho_n$ in (3.41), by the compactness of $\bar{\Theta}_\nu$, there exist finite subsets

$$\Theta_{\nu,n} = \{\theta_{\nu,n;1}, \cdots, \theta_{\nu,n;q_{\nu,n}}\} \subset \bar{\Theta}_\nu \subset \Theta$$

such that for all $n, \nu > 1$,

$$\sup_{\zeta \in \bar{\Theta}_\nu} \inf \{d(P_\zeta, P_\theta) | \theta \in \Theta_{\nu,n}\} \leq \rho_n.$$

Define $T_{\nu,n} : \mathcal{M}_1(\mathcal{A}) \to \Theta_{\nu,n}$ so that $T_{\nu,n}(Q)$ denotes the first element $\theta_{\nu,n;j}$ of $\Theta_{\nu,n}$ to achieve $\inf\{d(Q,P_\theta)|\theta \in \Theta_{\nu,n}\}$. This means that

$$T_{\nu,n}(Q) = \theta_{\nu,n;j} \Longleftrightarrow \begin{cases} d(Q,P_{\theta_{\nu,n;i}}) > d(Q,P_{\theta_{\nu,n;j}}), & i < j \\ d(Q,P_{\theta_{\nu,n;i}}) \geq d(Q,P_{\theta_{\nu,n;j}}), & i \geq j. \end{cases} \tag{3.42}$$

Given any $Q \in \mathcal{M}_1(\mathcal{A})$, by (2) and $\bar{\Theta}_\nu$ compact, there is some $\zeta \in \bar{\Theta}_\nu$ which minimizes $d(Q,P_\theta)$ for $\theta \in \bar{\Theta}_\nu$. For this ζ choose $\theta_{\nu,n;i} \in \Theta_{\nu,n}$ according to (3.5) and (3.42),

$$d(Q,P_{\theta_{\nu,n;j}}) \leq d(Q,P_{\theta_{\nu,n;i}}) \leq d(Q,P_\zeta) + \rho_n.$$

Thus, for all $Q \in \mathcal{M}_1(\mathcal{A})$ and all $n,\nu \geq 1$,

$$d(Q,P_{T_{\nu,n}(Q)}) \leq \inf_{\theta\in\bar{\Theta}_\nu} d(Q,P_\theta) + \rho_n.$$

Let $B_d(P_\theta,r) = \{Q \in \mathcal{M}_1(\mathcal{A})|d(Q,P_\theta) \leq r\}$.

We have the following lemma.

Lemma 3.5.1 *Assume (1) and (2), and let $T_{\nu,n}$ be defined by (3.42). Then for every $\nu \in \mathbb{N}$ there exists some $m_\nu \in \mathbb{N}$ such that for all $n \geq m_\nu$,*

$$\sup_{\theta\in\bar{\Theta}_\nu} \sup\{|T_{\nu,n}(Q) - \theta||Q \in B_d(P_\theta, r_\nu/\sqrt{n})\} \leq z_\nu.$$

We can arrange that $m_\nu < m_{\nu(n)+1}$ for all $n \geq 1$. The MD functional $T_d = (T_{d,n})$ is now obtained from $(T_{\nu,n})$ by a certain diagonalization:

$$T_{d,n} = T_{\nu(n),n} : \mathcal{M}_1(\mathcal{A}) \to \Theta_{\nu(n),n} \subset \Theta. \tag{3.43}$$

This construction achieves bounded infinitesimal oscillation of T_d on the neighborhood system $\mathcal{U}_d$.

Theorem 3.5.1 *Assume (1), (2) and (3), and let $T_d = (T_{d,n})$ be defined by (3.43). Then, for all $\theta \in \Theta$ and $r \in (0,\infty)$,*

$$\limsup_{n\to\infty} \sup\{\sqrt{n}|T_{d,n}(Q) - \theta|\,|Q \in B_d(P_\theta, r/\sqrt{n})\} < \infty.$$

Kolmogorov MD Estimate

Minimum distance estimates $S_* = (S_{*,n})$ are obtained by evaluating minimum distance functionals $T_* = (T_{*,n})$ at the empirical measure,

$$S_{*,n}(x_1,\cdots,x_n) = T_{*,n}(\hat{P}_n), \quad \hat{P}_n(x_1,\cdots,x_n) = \frac{1}{n}\sum_{i=1}^n \mathbf{I}_{x_i}.$$

For the Hellinger distance d_h, some (kernel) smoothing of the empirical is required as in Beran(1977), and likewise for total variation distance d_v, since in these two metrics $d_*(\hat{P}_n, \bar{Q}_n) = 1$, if $\bar{Q}_n$ has no atoms. We concentrate on the Kolmogorov estimates S_k. The sample space is assumed to be some finite-dimensional Euclidean $(\mathbb{R}^m, \mathbb{B}^m)$.

Theorem 3.5.2 *Let $\mathcal{P}$ be identifiable, and L_1 differentiable at all $\theta \in \Theta$. Then $S_k = (S_{k,n})$ is at all sample sizes $n \geq 1$ a random variable,*

$$S_{k,n} : (\mathbb{R}^{mn}, \mathbb{B}^{mn}) \to (\Theta, \mathbb{B}^k).$$

Moreover, for all $\theta \in \Theta$, all $r \in (0, \infty)$ and all arrays $Q_{n,i} \in B_k(P_\theta, r/\sqrt{n})$, the sequence of laws $\sqrt{n}(S_{k,n} - \theta)(Q_n^{(n)})$ is tight in $\mathbb{R}^k$.

For more discussion on MDE, we refer to Beran(1977, 1981, 1982), Millar(1981, 1983, 1984) and Koul(1985).

Chapter 4

Generalized and Feasible Generalized Median Estimators for the Linear Regression with AR(1) Error Model

One of standard assumptions in the regression model is that the error term ϵ_i and ϵ_j, associated with the ith and jth observations, are uncorrelated. Correlation in the error terms suggests that there is additional information in the data that has not been exploited in the current model. When the observations have a natural sequential order, the correlation is referred as autocorrelation or serial correlation.

The symptoms of autocorrelation may appear as the result of a variable having been omitted from the right-hand side of the regression equation. If successive values of the omitted variable are correlated, the errors from the estimated model will appear to be correlated. The presence of autocorrelation has several effects on the analysis. These are summarized as follows:

1. Least squares estimates of the regression coefficients are unbiased but are not efficient in the sense that they are no longer have minimum variance.

2. The estimate of σ^2 and the standard errors of the regression coefficients may be seriously understated; that is, from the data the estimated standard errors would be much smaller than they actually are, giving a spurious impression of accuracy.

3. The confidence intervals and the various tests of significance commonly employed would no longer be strictly valid.

Thus the presence of auto correlation can be a problem of serious concern for the preceding reasons and should not be ignored.

4.1 Introduction

The generalized least squares estimator (GLSE) and the feasible generalized least squares estimator (FGLSE) are, separately, extended to the generalized and the feasible generalized median estimators for the linear regression with AR(1) error model. The large sample theory for these estimators is developed. Furthermore results of Monte Carlo studies and an example of real data analysis are provided for the feasible generalized median estimator.

Consider the linear regression model

$$y_i = x_i'\beta + \epsilon_i, i = 1, ..., n \tag{4.1}$$

where, for each i, x_i is a known design p-vector with value 1 in its first element and $\epsilon_i, i = 1, ..., n$ are random error variables. Suppose that the error vector $\epsilon = (\epsilon_1, ..., \epsilon_n)'$ has the covariance matrix structure

$$\text{Cov}(\epsilon) = \sigma^2\Omega \tag{4.2}$$

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

$$\delta(X'\Omega^{-1}X)^{-1} \tag{4.3}$$

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

$$\delta(X'X)^{-1}(X'\Omega X)(X'X)^{-1} \tag{4.4}$$

where δ is some positive constant. In the least squares estimation when the matrix Ω is known, Aitken (1935) introduced the GLSE and showed that it has a covariance matrix of the form (4.3) and the least squares estimator (LSE) has a covariance matrix of the form(4.4) with $\delta = \sigma^2$. It is also well known that, when Ω is unknown, the FGLSE has the asymptotic covariance matrix of the form (4.3). Then these two generalized type estimators are more efficient than the LSE.

Although the GLSE and FGLSE 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 are interesting. The concept of developing robust type generalized estimators in regression analysis is not new. For the multivariate regression model, one of linear regression (4.1) with errors of a case of (4.2), Koenker and Portnoy (1990) introduced the concept of generalized M-estimators for the estimation of regression parameters. In this article, we consider the linear regression model of (4.1) with AR(1) errors in the sense that ϵ_i follows

$$\epsilon_i = \rho\epsilon_{i-1} + e_i \tag{4.5}$$

where $e_1, ..., e_n$ are i.i.d. random variables, is one of the most popular models. Suppose that $|\rho| < 1$ and e_i has a distribution function F. We introduce a generalized type and also a feasible generalized type median estimators (i.e. the ℓ_1-norm estimator) and derive their asymptotic properties for this linear regression with AR(1) error model.

We introduce the generalized and feasible generalized type median estimators in Section 2. The theory of these two median estimators is given in Section 3. We provide a Monte Carlo study and an example of real data analysis for the feasible generalized median estimator in Section 4. Finally, the proofs of the theorems are provided in Section 5.

4.2 Generalized and Feasible Generalized Median Estimator

Assume that F has a median 0. The population median of the i-th dependent variable y_i given both independent variables x_i and error variable ϵ_{i-1} is

$$F^{-1}_{y_i|x_i,\epsilon_{i-1}}(0.5) = X_i'\beta + \rho\epsilon_{i-1}. \tag{4.6}$$

Inserting the relation of (4.5) for y_{i-1}, we have $y_i \leq F^{-1}_{y_i|x_i,\epsilon_{i-1}}(0.5)$ iff $y_i - \rho y_{i-1} \leq (x_i - \rho x_{i-1})'\beta$ iff $e_i \leq 0$. This discussion also implies that

$$F^{-1}_{y_i|x_i,\epsilon_{i-1}}(0.5) = \rho y_{i-1} + (x_i - \rho x_{i-1})'\beta \tag{4.7}$$

which is conditional on independent variables x_i and preceeding dependent variable y_{i-1}. We can reformulate the conditional median of (4.7) as

$F^{-1}_{y_i|x_i,\epsilon_i}(0.5) = \rho y_{i-1} + (x_i - \rho x_{i-1})'\beta$. Then, once we have known ρ or an estimator $\hat{\rho}$ when it is unknown, we can estimate β through this formulation. For the estimation of β, we consider this transformation in a matrix form as

$$y = X\beta + \epsilon \tag{4.8}$$

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

$$\Omega = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}. \tag{4.9}$$

Define the half matrix of Ω^{-1} as

$$(\Omega^{-1/2})' = \begin{bmatrix} (1-\rho^2)^{1/2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix}. \tag{4.10}$$

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

$$u = Z\beta + ((1-\rho^2)^{1/2}\epsilon_1, e_2, e_3, ..., e_n)' \tag{4.11}$$

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 (4.8) can be carried over straightforwardly to transformed model (4.11) 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 GLSE is simply the LSE of β for model (4.11), we may consider the median estimator (generalized or feasible generalized) defining on this transformed model. To validate the terminology calling the generalized and the feasible generalized median estimators, we will show that they are asymptotically equivalent in the sense of having the same asymptotic covariance matrix of the form of (4.3).

