The Canadian Journal of Statistics
Vol. 22, No. 2, 1994, Pages 219–231
La Revue Canadienne de Statistique

One-step *M*-estimators in the linear model, with dependent errors

Christopher A. FIELD* and Douglas P. WIENS*

Dalhousie University and *University of Alberta*

Key words and phrases: *M*-estimation, robustness, regression, dependence.
AMS 1985 subject classifications: Primary 62F35, 62M10; secondary 62F11, 62J05.

ABSTRACT

We consider the problem of robust *M*-estimation of a vector of regression parameters, when the errors are dependent. We assume a weakly stationary, but otherwise quite general dependence structure. Our model allows for the representation of the correlations of any time series of finite length.

We first construct initial estimates of the regression, scale, and autocorrelation parameters. The initial autocorrelation estimates are used to transform the model to one of approximate independence. In this transformed model, final one-step *M*-estimates are calculated.

Under appropriate assumptions, the regression estimates so obtained are asymptotically normal, with a variance-covariance structure identical to that in the case in which the autocorrelations are known *a priori*. The results of a simulation study are given. Two versions of our estimator are compared with the L_1-estimator and several Huber-type *M*-estimators. In terms of bias and mean squared error, the estimators are generally very close. In terms of the coverage probabilities of confidence intervals, our estimators appear to be quite superior to both the L_1-estimator and the other estimators. The simulations also indicate that the approach to normality is quite fast.

RÉSUMÉ

Nous considérons le problème de *M*-estimation robuste pour un vecteur de paramètres de régression, lorsque les erreurs sont dépendantes. Nous supposons une stationnarité faible, mais autrement une structure de dépendance plutôt générale. Notre modèle permet la représentation des corrélations de n'importe quelle série chronologique de longueur finie.

Tout d'abord, nous construisons des estimateurs initiaux des paramètres de régression, d'échelle et d'autocorrélation. Les estimateurs initiaux d'autocorrélation sont utilisés afin de transformer le modèle en un modèle avec indépendance approximative. Les *M*-estimateurs finaux sont calculés avec ce modèle transformé.

Sous des hypothèses appropriées, les estimateurs de régression ainsi obtenus sont asymptotiquement normaux, avec une structure de variance/covariance identique à celle lorsque les autocorrélations sont connues a priori. Nous donnons les résultats d'une étude de simulation. Nous comparons deux versions de notre estimateur avec l'estimateur L_1 et plusieurs *M*-estimateurs de type Huber. Les estimateurs sont généralement très proches sur le plan du biais et de l'erreur quadratique moyenne. Nos estimateurs sont supérieurs à l'estimateur L_1 et aux autres estimateurs sur le plan des probabilités de couverture des intervalles de confiance. Nos simulations indiquent aussi que la convergence vers la normale est très rapide.

1. INTRODUCTION

We consider the following problem. One observes data $\mathbf{y} = (y_1, \ldots, y_n)^T$ following the

*Research supported by grants from the Natural Sciences and Engineering Research Council of Canada.

regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta}_0 + \mathbf{U}\boldsymbol{\epsilon}. \tag{1.1}$$

Here, $\mathbf{X}$ is a matrix of nonrandom regressors, of full column rank p. The p-vector $\boldsymbol{\theta}_0$ of unknown parameters is to be estimated. The vector $\boldsymbol{\epsilon}$ has components $\epsilon_1, \ldots, \epsilon_n$ which are i.i.d., with common, symmetric distribution function $\mathbf{F}$. The matrix $\mathbf{U}$ is a square root of an autocorrelation matrix:

$$\mathbf{U}\mathbf{U}^{\mathsf{T}} = \mathbf{P}, \qquad \operatorname{corr}[Y_i, Y_j] = P_{ij} = \rho_{|i-j|}, \qquad \rho_0 = 1. \tag{1.2}$$

The model (1.1), (1.2) thus describes, in a very general way, a situation commonly encountered in practice. One is often unwilling or unable to make the usual regression assumption of independent errors, but cannot justify any particular model of dependence, apart from weak stationarity.

In this paper we define a class of one-step M-estimators of $\boldsymbol{\theta}_0$ in the above model. Under certain assumptions, the estimates are asymptotically unbiased, and are asymptotically normally distributed with a variance-covariance matrix proportional to that of the BLUE. The constant of proportionality is a functional of the ψ-function defining the estimate.

The asymptotic properties of the estimate are in fact the same as if the matrix $\mathbf{U}$ in (1.2) were *known*, and one then transformed the model to one with independent observations, before calculating the estimates. That this should be the case is perhaps intuitively clear from the following brief description of the estimator. We first determine preliminary estimates of $\boldsymbol{\theta}_0$, $\mathbf{P}$, and $\mathbf{U}$. These are used to transform (1.1) to a model of approximate independence. In this model, the final one-step M-estimates are calculated.

The problem of robust estimation, when the observations are dependent, has received attention from a number of authors. The literature can, for our purposes, be roughly classified into three groups. In the first group are papers in which the main purpose seems to be to determine the behaviour of estimators, *calculated as if the observations were independent*, under various models of dependence. See in particular Gastwirth and Rubin (1975), Koul (1977), Deniau, Oppenheim, and Viano (1977), Rao (1981), Sadowsky (1986), Roussas (1989, 1990), Englund, Holst, and Ruppert (1988), and Kulperger (1990).

In the second group one finds robust estimation procedures for time series. If one is able to assume a dependence structure following one of the standard ARMA processes, then one may consult Bustos (1982), Martin (1981, 1982), Franke (1985), Martin and Yohai (1985, 1986), Bustos and Yohai (1986), or Masarotto (1987).

The third, and smallest, group of papers in the literature is that to which this work belongs. Here, one finds estimation methods which address the dependence problem in a manner which is at least partially adaptive. Portnoy (1977) considers estimation of location, with errors following a moving-average scheme. He shows that, to terms of first order in a serial correlation parameter ρ, an M-estimate with redescending ψ-function is optimal in a minimax-variance sense. The parameter ρ is assumed known, and controls the rate at which ψ redescends. See Portnoy (1979) for extensions.

