Test (1997)
Vol. 6, *No.* 1, *pp.* 187–203

Robust Estimation in the Errors Variables Model via Weighted Likelihood Estimating Equations

A. BASU and S. SARKAR
Applied Statistics Unit, Indian Statistical Institute,
203 B.T. Road, Calcutta 700 035, India

Department of Statistics, Oklahoma State University,
Stillwater, OK 74078, USA

SUMMARY

Parameters estimates for the errors-in-variables model are obtained by solving weighted likelihood estimating equations. They are consistent, asymptotically normal and asymptotically fully efficient, and exhibit robustness properties similar to the minimum disparity estimators (Basu and Sarkar 1994a) but are immensely simpler to compute and have some theoretical advantages over the latter. We illustrate the robustness properties through some numerical studies similar to those of Zamar (1989).

Keywords: MEASUREMENT ERROR MODEL; DISPARITY; HELLINGER DISTANCE; ROBUSTNESS; WEIGHTED LIKELIHOOD ESTIMATOR.

1. INTRODUCTION

Consider the classical errors-in-variables model:

$$y_i = \beta' \boldsymbol{x}_i, \quad Y_i = y_i + e_i, \quad \boldsymbol{X}_i = \boldsymbol{x}_i + \boldsymbol{u}_i, \tag{1}$$

for $i = 1, \ldots, n$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, y_i and $\boldsymbol{x}_i$ denote the true values of the variables observed as Y_i and $\boldsymbol{X}_i = (X_{1i}, X_{2i}, \ldots, X_{pi})'$

Received March 1995; Revised January 1997.

containing measurement errors e_i and $\boldsymbol{u}_i = (u_{1i}, u_{2i}, \ldots, u_{pi})'$ respectively, and $\{\boldsymbol{x}_i\}$ and $\{\boldsymbol{\epsilon}_i = (e_i, \boldsymbol{u}_i')'\}$ are independent sequences of i.i.d. random vectors such that $\boldsymbol{x}_i \sim N_p[\boldsymbol{\mu}_x, \Sigma_{xx}]$ and $\boldsymbol{\epsilon}_i \sim N[\boldsymbol{0}, \Sigma_{\epsilon\epsilon}]$. The common distribution of the vectors $(Y_i, \boldsymbol{X}_i')$ is a multivariate normal distribution. To make model (1) identifiable, we assume $\Sigma_{\epsilon\epsilon}$ is known up to a multiple (see Fuller 1987, Sec 2.3.2). The literature on robust estimation in errors-in-variables models is fairly small. It includes the works of Brown (1982), Zamar (1989), Cheng and Van Ness (1990, 1992) and Croos and Fuller (1991) on the classical model. Brown studied the iteratively reweighted orthogonal regression and Zamar investigated robust orthogonal regression M-estimators. Cheng and Van Ness examined the classical and generalized M-estimators. Carroll and Gallo (1982) discussed the classical independent error model with replication of the predictors. Carroll, Eltinge and Ruppert (1993) studied the case where there are replicates of the observed predictors and the errors in the response and predictor variables are correlated.

Basu and Sarkar (1994a), hereafter referred to as B&S, considered minimum disparity estimators (MDEs), defined in Section 2, of the parameters of model (1). Unlike the robust estimators previously studied for this model, the MDEs have been observed to exhibit strong robustness properties while at the same time having *full asymptotic efficiency* at the model if the kernel can be chosen appropriately relative to the model. Such kernels are known as *transparent* kernels (Basu and Lindsay 1994). In a modest Monte Carlo study reported in B&S the minimum Hellinger distance estimator performed better than the orthogonal regression M-estimator of Zamar (1989) with contaminated data for model (1).

While the MDEs have several attractive properties, their asymptotic efficiency depends on the choice of transparent kernels which may not be available if the model distribution is not normal. In addition, the evaluation of the MDEs requires an enormous amount of computation. In this paper, following the approach of Basu, Markatou and Lindsay (1993) we consider weighted likelihood estimators (WLEs), defined in Section 2, which remove the above mentioned limitations of the MDEs. The disparity based WLEs are related to the MDEs and have similar efficiency and robustness properties but are far more practical in terms of the computational ease. In some numerical studies similar to those of Zamar (1989), the Hellinger distance based WLE appears to be competitive to the minimum Hellinger distance estimator in terms of robustness; in a Monte Carlo study of B&S, the latter was seen to outperform the orthogonal regression M-estimator of Zamar (1989). We emphasize, however, that the purpose of this paper is not to find just another robust estimator; the proposed WLEs, like the MDEs, also achieve full asymptotic efficiency at the model.

The remainder of the paper is organized as follows. We discuss the MDEs and the WLEs in Section 2. Application of the weighted likelihood estimation to model (1) is discussed in Section 3. Section 4 presents results of a simulation study showing robustness of the Hellinger distance based WLE under an errors-in-variables model with nonnormal error distributions. In Section 5 we demonstrate the performance of the Hellinger distance based WLE through a real dataset; and in addition, we illustrate the computation of the Hellinger distance based WLE for a trivariate real dataset with two (non-intercept) explanatory variables.