Definition 4.1. The generalized median estimator for the linear regression with AR(1) error model is defined as

$$\hat{\beta}_G = \arg_{b \in R^p}\min \sum_{i=1} |u_i - z_i'b|$$

where u_i and z_i' are the i-th rows of u and Z respectively.

After the establishment of the generalized median estimator for the known ρ, question will be to find if there is a similar estimator for the unknown ρ. Specifically, can the estimators of the regression median of Definition 4.1 with replacing ρ by an consistent estimator $\hat{\rho}$ have asymptotic behavior exactly the same as it displayed for $\hat{\beta}_G$ when ρ is known. If yes, the theory of feasible generalized least squares estimation is then carried over to the theory of feasible generalized median estimator in this specific linear regression model. Let $\hat{\Omega}$ be the matrix Ω replacing its ρ by $\hat{\rho}$ and we define matrices $\hat{u} = (\hat{\Omega}^{-1/2})'y$, $\hat{Z} = (\hat{\Omega}^{-1/2})'X$ and $\hat{e} = (\hat{\Omega}^{-1/2})'\epsilon$.

Definition 4.2. The feasible generalized median estimator for the linear regression with AR(1) error model is defined as

$$\hat{\beta}_{FG} = \arg_{b \in R^p} \min \sum_{i=1}^{n} |\hat{u}_i - \hat{z}_i' b|$$

where $\hat{u}_i$ and $\hat{z}_i'$ are i-th rows of $\hat{u}$ and $\hat{Z}$ respectively.

4.3 Asymptotic theory of generalized and feasible generalized median estimators

Without examining these generalized and feasible generalized median estimators, we have still not known if they do play the role of generalized or feasible generalized robust estimators. A set of assumptions related to the design matrix X and the distribution of the error variable e in the Section 4.5 are assumed to be true throughout the paper. The following theorem states that the generalized and feasible generalized median estimators have the same Bahadur representation and the same asymptotic distribution.

Theorem 4.3.1. *Assume that* $Q_\rho = lim_{n \to \infty} n^{-1} X' \Omega^{-1} X$*, a positive definite matrix, and* $\hat{\rho}$ *satisfies that* $n^{1/2}(\hat{\rho} - \rho) = O_p(1)$*. Then both* $n^{1/2}(\hat{\beta}_G - \beta)$ *and* $n^{1/2}(\hat{\beta}_{FG} - \beta)$ *have the same representation,*

$$0.5(f(0)Q_\rho)^{-1} n^{-1/2} \sum_{i=1} z_i sgn(e_i) + o_p(1) \tag{4.12}$$

where f *is the p.d.f. of the i.i.d. variables* $e_1, ..., e_n$ *and* sgn *is the sign function with value* 1 *if* $e_i \geq 0$ *and* -1 *if* $e_i < 0$*. Furthermore,* $n^{1/2}(\hat{\beta}_G - \beta)$

and $n^{1/2}(\hat{\beta}_{FG} - \beta)$ *have the same asymptotic distribution that is normal with mean zero vector and the covariance matrix*

$$0.25 f^{-2}(0) Q_\rho^{-1}.$$

We have several aspects to explain the above theorem.

(a). According to (4.12), the median estimators $\hat{\beta}_G$ and $\hat{\beta}_{FG}$ have the same asymptotic covariance matrix of the form $\gamma(X'\Omega X)^{-1}$ with $\gamma = (0.5)^2 f^{-2}(0)$, and thus $\hat{\beta}_G$ and $\hat{\beta}_{FG}$ are asymptotically generalized and feasible generalized, respectively, estimators of the population regression parameter β. In contrast with the questions raised in Section 1, from a large sample point of view on the linear regression with AR(1) error model, we have extended the concept of generalized and feasible generalized estimators from least squares estimation to the median estimator.

(b).If we let $\rho = 0$, these representations for $\hat{\beta}_G$ and $\hat{\beta}_{FG}$ are exactly the same as the median estimator for usual linear regression model (see this in Ruppert and Carroll (1980)). Theorem 4.3.1 indicates that we have generalized the median estimation theory from the linear regression model with i.i.d. errors to that with AR(1) errors.

(c). It is interesting that the representations for these two median estimators are free of the representation of $\hat{\rho}$. Available estimators (see the two examples in next section) of ρ that are asymptotically normal may be seen in Fomby, Hill and Johnson (1980, 211-213), for the proof of asymptotic normality in detail, see Theil (1971).

4.4 Monte Carlo Study and Example

For the feasible generalized median estimator, there are two questions worth to be answered through the simulation study. We know that, as indicated from Theorem 4.3.1, all estimators of ρ that are asymptotically normal make the feasible generalized median estimators converging to the same normal distribution. Among the choices of asymptotically normal estimator $\hat{\rho}$, the Cochran-Orcutt (C-O) method and the Theil's method are most popular in application. Then, the first question is that if these two estimators of ρ make the two corresponding feasible generalized median estimators performing equivalently in simulation. 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 $\frac{\sum_{i=2}^n \hat{\epsilon}_i\hat{\epsilon}_{i-1}}{\sum_{i=2}^n \hat{\epsilon}_i^2}$ and the Theil's method defines $\hat{\rho}$ by $\frac{(\sum_{i=2}^n \hat{\epsilon}_i\hat{\epsilon}_{i-1})/(n-1)}{(\sum_{i=1}^n \hat{\epsilon}_i^2)/(n-p)}$. We perform a simulation to study this problem. 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. The error variable e is standard normal and x_i are independent normal random variables with mean $i/2$ and variance 1. A total of 1000 replications were performed where parameter values of ρ are $-0.9, -.07, -0.5, -0.3, 0$. We compute the mean squares errors for these two types of feasible generalized median estimators where the total mean squared error is the square of the Euclidean distance between the median estimator and regression parameter β. The mean squares errors are listed in Table 1.

ρ	-0.9	-0.7	-0.5	-0.3	0
C_O	0.0672	0.0830	0.1024	0.1274	0.2129
Theil	0.0669	0.0828	0.1021	0.1279	0.2129

Table 1. MSE's of feasible generalized median estimator using C-O and Theil's methods

The feasible generalized median estimators of using C-O and Theil methods to estimate ρ are nearly indifferent in MSE's. With this result, we then further concern the question if this feasible generalized median estimator is robust comparing with FGLSE. To answer this question, we also conduct a simulation.

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 $\delta = 0.1, 0.2$ and $\sigma = 1, 3, 5, 10, 25$. We compute the MSE's for FGLSE and the feasible generalized median estimator. We display the MSE's in Table 2.

$\rho = -0.9$	FGLSE	$\hat{\beta}_{FG}$	$\rho = -0.5$	FGLSE	$\hat{\beta}_{FG}$
$\delta = 0$	0.040	0.062			
$\delta = 0.1$			$\delta = 0.1$		
$\sigma = 3$	0.073	0.083	$\sigma = 3$	0.176	0.122
5	0.116	0.091	5	0.249	0.183
10	0.595	0.072	10	0.571	0.114
25	1.559	0.108	25	2.350	0.133
$\delta = 0.2$			$\delta = 0.2$		
$\sigma = 3$	0.137	0.076	$\sigma = 3$	0.135	0.149
5	0.192	0.089	5	0.364	0.110
10	1.207	0.178	10	1.299	0.186
25	5.129	0.259	25	8.811	0.336

Table 2. MSEs for GLSE and feasible generalized median estimator

We have several conclusions drawn from Table 2:

(a). The case $\delta = 0$ indicates that e_i follows a normal distribution. Then the results in table 2 full fill the statistical theory that the FGLSE is more efficient than other consistent estimators.

(b). In cases $\delta > 0$, almost all median estimates are with smaller MSE's relative to their corresponding FGLSE's. This result shows that the feasible generalized median estimator is indeed, among the class of feasible generalized estimators, a robust one.

The data described in Dielman (1996) can be used as an example to examine these methods. With sample size $n = 16$, this data set included corporate profits (in billion dollars) and gross national product (GNP) (in billion dollars). The regression model proposed was

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

here y's and x's represent the corporate profit and GNP respectively. Following the theory of economics, we anticipate the parameter β_1 to be positive. Dielman (1996) has shown that the disturbances, ϵ_i, have first-order autocorrelation $\epsilon_i = \rho\epsilon_{i-1} + e_i$ by rejecting the null hypothesis $\rho = 0$. Since both the FGLSE and the generalized median estimator have estimates of right sign on β_1, our main purpose for analyzing this data set will focus on comparing the performace of their corresponding approximated confidence intervals.

By applying normal theory, we construct the asymptotic confidence intervals of β_0 and β_1 by using the estimators of the asymptotic variances of elements of $\hat{\beta}'_{FG} = (\hat{\beta}_0, \hat{\beta}_1)$. We let the estimator of Q_ρ be $\hat{Q}_\rho = n^{-1}X'\hat{\Omega}^{-1}X$ and the estimator of $f(0)$ be $\hat{f}(0) = \frac{1}{nh_0}\sum_{i=1} I(-\frac{h_0}{2} < e_i \leq \frac{h_0}{2})$, using the uniform kernel, where $h_0 = 1.740 \times 1.059 \times \hat{\sigma} \times n^{-1/5}$ as suggeted by Simonff (1996) and $\hat{\sigma}^2 = \frac{1}{n-p}\hat{e}'\hat{e}$. The estimates of ρ based on C-O and Theil's methods are 0.7523 and 0.7155 respectively. The 90% and 95% confidence intervals for β_0 and β_1 based on these methods are presented in Table 3.

Parameter	FGLSE	$\hat{\beta}_{FG}$
$\gamma = 0.95$		
$C-O:\ \beta_0$	$(1.2665, 17.587)$	$(29.380, 41.510)$
	16.320	12.130
β_1	$(0.0265, 0.0313)$	$(0.0236, 0.0271)$
	0.0047	0.0035
$Theil:\ \beta_0$	$(4.7028, 16.307)$	$(21.086, 30.531)$
	11.604	9.4446
β_1	$(0.0273, 0.0311)$	$(0.0264, 0.0295)$
	0.0037	0.0030
$\gamma = 0.90$		
$C-O:\ \beta_0$	$(2.4689, 16.385)$	$(30.273, 40.616)$
	13.696	10.180
β_1	$(0.0269, 0.0309)$	$(0.0238, 0.0268)$
	0.0039	0.0029
$Theil:\ \beta_0$	$(5.6357, 15.374)$	$(21.846, 29.772)$
	9.7390	7.9262
β_1	$(0.02767, 0.0308)$	$(0.0267, 0.0293)$
	0.0031	0.0025

Table 3. Approximated confidence intervals and length for GNP data

There are two conclusions can be drawn from Table 3:

(a). Confidence intervals generated from the Theil's method are shorter than the corresponding ones generated from the C-O method.

(b). Confidence intervals generated from the proposed feasible generalized median estimator are shorter than the corresponding ones generated from the feasible generalized least squares estimator.

4.5 Proof of Theorem 4.3.1

The following conditions concerning the design matrix X 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 (1987):

(a1) $n^{-1}\sum_{i=1} x_{ij}^4 = O(1)$ for all j.

