ROBUST PROCEDURES FOR ESTIMATING PARAMETERS IN AN AUTOREGRESSION MODEL

Vickey C. C. Lin

林 秋 瑾*

摘 要

本文係以自我迴歸模式爲研究對象，探討最佳迴歸參數估計值及其有效性，引用‘Winsorized’的方法，修正後的‘Winsorized’的方法及Tiku的修正後最大概似估計方法，用於自我迴歸的模式中，以求得最佳迴歸參數估計值。

本研究以五種不同的分配樣本，在蒙地卡羅的實證中探討自我迴歸參數估計值，分析結果得知，修正後的‘Winsorized’估計值及Tiku的修正後最大概似估計值，較‘Winsorized’估計值更具有效性。

自我迴歸的模式較常應用於估計參數的方法有‘Durbin’方法及‘L_1’模估計方法，本研究所引用之三種‘最佳’參數估計方法在樣本爲非常態分配時，較前述二種估計方法更具有效性，但在樣本爲常態分配條件下，則本研究之方法有效性較低，爲了提高‘最佳’估計值之有效性，本研究採用‘可適性估計方法’，以增進‘最佳’參數估計式在樣本常態分配下之有效性，在樣本常態分配下，可適性的估計方法與一般自我迴歸模式估計方法—‘Durbin’及‘L_1’模參數估計值都具有效性，但在樣本爲非常態分配下，本文所提出之‘最佳’可適性的三種估計值前述二種估計值更具有效性。

Abstract

In this paper we develop some robust estimators for estimating regression coefficients in a simple regression model with autocorrelated erros. These robust estimators were derived by using the Winsorized method, the modified Winsorized method and the Tiku's MMLE method. Some Monte Carlo studies involving 5 different distributions indicate clearly that the modified Winsorized estimator and the Tiku's MMLE are more efficient than the Winsorized estimator in all cases. The robust estimators are considerably more efficient than the Durbin estimator and the L_1 norm estimator when the universe is not normal, but are less efficient when the universe is normally distributed. To improve on the efficiency of the robust estimators under normal distribution, some adaptive estimators were derived. These adaptive estimators are almost as efficient as the Durbin estimator and the L_1 norm estimator when the universe is normally distributed but are considerably more efficient when the universe is not normal.

*作者為本校地政系副教授

The Journal of National Chengchi University, Vol. 67, 1993

Key Words: Adaptive robust estimator, Autocorrelated error, Censoring, Durbin estimator, L_1 norm estimator, Modified Winsorized estimator, Monte Carlo studies, Robust estimator, Tiku's MMLE, Winsorized estimator.

1. INTRODUCTION

In predicting short term furture outcomes, it is often the case that the errors are correlated with one another. This is due to the fact that often times the random shocks may persist over some time. For example, in predicting the future stock price, the disturbances at any time may persist over some time to affect disturbances in the immediate future. In such cases, one is usually entertaining an autoregressive model. For analyzing data, autoregressive models have also been used in many problems in agricultural as well as biological and biomedical research besides business and economics; see for examples, Anderson (1949), Cochrane and Orcrrtt (1949), Dubin (1960), Tiao and Tan (1966), Rao and Griliches (1969), Gallant and Goebel (1976), Beach and Mackinnon (1978), Kramer (1980), and Dielman and Pfuffenberger (1989), Cogger (1990) and Weiss (1990).

In this paper we will consider an autoregressive regression model and procced to develop some efficient robust procedures for estimating the regression coefficient as well as other parameters. For analyzing such models, traditionally it is conveniently assumed that the random errors are normally distributed. In many practical problems, however, the normality assumption can hardly be justified, especially when the sample size is not very large. Thus, it is important to develop robust estimators which are considerably more efficient than the normal theory estimators when the errors are not normally distributed and are almost as efficient as normall theory estimators when the errors are in fact normally distributed. As shown by numberous examples in Tiku, Tan and Balakrishnan (1986), such procedures can be developed by using the MMLE method due to Tiku (1980), Alternatively, robust estimators may also be derived by a Winsorized method due to Yale and Forsythe (1976) or by a Modified Winsorized method due to Tan an Tabatabai (1988). In this paper we thus proceed to develop some efficient robust procedures by using these three methods.

In Section 2, we describe some optimal normal theory procedures for estimating procedures in regression models with autocorrelated errors. By using these normal theory procedures as bases, in Section 3 we proceed to develop some efficient robust estimators by using the above three methods. To improve on the robust estimators developed in Section 3, in Section 4 we proceed to develop some adaptive robust procedures. Finally in Section 5, we generate some Monte Carlo studies to demonstrate the usefulness of these robust estimators.

2. THE MODEL AND SOME NORMAL THEORY OPTIMAL ESTIMATION PROCEDURES

Consider a simple regression model with autocorrelated errors:

$$y_t = \alpha+\beta x_t+e_t \tag{2.1}$$

$$e_t = \sum_{i=1}^{p}\phi_i e_{t-i}+\epsilon_t \qquad t = 1,2, \ldots, n$$

where

y_t = Observed value at time t,

x_t = Nonstochastic independent variable at time t,

ϕ = Autoregressive coefficient.

In model (2.1), it is assumed that $0 < \phi < 1$ and that the ϵ_t', are independently and identically distributed random disturbances with mean zero and variance σ^2. The problem this paper addresses to is to estimate α and β when both ϕ and σ^2 are unknown. When the ϵ_t's are normally distributed, many procedures are available for estimating α and β. In the literature, these estimators have been referred to as ordinary least square estimator, the generalized least squares estimator, the maximum likelihood estimator, and the estimated generalized least squares estimators which include the Cochrane-Orcutt estimator, the Durbin estimator, and the Hildreth-Lu estimator; for a review see Judge et. al. (1984). Through Monte Carlo studies, Rao and Griliches (1969) showed that the Durbin method was more efficient than all other estimators when $\phi>0$.