2. MINIMUM DISPARITY AND WEIGHTED LIKELIHOOD ESTIMATION

2.1. *Minimum Disparity Estimation*

B&S have reviewed minimum disparity estimation (Lindsay 1994; Basu and Lindsay 1994; Basu and Sarkar 1994b) for a family of continuous models $m_\beta(\cdot)$, indexed by $\boldsymbol{\beta} \in \mathbb{R}^p$. Let $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n$ be k-dimensional i.i.d. observation vectors from $m_\beta(\boldsymbol{z})$. Minimum disparity estimators are obtained by minimizing disparities $\rho_G(f^*, m^*_\beta)$ which are density based distances with the special structure

$$\rho_G(f^*, m^*_\beta) = \int \ldots \int G(\delta^*(\boldsymbol{z}))m^*_\beta(\boldsymbol{z})d\boldsymbol{z}$$

where $f^*(\boldsymbol{z})$ is a nonparametric kernel density estimator, $m^*_\beta(\boldsymbol{z})$ is the model density smoothed with the same smooth kernel function and G is a real valued thrice differentiable convex function with $G(0) = 0$. Also $\delta^*(\boldsymbol{z}) = [(m^*_\beta(\boldsymbol{z}))^{-1} f^*(\boldsymbol{z}) - 1]$ which is called the *Pearson residual* at $\boldsymbol{z}$ by Lindsay (1994). The squared Hellinger distance corresponds to $G(x) = [(x+1)^{1/2} - 1]^2$. The minimizer of the likelihood disparity which corresponds to $G(x) = (x+1)\log_e(x+1)$ is called the "MLE*".

The estimating equation of the MLE* is

$$\begin{aligned}\int \cdots \int \frac{f^*(\boldsymbol{z})}{m^*_\beta(\boldsymbol{z})} \frac{\partial}{\partial\beta} m^*_\beta(\boldsymbol{x}) d\boldsymbol{z} \\ = \int \cdots \int \left[\frac{f^*(\boldsymbol{z})}{m^*_\beta(\boldsymbol{z})} - 1 \right] \frac{\partial}{\partial\beta} m^*_\beta(\boldsymbol{x}) d\boldsymbol{z} = \mathbf{0}.\end{aligned} \tag{2}$$

If the kernel chosen is transparent (see Basu and Lindsay 1994 and B&S), then MLE* is equal to the usual maximum likelihood estimator.

Computation of the MDE requires iterative techniques including numerical evaluation of integrals. For multivariate observations this entails numerical evaluation of multiple integrals. Computation becomes more and more time consuming as k, the dimension of the observation vectors, grows. We next describe a modification of the above estimation method which avoids this problem without sacrificing efficiency or robustness properties. More details of this modification are given in Basu, Markatou and Lindsay (1993).

2.2. *Weighted Likelihood Estimator*

The minimum disparity estimating equations can be expressed as

$$-\frac{\partial}{\partial\boldsymbol{\beta}} \rho_G = \int \cdots \int A(\delta^*(\boldsymbol{z})) \left[\frac{\partial}{\partial\boldsymbol{\beta}} m^*_\beta(\boldsymbol{z}) \right] d\boldsymbol{z} = 0, \tag{3}$$

where $A(x) = (1+x)\left[\frac{d}{d\delta} G(x)\right] - G(x)$. The function $A(x)$ is the residual adjustment function corresponding to a disparity G. Usually, $A(x)$ is redefined such that $A(0) = 0$ and $\frac{d}{dx} A(x)|_{x=0} = 1$. This means for the Hellinger distance $A(x) = 2[(x+1)^{1/2} - 1]$. Most of the theoretical properties of the MDE is governed by the form of the residual adjustment function. For disparities like the Hellinger distance the residual adjustment function can strongly downweight observations with large Pearson residuals. The quantity $A_2 = \frac{d^2}{dx^2} A(x)|_{x=0}$ can be used as a descriptor of the robustness of the MDE; large negative values of A_2 lead to greater robustness. For the Hellinger distance $A_2 = -1/2$, and but for the likelihood disparity $A_2 = 0$ See Basu and Lindsay (1994) for more details.

The estimating equation (3) can be rewritten as

$$\begin{aligned} \int \cdots \int \left[\frac{A(\delta^*)+1}{\delta^*+1}\right](\delta^*+1)\left[\frac{\partial}{\partial\beta}m^*_\beta(\boldsymbol{z})\right]d\boldsymbol{z} \\ = \int \cdots \int w(\delta^*)\left[\frac{\frac{\partial}{\partial\beta}m^*_\beta}{m^*_\beta}\right]d\boldsymbol{F}^*(\boldsymbol{z}) = 0, \end{aligned} \tag{4}$$

where F^* is the distribution function corresponding to f^* and $w(x) = (x+1)^{-1}[A(x)+1]$ with $w(0) = 1$, $\frac{d}{dx}w(x)|_{x=0} = 0$ and $\frac{d^2}{dx^2}w(x)|_{x=0} = A_2$. But equation (4) is a weighted version of the estimating equations of the MLE* in equation (2) with weights $w(\cdot)$. In analogy to equation (4) one can define the following estimating equation for $\boldsymbol{\beta}$