(a2) $n^{-1}X'\Omega^{-1}X = Q_\rho + o(1)$, where Q_ρ is a positive definite matrix.

(a3) The probability density function f and its derivative are both bounded and bounded away from 0 in a neighborhood of $F^{-1}(\alpha)$ for $\alpha \in (0, 1)$.

(a4) $n^{1/2}(\hat{\rho} - \rho) = O_p(1)$.

Proof of Theorem 3.1 : Consider first the representation of the feasible generalized median estimator $\hat{\beta}_{FG}$. According to Koenker and Bassett (1978), $\hat{\beta}_{FG}$ can be expressed as

$$\hat{\beta}_{FG} = \arg_{b\in R^p}\min \sum (\hat{u}_i - \hat{z}_i'b)(0.5 - I(\hat{u}_i \leq \hat{z}_i'b)).$$

From (4.1) and (4.5), we have

$$\hat{u}_i - \hat{z}_i'\hat{\beta}_{FG} = e_i - n^{-1/2}T_2\epsilon_{i-1} + n^{-1/2}(-z_i'T_1 + n^{-1/2}T_2x_{i-1}'T_1)$$

where $T_1 = n^{1/2}(\hat{\beta}_{FG} - \beta)$ and $T_2 = n^{1/2}(\hat{\rho} - \rho)$. Let

$$M(t_1, t_2) = n^{-1/2}\sum_{i=1}^{n} z_i(0.5 - I(e_i - n^{-1/2}t_2\epsilon_{i-1} \leq n^{-1/2}z_i't_1 - n^{-1}t_2x_{i-1}'t_1)).$$

Following the proof of Theorem 5.5 of Chen, Welsh and Chan (2001), we can derive

$$\sup_{\|t_1\|\leq k, |t_2|\leq k'}|M(t_1, t_2) - M(0, 0) - E(M(t_1, t_2) - M(0, 0))| = o_p(1) \quad (4.13)$$

and

$$\sup_{\|t_1\|\leq k, |t_2|\leq k'} |E(M(t_1,t_2) - M(0,0)) + f(0)n^{-1}\sum_{i=1} z_i z_i' t_1| = o_p(1) \quad (4.14)$$

for $k, k' > 0$. Moreover, Ruppert and Carroll (1980) provides that

$$n^{-1/2}\sum_{i=1}^{n} z_i(0.5 - I(\hat{u}_i \leq \hat{z}_i'\hat{\beta}_{FG})) = o_p(1). \quad (4.15)$$

From (4.13)-(4.15), we can derive

$$n^{1/2}(\hat{\beta}_{FG} - \beta) = O_p(1). \quad (4.16)$$

Then the representation of $\hat{\beta}_{FG}$ is followed from (4.15), (4.16) and the following result, induced from (4.13) and (4.14),

$$M(T_1^*, T_2^*) - M(0,0) + f(0)n^{-1}\sum_{i=1} z_i z_i' T_1^* = o_p(1)$$

for any sequences $T_1^* = O_p(1)$ and $T_2^* = O_p(1)$.

Letting $t_2 = 0$, the proof for the representation of $\hat{\beta}_G$ is exactly the same as for $\hat{\beta}_{FG}$.

Chapter 5

Generalized and Feasible Generalized Trimmed Means for the Linear Regression with AR(1) Error Model

In this chapter we propose the generalized and the feasible generalized trimmed means for the linear regression with AR(1) errors model. These play the role of robust type generalized and feasible generalized estimators for this regression model. Their asymptotic distributions are developed. We also show that the Gauss-Markov theorem holds for these two trimmed means in the sense that they are asymptotically the best in two corresponding classes of linear trimmed means.

5.1 Introduction

For some regression models such as linear regression with AR(1) errors, with uncorrelated but unequal variances errors or the seemingly unrelated regression model, the generalized least squares estimator (GLSE) and feasible generalized least squares estimator (FGLSE) have some advantages such as with variances (or asymptotic variance) smaller than the least squares estimator (LSE) and being best (or asymptotically) linear unbiased estimator. However, the GLSE and the FGLSE 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 concept 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 the interest of robust type generalized and feasible generalized estimators for estimation of regression parameters. Unlike the multivariate regression, we consider the linear regression with AR(1) error model

$$\begin{aligned} y_i &= x_i'\beta + \epsilon_i, i = 1, ..., n \\ \epsilon_i &= \rho\epsilon_{i-1} + e_i \end{aligned} \tag{5.1}$$

where $|\rho| < 1$, $e_i, i = 1, ..., n$ 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 of the estimation of β, it is known that, when ρ is known, the GLSE and, when ρ is unknown, the FGLSE are with (or asymptotically with) the same covariance matrix which is smaller than it of the LSE. To see the sensitivity of the GLSE and the FGLSE, by letting $X' = (x_1, ..., x_n)$ and $\Omega = Cov(\epsilon)$ with $\epsilon = (\epsilon_1, ..., \epsilon_n)'$, the GLSE and the FGLSE all have (asymptotic) covariance matrix of the form

$$\sigma^2(X'\Omega^{-1}X)^{-1} \tag{5.2}$$

where the term σ^2 is the variance of e_i. The sensitivity is revealing from that σ^2 could be arbitrary large when e_i obeys heavy tail distribution.

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

$$\gamma(X'\Omega^{-1}X)^{-1} \tag{5.3}$$

where robustness means that it has γ insensitive in heavy tail distribution. Based on the regression quantiles of Koenker and Bassett (1978), we will introduce the generalized trimmed mean (GTM) and feasible generalized trimmed mean (FGTM) to play the role of robust type generalized and feasible generalized estimators for the linear regression with AR(1) errors model. For advancing study of their properties, we will also show that the theory of robust type Gauss-Markov theorem holds asymptotically for the GTM and FGTM in the sense that they are the best in their corresponding classes of trimmed means linear in trimmed observations.

We introduce the concepts of GTM and FGTM in Section 2 and establish their large sample theory in Section 3. In Section 4, we introduce a best asymptotic linear estimation property for the GTM and FGTM in Section 4. Finally the proofs of theorems are displayed in Section 5.

5.2 Generalized and Feasible Generalized Trimmed Means

For the linear regression with AR(1) error model (5.1), to obtain a GTM we need to specify quantile for determining the observation trimming and to make a transformation for the linear model for obtaining generalized estimators. For given i-th dependent variable for model (5.1), assuming that $i \geq 2$, one way to derive a generalized estimator is to consider the transformation by the Cochrane and Orcutt (C-O, 1949) as

$$y_i = \rho y_{i-1} + (x_i - \rho x_{i-1})'\beta + e_i.$$

We assume that error variable e has distribution function F with probability density function f. With the transformation for generalized estimation, a qantile 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

$$e_i \leq F^{-1}(\alpha) \tag{5.4}$$

and

$$(-\rho, 1)\begin{pmatrix} y_{i-1} \\ y_i \end{pmatrix} \leq (-\rho, 1)\begin{pmatrix} x'_{i-1} \\ x'_i \end{pmatrix} \beta(\alpha) \tag{5.5}$$

with

$$\beta(\alpha) = \beta + \begin{pmatrix} \frac{1}{1-\rho} F^{-1}(\alpha) \\ 0_{p-1} \end{pmatrix}.$$

Here $\beta(\alpha)$ is called the population regression quantile by Koenker and Bassett (1978). With specification of quantiles and transformation, we will 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 (5.1) which is

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

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

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

Define the half matrix of Ω^{-1} as

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

The C-O transformation is

$$u = Z\beta + ((1-\rho^2)^{1/2}\epsilon_1, e_2, e_3, ..., e_n)' \tag{5.6}$$

where $u = (\Omega^{-1/2})'y$ and $Z = (z_1, ..., z_n)' = (\Omega^{-1/2})'X$. It is known that GLSE is simply the LSE of β for model (5.6).

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

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

where u_i and z_i' are the i-th rows of u and Z, respectively. We then define a generalized trimmed mean based on regression quantiles.

Definition 5.1. Define the trimming matrix as

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

The Koenker and Bassett's type GTM is defined as

$$L_G(\alpha_1, \alpha_2) = (Z'A_nZ)^{-1}Z'A_nu. \tag{5.7}$$

The next interesting problem is then that when parameter ρ is unknown, can the trimmed mean of (5.7) with replacing ρ by an consistent estimator $\hat{\rho}$ have the asymptotic behavior exactly the same as it displayed for $L_G(\alpha_1, \alpha_2)$. 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 $\hat{e} = (\hat{\Omega}^{-1/2})'\epsilon$. Let the regression quantile when the parameter ρ is unknown be defined as

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

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

Definition 5.2. Define the trimming matrix as

$$\hat{A}_n = \text{diag}\{a_i = I(\hat{z}_i'\hat{\beta}_{FG}(\alpha_1) \leq \hat{u}_i \leq \hat{z}_i'\hat{\beta}_{FG}(\alpha_2)) : i = 1, ..., n\}.$$

The Koenker and Bassett's type FGTM is defined as

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

With the C-O transformation, the half matrix $(\Omega^{-1/2})'$ has rows with only a finite number of elements that depend on the unknown parameter ρ. This trick makes the study of asymptotic theory for $\hat{\beta}_F G(\alpha)$ and FGTM $L_{FG}(\alpha_1, \alpha_2)$ similar to what we have for the classical regression quantile and trimmed mean for linear regression.

Large sample representations of the GTM and the FGTM and their role playing as generalized and feasible generalized robust estimators will be introduced in the next section.

5.3 Asymptotic theory of GTM and FGTM

The following conditions concerning the design matrix X and the distribution of error variable e are assumed to be true throughout the following study.

In the following we give a Bahadur representation for the generalized regression quantile which is followed straightforwardly from Theorem 3 of Ruppert and Carroll (1980).

Lemma 5.3.1. *The generalized regression quantile has the representation,*

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

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

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

According to (5.3), the quantile estimator $\hat{\beta}_G(\alpha)$ has asymptotic covariance of the form $\gamma(X'\Omega X)^{-1}$ with $\gamma = \alpha(1-\alpha)f^{-2}(F^{-1}(\alpha))$ 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 5.3.2. *The GTM has the following representation*

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

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

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

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

Corollary 5.3.3. *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

$$\begin{aligned}\sigma^2(\alpha_1,\alpha_2) =& (\alpha_2-\alpha_1)^{-2}[\int_{F^{-1}(\alpha_1)}^{F^{-1}(\alpha_2)}(e-\lambda)^2dF(e) + \alpha_1(F^{-1}(\alpha_1)-\lambda)^2 \\ &+ (1-\alpha_2)(F^{-1}(\alpha_2)-\lambda)^2 - (\alpha_1F^{-1}(\alpha_1)+(1-\alpha_2)F^{-1}(\alpha_2))^2].\end{aligned}$$

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 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 special case of symmetric distribution, the asymptotic distribution of the GTM can be simplified.

Corollary 5.3.4. *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*

$$\sigma^2(\alpha, 1-\alpha) = (1-2\alpha)^{-2}[\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} e^2 dF(e) + 2\alpha(F^{-1}(\alpha))^2].$$

How efficient is the GTM comparing 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 (1980) and Welsh (1987), the Huber's M-estimator and the Welsh's trimmed mean defined on model (5.6) are expected to have the same asymptotic distribution as it in Corollary 5.3.3. These then serve as other types of generalized robust estimators. In general, the parameter ρ is unknown. The interest is then if the FGTM has a representation as the same as it of the GTM? Before to state this result, we need to give a representation of the regression quantile $\hat\beta_{FG}(\alpha)$.