2.1 The Durbin Estimation Procedure

This procedure consists basically of two stages: (i) the estimation of ϕ and

(ii) the estimation of α and β. To implement this method, one needs to transform the dependent variables to yield variables with independent errors. This is achieved by taking the first order difference of the observations to yield:

$$y_t^* = \alpha^* + \beta x_t^* + \epsilon_t \qquad t = 2, \ldots, n \tag{2.2}$$

where $y_t^* = (y_t - \phi y_{t-1})$, $x_t^* = (x_t - \phi x_{t-1})$ and $\alpha^* = \alpha(1-\phi)$; or equivalently,

$$y_t = \phi y_{t-1} + \alpha^* + \beta x_t - \gamma x_{t-1} + \epsilon_t \qquad t = 2, \ldots, n \tag{2.3}$$

Note that y_t^* and x_t^* are functions of the unknown parameter ϕ. However, conditional on ϕ, equation (2.2) is a simple ordinary regression model with independent error and constant variance σ^2 so that one uses ordinary least squares method to obtain estimators $\hat{\beta}(\phi)$ and $\hat{\alpha}(\phi)$. Conditional on ϕ, these estimators are the linear, unbiased and minimum varianced estimators of α and β, respectively, by the Gauss-Markov theorem (see Tan 1971). The Durbin approach is to derive $\hat{\beta}$ and $\hat{\alpha}$ by substituting ϕ by $\hat{\phi}$ to obtain $\hat{\alpha} = \alpha(\hat{\phi})$ and $\hat{\beta} = \beta(\hat{\phi})$. The ϕ is obtained by minimizing G_D ignoring the relation $\gamma = \beta\phi$, where

$$G_D = \sum_t [y_t - \phi y_{t-1} - \alpha^* - \beta x_t + \gamma x_{t-1}]^2$$

This is done iteratively by first estimating ϕ, say $\hat{\phi}_1$, then $\hat{\alpha}_1$, $\hat{\beta}_1$ and etc. The final estimator $\hat{\theta} = (\alpha, \beta)'$ of $\theta = (\alpha, \beta)'$ is, in matrix form

$$\hat{\theta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}Y. \tag{2.4}$$

where $\hat{V}^{-1}$ is obtained from V^{-1} by replacing ϕ by the Durbin estimator $\hat{\phi}$,

$$V^{-1} = \begin{pmatrix} 1 & -\phi & 0 & \cdots & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & \cdots & 0 & 0 \\ 0 & -\phi & 1+\phi^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & & \\ \vdots & \vdots & \vdots & \cdots & 1+\phi^2 & -\phi \\ 0 & 0 & 0 & \cdots & -\phi & 1 \end{pmatrix} \frac{1}{1-\phi^2} \tag{2.5}$$

2.2 The L_1 Norm Estimation Procedure

In the above Durbin procedure, the method is basically based on least squares method which in essence is a method involving second moment. Intuitively, the higher the order of the moments, the less the procedure is insensitive to violation of possible assumptions. In order to develop procedures which may be more robust to departure from normality than the least square procedures, attempts have been made by Cogger (1990) and Weiss (1990) to employ the L_1 norm as a possible alternative to least squares method. The method proposed by Cogger (1990) and Weiss (1990) are quite similar to the Durbin procedure except that the method by Cogger (1990) and Weiss (1990) is based on the least absolute value (i.e. L_1 norm). Thus, conditional on ϕ, the L_1 norm procedure is to minimize H_L to obtain estimates of (α,β) and to obtain ϕ by minimizing G_L, where

$$H_L = n^{-1} \sum_t |y_t^* - \alpha^* - \beta x_t^*|$$

and

$$G_L = n^{-1} \sum_t |y_t - \phi y_{t-1} - \alpha^* - \beta x_t + \gamma x_{t-1}| \qquad t=2, \ldots, n$$

Again, the procedure is implemented iteratively by first starting with some initial value of (α^*,β), estimating ϕ and then (α^*,β), and so on. Through Monte Carlo studies, Weiss (1990) showed that if $\phi>0$, then L_1 norm estimator was more efficient than the Durbin estimator when the errors are contaminated normal or lognormal and was almost as efficient as the Durbin estimator when the errors are normally distributed.

3. SOME NEW ROBUST ESTIMATORS

Tan and Tabatabai (1988) and Tiku, Tan and Balakrishnan (1986) have shown that given a procedure based on normality assumption for the random disturbances, in many cases it is possible to construct another procedure which is more efficient than the normal theory procedure when the universe is not normal and is almost as efficient as the normal theory procedure when the universe is normal. Motivated by these results, we proceed in this section to develop some robust procedures for the autoregressive regression model (2.1), by using the Durbin procedure and L_1 norm procedure as the base distribution. The robustifying methods we will use include the Winsorized method (Yale and Forsythe (1976)), the MMLE method of Tiku (Tiku, Tan and Balakrishnan (1986)) and the modified Winsorized method of Tan and Tabatabai (1988). These methods first order the estimated residuals to censor r_1 observations from the left, r_2 observations from the right and then proceed to drive estimators of the unknown parameters from the censored samples in connection with some methods to recover partial information from censored data. To illustrate, let $(y_{(t)},x_{(t)})$ be the (y_t,x_t) associated with the t-th ordered residual ϵ_t then, these estimators are obtained as follows.