$$\int \cdots \int w(\delta^*)\left[\frac{\frac{\partial}{\partial\beta}m_\beta}{m_\beta}\right]dF_n(\boldsymbol{z}) = 0$$

i.e.,

$$n^{-1}\sum_{i=1}^{n} w(\delta^*(\boldsymbol{Z}_i))u(\boldsymbol{Z}_i,\boldsymbol{\beta}) = 0 \tag{5}$$

where

$$u(\boldsymbol{z},\boldsymbol{\beta}) = \frac{\partial}{\partial\boldsymbol{\beta}}\log_e m_\beta(\boldsymbol{z})$$

is the usual maximum likelihood score function and F_n is the empirical distribution function. Note that the kernel smoothing is now *only* in the weight part and not in the score part of the above equation. Just as one does in iteratively reweighted least squares estimation method (Beaton and Tukey 1974; Holland and Welsch 1977; Birch 1980), equation (5) can be solved iteratively by updating the weights $w(\cdot)$ at every stage. The solution $\hat{\boldsymbol{\beta}}$ of the above estimating equation is called the WLE, which depends on the choice of disparity G.

By the results of Basu, Markatou and Lindsay (1993), the WLEs are asymptotically fully efficient at the model. Under the model, asymptotically the weights $w(\delta^*(Z_i))$ tend to one and equation (5) behaves like the maximum likelihood score equation for all the disparities. In addition, WLEs generated by disparities like the Hellinger distance may have good robustness properties since the weight function $w(\cdot)$ downweights observations with large Pearson residuals. Unlike the MDEs, the WLEs do not require a transparent kernel to achieve full asymptotic efficiency.

2.3. *A Root Selection Criterion under Multiple Roots*

Like the quasi likelihood estimates (Wedderburn 1974) the weighted likelihood estimates are obtained as roots of a set of estimating equations and the computation is not based on any specific optimization criterion. Any nonlinear equation can potentially have more than one root. While computing the MLE by solving the ordinary likelihood equations, the multiple roots problem can be resolved by considering the root at which the likelihood function is maximized. This approach is not possible with the weighted likelihood estimation method when equation (5) has multiple roots.

However, since the estimating equation (5) is obtained by simply replacing the smoothed model and smoothed empiricals in equation (4) with the corresponding unsmoothed versions, one approach to root selection may be to compare the roots against the *parallel disparity measure* whose minimizer is obtained as a solution of equation (4). While it can not be guaranteed that this method will always solve the multiple roots problem, we expect that this will help us identify the good roots and reject the bad roots most of the time. To illustrate this, consider m_β to be $N(\beta, 1)$ model, kernel function to be $N(0, h^2)$ with $h = 0.5$, observed empirical distribution $F_n(z)$ to be the mixture distribution $[0.5N(0,1) + 0.5N(10,1)]$ and the disparity to be the Hellinger distance. Then, m_β^* is $N(\beta, 1 + h^2)$ and the kernel density estimator f^* is the $[0.5N(0, 1 + h^2) + 0.5N(10, 1 + h^2)]$ density. In this case, using the Hellinger distance based weights three roots for the estimating equation (5), i.e., for $\int w(\delta^*)[\partial m_\beta/\partial\beta]dF_n(z) = 0$, are observed —one close to 0, one close to 10 and one exactly at 5. An investigation of the Hellinger distance (parallel disparity measure) between f^* and m_β^* as a function of β reveals that the distance has one local *maximum* at 5, and two local minima – one close to 0 and one close to 10. Thus, the parallel disparity measure allows us to isolate the bad root of the estimating equation in this case.

2.4. *Computation of the WLE*

Finding the WLEs requires the construction of the appropriate weights and solving the weighted likelihood estimation equation (5) iteratively. Note that the estimating equation (5) of the WLE is a sum over the data points rather than an integral over the entire support of $\boldsymbol{Z}$; consequently, the evaluation of the WLE requires no numerical integration. We illustrate the computation of the WLE for the normal model, i.e., when $m_\beta(\boldsymbol{z})$ is the $N_k(\boldsymbol{\mu}, \Sigma)$ density. Given the data $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_n$, where $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{ik})'$, it is convenient to use the multivariate normal kernel $N(0, h^2\boldsymbol{I})$ with covariance matrix $h^2\boldsymbol{I}$, since for the multivariate normal model smoothed density m_β^* itself becomes the multivariate normal density of $N(\boldsymbol{\mu}, \Sigma + h^2\boldsymbol{I})$. The kernel density estimate is calculated as

$$f^*(\boldsymbol{z}) = n^{-1}\sum_{i=1}^{n} \frac{1}{(2\pi)^{p/2}} \exp\left[-\frac{1}{2h^2}(\boldsymbol{z} - \boldsymbol{Z}_i)'(\boldsymbol{z} - \boldsymbol{Z}_i)\right], \boldsymbol{z} \in \mathbb{R}^k.$$