Lemma 5.3.5. *The regression quantile $\hat\beta_{FG}(\alpha)$ has the representation,*

$$n^{1/2}(\hat\beta_{FG}(\alpha) - \beta(\alpha)) = Q_\rho^{-1} f^{-1}(F_\rho^{-1}(\alpha))[n^{-1/2}\sum_{i=1}^n z_i(\alpha - I(e_i \le F_\rho^{-1}(\alpha))) + f(F_\rho^{-1}(\alpha))\theta_z n^{1/2}(\hat\rho - \rho)F_\rho^{-1}(\alpha)] + o_p(1),$$

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

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

Theorem 5.3.6. *The FGTM has the same representation as that expressed for GTM in Theorem 5.3.2.*

From Theorem 5.3.6, the FGTM indeed plays the role of feasible generalized estimator for estimating regression parameter β.

5.4 Best asymptotically linear in K-B trimmed observations

Besides the properties of generalized and feasible generalized estimations that the GTM and FGTM have, do they also have analogous property such as the best linear unbiased estimation that the GLSE and FGLSE have? We will show that the GTM and FGTM are asymptotically the best in classes of estimators linear in Koenker and Bassett's trimmed observations. The design of this linear trimmed means follows the idea of estimators linear in Welsh's trimmed observations by Chen and Welsh (2002).

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

$$HH_0'y. \tag{5.8}$$

To make generalization of the linear unbiased estimators to the trimmed estimators, we consider linear function of trimmed observations $A_n\Omega^{-1/2'}y$ and $\hat{A}_n\hat{\Omega}^{-1/2\prime}y$, respectively, for cases of known and unknown ρ.

Definition 4.1. A statistic $\hat{\beta}_{lg}$ is asymptotically linear in the generalized Koenker and Bassett's trimmed observations (ALGKB) y if

$$\hat{\beta}_{lg} = My^*, \tag{5.9}$$

where $y^* = \Omega^{-1/2} A_n \Omega^{-1/2\prime} y$ and M can be decomposed as $M = HH_0'$ with H a $p \times p$ stochastic or non-stochastic matrix and H_0 a $n \times p$ matrix which is independent of the error variables ϵ, satisfying the following two conditions:

(a6) $nH \to \tilde{H}$ in probability, where $\tilde{H}$ is a full rank $p \times p$ matrix.

(a7) $HH_0'\Omega^{-1/2\prime}X = (\alpha_2 - \alpha_1)^{-1} I_p + o_p(n^{-1/2})$, 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. A question arises for the class of ALGKB estimators. Besides the generalized trimmed mean, does this class of estimators contain interesting estimators? To answer this question, we consider a generalization of Mallows-type bounded influence trimmed means for linear regression model by De Jongh, De Wet and Welsh (1988).

Definition 4.2. The Mallows-type bounded influence generalized trimmed mean is defined as

$$L_{MG}(\alpha_1, \alpha_2) = (Z'WA_nZ)^{-1} Z'WA_n u. \tag{5.10}$$

with W a diagonal matrix of weights. Mallows-type bounded influence generalized trimmed means in terms of weighted matrix W is a subclass of ALGKB estimators seen by letting $H = (Z'WA_nZ)^{-1}$ and $H_0 = ZW$.

Some assumptions related to the design of matrix H_0 are contained in (a1)-(a3). The following theorem gives a Bahadur representation for ALGKB estimators.

Theorem 5.4.3. *Under conditions (a1)-(a7), we have*

$$n^{1/2}(\hat{\beta}_{lg} - (\beta + \gamma_{lg})) = n^{-1/2} \tilde{H} \sum_{i=1}^{n} h_i(\phi(e_i - E(\phi(e_i))) + o_p(1)$$

with $\gamma_{lg} = \lambda \tilde{H} \theta_h$.

The limiting distribution of ALKB estimators is then followed.

Corollary 5.4.4. *Under the conditions of Theorem 5.4.3, the normalized ALGKB estimator* $n^{1/2}(\hat{\beta}_{lg} - (\beta + \gamma_{lg}))$ *has an asymptotic normal distribution with zero mean vector and asymptotic covariance matrix* $(\alpha_2 - \alpha_1)^2 \sigma^2(\alpha_1, \alpha_2) \tilde{H} Q_h \tilde{H}'$.

If we further assume that F is symmetric at 0 and let $\alpha_1 = 1 - \alpha_2 = \alpha$, $0 < \alpha < 0.5$, then $\gamma_{lg} = 0$ and $\hat{\beta}_{lg}$ is a consistent estimator of β. In general, when F is asymmetric, $\hat{\beta}_{lg}$ is asymptotically biased for β and the asymptotic bias is given by γ_{lg}. Again, 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.

Lemma 5.4.5. *For any matrices* $\tilde{H}$ *and* Q_h *induced from conditions (a1) and (a4), the difference* $(\alpha_2 - \alpha_1)^2 \tilde{H} Q_h \tilde{H}' - Q_\rho^{-1}$ *is positive semidefinite.*

Put $H = (Z' A_n Z)^{-1}$ and $H_0 = Z$, we have $n^{-1} Z' A_n Z \to (\alpha_2 - \alpha_1) Q_\rho$ so we can see that conditions (a6) and (a7) hold for $L_G(\alpha_1, \alpha_2)$, and the generalized K-B trimmed mean is an ALGKB estimator. Moreover, Corollary 5.3.3 proved that $n^{1/2}(L_G(\alpha_1, \alpha_2) - (\beta + \gamma_g))$ has an asymptotic normal distribution with zero mean and covariance matrix $\sigma^2(\alpha_1, \alpha_2) Q_\rho^{-1}$.

Theorem 5.4.6. *Under conditions (a1)-(a7), generalized K-B trimmed mean* $L_G(\alpha_1, \alpha_2)$ *defined in (5.6) is a best ALGKB estimator.*

In the class of linear estimators based on the trimmed observations, we have shown that for estimating the parameter vector β, the GTM is a best ALGKB estimator. This proves that the robust type Gauss-Markov theorem holds for the GTM. Since the class of Mallow-type bounded influence generalized trimmed means is a subclass of linear estimators based on the trimmed observations and the generalized K-B trimmed mean is one in this subclass, we then have the following theorem.

Theorem 5.4.7. *The GTM is also the best Mallow-type bounded influence generalized trimmed mean.*

The Gauss-Markov theorem for the FGTM is similar to it of the GTM. We list them in the followings.

Definition 4.8. A statistic $\hat{\beta}_{lfg}$ is asymptotically linear in the generalized Koenker and Bassett's trimmed observations (ALFGKB) y if

$$\hat{\beta}_{lfg} = My^*,$$

where $y^* = \hat{\Omega}^{-1/2}\hat{A}_n\hat{\Omega}^{-1/2\prime}y$ and M satisfies the conditions in Definition 4.1.

Definition 4.9. The Mallows-type bounded influence feasible generalized trimmed mean is defined as

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

Theorem 5.4.10. *$\hat{\beta}_{lfg}$ and $\hat{\beta}_{lg}$ have the same asymptotic distributions.*

Theorem 5.4.11. *The feasible FGTM is also the best Mallow-type bounded influence feasible generalized trimmed mean.*

This establishes the robust version of the Gauss-Markov theorem for the estimation of the linear regression with AR(1) error model.

5.5 Proofs

The following conditions concerning design matrices X and H_0 and 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 (1987):

(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 a positive definite matrix.
(a3) $n^{-1}\sum_{i=1} x_i = \theta_x + o(1)$, 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 $F^{*-1}(\alpha)$ for $\alpha \in (0, 1)$.
(a5) $n^{1/2}(\hat{\rho} - \rho) = O_p(1)$.

Proof of Lemma 5.3.5 : Let

$$M(t_1, t_2) = n^{-1/2}\sum_{i=1}^{n} z_i\{\alpha - I(e_i - n^{-1/2}t_1\epsilon_{i-1} \leq (z_i - n^{-1/2}t_1x_{i_1})'(n^{-1/2}t_2 + F^{*-1}(\alpha)))\}.$$

We want to show that

$$\sup_{||(t_1,t_2)||\leq k}|M(t_1, t_2) - M(0,0) - F^{*-1}(\alpha)f(F^{*-1}(\alpha))n^{-1/2}\sum_{i=1}^{n} z_i(z_i't_2 - t_1F^{*-1}(\alpha))| = o \tag{5.11}$$

By letting, for $k > 0$, $S_n(t_1, t_2) = M(t_1, t_2) - M(0, 0)$, we will prove (5.11) in two steps. In the first step, we will show that

$$\sup_{||(t_1,t_2)||\leq k}|S_n(t_1, t_2) - ES_n(t_1, t_2)| = o_p(1) \tag{5.12}$$

based on Lemma 3.2 in Bai and He (1998).

Now we prove (5.12) by checking the three conditions L_1, L_2 and L_3 in the hypothesis of Lemma 3.2 in Bai and He (1998). First we prove

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} z_i'z_iE|I(e_i - n^{-1/2}t_1\epsilon_{i-1} \leq (z_i - n^{-1/2}t_1x_{i-1})'(n^{-1/2}t_2 + F^{*-1}(\alpha)) \\ - I(e_i - n^{-1/2}t_1^*\epsilon_{i-1} \leq (z_i - n^{-1/2}t_1^*x_{i-1})'(n^{-1/2}t_2^* + F^{*-1}(\alpha))| \\ \leq M(||t_1 - t_1^*|| + ||t_2 - t_2^*||), \text{ for some } M > 0. \end{aligned} \tag{5.13}$$

Define

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

and

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

Represent $A = A_1 + A_2$ as follows,

$$A = n^{-1}\sum_{i=1}^{n} z_i' z_i EI(e_i - n^{-1/2}t_1\epsilon_{i-1} \le (z_i - n^{-1/2}t_1x_{i-1})'(n^{-1/2}t_2 + F^{*-1}(\alpha)),$$
$$e_i - n^{-1/2}t_1^*\epsilon_{i-1} > (z_i - n^{-1/2}t_1^*x_{i-1})'(n^{-1/2}t_2 + F^{*-1}(\alpha)))$$
$$+ n^{-1}\sum_{i=1}^{n} z_i' z_i EI(e_i - n^{-1/2}t_1\epsilon_{i-1} > (z_i - n^{-1/2}t_1x_{i-1})'(n^{-1/2}t_2 + F^{*-1}(\alpha)),$$
$$e_i - n^{-1/2}t_1^*\epsilon_{i-1} \le (z_i - n^{-1/2}t_1^*x_{i-1})'(n^{-1/2}t_2 + F^{*-1}(\alpha)))$$
$$= A_1 + A_2.$$

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

$$A_1 = n^{-1}\sum_{i=1}^{n} z_i' z_i EI(e_i \le z_i'\xi_n - n^{-1/2}t_1U_{i-1}, e_i > z_i'\xi_n - n^{-1/2}t_1^*U_{i-1})$$
$$= n^{-1}\sum_{i=1}^{n} z_i' z_i E\{f(z_i'\xi_n)n^{-1/2}||t_1 - t_1^*||U_{i-1}\}$$
$$\le Mn^{-1/2}||t_1 - t_1^*||.$$

Similarly, $A_2 \le Mn^{-1/2}||t_1 - t_1^*||$ and $B \le Mn^{-1/2}||t_2 - t_2^*||$. Hence (5.13) holds and so does the condition (L1) in the hypothesis of Lemma 3.2 in Bai and He (1998). The condition (L2) is satisfied automatically since the indicator function is bounded.