3.1 The Winsorized Estimator

The Winsorized procedure for regression models was proposed by Yale and Forsythe (1976). Applying this procedure to the Durbin procedure, the Winsorized estimator $(\hat{\alpha}_w,\hat{\beta}_w,\hat{\phi}_w)$ is then obtained by the following steps:

(1) Obtain ϕ by minimizing H_{M1} where

$$\begin{aligned} H_{M1} &\\ = &\sum_{t=r_1+1}^{n-r_2} [y_{(t)}-\phi y_{(t-1)}-\alpha^*-\beta x_{(t)}+\gamma x_{(t-1)}]^2 \\ &+r_1[y_{(r_1+1)}-\phi y_{(r_1)}-\alpha^*-\beta x_{(r_1+1)}+\gamma x_{(r_1)}]^2 \\ &+r_2[y_{(n-r_2)}-\phi y_{(n-r_2-1)}-\alpha^*-\beta x_{(n-r_2)}+\gamma x_{(n-r_2-1}]^2 \end{aligned}$$

(2) Give ϕ, estimate α^* and β by minimizing

$$\begin{aligned} H_{M2} &\\ = &\sum_{t=r_1+1}^{n-r_2} [y^*_{(t)}-\alpha^*-\beta x^*_{(t)}]^2 \\ &+r_1\,[y^*_{(r_1+1)}-\alpha^*-\beta x^*_{(r_1+1)}]^2 \\ &+r_2\,[y^*_{(n-r_2)}-\alpha^*-\beta x^*_{(n-r_2)}]^2 \end{aligned}$$

where $\quad y^*_{(t)}=(y_{(t)}-\phi y_{(t-1)}), \qquad \alpha^*=(1-\phi)\alpha,$

$$x^*_{(t)}=(x_{(t)}-\phi x_{(t-1)}).$$

(3) Use the estimates α, β, ϕ as new initial parameter values for ordering the residuals and repeat above until estimates stabilize.

Put:

$$y_s = [y^*_{(r_1+1)} \cdot \cdot \cdot y^*_{(n-r_2)}]$$

$$n\hat{x}\omega = \sum_{t=r_1+1}^{n-r_2} x^*_{(t)} + r_1 x^*_{(r_1+1)} + r_2 x^*_{(n-r_2)}$$

$$n\hat{y}\omega = \sum_{t=r_1+1}^{n-r_2} y^*_{(t)} + r_1 y^*_{(r_1+1)} + r_2 y^*_{(n-r_2)}$$

$$Z_W = [x^*_{(r_1+1)} - \hat{x}_w, \ . \ . \ ., \ x^*_{(n-r_2)} - \hat{x}_w]^T$$

Let D_W be the $(n-r_1-r_2) \times (n-r_1-r_2)$ diagonal matrix with diagonal elements

$$[1+r_1, 1, \ . \ . \ ., \ 1, \ 1+r_2].$$

Then the Winsorized estimator are given by

$$\hat{\alpha}^*_w = \hat{y}_w - \hat{x}_w{}'\hat{\beta}_w$$

$$\hat{\beta}_w = (Z_w D_w Z'_w)^{-1} Z_w D_w y_s$$

$$\hat{\sigma}^2_w = \frac{1}{n-r_1-r_2-3} H_{M2}(\hat{\alpha}^*_w, \hat{\beta}_w).$$

If we replace the sums of squares by the sums of absolute values in the above procedures, then we have the Winsorized estimator of (α,β,ϕ) based on L_1 norm.

3.2 The Modified Winsorized Estimator

The modified winsorized procedure was proposed by Tan and Tabatabai (1988). In this procedure on first derives δ_1 and δ_2 as given in Tiku, Tan and Balakrishnan (1986). The procedure for deriving δ_1 and δ_2 consists basically of two steps:

(1) Define $q_i = r_i / n$, $i=1,2$ and obtain (k_i, h_i), i=1,2 from the following set of equations,

$$1-F(h_i) = q_i + [\frac{q_i(1-q_i)}{n}]^{1/2}$$

$$1-F(k_i) = q_i - [\frac{q_i(1-q_i)}{n}]^{1/2}$$

where F(x) is the cumulative distribution function of the standard normal variate N(0,1).

(2) Put $g_1(x) = \frac{dlogF(x)}{dx}$ and $g_2(x) = -\frac{dlog[1-F(x)]}{dx}$ and obtain (δ_i, η_i), $i=1,2$ from the following equations:

$$\delta_1 = \frac{-[g_1(k_1)-g_1(h_1)]}{k_1-h_1}, \quad \eta_1 = g_1(h_1)+\delta_1 h_1$$

$$\delta_2 = \frac{-[g_2(k_2)-g_2(h_2)]}{k_2-h_2}, \quad \eta_2 = g_2(h_2)+\delta_2 h_2$$

Once δ_1 and δ_2 are calculated, the modified estimator $(\hat{\alpha}_m, \hat{\beta}_m, \hat{\phi}_m)$ of (α, β, ϕ) is then derived by the following steps:

(1) Give ϕ, estimate α^* and β by minimizing

$$H_{MM2}$$

$$= \sum_{t=r_1+1}^{n-r_2} [y^*_{(t)} - \alpha^* - \beta x^*_{(t)}]^2$$

$$+ r_1 \delta_1 [y^*_{(r_1+1)} - \alpha^* - \beta x^*_{(r_1+1)}]^2$$

$$+ r_2 \delta_2 [y^*_{(n-r_2)} - \alpha^* - \beta x^*_{(n-r_2)}]^2$$

(2) Estimate ϕ by minimizing

$$H_{MM1}$$

$$= \sum_{t=r_1+1}^{n-r_2} [y_{(t)}-\phi y_{(t-1)}-\alpha^*-\beta x_{(t)}+\gamma x_{(t-1)}]^2$$

$$+r_1\delta_1[y_{(r_1+1)}-\phi y_{(r_1)}-\alpha^*-\beta x_{(r_1+1)}+\gamma x_{(r_1)}]^2$$

$$+r_2\delta_2[y_{(n-r_2)}-\phi y_{(n-r_2-1)}-\alpha^*-\beta x_{(n-r_2)}+\gamma x_{(n-r_2-1)}]^2$$

(3) Use estimated (α,β,ϕ) as initiate values for ordering the residuals, repeat above until estimates stabilize.