Grossman (1979) considers (approximate) *maximum-likelihood estimation* in a non-linear regression model, with an error structure similar to that in (1.1), (1.2).

Poetscher and Prucha (1986) study one-step M-estimators defined by ψ-functions which correspond to MLEs for Student's t-distributions. The relevant degrees-of-freedom parameter is determined by the data. Such estimators are only optimal for i.i.d. errors from the correct Student's-t family. They are, however, seen to be robust when the errors instead obey one of a variety of mixing conditions, and have marginal distributions outside of the family of t-distributions.

Samarov (1987) investigates *generalized least squares* in the linear model, with serial correlations between the observations. He obtains the estimate in this class which minimizes the maximum asymptotic variance as the spectral density varies over a specified neighbourhood.

Zamar (1989) again considers the location problem, with errors satisfying a mixing-like condition. He shows that a ψ-function of the Huber type yields a minimax-variance *M*-estimator. The truncation point of this ψ-function approaches zero, so that the estimate approaches the median, quite quickly as the strength of the dependence increases.

The organization of the remainder of the paper is as follows. In Section 2 below, the calculation of the estimator will be described, and its asymptotic properties stated. Proofs are in the technical report Field and Wiens (1990), available from the authors. Section 3 contains details of some numerical studies carried out to assess the manner in which our estimator, and some competitors, perform in small samples.

2. CONSTRUCTION AND ASYMPTOTIC PROPERTIES OF THE ESTIMATOR

The construction of the estimator of $\boldsymbol{\theta}_0$ in the model defined by (1.1), (1.2) requires a preliminary, $\sqrt{n}$-consistent estimator $\boldsymbol{\theta}_*$ of $\boldsymbol{\theta}_0$. In the simulations detailed in Section 3 we have used the minimum-L_1-norm estimator, whose asymptotic properties are discussed in Basset and Koenker (1978) and Mehra and Rao (1988). See also Dodge (1987).

Given $\boldsymbol{\theta}_*$, we construct an estimate $\mathbf{P}_* = \hat{\mathbf{P}}(\boldsymbol{\theta}_*)$ of $\mathbf{P}$, and then an estimate $\mathbf{U}_* = \hat{\mathbf{U}}(\boldsymbol{\theta}_*)$ of $\mathbf{U}$. As $\hat{\mathbf{U}}(\boldsymbol{\theta})$ one may take any $n \times n$ matrix satisfying $\hat{\mathbf{U}}(\boldsymbol{\theta})\hat{\mathbf{U}}^{\mathsf{T}}(\boldsymbol{\theta}) = \hat{\mathbf{P}}(\boldsymbol{\theta})$. Now define

$$\mathbf{A} = \mathbf{U}^{-1}\mathbf{X}, \qquad \hat{\mathbf{A}}(\boldsymbol{\theta}) = \hat{\mathbf{U}}^{-1}(\boldsymbol{\theta})\mathbf{X}, \qquad \mathbf{A}_* = \hat{\mathbf{A}}(\boldsymbol{\theta}_*) = \mathbf{U}_*^{-1}\mathbf{X}.$$

Define also transformed observations

$$\mathbf{z} = \mathbf{U}^{-1}\mathbf{y}, \qquad \mathbf{z}_* = \mathbf{U}_*^{-1}\mathbf{y}$$

and associated residual vectors

$$\boldsymbol{\epsilon}_+ = \mathbf{z} - \mathbf{A}\boldsymbol{\theta}_*, \qquad \boldsymbol{\epsilon}_* = \mathbf{z}_* - \mathbf{A}_*\boldsymbol{\theta}_*.$$

Let the elements of $\boldsymbol{\epsilon}_+, \boldsymbol{\epsilon}_*$ have empirical distribution functions (e.d.f.'s) $\mathbf{F}_n^+, \mathbf{F}_n^*$ respectively. Recall that $\mathbf{F}$ is the distribution function of the i.i.d. errors $\epsilon_1, \ldots, \epsilon_n$. Let $\sigma(\mathbf{F})$ be a positive, scale-equivariant, shift-invariant functional of $\mathbf{F}$, and define

$$\sigma_+ = \sigma(\mathbf{F}_n^+), \qquad \sigma_* = \sigma(\mathbf{F}_n^*).$$

For an odd, bounded, piecewise continuously differentiable function ψ, define

$$\begin{aligned} &\psi_\sigma(\mathbf{x}) = \psi(\mathbf{x}/\sigma) \qquad (\sigma > 0), \\ &\boldsymbol{\psi}_\sigma(\mathbf{x}) = (\psi_\sigma(x_1), \ldots, \psi_\sigma(x_n))^{\mathsf{T}} \qquad \text{if} \quad \mathbf{x} = (x_1, \ldots, x_n)^{\mathsf{T}}, \\ &D(\psi_\sigma, \mathbf{F}) = \mathcal{E}_{\mathbf{F}}[\psi'_\sigma(\boldsymbol{\epsilon})]. \end{aligned} \tag{2.1}$$

Were $\mathbf{P}$ known, we would have the exact model

$$\mathbf{z} = \mathbf{A}\boldsymbol{\theta}_0 + \boldsymbol{\epsilon}, \tag{2.2}$$

with independent errors. In this situation Bickel (1975) studied the one-step M-estimator

$$\boldsymbol{\theta}_B = \boldsymbol{\theta}_* + \frac{(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\psi}_{\sigma_+}(\boldsymbol{\epsilon}_+)}{\hat{D}_n(\psi_{\sigma+}, \mathbf{F})},$$

where $\hat{D}_n(\psi_\sigma, \mathbf{F})$ estimates $D(\psi_\sigma, \mathbf{F})$. Under a set of assumptions contained in ours, Bickel showed that

$$\sqrt{n}(\boldsymbol{\theta}_B - \boldsymbol{\theta}_0) \xrightarrow{w} \mathsf{N}_p(0, V(\psi_\sigma, \mathbf{F})\mathbf{A}_0^{-1}), \tag{2.3}$$