Next, similar arguments to those used to prove (5.11) can be used to prove that the following

$$n^{-1}\sum_{i=1}^{n} z_i' z_i E\{\sup_{||t_1-t_1^*||+||t_2-t_2^*||\le d}|I(e_i - n^{-1/2}t_1\epsilon_{i-1} \le (z_i - n^{-1/2}t_1x_{i-1})'(n^{-1/2}t_2$$
$$+ F^{*-1}(\alpha))) \qquad - I(e_i - n^{1/2}t_1^*\epsilon_{i-1} \le (z_i - n^{-1/2}t_1^*x_{i-1})'(n^{-1/2}t_2^* + F^{*-1}(\alpha))\}$$

is bounded by $Mn^{-1/2}d$ which implies that condition (L3) holds. Therefore, from Lemma 3.2 in Bai and He (1998), we obtain

$$\sup_{||(t_1,t_1)||\le K}|S_n(t_1,t_2) - ES_n(t_1,t_2)| = o_p(1). \tag{5.14}$$

On the other hand, through the technique of Chen, Welsh and Chan (2001), we can developed the following,

$$\sup_{\|(t_1,t_2)\|\le k}|E(S_n(t_1,t_2)) - F^{*-1}(\alpha)f(F^{*-1}(\alpha))n^{-1/2}\sum_{i=1}^{n} z_i(z_i't_2 - t_1F^{*-1}(\alpha))| = o_p(1), \tag{5.15}$$

joining (5.12) and (5.15), statement (5.11) holds. Using the method of Jureckova (1977, Lemma (5.12) and (5.11)) again, $n^{1/2}(\hat{\beta}(\alpha)-\beta(\alpha)) = O_p(1)$ is obtained. Thus, the theorem is proved.

Proof of Theorem 5.3.6. The FGTM can be formulated as

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

Since $n^{1/2}(\hat{\rho}-\rho) = O_p(1)$, we have $n^{-1/2}\hat{Z}'A_n\hat{e} = n^{-1/2}\hat{Z}'A_n e + o_p(1)$. By letting $M(t_1,t_2,\alpha) = n^{-1/2}\sum_{i=1}^n z_i e_i I(e_i - n^{-1/2}t_1\epsilon_{i-1} \leq F^{*-1}(\alpha) + n^{-1/2}(z_i + n^{-1/2}t_1 x_{i-1})'t_2 + n^{-1/2}t_1 F^{*-1}(\alpha))$, we see that

$$n^{-1/2}\hat{Z}'A_n e = M(T_1^*(\alpha_2), T_2^*, \alpha_2) - M(T_1^*(\alpha_1), T_2^*, \alpha_1) \tag{5.16}$$

with $T_1^*(\alpha) = n^{1/2}(\hat{\beta}(\alpha)-\beta(\alpha))$ and $T_2^* = n^{1/2}(\hat{\rho}-\rho)$. However, using the similar techniques of the proof for Lemma 5.3.5, we can see that

$$M(T_1,T_2,\alpha)-M(0,0,,\alpha) = F^{*-1}(\alpha)f(F^{*-1}(\alpha))n^{-1/2}\sum_{i=1}^n z_i(z_i'T_2 - T_1F^{*-1}(\alpha)) + o_p(1) \tag{5.17}$$

for any sequences $T_1 = O_p(1)$ and $T_2 = O_p(1)$. Then, from Lemma 5.3.1, (5.16) and (5.17), we have

$$n^{-1/2}\hat{Z}'A_n e = n^{-1/2}\sum_{i=1}^n z_i[e_i I(F^{*-1}(\alpha_1) \leq e_i \leq F^{*-1}(\alpha_2)) + F^{*-1}(\alpha_2)(\alpha_2 - I(e_i \leq F^{*-1}(\alpha_2))) - F^{*-1}(\alpha_1)(\alpha_1 - I(e_i \leq F^{*-1}(\alpha_1)))] + o_p(1). \tag{5.18}$$

Also, similar discussion of the proof for Lemma 5.3.5 provides the result

$$n^{-1}\hat{Z}'A_n\hat{Z} = Q_\rho + o_p(1). \tag{5.19}$$

Then (5.18) and (5.19) imply the theorem.

Proof of Lemma 5.4.5. Write $\text{plim}(B_n) = B$ if B_n converges to B in probability. Let $C = HH_0 - (Z'AZ)^{-1}Z'$.

Now $\text{plim}(CAZ) = \text{plim}(HH_0'AZ) - \text{plim}(Z'AZ)^{-1}\hat{Z}'AZ = 0$. Hence

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

The following assumptions are needed for the proof of Theorems in Section 4:

(b3) $n^{-1}\sum_{i=1}^n z_{ij}^4 = O(1)$ for $z = x$ or h and all j,

(b4) $n^{-1}X'X = Q_x + o(1)$, $n^{-1}H_0'X = Q_{hx} + o(1)$ 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.

(b5) $n^{-1}\sum_{i=1}^n h_i = \theta_h + o(1)$.

The proof of Theorem 5.4.3 may be simplified from the proof of Theorem 5.4.9, so it is skipped.

Proof of Theorem 5.4.9. From condition (a2) and (A.10) of Ruppert and Carroll (1980), $HH_0'AX\beta = \beta + o_p(n^{-1/2})$. Inserting (5.1) in equation (5.9), we have

$$n^{1/2}(\hat{\beta}_{lfg} - \beta) = n^{1/2}HH_0'A_ne + o_p(1).$$

Now we develop a representation of $n^{-1/2}H_0'A_ne$. Let

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

and

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

Also, let

$$T_n^*(\alpha) = n^{1/2}[\hat{\beta}_{FG}(\alpha) - \beta(\alpha)].$$

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

$$| U_j(\alpha, T_n) - U_j(\alpha, 0) - n^{-1}F^{-1}(\alpha)f(F^{-1}(\alpha))\sum_{i=1}^{n} h_{ij}x_i'T_n |= o_p(1), \quad (5.20)$$

for $j = 1, ..., p$ and $T_n = O_p(1)$. From (5.20) and Definition 4.1,

$$\begin{aligned} n^{-1/2}H_0'Ae &= (U(\alpha_2, T_n^*(\alpha_2)) - U(\alpha_2, 0)) - (U(\alpha_1, T_n^*(\alpha_1)) - U(\alpha_1, 0)) \\ &+ (U(\alpha_2, 0) - U(\alpha_1, 0)) \\ &= n^{-1/2}\sum_{i=1}^{n}[h_i e_i I(F^{-1}(\alpha_1) \leq e_i \leq F^{-1}(\alpha_2)) + (F^{-1}(\alpha_2)I(e_i > F^{-1}(\alpha_2)) \\ &+ F^{-1}(\alpha_1)I(e_i < F^{-1}(\alpha_1))h_i - ((1-\alpha_2)F^{-1}(\alpha_2) + \alpha_1 F^{-1}(\alpha_1))h_i] + o_p(1) \end{aligned} \quad (5.21)$$

Then (5.21) and Condition (a6) indicate that $\hat{\beta}_{lfg}$ has a representation as the same as the one in Theorem 5.4.3 for $\hat{\beta}_{lg}$ which implies the theorem.

Chapter 6

Generalized Trimmed Means for the Nonlinear Regression With AR(1) Error Model

In this work, we apply the idea of trimmed mean of Welsh (1987) to introduce generalized and feasible generalized trimmed means for the nonlinear regression with AR(1) error model. We show that these estimators are asymptotically more efficient than the trimmed means. These results then extend the concept of generalized and feasible generalized least squares estimators for linear regression with AR(1) error model to the robust estimators for nonlinear regression models.

6.1 Introduction

We consider the general nonlinear regression model

$$y_i = g(x_i, \beta) + \epsilon_i, i = 1, ..., n \tag{6.1}$$

where y_i and x_i are, respectively, the response variables and vectors of independent variables, and ϵ_i are error variables. Concerning with estimating regression parameter vector β in the nonlinear regression model with i.i.d. errors, the least squares estimators have been extensively studied. For instance, Hartley and Booker (1965), Jennrich (1969) and Wu (1981) demonstrated the asymptotic normality property. Whereas Ivanov (1976), and Ivanov and Zwanzig (1983) derived an asymptotic expansion for its distribution. Under some regularity conditions and assuming that the errors have common mean 0 and variance σ^2, then, if we let $\hat{\beta}_{lse}$ represents the least squares estimator (LSE), it has the asymptotic covariance matrix of the form

$$\sigma^2(X(\beta)'X(\beta))^{-1}$$

where $X(b) = (d_1(b), ..., d_n(b))'$ with $d_i(b) = \frac{\partial g(x_i,b)}{\partial b}$. However, outliers or heavy tail error distribution may makes σ^2 large that heavily decreases the efficiency of the LSE.

For increasing the efficiency of the nonlinear least squares estimator, robust estimation aims to develop estimators that have asymptotic covariance matrices of the form

$$\delta(X(\beta)'X(\beta))^{-1}$$

where δ is positive and bounded in error distribution. Among the robust approaches, several authors have proposed and studied some L-estimators. Oberhofer (1982), Richardson and Bhattacharyya (1987) and Wang (1995) studied the ℓ_1-norm estimators, whereas Liese and Vajda (1994) studied the theory of M-estimator. Additionally, from a computational aspect, Procházka (1988) and Koenker and Park (1992) studied the trimmed least squares estimator based on regression quantiles of Koenker and Bassett (1978). From a theoretical aspect, Jurečková and Procházka (1994) studied it for that model (1.1) includes an intercept term. This trimmed mean is nice to have representation of the form of location trimmed mean. Recently, Huang, Yang and Chen (2004) studied a trimmed mean of Welsh (1987) that has the advantage of easy computation but has the representation in Jurečková and Procházka (1994).

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

$$\sigma^2\Omega \tag{6.2}$$

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

$$\delta(X(\beta)'\Omega^{-1}X(\beta))^{-1} \tag{6.3}$$

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

$$\delta(X(\beta)'X(\beta))^{-1}X(\beta)'\Omega X(\beta)(X(\beta)'X(\beta))^{-1}. \tag{6.4}$$

In the linear regression model with error structure of (6.2), Aitken (1935) call estimators with covariance matrices of the form of (6.3) the generalized estimators. The question we may be interesting is if we have a robust estimator for the nonlinear regression model of (6.1) with error structure of (6.2) that has asymptotic covariance matrix in the form of (6.3).

In this article, we consider the nonlinear regression model of (6.1) with AR(1) errors in the sense that ϵ_i follows

$$\epsilon_i = \rho\epsilon_{i-1} + e_i \tag{6.5}$$

where $e_1, ..., e_n$ are i.i.d. random variables, is one of the most popular models. Suppose that $|\rho| < 1$ and e_i has a distribution function F. We introduce a generalized trimmed mean and derive their asymptotic properties for the regression parameter vector β.