The Pearson residuals $\delta(\boldsymbol{Z}_i), i = 1, \ldots, k$, can then be computed as $\delta(\boldsymbol{Z}_i) = [(m_{\boldsymbol{\beta}}^*(\boldsymbol{Z}_i))^{-1} f^*(\boldsymbol{Z}_i) - 1]$ for any set of initial values of the parameter vector $\boldsymbol{\beta}$ consisting of the distinct parameters in $\boldsymbol{\mu}$ and Σ, and one can construct the initial weights as

$$w(\delta^*(\boldsymbol{Z}_i)) = (\delta^*(\boldsymbol{Z}_i) + 1)^{-1}[A(\delta^*(\boldsymbol{Z}_i)) + 1].$$

Let μ_j and Σ_{ij} denote the j-th and (i, j) the element of $\boldsymbol{\mu}$ and Σ respectively. The weighted likelihood estimating equations given by

$$\sum_{1=1}^{n} w(\delta^*(\boldsymbol{Z}_i))(Z_{ij} - \mu_j) = 0, \quad j = 1, \ldots, p,$$

and

$$\sum_{1=1}^{n} w(\delta^*(\boldsymbol{Z}_i))[(Z_{ij} - \mu_j)(Z_{ik} - \mu_k) - \Sigma_{ij}] = 0, \quad i, j = 1, \ldots, p$$

can be iteratively solved for μ_j's and Σ_{ij}'s, by constructing new weights at every iterative stage.

2.5. *On the Asymptotic Efficiency and Robustness of the WLE*

The WLEs, like the MDEs, combine certain robustness properties with that of full asymptotic efficiency. The influence functions of the MDEs and the WLEs are exactly the same as that of the MLE and can be unbounded. See Beran (1977), Simpson (1987), Lindsay (1994) and Basu and Lindsay (1994). All these authors have recognized that the influence function is not an adequate measure of robustness for the MDEs such as the Hellinger distance estimator and its relatives. Below we show that the same can be true for the WLEs. For the MDEs, the outlier shrinking effect is provided by the residual adjustment function $A(\delta)$ (see Lindsay 1994), whereas for the WLEs this downweighting effect is provided by the weight functions $w(\delta)$. For a robust disparity like the Hellinger distance, $A(\delta) << \delta$ for large positive δ values representing large outliers (see Figure 3 of Lindsay 1994), and the corresponding weight function $w(\delta) = [A(\delta) + 1]/(\delta + 1)$ will be significantly downweighted from 1.

We now examine the influence function of the WLEs. For simplicity, consider the case $k = 1$. Let M_β denote the distribution function of the true density m_β. For $z \in \mathbb{R}$, let $M_\epsilon(z) = (1-\epsilon)M_\beta(z)+\epsilon\chi_y(z), \epsilon > 0$, be an ϵ-contaminated version of $M_\beta(z)$, where $\chi_y(z)$ is the distribution function corresponding to the density which puts mass 1 at $z = y$. Let $T(\cdot)$ be the WLE functional defined on the space of distributions. From equation (5) it can be seen that $T(M_\beta) = \boldsymbol{\beta}$, so that the WLE functional is Fisher consistent. Define the influence function $IF(y)$ to be

$$IF(y) = T^{(1)}(y) = \frac{\partial}{\partial\epsilon}T(M_\epsilon)|_{\epsilon=0}.$$

Taking the derivatives of both sides of equation (5) under the contaminated distribution M_ϵ and using $w(0) = 1, \frac{d}{dx}w(x)|_{x=0}$ one gets, after some simple algebra, $T^{(1)}(y) = (\boldsymbol{I}(\boldsymbol{\beta}))^{-1}\boldsymbol{u}(y, \boldsymbol{\beta})$ where

$$\boldsymbol{u}(z, \boldsymbol{\beta}) = \frac{\partial \log_e m_\beta(z)}{\partial\boldsymbol{\beta}}, \quad \boldsymbol{I}(\boldsymbol{\beta}) = \int [\boldsymbol{u}(Z, \boldsymbol{\beta})][\boldsymbol{u}(Z, \boldsymbol{\beta})]' m_\beta(z)dz,$$

the Fisher information about parameter $\boldsymbol{\beta}$ in model m_β. In particular, this is satisfied by the WLE generated by the Hellinger distance based weights, and the MLE (which is also a WLE with weights identically equal to 1). Under standard regularity conditions the asymptotic distribution of a Fisher consistent functional $T(\cdot)$ at the model can be derived using the expression

$$n^{1/2}(T(F_n) - \beta) = n^{-1}\Sigma_{i=1}^{n} IF(Z_i) + o_p(1),$$

with F_n denoting the empirical distribution function. See, for example, Fernholz (1983). This implies that the asymptotic distribution of the WLEs and the MLE are the same, indicating the full asymptotic efficiency of the WLEs and the potential of their influence function to be unbounded.