where

$$\mathbf{A}_0 = \lim_{n\to\infty} \frac{\mathbf{A}^T\mathbf{A}}{n},$$

$$V(\psi, \mathbf{F}) = \frac{\mathcal{E}_\mathbf{F}[\psi^2(\boldsymbol{\epsilon})]}{D^2(\psi, \mathbf{F})},$$

$$\sigma = \sigma(\mathbf{F}).$$

We are proposing the esitmator

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_* + \frac{(\mathbf{A}_*^T\mathbf{A}_*)^{-1}\mathbf{A}_*^T\boldsymbol{\psi}_{\sigma*}(\boldsymbol{\epsilon}_*)}{\hat{D}_n(\psi_{\sigma*}, \mathbf{F})},$$

where $\hat{D}_n(\psi_{\sigma*}, \mathbf{F})$ is any consistent estimator of $D(\psi_\sigma, \mathbf{F})$ for $\sigma = \sigma(\mathbf{F})$. Under the assumptions of Field and Wiens (1990), $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_B$ are $\sqrt{n}$-equivalent:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_B) \xrightarrow{pr} \mathbf{0},$$

and hence by (2.3)

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{w} \mathsf{N}_p(0, V(\psi_\sigma, \mathbf{F})\mathbf{A}_0^{-1}). \tag{2.4}$$

Revised estimates of $\mathbf{P}$ and $\sigma(\mathbf{F})$ are then given by $\hat{\mathbf{P}}(\hat{\boldsymbol{\theta}})$ and $\sigma(\hat{\mathbf{F}}_n)$, where $\hat{\mathbf{F}}_n$ is the e.d.f. of the elements of $\hat{\boldsymbol{\epsilon}} = \hat{\mathbf{U}}^{-1}(\hat{\boldsymbol{\theta}})(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}})$. We have generally found it desirable to repeat this process at least once, using $\hat{\boldsymbol{\theta}}$ as initial value for a further iteration.

The most difficult issue is that of the estimation of $\mathbf{P}$. In the aforementioned paper of Grossman (1979), this was dealt with, in the case that the errors follow an autoregressive process, by using a stepwise procedure which involved estimating the parameters and spectral density iteratively. Hannan (1971) and Goebel (1974) use a circular symmetric matrix to approximate $\mathbf{P}$, and estimate the approximating matrix by first estimating the spectral density of the generating series. The method requires only mild assumptions on this series.

Here, we use a computationally efficient method described by Gallant (1987, pp. 127–139). The method is derived for AR(q) errors; hence it is not exact for the model (1.1). However, simulations reported in Gallant (1987) (see also Gallant and Goebel 1976), as well as our own simulations, indicate that the method is quite robust against process misspecifications if the true process is one, such as a moving-average process, that can be approximated by an autoregression. To apply the method, the residuals from the preliminary fit are used to estimate the autocovariances, up to lag q. Using these estimates, the coefficients of the Yule-Walker equations are estimated, and the leading $q \times q$ submatrix of $\mathbf{P}_*^{-1}$ is factored, using a Choleski decomposition. The factor is then augmented by the estimated Yule-Walker coefficients to yield $\mathbf{U}_*^{-1}$. See Gallant (1987) for details.

Note that the asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is $V(\psi_\sigma, \mathbf{F})$ times

$$\mathbf{A}_0^{-1} = \lim_n n(\mathbf{X}^\mathsf{T}\mathbf{P}^{-1}\mathbf{X})^{-1}.$$

Thus, we do as well, asymptotically, by estimating the autocorrelations as we would if the autocorrelations were known and we estimated $\boldsymbol{\theta}_0$ by $\boldsymbol{\theta}_B$ in the model (2.2).

There is a further interesting consequence to the asymptotic covariance structure of $\hat{\boldsymbol{\theta}}$. This concerns the optimal choice of ψ. since the asymptotic covariance of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ depends on ψ only through $V(\psi, \mathbf{F})$, a ψ-function which is optimal for independent observations retains this optimality when applied to dependent observations. For choices of ψ yielding minimax estimators, see Huber (1981), Collins and Wiens (1985), Wiens (1986).

3. SIMULATION STUDY

We report here the details of a simulation study, in which two versions of the estimator described in Section 2 are compared with the L_1-estimator and with three ordinary M-estimators.

The fitted model was $y_t = \theta_0 + x_t\theta_1 + \delta_t$, with $\theta_0 = \theta_1 = 1$, and with x_t assuming n equally spaced values in $[0, 5]$. For each of $n = 20$ and $n = 40$, and each combination of those marginal error distributions and dependence structures given below, we simulated 1000 series of errors $\{\delta_t\}_{t=1}^n$. The error distributions used, with their mnemonics in brackets, were:

(1) standard normal ("Gaussian"),

(2) standard normal with probability 0.8, and normal, with $\mu = 0$ and $\sigma^2 = 25$, with probability 0.2 ("mixture").

The dependence structures were:

(1) independence ("Ind"),

(2) an MA(1) moving average process, with a first-order autocorrelation of $\frac{1}{3}$ ("ma.mild"),

(3) an MA(7) process, with parameters (1, 0.8, −0.6, 0.4, 0.2, 0.1, −0.1) and autocorrelations $\rho(1) = -0.26$, $\rho(2) = -0.37$, $\rho(3) = 0.33$, $\rho(4) = -0.02$, $\rho(5) = -0.06$, $\rho(6) = -0.06$, $\rho(7) = -0.03$, ("ma.strong"),

(4) an AR(1) autoregressive process with parameter 0.5 and autocorrelations $\rho(j) = 2^{-j}$, $j = 0, 1, 2, \ldots$ ["ar(1)"]. Note that this process is not representable as at (1.1). It thus acts as a test of the robustness of the proposed new estimators to process misspecifications.