6.2 Generalized Trimmed Means

Consider the nonlinear regression model (6.1) where its errors follow the structure of (6.5). For simplification, denote $D_i(b) = \frac{\partial^2 g(x_i,b)}{\partial b \partial b'}$, the second order partial derivative of the regression function with respect to vector b. The least squares estimate, using the quadratic approximation, is defined as the convergent estimator of the sequence defined by

$$\begin{aligned} b_j =& b_{j-1} + [\sum_{i=1}^{n}(d_i(b_{j-1})d_i'(b_{j-1}) - (y_i - g(x_i, b_{j-1}))D_i(b_{j-1}))]^{-1} \\ & \sum_{i=1}^{n} d_i(b_{j-1})(y_i - g(x_i, b_{j-1})) \end{aligned} \tag{6.6}$$

where b_0 is a fixed vector.

It is seen that $\text{Cov}(\epsilon) = \sigma^2\Omega$ with

$$\Omega = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}. \tag{6.7}$$

Define the half matrix of Ω^{-1} as

$$(\Omega^{-1/2})' = \begin{pmatrix} (1-\rho^2)^{1/2} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}. \tag{6.8}$$

Define matrix $Z(b) = (\Omega^{-1/2})'X(b)$. Let $\hat{\beta}_I$ be a predetermined estimator of β. We also define vector $V = (\Omega^{-1/2})'(y - G_n(\hat{\beta}_I))$ where $G_n(b) = (g(x_1, b), \cdots, g(x_n, b))'$ and residuals $e_i = y_i - g(x_i, \hat{\beta}_I), i = 1, ..., n$. Denote the α-th residual quantile as $\eta_n(\alpha)$ and we let $z_i'(b)$ and v_i be, respectively, $i - th$ row of $Z(b)$ and $i - th$ element of V. Combining the quadratization method (6.7) with the construction of Welsh's trimmed mean allows us to define the generalized trimmed mean as the convergent estimator of the sequence defined in the following.

Definition 6.1. The generalized trimmed mean for the nonlinear regression model is

$$\begin{aligned} L_G(\alpha_1, \alpha_2) =& \hat{\beta}_I + [\sum_{i=1}^{n}(z_i(\hat{\beta}_I)z_i'(\hat{\beta}_I) - v_i M_i(\hat{\beta}_I))I(\eta_n(\alpha_1) \leq v_i \leq \eta_n(\alpha_2))]^{-1} \\ & \sum_{i=1}^{n} z_i(\hat{\beta}_I)[v_i I(\eta_n(\alpha_1) \leq v_i \leq \eta_n(\alpha_2)) + \eta_n(\alpha_1)(I(v_i \leq \eta_n(\alpha_1)) - \alpha_1) \\ & + \eta_n(\alpha_2)(I(v_i \geq \eta_n(\alpha_2)) - (1-\alpha_2))]. \end{aligned} \tag{6.9}$$

After the development of the generalized trimmed mean, the next interesting problem is whether when the parameter ρ is unknown, the trimmed mean of (6.9) 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 nonlinear regression model. Let $\hat{\Omega}$ be the matrix of Ω with ρ replaced by its consistent estimator $\hat{\rho}$. Define matrices $\hat{Z}(b) = (\hat{\Omega}^{-1/2})'X(b)$ and $\hat{V} = (\hat{\Omega}^{-1/2})'(y - G_n(\hat{\beta}))$. We also let $\hat{z}_i'(b)$ and $\hat{v}_i$ be, respectively, $i - th$ row of $\hat{Z}(b)$ and $i - th$ element of $\hat{V}$.

Definition 6.2. The feasible generalized trimmed mean for the nonlinear regression model is

$$\begin{aligned} L_{FG}(\alpha_1,\alpha_2) =& \hat{\beta}_I + [\sum_{i=1}^{n}(\hat{z}_i(\hat{\beta}_I)\hat{z}_i'(\hat{\beta}_I) - \hat{v}_i M_i(\hat{\beta}_I))I(\eta_n(\alpha_1) \le \hat{v}_i \le \eta_n(\alpha_2))]^{-1} \\ & \sum_{i=1}^{n} \hat{z}_i(\hat{\beta}_I)[\hat{v}_i I(\eta_n(\alpha_1) \le \hat{v}_i \le \eta_n(\alpha_2)) + \eta_n(\alpha_1)(I(\hat{v}_i \le \eta_n(\alpha_1)) - \alpha_1) \\ & + \eta_n(\alpha_2)(I(\hat{v}_i \ge \eta_n(\alpha_2)) - (1-\alpha_2))]. \end{aligned} \tag{6.10}$$

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 $\hat{\beta}_{PG}(\alpha)$ and PGTM $L_{PG}(\alpha_1,\alpha_2)$ 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.

6.3 Large Sample Properties of Generalized Trimmed Mean

We state a set of assumptions (a1-a5) related to the design matrix X and the distribution of the error variable e in the Section 6.4 that are assumed to be true throughout the paper.

Lemma 6.3.1. The quantile η_n has the following representation

$$n^{1/2}(\eta_n(\alpha) - F^{-1}(\alpha)) = f^{-1}(F^{-1}(\alpha))n^{-1/2}\sum_{i=1}^{n}(\alpha - I(e_i \le F^{-1}(\alpha)))) + o_p(1).$$

Theorem 6.3.2. The generalized trimmed mean has the following representation

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

where $\gamma = \lambda(\alpha_2-\alpha_1)^{-1}Q_\rho^{-1}\theta$ with

$$\lambda = \frac{1-\rho}{\alpha_2-\alpha_1}\int_{F^{-1}(\alpha_1)}^{F^{-1}(\alpha_2)} ef(e)de,$$

and

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

.

For statistical inference, we need an asymptotic distribution of the generalized trimmed mean which is stated in the following.

Corollary 6.3.3

(a)

$$n^{1/2}(L_G(\alpha_1,\alpha_2)-(\beta+\gamma)) \to N(0,\sigma^2(\alpha_1,\alpha_2)Q_\rho^{-1})$$

where

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

(b) If F is further assumed to be symmetric and we let $\alpha_1 = \alpha = 1-\alpha_2$, $0 < \alpha < 0.5$, then

$$n^{1/2}(L_G(\alpha,1-\alpha)-\beta) \to N(0,\sigma^2(\alpha,1-\alpha)Q_\rho^{-1})$$

where in this situation

$$\sigma^2(\alpha,1-\alpha) = (1-2\alpha)^{-2}(2\alpha(F^{-1}(1-\alpha))^2 + \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} e^2 dF).$$

The asymptotic covariance matrix of $L_G(\alpha_1, \alpha_2)$ is also of the form γQ_ρ^{-1} with $\gamma = \sigma^2(\alpha_1, \alpha_2)$ which is the asymptotic variance of the trimmed mean for the location model. 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.

Lemma 6.3.4. The quantile function $\hat{\eta}_{FG}$ has the following representation

$$\begin{aligned}\sqrt{n}(\hat{\eta}_{FG} - F^{-1}(\alpha)) =& f^{-1}(F^{-1}(\alpha))n^{-1/2}\sum_{i=1}^{n}(\alpha - I(e_i \leq F^{-1}(\alpha))) - \\ &+ \theta'_\rho n^{1/2}(\hat{\beta}_I - \beta) + o_p(1).\end{aligned}$$

An interesting question is then whether the PGTM has the same representation as that of the GTM.

Theorem 6.3.5. The PGTM has the same representation as that expressed for the GTM in Theorem 6.3.2.

6.4 Proofs

The following are a set of assumptions regarding the design vectors and the distribution function that are assumed to be true throughout this paper:

(a.1) $n^{-1}\sum_{i=1}^n z_i(\beta)z_i'(\beta) = Q_\rho + o(1)$ where Q is a positive definite.

(a.2) $n^{-1}\sum_{i=1}^n z_i(\beta) = \theta_\rho + o(1)$ where θ_ρ is a finite vector depending on ρ.

(a.3) $n^{-1}\sum_{i=1}^n z_{ij}^4(\beta) = O(1)$, $n^{-1}\sum_{i=1}^n D_{ij}^2(\beta) = O(1)$.

(a.4) For $b > 0$

$$\begin{aligned} &n^{-1}\max_{\|\beta\|\le b}\sum_{i=1}^{n}|d_{ij}(\beta)|^2 = O(1),\\ &n^{-1/4}\max_{\|\beta\|\le b}|d_{ij}(\beta)| = O(1),\\ &n^{-1/2}\max_{\|\beta\|\le b}|D_{jk}(\beta)| = O(1),\\ &n^{-1/2}\max_{\|\beta\|\le b}|G_{jkh}(\beta)| = O(1). \end{aligned}$$

(a.5) $n^{1/2}(\hat{\beta}_I - \beta) = O_p(1)$.

(a.6) The probability density function f is bounded away from 0 in a neighborhood of $F^{-1}(\alpha)$, for $0 < \alpha < 1$, and its fourth population moment is finite.

Proof of Theorem 6.3.1
Following Ruppert and Carroll (1980), we have

$$n^{-1/2}\sum_{i=1}^{n}\psi_\alpha(v_i \le \eta_n(\alpha)) = o_p(1). \tag{6.11}$$

From the Taylor expansion of $g(x_i, \beta)$ at $\hat{\beta}_I$ up to order 2, for $t = \begin{pmatrix} t_0 \\ t_1 \end{pmatrix} \in R^{p+1}$, we let

$$\begin{aligned} S(t) =&n^{-1/2}\sum_{i=1}^{n}\psi_\alpha(e_i - F^{-1}(\alpha) - n^{-1/2}t_0 - n^{-1/2}z_i'(\beta + t_1)t_1\\ &+ 0.5n^{-1}t_1'M_i(\beta + t_1)t_1). \end{aligned}$$

By adopting the method of Jurečková (1977, Proof of Lemma 5.2, see also Chen (1988, p72-75)), we can have

$$\max_{\|t\|\le b}|S(t) - S(0) + f(F^{-1}(\alpha))(t_0 - n^{-1}\sum_{i=1}^{n} z_i'(\beta)t_1)| = o_p(1), \text{for } b > 0. \tag{6.12}$$

By using (6.11) and (6.12), we see that, from Jurečková (1984), we have

$$n^{1/2}(\eta_n(\alpha) - F^{-1}(\alpha)) = O_p(1). \tag{6.13}$$

Then, from (6.11)-(6.13),

$$f(F^{-1}(\alpha))(n^{1/2}(\eta_n(\alpha) - F^{-1}(\alpha)) - n^{-1}\sum_{i=1}^{n} d_i'(\beta)n^{1/2}(\hat{\beta}_I - \beta))$$
$$= n^{1/2}\sum_{i=1}^{n} d_i(\beta)\psi_\alpha(\epsilon_i - F^{-1}(\alpha)) + o_p(1).$$

This further implies the theorem.