Put:

$$n\hat{x}M = \sum_{t=r_1+1}^{n-r_2} x^*_{(t)}+r_1\delta_1 x^*_{(r_1+1)}+r_2\delta_2 x^*_{(n-r_2)}$$

$$n\hat{y}M = \sum_{t=r_1+1}^{n-r_2} y^*_{(t)}+r_1\delta_1 y^*_{(r_1+1)}+r_2\delta_2 y^*_{(n-r_2)}$$

$$Z_M = [(x^*_{(r_1+1)}-\hat{x}_M),\ .\ .\ .,\ (x^*_{(n-r_2)}-\hat{x}_M)]^T$$

Let D_M be the $(n-r_1-r_2) \times (n-r_1-r_2)$ diagonal matrix with diagonal elements

$$[1+r_1\delta_1, 1,\ .\ .\ .,\ 1,\ 1+r_2\delta_2].$$

Then the modified Winsorized estimator are given by

$$\hat{\alpha}^*_M = \hat{y}_M-\hat{x}'_M{}'\hat{\beta}_M \quad = \hat{\alpha}_M(1-\hat{\phi}_M)$$

$$\hat{\beta}_M = (Z_M D_M Z'_M)^{-1} Z_M D_M y_s$$

$$\hat{\sigma}^2_M = \frac{1}{n-r_1-r_2-3} H_{MM2}(\hat{\alpha}^*_M,\hat{\beta}_M).$$

Again we can obtain the modified Winsorized estimator based on L_1 norm if we replace the sums of squares in the above steps by the sums of absolute values.

(3.3). The Tiku's MMLE Estimator.

This procedure was due to Tiku (1980a) and has been illustrated by many examples in Tiku, Tan and Balakrichnan (1986). In essence, the Tiku's MMLE estimator $(\hat{\alpha}_t,\hat{\beta}_t,\hat{\phi}_t)$ are obtained by the following steps:

(1) Put $V_t = (y^*_{(t)} - \alpha^* - \beta x^*_{(t)})/\sigma$ and use the approximation $g_1(V_{r_1+1}) \approxeq \eta_1 - \delta_1 V_{r_1+1}$ and $g_2(V_{n-r_2}) \approxeq \eta_2 + \delta_2 V_{n-r_2}$ to obtain approximate likelihood $L(\alpha,\beta,\phi,\sigma)$ as

$$L(\alpha,\beta,\phi,\sigma)$$
$$\approxeq \sigma^{n-r_1-r_2} exp\{-\frac{B}{\sigma} - \frac{1}{2\sigma^2}H_{MM2}(\alpha,\beta,\phi)\},$$

where

$B = r_2\eta_2(y^*_{(n-r_2)}) - \alpha^* - \beta x^*_{(n-r_2)}) - r_1\eta_1(y^*_{(r_1+1)} - \alpha^* - \beta x^*_{(r_1+1)})$ and $H_{MM2}(\alpha,\beta,\phi) = H_{MM2}$ as in the modified Winsorized method.

(2) Obtain MLE of α,β,ϕ,σ by maximizing the above approximate likelihood. These estimators are referred to as MMLE.
(3) Use the estimates in (2) as initial estimates for ordering the residuals and repeat above until the estimates stabilize.

Put

$$A = n - r_1 - r_2 - 3, \quad u_s = r_2\eta x^*_{(n-r_2)} - r_1\eta_1 x^*_{(r_1+1)}$$

$$\hat{B} = r_2\eta_2(y^*_{(n-r_2)} - \hat{\alpha}^*_M - \hat{\beta}_M x^*\ (\text{n-r}_2)) - r_1\eta_1(y^*_{(r_1+1)} - \hat{\alpha}^*_M - \hat{\beta}_M x^*_{(r_1+1)}),$$

$$C = H_{MM2}(\hat{\alpha}^*_M,\hat{\beta}_M).$$

Then

$$\hat{\alpha}_T = \hat{y}_M - \hat{x}'_M - \hat{\beta}_T = \hat{\alpha}_T(1-\hat{\phi}_T)$$

$$\hat{\beta}_T = \hat{\beta}_M - \hat{\sigma}_T\hat{\beta}_2$$

$$\hat{\sigma}_T = \{\hat{B} + (\hat{B}^2 + 4AC)^{1/2}\}/(2A)$$

$$\text{where } \hat{\beta}_2 = Z'_M D_M Z_M^{-1} u_s$$

As in Tan and Tabatabai (1988), it can easily be shown that the modified Winsorized estimators provide close approximations to Tiku's MMLE if the sample size n is not too small.

4. SOME ADAPTIVE ROBUST PROCEDURES

A general principle in making statistical inferences is that the more one takes into account available information about the model, the greater the chances that one would derive more efficient inference procedures. Underlying this principle, Hogg (1982) has derived more efficient procedures than the classical nonparametric inference procedures by taking into account information provided by the sample. In this paper, information provided by the sample to derive adaptive robust estimators for the AR(1) model given in (2.1) will be taken into account. Two types of adaptive procedures will be considered. One is concerned with deciding on the censoring procedures (symmetric versus asymmetric censoring) whereas the other is concerned with improving the efficiency of the robust estimators in cases in which the universe is normally distributed. To censor extreme observations, one has to decide whether symmetric censoring or asymmetric censoring should be adopted. As shown by Tan and Balakrishnan (1986), different censoring schemes have significant impact on the end results so that one has to use correct censoring schemes in deriving robust estimators. Information by means of which one would use to decide on the censoring scheme may be provided by the sample. To illustrate, suppose that one is sampling from a skew population such as a Chi-square distribution with 5 degrees of freedom. Then, if one plots the observed sample, it is very likely that the sample frequency distribution would be skewed toward the right. In such cases, it is more appropriate to use asymmetric censoring than symmetric censoring, or put more weight on asymmetric censoring by weighting these two types of censoring by the p-values of the statistical significant test. In our practice, whenever the sample suggests skew on the population distribution, we censor 20% on the skew side, and without censoring on the other side. To derive adaptive estimators for improving efficiency in cases in which the errors are normally distributed, note that it is common practice that one first checks the validity of the basic assumptions such as normality and others before applying some statistical procedures. One would normally use procedures that were derived under normality assumption if the test result against normality was not significant, the significance level usually being taken as 0.05. To improve the efficiency of our robust estimators so that they are almost as efficient as normal theory estimators (such as Durbin estimator) when the universe is normally distributed, this practice in our inference procedure will also be taken into account. Thus, by properly chosen significance level ($\alpha=0.1$ instead of $\alpha=0.05$ for example, see Section 5), we will use the normal theory estimators (Durbin estimators) when the significance test for normality is not significant and use our robust estimators (modified Winsorized for example) for otherwise. To implement this procedure, observe that there are several methods for testing normality of the sample. These methods include the Q-Q plot, the Shapiro and Wilk test (1965), the D'agostino test (1971), the directional test (the moment test, see Pearson 1963), the Kolmogorov-Simirnov test, and the Tiku's robust test (see Tiku 1980b). Our Monte Carlo studies on the comparison of powers of these tests suggest that the Tiku's Z test and the Shapiro-Wilk test appear to be far better than the other tests. Hence, in our adaptive procedure, adopting the Tiku's test or the Shapiro-Wilk test for testing normality will be proposed.