In total, 16,000 series were simulated. From each, we computed regression estimates, variance-covariance estimates, and a 90% confidence interval on $\theta_0 + \theta_1$, using each of six estimation methods. These methods were:

(1) L1: This, the minimum L_1-norm estimator, was used to obtain starting values for all five of the other methods. The asymptotic covariance matrix was estimated assuming the errors to be independent, in which case it is given by $\sigma^2(\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}/\{4f^2(0)\}$. We estimated σ^2 by the scaled MAD of the residuals $\mathbf{e}_t = \mathbf{y}_t - \mathbf{x}_t^\mathsf{T}\hat{\boldsymbol{\theta}}$:

$$\hat{\sigma}^2 = \frac{\text{MAD}(e_t)}{\Phi^{-1}(0.75)} \tag{3.1}$$

and, for simplicity, "estimated" $f(0)$ merely by substituting $\phi(0) = (2\pi)^{-\frac{1}{2}}$. (We took this step because the properties of this estimator were not of primary concern to us. We were then pleasantly surprised by its reasonable performance.)

(2) FW1: This is the estimator described in Section 2. For it, and all four of the other M-estimators, we employed Huber's score function $\psi(x) = \min(k, \max(x, -k))$, with $k = 1.345$. The scaled residuals $\mathbf{r}_t = \mathbf{e}_t/\hat{\sigma}$ from the L_1 fit are used to fit an AR model, as described in Section 2. The model is then transformed to one of approximate independence, and an ordinary M-estimate is computed in the transformed model. This last step consists of performing one least-squares regression on *pseudovalues*, as described in Street, Carroll, and Ruppert (1998). This two-stage process of fitting an AR model and then reestimating the regression coefficients is repeated until the order of the AR process, as determined by Akaike's information criterion, first becomes zero. Only rarely were more than two steps required. At each stage σ is estimated as at (3.1). The asymptotic covariance matrix is then estimated by

$$\hat{\sigma}^2 K^2 \frac{\frac{1}{n-p}\sum \psi^2(\mathbf{r}_t)}{\left(\frac{1}{n}\sum \psi'(\mathbf{r}_t)\right)^2}(\mathbf{A}_*^{\mathsf{T}}\mathbf{A}_*)^{-1}, \tag{3.2}$$

where K is the correction factor derived by Huber (1981, Section 7.6).

(3) FW2: This is identical to FW1, except that the scaled residuals $\mathbf{r}_t$ are replaced by $\psi(\mathbf{r}_t)$ before using them to fit the AR model. See Huber (1981, Section 8.3) for a discussion.

(4) Hb.ind: This is an ordinary M-estimator, with the variance-covariance estimates calculated under the assumption of independent errors. Thus, the assumed asymptotic covariance matrix is $V(\psi_\sigma, \mathbf{F})(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}$, estimated as at (3.2) but with $\mathbf{A}_*$ replaced by $\mathbf{X}$.

(5) Hb.dep: This estimator is identical to Hb.ind, but in calculating the variance-covariance estimates we incorporate an assumption that the errors are dependent, with the sequence $\{\psi(\epsilon_t)\}$ being weakly stationary. Under this assumption the influence function $\mathbf{IF}$ is easily calculated, and the asymptotic covariance matrix $(= \mathcal{E}_{\mathbf{F}}[\mathbf{IF}\cdot\mathbf{IF}^{\mathsf{T}}])$ of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ obtained as $V(\psi_\sigma, \mathbf{F})(\lim_n \mathbf{X}^{\mathsf{T}}\mathbf{X}/n))^{-1}(\lim_n \mathbf{X}^{\mathsf{T}}\mathbf{P}\mathbf{X})(\lim_n \mathbf{X}^{\mathsf{T}}\mathbf{X}/n)^{-1}$, where $\mathbf{P}$ is the autocorrelation matrix of $\{\psi(\epsilon_t)\}$. The well-known relationship between the BLUE and the GLSE then implies that the limiting covariance matrix of Hb.dep is larger, in the sense of positive semidefiniteness, than that of FW2. We estimate $V(\psi_\sigma, \mathbf{F})$ as for the other estimators [employing the correction factor K, as at (3.2)], and estimate $\mathbf{P}$ by substituting the sample autocorrelations of $\{\psi(\mathbf{r}_t)\}$.

(6) Hb.ma: This is identical to Hb.dep, except that $\mathbf{P}$ is estimated by first fitting a moving-average model, of order 2, to the sequence $\{\psi(\mathbf{r}_t)\}$, and then substituting the autocorrelations of the fitted process. Thus $\hat{\mathbf{P}}_{|s-t|} = 0$ if $|s-t| > 2$. The MA fit is carried out by applying Gaussian maximum likelihood to $\{\psi(\mathbf{r}_t)\}$. The order (two) of the MA process was found to be the most satisfactory choice in a related, small-scale simulation study.

In reporting the results of the simulations, we will be concentrating on the root mean square error (rmse) of the estimates and the coverage obtained by the confidence intervals. Table 1 gives the rmse for the estimates over the eight sampling situations for $n = 20$. In terms of rmse, the three variations of the Huber estimates are identical. They differ only in the estimates used for the asymptotic covariance. The results are for θ_1. The pattern is very similar for θ_0.

TABLE 1

Estimate density	Dependence	Rmse L1	Huber	FW1	FW2
Normal	Independent	0.191	0.157	0.159	0.160
	ma.mild	0.229	0.201	0.204	0.202
	ar(1)	0.295	0.280	0.283	0.283
	ma.strong	0.282	0.182	0.172	0.173
Mixture	Independent	0.253	0.248	0.248	0.268
	ma.mild	0.348	0.360	0.354	0.369
	ar(1)	0.535	0.55	0.520	0.536
	ma.strong	0.521	0.369	0.372	0.364

For the normal data, the L1 estimate has higher rmse, while the other three are quite close together except for the case of strong dependence. In this case both versions of the FW estimates outperform the Huber by a small margin. The pattern for the mixture data is less clear. Each of the four estimates performs best in one of the situations. L1 does well for all cases except the strong dependence, where it is substantially worse than the other estimates. Over the four dependence situations, FW1 shows the best average performance and is always close to the minimum achieved for any of the estimates. A similar pattern exists for θ_0 when $n = 20$ and for both estimates when $n = 40$. In all the cases the bias is negligible.