However, the first order influence function approach may give an unreliable prediction of the bias $\Delta T(\epsilon) = T(M_\epsilon) - T(M_\beta)$ of some of the WLEs under contamination. Here, for p = 1 we present an analysis of the bias function of the WLEs which parallels that of Lindsay (1994, Section 4). Up to the first order of approximation, the bias function of all the WLEs is the same:

$$\Delta T(\epsilon) \equiv \epsilon T^{(1)}(y),$$

so that the first order analysis fails to distinguish between the MLE and other WLEs that exhibit robustness properties. However, the second order analysis can be more informative and may provide more insight into the robustness of the WLEs. This is given by:

$$\Delta T(\epsilon) \equiv \epsilon T^{(1)}(y) + \frac{\epsilon^2}{2} T^{(2)}(y),$$

where $T^{(2)}(y) = \partial^2/\partial\epsilon^2 T(m_\epsilon)|_{\epsilon=0}$. One could consider the ratio of the quadratic to linear approximations of $\Delta T(\epsilon)$ as a measure of adequacy of the first order approximation. Note that this ratio is given by $(1 + [T^{(2)}(y)/T^{(1)}(y)](\epsilon/2))$. Provided the signs of $T^{(1)}(y)$ and $T^{(}2)$(y) are opposite, this indicates that the first and second order approximations to the bias of the estimator will differ by more than 50% whenever ϵ is larger than $\epsilon_{crit} = |T^{(1)}(y)/T^{(2)}(y)j$, with the second order approximation predicting smaller bias. A tedious but straightforward calculation for $T^{(2)}(y)$ gives

$$T^{(2)}(y) = (I(\beta))^{-1}[f_1(y) + A_2 f_2(y)]T^{(1)}(y) \tag{6}$$

and

$$\epsilon_{crit} = |(I(\beta))^{-1}[f_1(y) + A_2 f_2(y)]j^{-1},$$

where $f_1(y)$ is exactly as defined in Proposition 3 of Lindsay (1994), and

$$A_2 = \frac{d^2}{dx^2}A(x)|_{x=0} = \frac{d^2}{dx^2}w(x)|_{x=0}.$$

It follows from Lindsay (1994) that $f_1(y) = 0$ for the mean parameter of a one parameter exponential family model. However, in general $f_2(y)$ is a complicated function of y involving the kernel density estimate and the true model density, but we calculated the values of $f_2(y)$ at the normal $N(\beta, 1)$ model, using true $\beta = 0$, and normal kernel with variance h^2 for several values of h and y. The results are presented in Table 1, and are similar to those of Lindsay (1994, Table 1). Note that the values of $f_2(y)$ in Table 1 are all positive, indicating whenever A_2 is negative (e.g., in the Hellinger distance case), the second order approximation will predict smaller bias than the first order approximation. Thus, even though the bias predicted in the normal model for the robust WLEs due to the presence of a contaminating observation at y using the first order influence function approximation can be quite large, the true bias can actually be quite small. Table 1 also shows that $f_2(y)$ increases as h decreases. Therefore, choice of smaller h values for the $N(0, h^2)$ kernel will lead to greater robustness for disparities like the Hellinger distance. For multivariate data, one strategy may be to select h as a suitable small multiple of the average of the median absolute deviations (MADs) of the components of $\boldsymbol{Z}$. In Examples 1 and 2 in Section 5 below we have taken h to be approximately 1/5(average of the MADs) and used $h = 0.5$ in Example 1 and $h = 0.05$ in Example 2.

Table 1. *Values of $f_2(y)$ in the expression for $T^{(2)}(y)$ in Equation (6) for the normal example for various values of y and h.*

y / h	2.0	2.5	3.0	3.5	4.0	4.5
0.25	14.18	45.18	174.45	797.71	4435.26	30601.03
0.50	4.26	11.00	32.18	101.03	352.32	1410.95
0.75	1.75	2.84	5.28	9.91	18.70	36.41

3. WEIGHTED LIKELIHOOD ESTIMATION IN THE ERRORS-IN-VARIABLES MODEL

Since in model (1) the joint distribution of $\boldsymbol{Z}_i = (Y_i, \boldsymbol{X}_i')'$ is multivariate normal, the model parameters can be determined by a straightforward application of the weighted likelihood estimation method considered in Section 2.4. Here $m_\beta(\boldsymbol{z})$ is the $N_{p+1}(\boldsymbol{\mu}, \Sigma)$ density. In the weighted likelihood estimation procedure any smooth kernel can be used to obtain a nonparametric density estimator and to smooth the model density. In particular, if the $N_{p+1}(0, h^2\boldsymbol{I})$ density is used as the kernel, where $\boldsymbol{I}$ is the $(p+1) \times (p+1)$ identity matrix, then the smoothed model density $m_\beta^*(\boldsymbol{z})$ would be the $N_{p+1}(\boldsymbol{\mu}, \Sigma + h^2\boldsymbol{I})$ density. Then, given a particular disparity ρ_G, the corresponding WLE of the parameter vector can be calculated by iteratively solving equation (5). From the viewpoint of robustness some disparity measures are more desirable than others. In our study we have investigated the WLE with weights based on the Hellinger distance (with $A(x) = 2[(x+1)^{1/2} - 1]$) and the results are presented in Sections 4 and 5. In computing the WLEs, unlike the MDEs, we no longer have to choose a transparent kernel and any other smooth kernel would work as well.