5. SOME MONTE CARLO COMPARISONS

To compare the performance of our robust estimators with one another and with the Durbin estimators and the L_1 norm estimators over different error distributions, Monte Carlo data will be generated by assuming an AR(1) simple regression model as given in (2.1) and by assuming different error distributions. The parameter values of α and β were chosen arbitrarily as $\alpha=0$, $\beta=1.0$, $\phi=0$, 0.1, 0.2, 0.4, 0.5, 0.6, 0.8, 0.9, 0.98. The values x_t were generated by computer from a normal universe with mean 0 and variance 1.

(i) The Error Distributions.

To compare different estimators regarding departure from normality,five different error distributions will be chosen and standardized so that they all have zero means and unit variance. These error distributions are the standard normal distribution, the mixture of two normals distribution $0.05N(u_1, 1) + 0.95N(u_2\ 6)$, a t-distribution with 5 degrees of freedom, a chi-square distribution with 5 degrees of freedom, and the double exponential distribution. These distributions are denoted respectively by N(0,1), MIX ($p = 0.05, u_1, u_2$, 1,6) (or simply MIX), TDIS, XDIS and DP(0,2). Note that the nonnormal distributions cover a wide range of nonnormal universes including symmetric distributions (t-distribution, double exponential distribution), asymmetric distribution (Chi-square), leptokurtic distribution (double exponential) as well as plytokurtic distribution (t-distribution). The mixture of two normal distributions was chosen not only because of many available results on outlier models (mixtures of normals, etc.) but also because mixtures of normals are very common in practice; for example, see Box and Tiao (1968), Hyrenious (1950,1952), Lee and Gurland (1977), Subrahmaniam (1972), Subrahmaniam, Subrahmaniam and Messeri (1972), Tan and Chang (1972), Tiao and Ali (1971), etc.

(ii) The Censoring Scheme

There are two types of censoring schemes in deriving Winsorized estimators, modified Winsorized estimators as well as MMLE estimators. It has been shown by Tan and Balakrishnan (1986) that a different censoring scheme has significant impact on the end results. Hence, it is important to use a correct censoring scheme to achieve optimal results. Thus, for data that do not show obvious skew visually, symmetric censoring will be used; that is, an equal number of observations will be censored on both sides. In these cases, the size of censoring was taken as 10% on both sides since this size has been shown to yield optimal results in many cases; see Tiku, Tan and Balakrishnan (1988). If the data show obvious skew, then asymmetric censoring will be applied, with the skew side censoring 20% and the other side 0%. For example, if the frequency plot of the sample shows right skew, then 20% on the right side and 0% on the left side will be censored.

(iii) The Initial Estimates.

In deriving Winsorized estimators, modified Winsorized estimators as well as the MMLE estimators, we would need initial estimates of the parameters to start the ordering process. As in Yale and Forsythe (1976) and Tan and Tabatabai (1988), two approaches will be used: One method uses least squares estimates and the other the PRESS procedure due to Allen (1971) which is closely related to the Jacknife procedure. To improve these procedures, one may also follow Tan and Tabatabai (1988) to consider iterative procedures for obtaining initial $(\hat{\alpha},\hat{\beta})$, by using $(\hat{\alpha}_m, \hat{\beta}_m)$ in the mth iteration as initial values iteratively until the results stabilize.

(iv) The Comparison Criterium.

To compare the performance of the robust estimators with one another and with the Durbin estimator and the L_1 norm estimator, the sample mean squared errors of the estimators will be computed. Then, by using the sample mean squared errors, the sample relative efficiencies of the estimators relative to the Durbin estimator and to the L_1 norm estimator will be computed. To illustrate, assume that the experimental run will be repeated 1,000 times. Then, following Tan and Tabatabai (1988), the relative efficiencies of the estimators of α, β and σ^2 are defined respectively as:

$$eff(\hat{\alpha}_m) = \frac{\Sigma_{i=1}^{1000} (\hat{\alpha}_i - \alpha)^2}{\Sigma_{i=1}^{1000} (\hat{\alpha}_{mi} - \alpha)^2}$$

$$eff(\hat{\beta}_m) = \frac{\Sigma_{i=1}^{1000} (\hat{\beta}_i - \beta)^2}{\Sigma_{i=1}^{1000} (\hat{\beta}_{mi} - \beta)^2}$$

and

$$eff(\sigma_m^2) = \frac{\Sigma_{i=1}^{1000} \hat{\sigma}_i^2}{\Sigma_{i=1}^{1000} \sigma_{mi}^2}$$

Sample sizes are taken as $n=$ 20 and 50 each experiment is repeated 1,000 times.