The next step is to examine the normality of the distribution of the estimates. The approach to normality is assessed by looking at the quantile-quantile plot of the 1000 values of the estimate versus the quantiles of the standard normal. We expected the approach to normality to be slower for the mixture data than for the normal data. The results for the mixture data with ar(1) dependence and $n = 20$ is shown in Figure 1 for θ_1. As can be seen, all the estimates are very similar in the normality of their distribution. They all have distributions which are slightly longer tailed than the normal. Some previous work (cf. Field 1982, p. 686) has suggested a degrees-of-freedom correction of $0.6n$ in the case of the Huber estimate under independence. Figure 2 shows the results of a comparison of the distributions with a t-distribution with 12 df. As can be seen, the empirical distribution agrees reasonably well with the t12 distribution. Since all four estimates behave in a similar fashion, we can summarize their behaviour over the four dependence situations by looking at the empirical distribution of FW1 as compared with a normal. The results in Figure 3 show that the empirical distribution is similar for the three dependent situations and is shorter tailed than in the independent case. A comparison with the t-distribution with 12 df shows good agreement in the dependent cases but an indication of a longer-tailed distribution in the independent case.

It is important to evaluate the coverage of the confidence intervals computed using an estimate of the asymptotic variance. We now have three different results for the Huber estimate, corresponding to three different calculations of the asymptotic variance. In evaluating these results it is important to look at both the actual coverage and the variation in the coverage over the four dependence situations. The confidence interval used a normal quantile to compute the 90% confidence interval. We have evaluated the coverage by looking at both the lower and the upper coverage obtained by the intervals in the 1000 trials. The results for the lower coverage with $n = 20$ in the confidence interval for $\theta_0 + \theta_1$ are given in Table 2.

As can be seen, the coverage differs somewhat from the nominal level of 5%. To

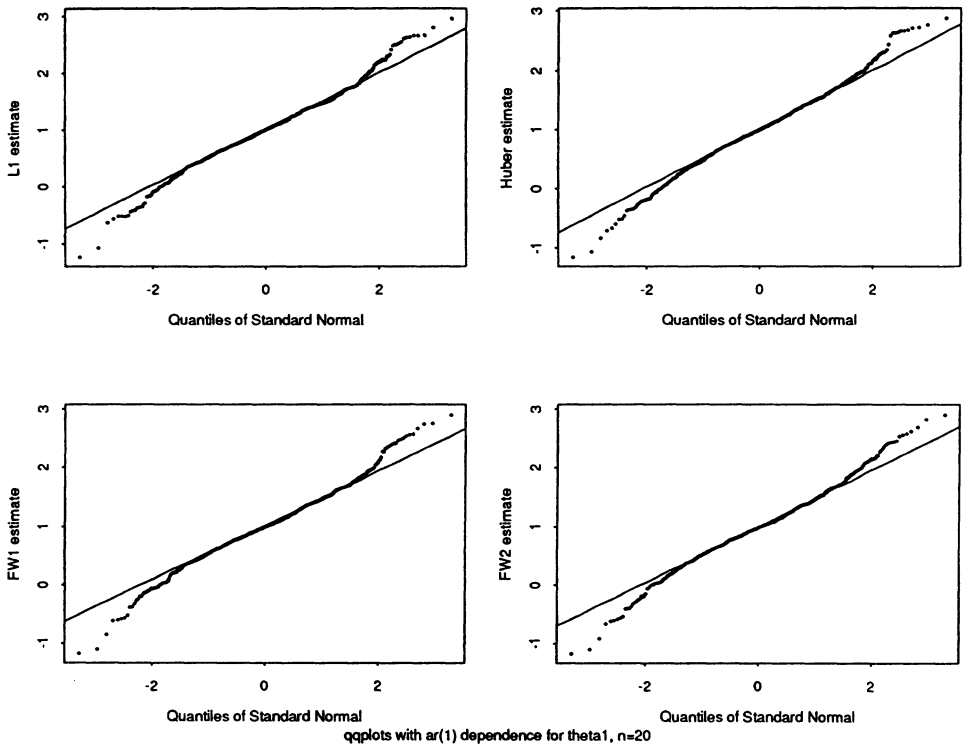

FIGURE 1: Quantile-quantile plots with ar(1) dependence for θ_1, $n = 20$.

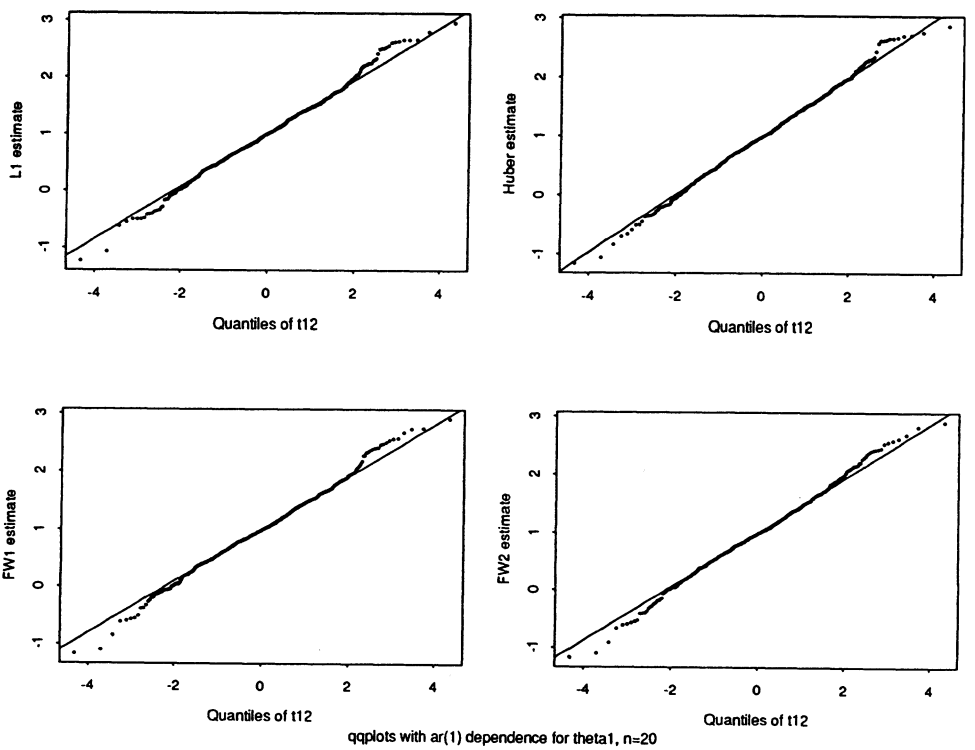

FIGURE 2: Quantile-quantile plots with ar(1) dependence for θ_1, $n = 20$.