We now discuss how the WLE relates to the estimators defined by Brown (1982) and Zamar (1989). The MLE of $\boldsymbol{\beta}$ in model (1) minimizes the sum of squared *orthogonal residuals* of the observations $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ with the residuals measured perpendicular to the estimated plane. Thus, the MLE is the orthogonal regression estimator. Generalizing the idea of Beaton and Tukey (1974) in constructing a robust estimator for the regression problem when there is no measurement error, Brown (1982, p. 75, eq. (6.4)) considers iteratively reweighted orthogonal regression estimator for the special case of model (1): $y_i = \beta x_i$, a straight line through the origin. Brown introduces two sets of weights in modifying the likelihood function each intended to downweight outlying coordinates (X_i, Y_i). Zamar (1989, p. 150, eq. (3)) defines the orthogonal regression M-estimator using a robust estimate S_n of the scale of the orthogonal residuals and a robustifying loss function that downweights large orthogonal residuals. In Zamar's approach an observation $\boldsymbol{Z}_i = (Y_i, \boldsymbol{X}_i')'$ is considered an outlier if its orthogonal residual is large in comparison to S_n. In case of the WLE, its robustness results from downweighting large Pearson residuals $\delta^*(\boldsymbol{Z}_i)$, in which case an observation Z_i is considered an outlier if the kernel density $f^*(Z_i)$ at Z_i is large relative to the smoothed model density $m^*_\beta(Z_i)$ at Z_i. Thus, the WLE downweights observations which represent large residuals in a probabilistic, rather than geometric, sense.

4. SIMULATION RESULTS

To investigate the small sample performance of the WLEs we carried out a Monte Carlo experiment for model (1) using weights based on the Hellinger distance. The experiment was designed exactly following Zamar (1989), as was done by B&S. Here we give a brief description. The model considered is

$$y_i = \beta_1 + \beta_2 x_i, \quad Y_i = y_i + e_i, \quad X_i = x_i + u_i,$$

with

$$x_i \sim N(0,1), \; u_i \sim CN(0.25, \sigma^2, 0.05), \; e_i \sim CN(0.25, \tau^2, 0.05)$$

where x_i, u_i and e_i are all independent and

$$CN(\sigma_1^2, \sigma_2^2, \epsilon) = (1-\epsilon)N(0, \sigma_1^2) + \epsilon N(0, \sigma_2^2).$$

For each of the (σ, τ) combinations considered by Zamar (1989) one hundred samples of size 20 were generated for the random variables Y_i, X_i under the above model. Computations were done using the same set of 100 samples of size 20 used by B&S. The true value of β_1 was set to zero and β_2 was chosen at random uniformly between -5 and 5. In B&S computations were presented for the two estimators of β_2: (i) T_1, the orthogonal regression M-estimator of Zamar; (ii) T_2, the minimum Hellinger distance estimator. In this study we computed (iii) T_3, the WLE using weights $w(\delta^*)$ corresponding to the Hellinger distance. In computing T_3 (as well as T_2) the $N_2(0, h^2 \boldsymbol{I})$ density with $h = 0.5$ was used as the kernel.

To evaluate T_2 and T_3 we computed the initial estimates of the means, variances and covariance $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ and σ_{XY} of X and Y variables as follows: $\hat{\mu}_X = \text{median}(X_1, X_2, \ldots, X_n), \hat{\mu}_Y = \text{median}(Y_1, Y_2, \ldots, Y_n), \hat{\sigma}_X = 1.48\times \text{median}(|X_1 - \hat{\mu}_X|, |X_2 - \hat{\mu}_X|, \ldots, |X_n - \hat{\mu}_X|), \hat{\sigma}_Y = 1.48\times$ median $(|Y_1 - \hat{\mu}_Y|, |Y_2 - \hat{\mu}_Y|, \ldots, |Y_n - \hat{\mu}_Y|)$ and $\hat{\sigma}_{XY}$ was computed using the covariance estimate formula as given in equation (2.7) of Huber (1981, p. 203).

The criterion m defined in equation (6) of Zamar (1989) and used by B&S is applied here to measure the performance of the WLE. For the i-th estimator T_i, it is defined by

$$m = \sum_{j=1}^{100} \left[1 - \frac{|1 + T_{ij}\beta_{2j}|}{(1 + T_{ij}^2)^{1/2}(1 + \beta_{2j}^2)^{1/2}} \right]$$

where T_{ij} and β_{2j} are respectively the i-th estimator and the true value of β_2 for the j-th replication. An estimator is better than another if its value for measure m is smaller.

The simulation results are summarized in Table 2. Numbers for estimators T_1 and T_2 given in B&S are presented in Table 2 for the convenience of comparisons. Under the contaminated distributions (i.e., for $(\sigma, \tau)\tau(0.5, 0.5)$ in Table 2), the WLE based on Hellinger distance and the minimum Hellinger distance estimator have performed equally well in terms of robustness, and both of them have outperformed the orthogonal regression M-estimator. The numbers show that the WLE and the MDE are very close; what they do not show, however, is that the Hellinger distance based WLE is by far much easier to calculate than the minimum Hellinger distance estimator.