Proof of Theorem 6.3.2

The representation of $L_G(\alpha_1, \alpha_2)$ is a linear combination of the representation of $\hat{\beta}_I$ and the following

$$[\sum_{\eta_n(\alpha_1)\le v_i \le \eta_n(\alpha_2)} (z_i(\hat{\beta}_I)z_i'(\hat{\beta}_I) - v_i M_i(\hat{\beta}_I))]^{-1} \sum_{\eta_n(\alpha_1)\le v_i \le \eta_n(\alpha_2)} z_i(\beta)(e_i + z_i'(\beta)(\hat{\beta}_I - \beta))$$
$$+ \sum_{\eta_n(\alpha_1)\le v_i \le \eta_n(\alpha_2)} z_i(\beta)[\eta_n(\alpha_1)(I(v_i \le \eta_n(\alpha_1)) - \alpha_1) + \eta_n(\alpha_2)(I(v_i \ge \eta_n(\alpha_2)) - (1 - \alpha_2 \tag{6.14}$$

The representation of $n^{-1/2}\sum_{i=1}^{n} z_i(\beta)e_i I(\eta_n(\alpha_1 \le v_i \le \eta_n(\alpha_2))$ is considered first. Let

$$U(t) = n^{-1/2}\sum_{i=1}^{n} z_i(\beta)e_i I(e_i \le F^{-1}(\alpha) + n^{-1/2}t_0 - n^{-1/2}z_i'(\beta + t_1)t_1$$
$$- 0.5n^{-1}t_1'M_i(\beta + t_1)t_1)$$

Again, the method of Jurečková (1977, Proof of Lemma 5.2) implies that

$$U(T) = U(0) + F^{-1}(\alpha)f(F^{-1}(\alpha))n^{-1}\sum_{i=1}^{n} d_i(\beta)(T_0 + z_i'(\beta)T_1) + o_p(1) \tag{6.15}$$

for any $O_p(1)$ sequences T_0 and T_1. Imposing the facts of assumption (a.5) and (6.15) leads to

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n} z_i(\beta)e_i I(\eta_n(\alpha_1 \leq v_i \leq \eta_n(\alpha_2) = \\
&n^{-1/2}\sum_{i=1}^{n} z_i(\beta)e_i I(F^{-1}(\alpha_1) \leq e_i \leq F^{-1}(\alpha_2)) \\
&- F^{-1}(\alpha_2)f(F^{-1}(\alpha_2))(\theta_\rho n^{1/2}(\eta_n(\alpha_2) - F^{-1}(\alpha_2)) \\
&+ Q_\rho n^{1/2}(\hat{\beta}_I - \beta)) + F^{-1}(\alpha_1)f(F^{-1}(\alpha_1))(\theta_\rho n^{1/2}(\eta_n(\alpha_1) - F^{-1}(\alpha_1)) \\
&+ Q_\rho n^{1/2}(\hat{\beta}_I - \beta)) + o_p(1).
\end{aligned} \tag{6.16}$$

Analogous discussion will see that

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n} z_i(\beta)[\eta_n(\alpha_1)(I(v_i \leq \eta_n(\alpha_1)) - \alpha_1) \\
&+ \eta_n(\alpha_2)(I(v_i \geq \eta_n(\alpha_2)) - (1 - \alpha_2))]I(\eta_n(\alpha_1) \leq v_i \leq \eta_n(\alpha_2)) \\
&= F^{-1}(\alpha_1)[n^{-1/2}\sum_{i=1}^{n} z_i(\beta)(I(e_i \leq F^{-1}(\alpha_1)) - \alpha_1) \\
&- f(F^{-1}(\alpha_1))(\theta_\rho(n^{1/2}(\eta_n(\alpha_1) - F^{-1}(\alpha_1)) - Q_\rho n^{1/2}(\hat{\beta}_I - \beta))] \\
&+ F^{-1}(\alpha_2)[n^{-1/2}\sum_{i=1}^{n} z_i(\beta)(I(e_i \geq F^{-1}(\alpha_2)) - (1 - \alpha_2)) \\
&+ f(F^{-1}(\alpha_2))(\theta_\rho n^{1/2}(\eta_n(\alpha_2) - F^{-1}(\alpha_2)) + Q_\rho n^{1/2}(\hat{\beta}_I - \beta))] + o_p(1)
\end{aligned} \tag{6.17}$$

The theorem is followed by imposing (6.16) and (6.17) in (6.14).

The proof of Lemma 6.3.4 is quite similar to those of Lemma 6.3.1 and Theorem 6.3.5 and then is skipped.

Proof of Theorem 6.3.5.
Note that we may represent $\hat{v}_i$ and $\hat{\eta}_n$ as

$$\begin{aligned}
\hat{v}_i =& e_i - n^{-1/2}z_i'(\beta + n^{-1/2}T_1)T_1 + n^{-1}T_1'M_i(\beta + n^{-1/2}T_1)T_1 \\
&+ n^{-1}T_2 d_{i-1}'(\beta + n^{-1/2}T_1)T_1 - n^{-3/2}T_2T_1'G_{i-1}(\beta + n^{-1/2}T_1)T_1),
\end{aligned} \tag{6.18}$$

$$\hat{\eta}_n = F^{-1}(\alpha) + n^{-1/2}T_0$$

with $T_0 = n^{1/2}(\hat{\eta}_n - F^{-1}(\alpha), T_1 = n^{1/2}(\hat{\beta}_I - \beta)$ and $T_2 = n^{1/2}(\hat{\rho} - \rho)$.

From (6.18) and by denoting

$$C_n = \sum_{i=1}^{n} (\hat{z}_i(\hat{\beta}_I)\hat{z}_i'(\hat{\beta}_I) - \hat{v}_i M_i(\hat{\beta}_I)) I(\eta_n(\alpha_1) \leq \hat{v}_i \leq \eta_n(\alpha_2)),$$

we may see that

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} \hat{z}_i(\hat{\beta}_I)\hat{v}_i I(\hat{\eta}_n(\alpha_1) \leq \hat{v}_i \leq \hat{\eta}_n(\alpha_2)) \\ & = -C_n n^{1/2}(\hat{\beta}_I - \beta) + n^{-1/2} \sum_{i=1}^{n} z_i(\beta) e_i I(\eta_n(\alpha_1) \leq \hat{v}_i \leq \eta_n(\alpha_2)) + o_p(1). \end{aligned} \tag{6.19}$$

The term $\hat{\beta}_I$ in (6.10) will be cancelled out with the first term on the right hand side of (6.19).

Consider a representation of the second term on the right hand side of (6.19). Let $t' = (t_0, t_1', t_2)$ and

$$\begin{aligned} U^*(t) = n^{-1/2} \sum_{i=1}^{n} z_i(\beta) e_i I(e_i \leq F^{-1}(\alpha) + n^{-1/2}T_0 + n^{-1/2} z_i'(\beta + n^{-1/2}T_1)T_1 \\ - n^{-1}T_1' M_i(\beta + n^{-1/2}T_1)T_1 - n^{-1}T_2 d_{i-1}'(\beta + n^{-1/2}T_1)T_1 + n^{-3/2}T_2 T_1' G_{i-1}(\beta + n^{-1/2}T \end{aligned}$$

With the analogous discussions for Theorem 5.5 of Chen, Welsh and Chan (2001) and Lai, Thompson and Chen (2004), we may see the following

$$U^*(T) = U^*(0) + F^{-1}(\alpha)(\theta_\rho T_0 + Q_\rho T_1) + o_p(1)$$

for any sequences $T_0 = O_p(1)$ and $T_1 = O_p(1)$.
This implies that

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} z_i(\beta) e_i I(\hat{\eta}_n(\alpha_1) \leq \hat{v}_i \leq \hat{\eta}_n(\alpha_2)) \\ & = n^{-1/2} \sum_{i=1}^{n} z_i(\beta) e_i I(F^{-1}(\alpha_1) \leq e_i \leq F^{-1}(\alpha_2)) + F^{-1}(\alpha_2)\theta_\rho n^{1/2}(\hat{\eta}_n(\alpha_2) - F^{-1}(\alpha_2)) \\ & - F^{-1}(\alpha_1)\theta_\rho n^{1/2}(\hat{\eta}_n(\alpha_1) - F^{-1}(\alpha_1)) + (F^{-1}(\alpha_2) - F^{-1}(\alpha_1)) Q_\rho n^{1/2}(\hat{\beta}_I - \beta) + o_p(1). \end{aligned} \tag{6.20}$$

Similarly, we may also derive the followings,

$$n^{-1/2}\sum_{i=1}^{n} z_i(\beta)(I(\hat{v}_i \leq \hat{\eta}_n(\alpha_1)) - \alpha_1)$$
$$= f(F^{-1}(\alpha_1))[\theta_\rho T_0(\alpha_1) + Q_\rho T_1] + n^{-1/2}\sum_{i=1}^{n} z_i(\beta)(I(e_i \leq F^{-1}(\alpha_1)) - \alpha_1) + o_p(1), \tag{6.21}$$

$$n^{-1/2}\sum_{i=1}^{n} z_i(\beta)(I(\hat{v}_i \geq \hat{\eta}_n(\alpha_2)) - (1-\alpha_2))$$
$$= f(F^{-1}(\alpha_2))[\theta_\rho T_0(\alpha_2) + Q_\rho T_1] + n^{-1/2}\sum_{i=1}^{n} z_i(\beta)(I(e_i \geq F^{-1}(\alpha_2)) - (1-\alpha_2)) + o_p(1) \tag{6.22}$$

and

$$n^{-1}C_n = Q_\rho + o_p(1). \tag{6.23}$$

The theorem is induced from combining the representations in (6.21)-(6.23) into (6.10).

References

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

Amemiya, T. (1985). *Advanced econometrics.* Harvard University Press: Cambridge, Massachusetts.

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). *Robust Estimates of Location: Survey and Advances.* Princeton, NJ: Princeton Univ. Press.

Antoch, J. (1984). *Collection of Programs for Robust Regression Analysis.* Charles Univ., Prague.

Bahadur, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 557-580.

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

Bassett, G., and Koenker, R. (1978). Asymptotic theory of least absolute error regression. *J. Amer. Statist. Assoc.* **73**, 618-622.

Bassett, G., and Koenker, R. (1982). An empirical quantile function for linear models with iid errors. *J. Amer. Statist. Assoc.* **77**, 407-415.

Bennett, C. A. (1952). Asymptotic properties of ideal linear estimators. Ph. D. Dissertation, Univ. of Michigan, Ann Arbor.

Beran, R. J. (1977). Minimum Hellinger distance estimates for parameteric models. *Ann. Sstatisti.* **5**, 445-463.

Beran, R. J. (1981). Efficient robust estimates in parametric models. *Z. Wahrsch. Verw. Gebiete* **55**, 91-108.

Beran, R. J. (1982). Robust estimation in models for independent non-identically distributed data. *Ann. Statist.* **10**, 415-428.

Beran, R. J. (1984). Minimum Distance Procedures. *Jandbook of Statistics* vol. 4 (P. R. Krishnaiah and P. K. Sen, eds.), 741-754. Elsevier, New York.

Bhattacharyya, S. P. (1987). Robust stabilization against structured perturbations. *Lectures Notes in Control and Information Sciences*, **99**, Springer-VBerlag, Berlin.

Bickel, P. J. (1973). On some analogues to linear combinations of order statistics in the linear models *Ann. Statist.* **1**, 597-616.

Birnbaum, A., and Laska, E. (1967). Optimal robustness: A general method with applications to linear estimates of location. *J. Amer. Statist. Assoc.* **62**, 1230-1240.

Boos, D. (1979). A differential for L-statistics. *Ann. Statist.* **7**, 955-959.