In what follows, let AR denote the autoregressive coefficient which was chosen between 0.0 and 0.98 WW denotes the estimators of Winsorized method, MW the estimators of modified Winsorized method; and TW the estimators of Tiku's MMLE. Because of space limitation, we will only display results from the mixture of two normal distributions and the t-distribution with 5 degrees of freedom. Given in

Figure 1 are the relative efficiencies of the three robust estimators (MW, WW, TW) relative to the Durbin estimator for $n = 20$. Corresponding results relative to the L_1 norm estimator are shown in Figure 2. Given in Figure 3 are the relative efficiencies of the adaptive estimators relative to the Durbin estimator for $n=20$.

From results given in these Figures and from results not shown here, the following observations are made:

(1) For sample size $n=20$, the modified Winsorized estimators (MW) and the Tiku's MMLE (TW) are more efficient than the Winsorized estimators (WW) which in turn are considerably more efficient then the Durbin's estimators and the L_1 norm estimators when the universe is not normal. If the universe is normal, however, the robust estimators are less efficient.

(2) For sample size $n=50$, and for $\phi \leq 0.5$, the sample conclusion as above holds. However, for $\phi > 0.6$, there are hardly any differences between these estimators.

(3) The adaptive estimators are as efficient as the Durbin estimators and the L_1 norm estimators, when the universes is normal; but are considerably more efficient than the Durbin estimators and the L_1 norm estimators when the universes are not normal.

REFERENCES

Allen, D.M. (1971), "Mean square error of prediction as criterion for selecting variables," *Technometric* 13, 469-476.

Anderson, R.L. (1949), "The problem of autocorrelation in regression analysis," *JASA* 44, 113-129.

Beach, C.M. and J.G. Mackinnon (1978), "A maximum Likelihood procedure for regression with autorcorrelated erros," *Econometrica* 46, 51-58.

Box, G.E.P. and Tiao, C.H. (1968), "A Bayesian approach to some outlier problems," *Biometrika* 55, 119-129.

Cochrane, D. and Orcutt, G.H. (1949), "Application of least squares regression to relationships containing autocorrelated error terms," *JASA* 44, 32-61.

Cogger, K.O. (1990), "Robust time series analysis — An L_1 approach," in *Robust Regression K. D. Lawrence and J. L. Arthur, eds.* Marcel Dekker, New York.

D'agostino, R. B., and Tietjen, G. L. (1971), "Simulation probability points of b_2 for small sample.," *Biometrika* 58,669-672.

Dielman. T. E. and Pfaffenberger, R. C. (1989), "Efficiency of ordinary least square for linear model with autocorrelation" *JASA* 84,248.

Figure 1.
The R.E. of the three robust estimators relative to the Durbin estimators