Table 2. *Simulated performance measure m for the estimators T_1, T_2 and T_3; T_1, Orthogonal M-Estimator; T_2, Minimum Hellinger Distance Estimator; T_3, Weighted Likelihood Estimator based on the Hellinger distance (using 100 samples of size 20).*

σ	τ	T_1	T_2	T_3
0.5	0.5	0.2194	0.1581	0.1558
0.5	2.0	0.3021	0.2273	0.2230
0.5	5.0	1.0172	0.2027	0.1893
2.0	0.5	0.2974	0.1952	0.1979
2.0	2.0	0.3526	0.2506	0.2561
2.0	5.0	0.8959	0.2360	0.2315
5.0	0.5	1.4596	0.1669	0.1659
5.0	2.0	1.4217	0.2354	0.2413
5.0	5.0	1.6589	0.1856	0.1815

5. APPLICATIONS

Example 1. We consider the data on simultaneous pairs of measurements of serum kanamycin levels in blood samples drawn from umbilical catheter and heel venapuncture of 20 babies given by Kelly (1984). This dataset was analyzed by Zamar (1989, Example 2, Table 3, pp. 154-155) who argued that it was reasonable to assume that both measurements were subject to random errors with equal variances, i.e., the identifiability condition of model (1) was satisfied. To illustrate the behavior of different estimates in the presence of outliers, Zamar changed the second observation vector (33.2, 26.0) to (39.2, 32.0), the latter is the outlying point in the upper right corner of Figure 1. Through three fitted lines, Zamar (1989, Figure 2) showed outlier-resistance of the orthogonal regression M-estimate and outlier-sensitivity of the classical orthogonal regression estimate. In Figure 1, in addition to reproducing Zamar's fitted lines, we display the WLE-fit for the contaminated data. In computing the WLE the initial estimates for the iteration were chosen as described in Section 4 and used the normal kernel with h = 0.5. Like the orthogonal regression M-estimator, the WLE line (estimated intercept –6.43, estimated slope 1.35) is insensitive to the presence of the outlier.

Example 2. We now illustrate weighted likelihood estimation for the trivariate dataset given in Table 2.3.1 of Fuller (1987, p. 131) on log crop (Y), log extension wood growth (X_1) and log girth increment (X_2). The dataset has 5 observations. Fuller computes the MLEs of the regression equation

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$$

using a 4×4 matrix $\boldsymbol{S}_{\epsilon\epsilon}$ as an estimate of $\Sigma_{\epsilon\epsilon}$, the covariance matrix of $\epsilon_i = (e_i, u_{0i}, u_{1i}, u_{2i})'$ where $u_{0i} = 0$ denotes the measurement error for the intercept variable. For this dataset we illustrate the computation of the WLEs of $\beta_0, \beta_1, \beta_2$.

First we obtain the Hellinger distance based WLEs of $\hat{\boldsymbol{\mu}}_Z$ of $\boldsymbol{\mu}_Z = (\mu_Y, \mu_{X_1}, \mu_{X_2})'$, the mean vector of $(Y, X_1, X_2)'$ to be 1.9481, 3.9954 and 0.5087 respectively, with the initial estimates chosen as described in Section 4. We use the $N_3(0, h^2\boldsymbol{I})$ kernel with h = 0.05. The final weights $w(\delta^*(\boldsymbol{Z}_i))$ attached to the observations $\boldsymbol{Z}_i = (Y_i, X_{1i}, X_{2i})'$,