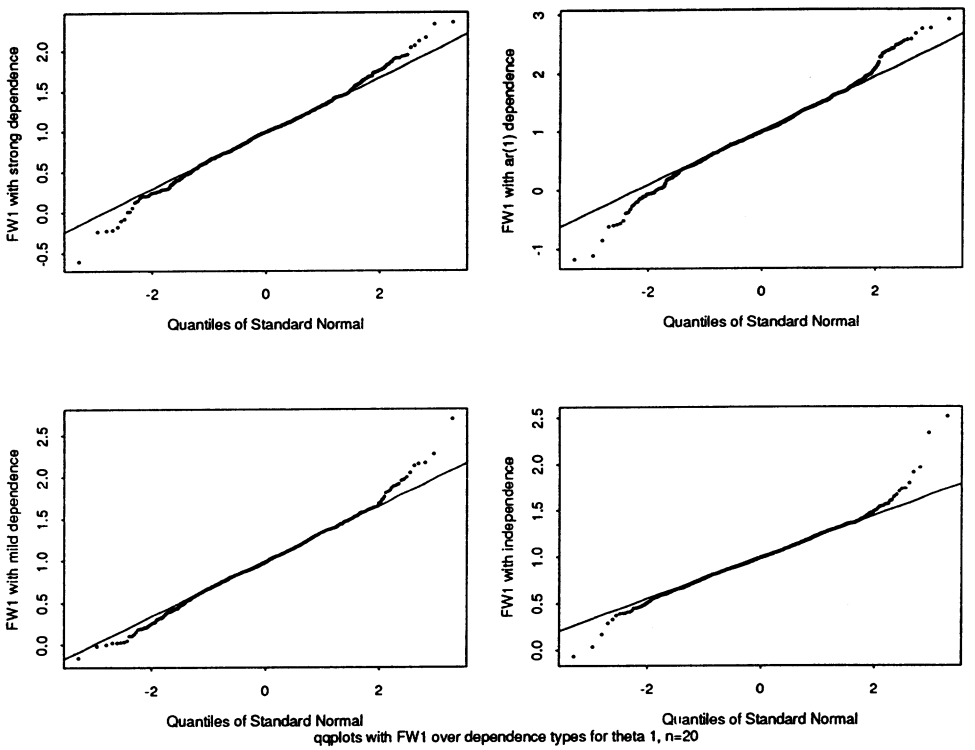

FIGURE 3: Quantile-quantile plots with FW1 over dependence types for θ_1, $n = 20$.

TABLE 2

Estimate density	Dependence	Lower coverage					
		L1	Hb.ind	Hb.dep	Hb.ma	FW1	FW2
Normal	Independent	0.077	0.064	0.194	0.163	0.097	0.098
	ma.mild	0.104	0.117	0.211	0.164	0.136	0.142
	ar(1)	0.177	0.182	0.236	0.156	0.167	0.167
	ma.strong	0.044	0.010	0.121	0.109	0.073	0.081
Mixture	Independent	0.067	0.049	0.195	0.159	0.062	0.087
	ma.mild	0.115	0.128	0.207	0.162	0.131	0.156
	ar(1)	0.162	0.187	0.241	0.152	0.163	0.175
	ma.strong	0.028	0.007	0.080	0.075	0.089	0.087

assess the various estimates, it is instructive to look at the coverage obtained over all eight combinations of density and dependence with $n = 20$. Figure 4 gives a summary of the overall coverage (sum of upper and lower coverage) via a boxplot for each estimate. Note that the target value is 0.10. The three estimates showing the least variation are the two FW estimates and huber.ma. Of these three the FW1 estimate is closest to giving coverage at the 5% level. The amount of variation across the cases is important in that it gives a measure of the stability of the intervals. If the levels are wrong, as they tend

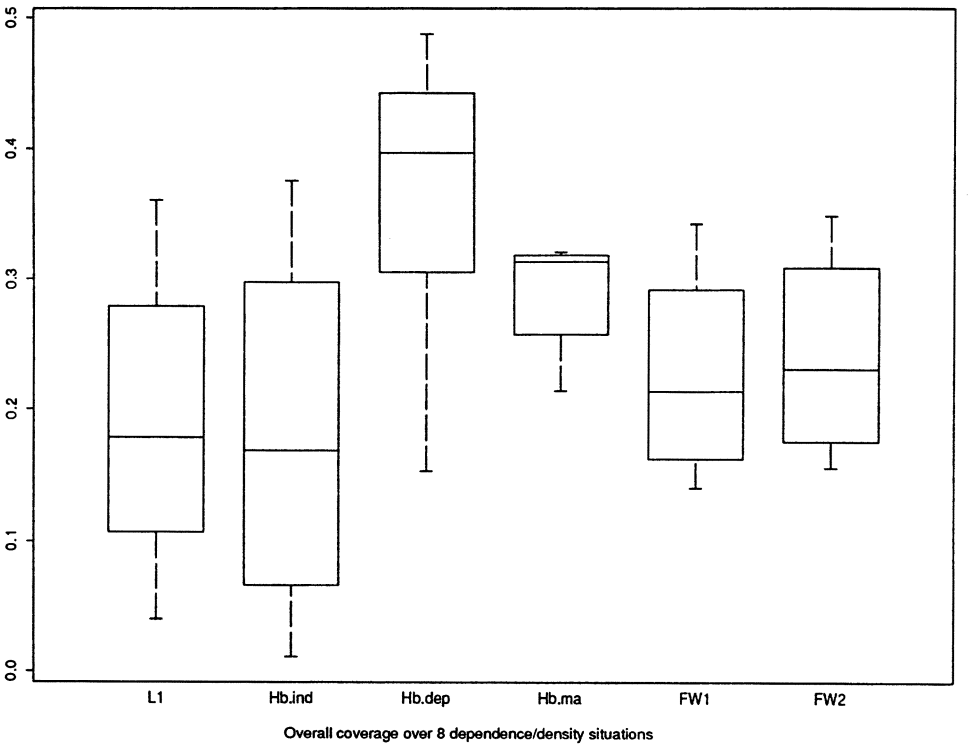

FIGURE 4: Overall coverage over eight dependence-density situations.