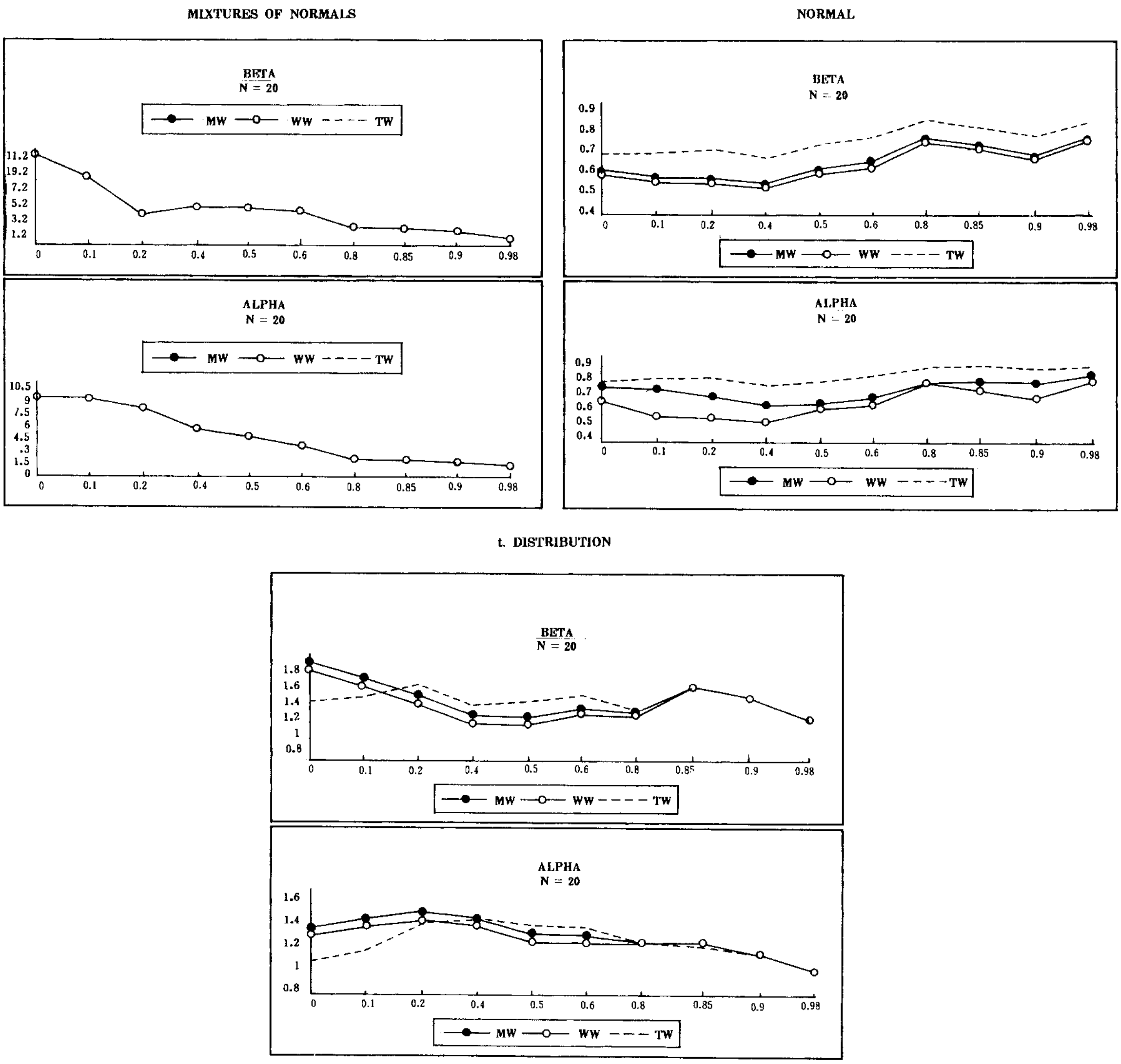


Figure 2.
The R.E. of the three roubst estimators relative to the L_1 estimators

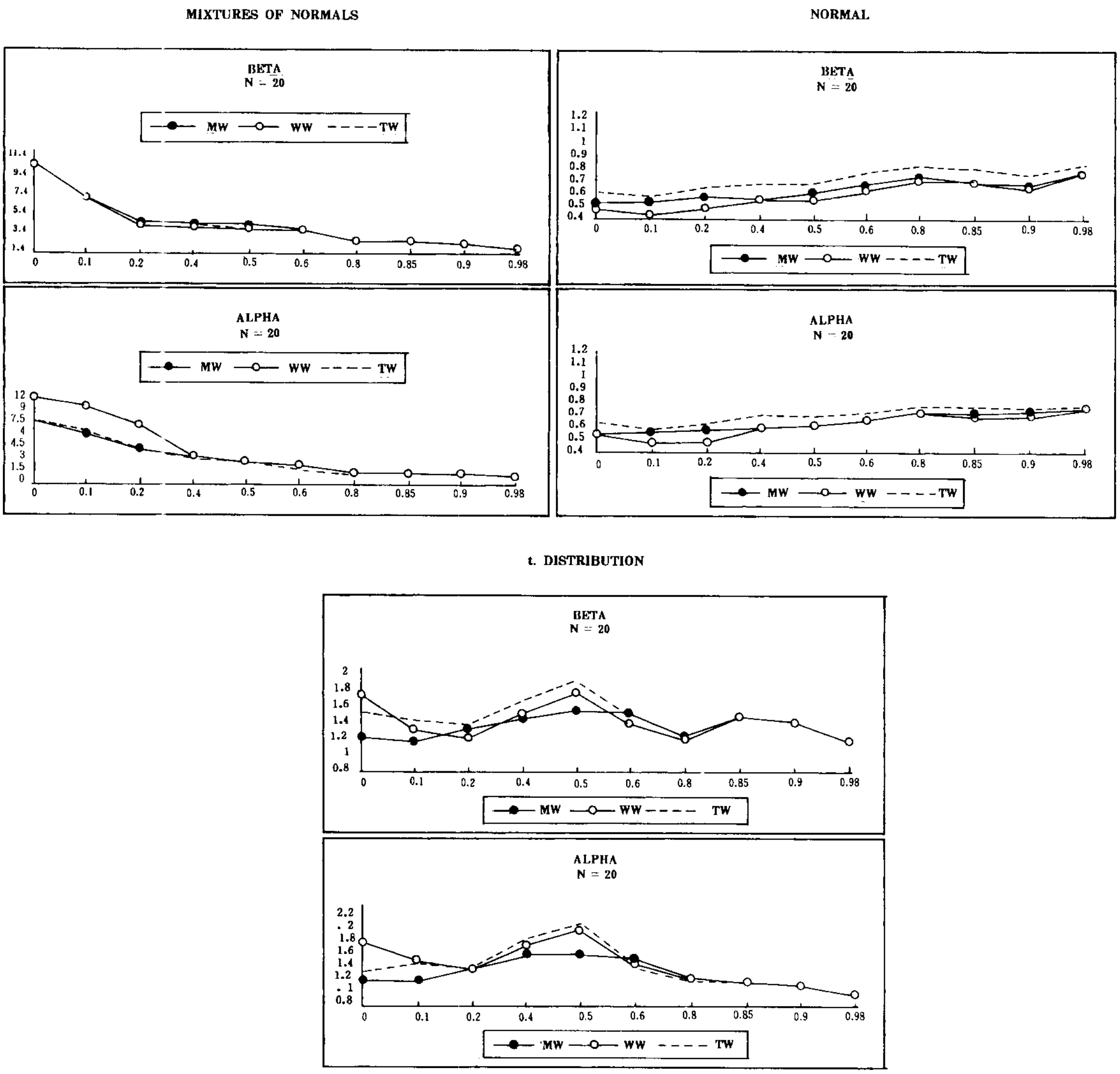
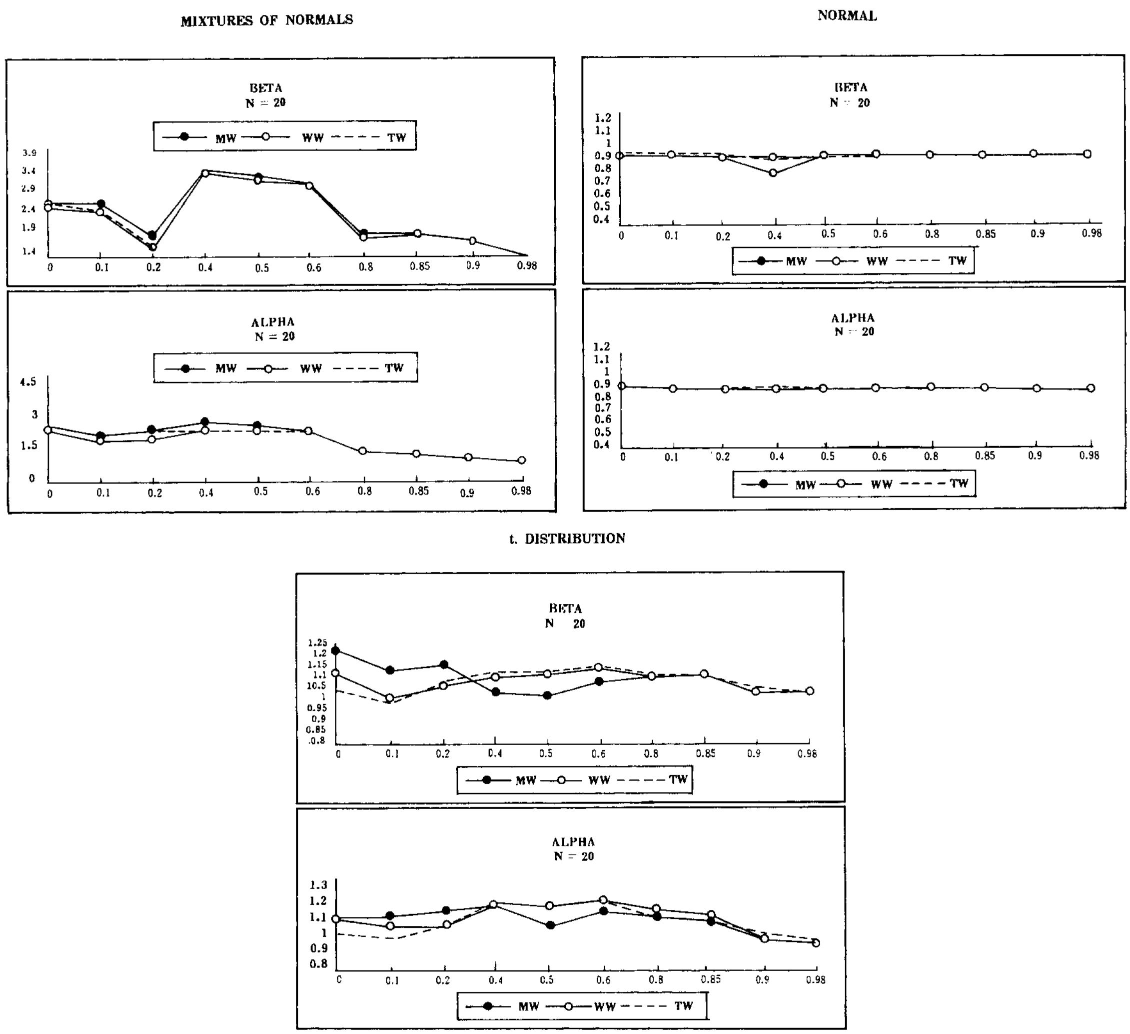