Boos, D., and Serfling, R. J. (1979). ON Berry-Esseen rates for statistical functions with application to L-estimates. *Technical report,* Florida State Univ. Tallahassee.

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

Chen, L.-A. and Welsh, A. H. (2002), Distribution-Function-Based Bivariate Quantiles. *Journal of Multivariate Analysis.* 83, 208-231.

Chernoff, H., Gastwirth, J. L., and Johns, M. V. (1967). Asymptotic distribution of linear combinations of order statistics. *Ann. Math. Statist.* **38**, 52-72.

Cochrane, D. and Orcutt, G. H. (1949). Application of least squares regressions to relationships containing autocorrelated error terms. *Journal of the American Statistical Association*, **4**, 32-61.

Daniell, P. J. (1920). Obervations weighted according to order. *Ann. Journ. Math.* **42**, 222-236.

David, H. A. (1981). *Order Statistics*, 2nd ed. Wiley, New York.

De Jongh, P. J., De Wet, T. and Welsh, A. H. (1988). Mallows-type bounded-influence-regression trimmed Means. *Journal of the American Statistical Association* **83**, 805-810.

Dielman, T. E. (1996). *Applied Regression Analysis for Business and Economics.* Duxbury Press.

Donoho, D. L., and Liu, R. C. (1988). The automatic robustness of minimum distance functional. *Ann. Statist.* **16**, 552-586.

Fomby, T. B., Hill, R. C. and Johnson, S. R. (1980). *Advanced Econometric Methods.* Springer-Verlag: New York.

Geary, R. C. (1947). Test for normality. *Biometrika* **34**, 209-220.

Ghosh, J. K. (1971). A new proof of the Bahadur representation of quantiles and an application. *Ann. Math. Statist.* **42**, 1957-1961.

Gutenbrunner, C., and Jurečková, J. (1992). Regression rank scores and regression quantiles. *Ann. Statist.* **20**, 305-330

Hample, F. R. (1968). Contributions to the Theory of Robust Estimation. *Ph. D. dissertation*, Univ. California, Berkeley.

Hample, R. R. (1971). A general qualitative definition of robustness. *Ann. Math. Stat.* **42**, 1887-1896.

Hampel, F. R. (1973). Robust Estimation: A Condensed Partial Survey, *Z. Wahrsch. Verw. Geb.* **27**, 87-104.

Hampel, F. R., rousseeuw, P. J., Ronchetti, E., and Stahel, W. (1986). *Robust Statistics- The Approach Based on Influence Functions.* Wiley, New York.

Hartley, H. O. and Booker, A. (1965). Nonlinear least squares estimation. *Annals of Mathematical Statistics*, **36**, 638-650.

Helmers, R. (1977). The order of normal approximation for linear combinations of order statistics with smooth weight function. *Ann. Probab.* **5**, 940-953.

Helmers, R. (1980). Edgeworth expansions for linear combinations of order statistics. *Math. Centre Tract* 105, Amsterdam.

Helmers, R. (1981). A Berry-Esseen theorem for linear combination of order statistics. *Ann. Probab.* **9**, 342-347.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (1983). *Understanding Robust and Esploratory Data Analysis.* New York: Wiley.

Hodges, J. L. Jr., and Lehmann, E. H. (1956). The efficiency of Some Nonparametric Competitiors of the t-test, *Ann. Math. Stat.* **27**, 324-335.

Huang, J.-Y., Yang, E. K. and Chen, L.-A. (2004). Welsh's trimmed mean for the nonlinear regression model. *Sankhya.* To appear.

Huber, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35**, 73-101.

Huber, P. J. (1969). *Thorie de linfrence Statistique Robuste.* Presses de IUniversit, Montreal.

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

Ivanov, A. V. (1976). An asymptotic expansion for the distribution of the least squares estimator of non-linear regression parameters. *Theory of Probability and its Applications*, **21**, 557-570.

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

Ivanov, A. V. and Zwanzig, S. (1983). An asymptotic expansion of the distribution of least squares estimators in the nonlinear regression model. *Mathematische Operationsforschung und Statistik, Series Statistics*, **14**, 7-27.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics*, **40**, 633-643.

Jurečková, J. (1983a). Robust estimators of location and regression parameters and their second order asymptotic relations. *Trans. 9th Prague conf. On Inform. Theor. Rand. Proc. And Statist. Dec. Func.* (ed. J. A. Visek), Reidel Dordrecht, 19-31.

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

Jurečková, J. (1983b). Winsorized least-squares estimator and its M-estimator counterpart. *Contributions to statistics: Essays in Honor of Norman L. Johnson (ed. P. K. sen)*, North Holland, Amsterdam, 237-245.

Jurečková, J. (1983c). Trimmed polynomial regression. *comment. Math. Univ. Carolinae,* **24**, 597-607.

Jurečková, J. (1984b). Regression quantiles and trimmed least squares estimator under a general design. *Kybernetika* **20**, 345-357.

Jurečková, J., and Procházda, B. (1994). Regression quantiles and trimmed least squares estimator in nonlinear regression model. *J. Nonparametr. Statist.* **3**. No. 3-4, 201-222.

Jurečková, J. and Sen, P. K. (1984). On adaptive sale-equivariant M-estimators in linear models. *Statist. Dec. Suppl*, **1**, 31-46.

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

Kiefer, J. (1967). On Bahadurs representation of sample quantils. *Ann. Math. Statist.* **38**, 1323-1342.

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

Koenker, R. and Park, B. J. (1992). An interior point algorithm for non-linear quantile regression. Faculty Working Paper 92-0127. College of

Commerce and Business Administration. University of Illinois at Urbana-Champaign.

Koenker, R. and Portnoy, S. (1987). L-estimators for linear models. *Journal of the American Statistical Association* 82, 851-857.

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

Koul, H. L. (1985). Minimum distance estimation in multiple linear regression. *Sankhya A* **47**, 57-74.

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

Lai, Y.-H., Thompson, P. and Chen, L.-A. (2004). Generalized and pseudo-generalized trimmed means for the linear regression with AR(1) error model. *Statistixs and Probability Letters.* **67**, 203-211.

Liese, F. and Vajda, I. (1994). Consistency of M-estimates in general regression models. *Journal of Multivariate Analysis*, **50**, 93-114.

Lloyd, E. H. (1952). Least squares estimation of location and sale parameters using order statistics. *Biometrika* **34**, 41-67.

Millar, P. W. (1981). Robust estimation via minimum distance methods. *A. Wahrsch. Verw. Bebiete* **55**, 73-89.

Millar, P. W. (1984). A general approach to the optimality of minimum distance estimators. *Trans. Amer. Mth. Soc.* **286**, 377-418.

Oberhofer, W. (1982). The consistency of nonlinear regression minimizing the ℓ_1-norm. *Annals of Statistics*, **10**, 316-319.

Parr, W. C. (1981). Minimum distance estimation: A bibliography. *Commun. Statist. Theor. Meth.* A **10**, no. 1, 316-319.

Parr, W. C., and Schucany, W. R. (1982). *Contributions to a Gerneral Asymptotic Statistical Theory*, Lecture Notes in Statistics, **13**, Springer-

Verlag, New York.

Portnoy, S. (1984). Tightness of the sequence of c.d.f. processes defined from regression fractiles. *Robust and Nonlinear Time Series Analysis* (eds. J. Franke et al.), Springer-Verlag, New York, 231-246.

Procházka, B. (1988). Regression quantiles and trimmed least squares estimator in the nonlinear regression model. *Computational Statistics and Data Analysis*, **6**, 358-391.

Richardson, G. D. and Bhattacharyya, B. B. (1987). Consistent ℓ_1-estimators in nonlinear regression for a noncompact parameter space. *Shankhya A*, **49**, 377-387.

Rieder, H. (1994). *Robust Asymptotic Statistics.* springer-Verlag, New York.

Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association* 75, 828-838.

Sen, P. K. (1981a). *Sequential Nonparametrics: Invariance Principles and Statistcal inference.* Wiley, New York.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Application to Statistics.* Wiley, New York.

Stigler, S. M (1969). Linear Functions of order statistics. *Ann. Math. Statist.* **40**, 770-784.

Stigler, S. M. (1973). The asymptotic distribution of the trimmed mean. *Ann. Statist.* **1**, 472-477.

Stigler, S. M. (1974). Linear functions of order statistics with smooth weight function. *Ann. Satist.* **2**, 676-693.

Theil, H. (1971). *Priciples of Econometrics.* Wiley: New York.

Vajda, I. (1994). Conditions for consistency of MLEs. *Asymptotic statistics (Prague, 1993),*, 459-4665, Contrib. Statist., Physica, Heidelberg, 1994.

Wang, J. (1995). Asymptotic normality of ℓ_1-estimators in nonlinear regression. *Journal of Multivariate Analysis*, **54**, 227-238.

Welsh, A. H. (1987). The trimmed mean in the linear model. *Annals of Statistics* **15**, 20-36.

Welsh, A. H., Carroll, R. J., and Ruppert, D. (1994). Fitting heteroscedastic regression models. *J. Amer. Statist. Assoc.* **89**, no. 425, 100-116.

Wolfowitz, J. (1957). The minimum distance method. *Ann. Math. Statist.* **28**, 75-88.

Wu, C. F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics*, **9**, 501-513.

Curriculum Vitae

Yi-Hsuan Lai
e-mail: laiyihsu@stat.nctu.edu.tw

Education :

Ph. D. (09/2000- 06/2004) Institute of Statistics, National Chiao Tung University

Ph. D. candidate (1993-1994, 1995-1999) Department of Probability and Statistics
Michigan State University, U. S. A.

M. S. (09/1991 V 06/1993) Department of Probability and Statistics
Michigan State University U. S. A.

B. S. (09/1987-06/1991) Department of Applied Mathematics, National Chiao Tung University

Publications:

1. Y.-H. Lai, P. Thompson and L.-A. Chen, Generalized and pseudo generalized trimmed means for the linear regression with AR(1) error model, Statistics and Probability Letters 67 (2004) 203-211.

2. Y. H. Lai, W.-Y. Chan and L.-A. Chen, Generalized and Feasible Generalized Median Estimators for the Linear Regression with AR(1) Error Model, Accepted by Sankhya.

Research Interest: Game theory, Linear Model, Robust Estimation.

Experience:

Research Assistant (since 2003, spring) National Chiao Tung University, Institute of Statistics Employer: Prof. Lin-An Chen

Visiting Lecturer (2002 fall) Yuanpei University of Science and Technology Department of Information Management

Teaching Assistant (1998 spring, 1999) Michigan State University, Department of Statistics and Probability Course: Introduction to Probability and Statistics for Business

Leading Teaching Assistant (1997 fall) Michigan State University, Department of Statistics and Probability Course: Statistical Methods

Consultant (1996 fall-1997 spring) Michigan State University, Department of Statistics and Probability Consulting Office

Lecturer (1996 summer) Michigan State University, Department of Statistics and Probability Course: Probability and Statistics for Engineering

Leading Teaching Assistant (1996 spring) Michigan State University, Department of Statistics and Probability Course: Introduction to Probability and Statistics for Business

Teaching Assistant (1993 spring-1994 fall, 1995 fall) Michigan State University, Department of Statistics and Probability Course: Introduction to Probability and Statistics