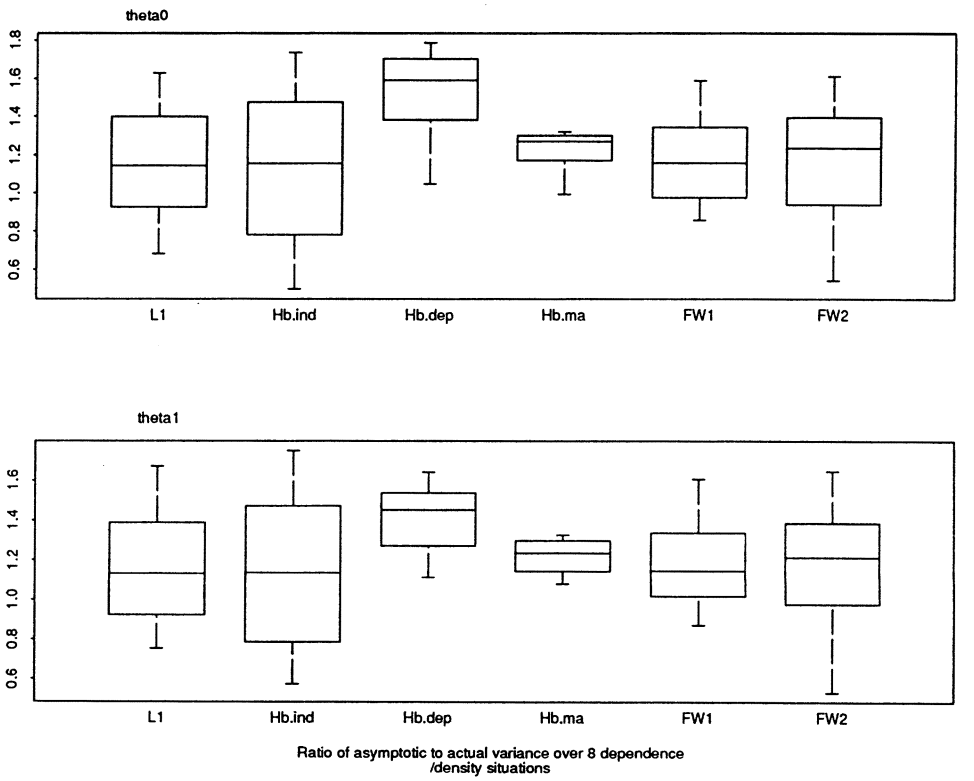

FIGURE 5: Ratio of asymptotic to actual variance over eight dependence-density situations.

to be for these estimates, but are stable, this opens the possibility of making a finite-sample correction to obtain more accurate coverage. To understand the reason for this it is worthwhile to compare the asymptotic variance with the empirical variance. Given the fact that the distribution of the estimates is reasonably well modelled by a t-distribution with $0.6n$ degrees of freedom, the coverage should be accurate if the asymptotic variance is close to the empirical variance. Specifically, Figure 5 compares the ratio of the mean asymptotic variance and the empirical variance over the 1000 replicates. For each estimate this ratio is computed for the eight combinations of density and dependence and the results summarized with a boxplot. As can be seen, the ratio varies the least for the huber.ma estimate, with the FW and the huber.dep showing more variation than the huber.ma but less than the other three estimates. A check of the plot of the ratio versus the coverage shows a reasonably good linear relationship with a correlation coefficient of 0.92. If the ratio of the asymptotic to the empirical variance stayed almost the same over all situations, it would be possible to make a "finite-sample correction" to the asymptotic variance to bring it in line with the empirical variance and hence give confidence values much closer to the nominal level, simply by using a t-interval. In our case, such a correction would work to some extent on FW1, huber.dep, and huber.ma, although we cannot bring the coverage in line with the nominal level for all dependence-density situations. It should be noted that for both huber.ma and huber.dep, we attempted to correct the asymptotic variance, but the corrections have not had the desired effect. This suggests that finite-sample corrections to the asymptotic variance to bring it in line with the empirical variance would give coverage closer to the nominal level for the estimates where the ratio does not change too much from one situation to the next.

One specific adjustment would be to replace the normal z-value used in the coverage calculations by a t-value. As suggested above, for $n = 20$ using a t-density with 12 df seems reasonable. For the situation with mixture data and mild.ma dependence, the changes in the coverage of the lower bounds are by a factor between 0.8 and 0.9 in going from the normal to t12. This matches the ratio of the 95th quantile of the t12 to the normal. This suggests that the P-values for the noncovering intervals may well be approximately uniformly distributed and that a t-correction is not enough to bring the coverage in line with the nominal values. Similar behaviour is observed for the mixture density and strong dependence.

There still remains work to be done in order to achieve more accuracy in the confidence-interval coverage. One reasonable approach for the FW estimates is to bootstrap the ϵ_*'s and construct bootstrap intervals based on the results. This should give accurate approximations to the actual standard deviation of the estimates, and these estimates can be used to get reasonable coverage levels using a t-distribution result as noted above. It is not clear how to do this with the Huber or L1 estimate.

In an overall comparison of the estimates, FW1 is consistently a good performer over the situations considered. No one estimate consistently dominates, and FW always is reasonably close to the best in each situation. By estimating the dependence structure and using it in the construction of the estimate, FW has the ability to adapt to the dependence in the data.