Figure 3.
The R.E. of the three adaptive roubst estimators relative to the Durbin estimators



Durbin, J. (1960), "Estimation of parameters in time-series regression models," *JRSS (B)* 22, 139-153.

Gallant, A. R. and Goebel J. J. (1976), "Nonlinear Regression with autocorrelated error.," *JASA* 71, 961-967.

Griliches, Z. and Rao P. (1969), "Small-Sample Properties of several Two-Stage Regression Methods in the context of Autocorrelated Errors.," *JASA* 71, 961-967.

Hogg, R. V. (1982), "On adaptive statistical inferences," *Commun. Statist. Theor, Meth.* 11, 2531-2542.

Hyrenius, H. (1950), "Distribution of 'student'-Fisher's t in samples from compound normal distribution,," *Biometrika* 37, 429-442.

Hyrenius, H. (1952), "Sampling from bivariance nonnormal universes by means of compound normal distribution," *Biometrika* 39,238-246.

Judge, G. G. Hill, R. C. Griffths, W. E. Lutkepohl, H. and Lee, T. C. (1984), *The Theory and practice of Econometrics* Wiley, New York.

Kramer, W. (1980), "Finite sample efficiency of ordinary least square in the linear regression model with autocorrelated errors," *JASA* 5, 1005-1009.

Lee, A. F. S. and Gurland, J. (1977), "One sample t — test when sampling from a mixture of normal distributions.," *Amer. Statist.* 5, 803-807.

Levenberg, K. (1944), "A method for the solution of certain problems in least squares," *Quarterly of Applied Mathematics* 2, 164-168.

Pearson, E. S. (1963), "Some problems arising in approximating to probability distributions using moments," *Biometrika* 50, 95-112.

Shapiro, S. S. and Wilk, M. B. (1965), "An analysis of variance test for normality (complete samples)., *Biometrika* 52, 591-616.

Subrahmaniam, K. (1972), "On quadratic forms from mixtures of two normal populations.," *S. Afr. Statist. J.* 6, 103-120.

Subrahmaniam, K. Subrahmaniam. K. and Messeri, J. Y. (1975), "On the robustness of some tests of significance," *JASA* 70, 435-438.

Tan, W. Y. (1971), "Note on an extension of Gauss-Markov theorem to multivariate regression models," *SIAM J. Appl. Math.* 20,24-29.

Tan, W. Y. and Balakrishnan N. (1986), "Bayesian insight into Tiku's robust procedures based on asymmetric censored samples," *J. Statist. Comput. Simul.* 24,17-31.

Tan, W. Y. and Chang, W. C. (1972), "Convolution approach to the genetic analysis of quantitive characters of self-fertilized populations," *Biometrics* 28, 1073-1090.

Tan, W. Y. and Tabatabai, M. A. (1988), "A modified Winsorized regression procedure for linear models," *J. Statist. Comput. Simul* 30, 299-313.

Tiao, C. H. and Ali, M. M. (1971), "Effect of nonnormality on inference about variance components," *Technometrics* 13,625-650.

Tiao, C. H. and Tan, W. Y. (1966), "Bayesian analysis of random-effect models in the analysis of variance, II: effect of autocorrelated error," *Biometrika* 53, 477-495.

Tiku, M. L. (1980a), "Robustness of MMLE estimators based on censored sample and robust test statistics," *J. Stat. planning and Inference* 4, 123-143.

Tiku, M. L. (1980b), "Robust two-sample variance via MML estimators," in *Proceedings of International Symposium on Statistics and Related Topics,* (M. Csorgo, D. A. Dawson, J. N. K. Rao, and A. K. M. E. Saleh eds.)., North-Holland, Amsterdam.