REFERENCES

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

Bickel, P.J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.*, 70, 428–434.

Bustos, O.H. (1982). General *M*-estimates for contaminated pth-order autoregressive processes: Consistency and asymptotic normality. *Z. Wahrsch. Verw. Geb.*, 59, 491–504.

Bustos, O.H., and Yohai, V.J. (1986). Robust estimates for ARMA models. *J. Amer. Statist. Assoc.*, 81, 155–169.

Collins, J.R., and Wiens, D.P. (1985). Minimax variance *M*-estimates in ϵ-contamination models. *Ann. Statist.*, 13, 1078–1096.

Deniau, C., Oppenheim, G., and Viano, M.C. (1977). Asymptotic normality of *M*-estimators on weakly dependent data. *Recent Developments in Statistics.* (J.R. Barra *et al., eds.*), North Holland, Amsterdam.

Dodge, Y. (1987), *ed. Statistical Data Analysis Based on the L_1-Norm and Related Methods*. North-Holland, 415–420.

Englund, J.-E., Holst, U., and Ruppert, D. (1988). Recursive *M*-estimates of location and scale for dependent sequences. *Scand. J. Statist.*, 15, 147–159.

Field, C.A. (1982). Small sample asymptotic expansions for multivariate *M*-estimates. *Ann. Statist.*, 10, 672–689.

Field, C.A. and Wiens, D.P. (1990). One-step *M*-estimators in the linear model, with dependent errors. Technical Report 90.01, Department of Statistics and Applied Probability, University of Alberta.

Franke, J. (1985). Minimax-robust prediction of discrete time series. *Z. Wahrsch. Verw. Geb.*, 68, 337–364.

Gallant, A.R. (1987). *Nonlinear Statistical Model.* Wiley, New York.

Gallant, A.R., and Goebel, J.J. (1976). Nonlinear regression with autocorrelated errors. *J. Amer. Statist. Assoc.*, 71, 961–967.

Goebel, J.J. (1974). Nonlinear regression in the presence of autocorrelated errors. Ph.D. Dissertation, Department of Statistics, Iowa State University.

Gastwirth, J.L., and Rubin, H. (1975). The behavior of robust estimators on dependent data. *Ann. Statist.*, 3, 1070–1100.

Grossman, W. (1979). Efficiency of estimators for nonlinear time series models. *Proc. Second Prague Symposium on Asymptotic Statistics.* (P. Mandl and M. Huskova, *eds.*), 199–210.

Hannan, E.J. (1971). Nonlinear time series regression. *J. Appl. Probab.*, 8, 767–780.

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

Koul, H.L. (1977). Behavior of robust estimators in the regression model with dependent errors. *Ann. Statist.*, 5, 681–699.

Kulperger, R. (1990). On the asymptotic normality of trimmed means and *M*-estimates of location for stationary mixing processes. Technical Report TR 90-06, Department of Statistics and Actuarial Sciences, University of Western Ontario.

Martin, R.D. (1981). Robust methods for time series. *Applied Time Series Analysis.* Volume 2 (D.F. Findley, *ed.*), Academic Press, New York, 683–760.

Martin, R.D. (1982). The Cramér-Rao bound and robust *M*-estimates for autoregressions. *Biometrika*, 69, 437–442.

Martin, R.D., and Yohai, V.J. (1985). Robustness in time series and estimating ARMA models. *Handbook of Statistics.* Volume 5 (E.J. Hannan, D.R. Brillinger, and P.R. Krishnaiah, *eds.*), Elsevier, New York, 119–155.

Martin, R.D., and Yohai, V.J. (1986). Influence functionals for time series. *Ann. Statist.* 14, 781–818.

Masarotto, G. (1987). Robust and consistent estimates of autoregressive-moving-average parameters. *Biometrika*, 74, 791–797.

Mehra, K.L., and Rao, M.S. (1988). On H-L type estimators of multiple regression parameters under strong mixing. Technical Report 89.02, Department of Statistics and Applied Probability, University of Alberta.

Poetscher, B.M., and Prucha, I.R. (1986). A class of partially adaptive one-step *M*-estimators for the non-linear regression model with dependent observations. *J. Econometrics*, 32, 219–251.

Portnoy, S.L. (1977). Robust estimation in dependent situations. *Ann. Statist.*, 5, 22–43.

Portnoy, S.L. (1979). Further remarks on robust estimation in dependent situations. *Ann. Statist.*, 7, 224–231.

Rao, B.L.S.P. (1981). Asymptotic behavior of *M*-estimators for the linear model with dependent errors. *Bull. Inst. Math. Acad. Sinica*, 9, 367–375.

Roussas, G.R. (1989). Consistent regression estimation with fixed design points under dependence conditions. *Stat. Prob. Letters*, 8, 41–50.

Roussas, G.R. (1990). Nonparametric regression estimation under mixing conditions. *Stoch. Proc. Appl.* 36, 107–116.

Sadowsky, J.S. (1986). A maximum variance model for robust detection and estimation with dependent data. *IEEE Trans. Inform. Theory*, 32, 220–226.

Samarov, A.M. (1987). Robust spectral regression. *Ann. Statist.*, 15, 99–111.

Street, J.O., Carroll, R.H., and Ruppert, D. (1988). A note on computing robust regression estimates via iteratively reweighted least squares. *Amer. Statist.*, 42, 152–154.

Wiens, D.P. (1986). Minimax variance *M*-estimators of location in Kolmogorov neighbourhoods. *Ann. Statist.*, 14, 724–732.

Zamar, R.H. (1989). Robustness against unexpected dependence in the location model. *Statist. Probab. Lett.*, 9, 367–374.

Received 1 October 1992
Revised 15 June 1993
Accepted 1 November 1993

Department of Mathematics, Statistics and Computing Science
Dalhousie University
Halifax, Nova Scotia
Canada B3H 3J5

Department of Statistics and Applied Probability
University of Alberta
Edmonton, Alberta
T6C 2G1