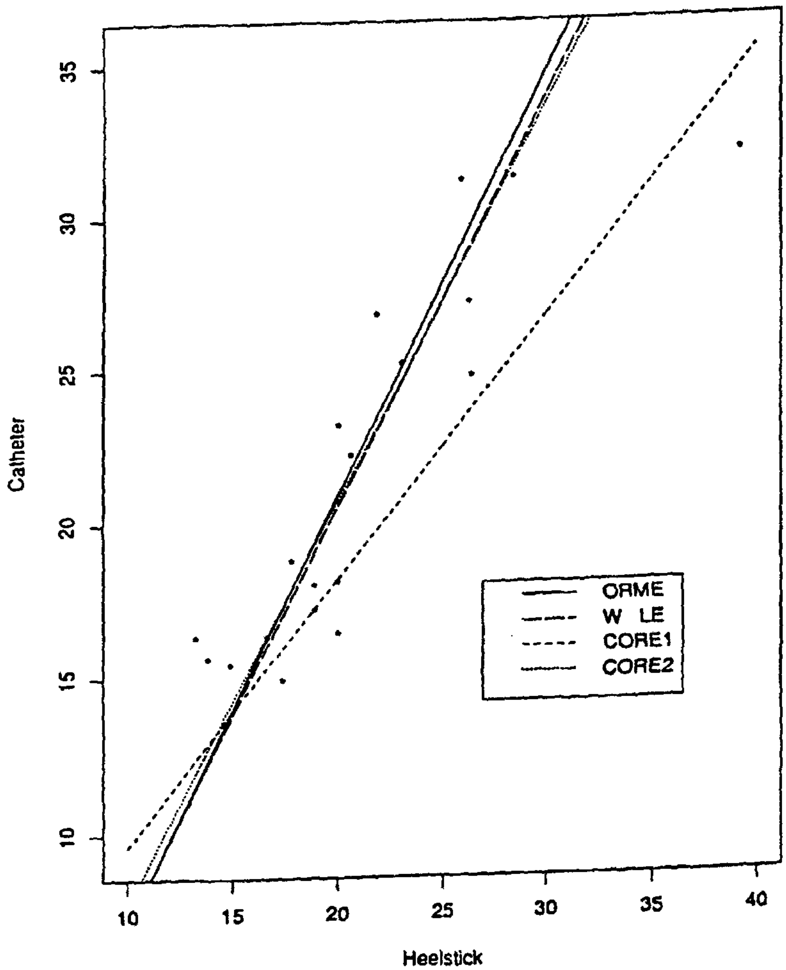

Figure 1. *Four different fits for data in Table 3 of Zamar (1989). ORME, WLE, CORE1 and CORE2 represent the orthogonal regression M-estimator (Zamar 1989), the Hellinger distance based weighted likelihood estimator, the classical orthogonal regression estimator using all data, and the classical orthogonal regression estimator with second observation deleted respectively.*

$i = 1, \ldots, 5$, are observed to be 0, 0, 0.9794, 0.9041 and 0.8727 respectively. This indicates that the first two observations are possible outliers and the Hellinger distance based robust WLEs are likely to be better than the MLEs obtained by Fuller (1987, p. 132).

Next to compute robust estimates of β_1 and β_2 using the WLE of μ_Z we modify Fuller's method of MLE computation as described in Theorem 2.3.1 of Fuller (1987, p. 124). We calculate the 3×3 matrix

$$M^*_{ZZ} = \left[\sum_{i=1}^{n} w(\delta^*(Z_i)\right]^{-1} \sum_{i=1}^{n} w(\delta^*(Z_i))(Z_i - \hat{\mu}_Z)(Z_i - \hat{\mu}_Z)'$$

and find the smallest root of the determinantal equation $|M^*_{ZZ} - \lambda S^*_{\epsilon\epsilon}| = 0$ where $S^*_{\epsilon\epsilon}$ is the 3×3 submatrix of $S_{\epsilon\epsilon}$ obtained by removing the 2nd row and 2nd column of the latter. Let M^*_{XX} be the 2×2 lower submatrix of M^*_{ZZ}, let S^*_{uu} denote the 2×2 lower submatrix of $S^*_{\epsilon\epsilon}$ and let S^*_{ue} denote the 2×1 vector obtained by deleting the first component of the first column of $S^*_{\epsilon\epsilon}$. The smallest root $\hat{\lambda}$ is found to be 2.982×10^{-9}. Then the Hellinger distance based WLEs of β_1 and β_2 obtained by using $(\hat{\beta}_1, \hat{\beta}_2)' = (M_{XX} - \hat{\lambda} S^*uu)^{-1}(M^*_{XY} - \hat{\lambda} S^*_{ue})$ are observed to be 0.0567 and 2.1801 respectively, and the Hellinger distance based WLE $\hat{\beta}_0 = \hat{\mu}_Y - \hat{\beta}_1 \hat{\mu}_{X_1} - \hat{\beta} \hat{\mu}_{X_2}$ of β_0 is found to be 0.6125.

ACKNOWLEDGEMENTS

The research of the first author was partially supported by a Mathematics Department Research Award, University of Texas at Austin, while the author was a faculty member at the University of Texas. The research of the second author was supported in part by a Grant from the College of Arts and Sciences at Oklahoma State University. The authors wish to thank Professor Bruce G. Lindsay for some helpful comments, a referee and the Associate Editor for many helpful suggestions that led to the present improved version of this paper.

REFERENCES

Basu, A. and Lindsay, B. G. (1994). Minimum disparity estimation for continuous models: Efficiency, distributions and robustness. *Ann. Inst. Statist. Math.* **46**, 683–705.

Basu, A., Markatou, M. and Lindsay, B. G. (1993). likelihood estimating equations. *Tech. Rep.* **93-01**, Department of Statistics, Columbia University, New York, NY.

Basu, A. and Sarkar, S. (1994a). Minimum disparity estimation in the errors-in-variables model. *Statist. Probab. Lett.* **20**, 69–73.

Basu, A. and Sarkar, S. (1994b). The trade-off between robustness and efficiency and the effect of model smoothing in minimum disparity inference. *J. Statist. Comput. Simul.* **50**, 173–185.

Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band spectroscopic data. *Technometrics* **16**, 1470–1485.

Birch, J. B. (1980). Some convergence properties of iterated least squares in the location model. *Commun. Statist. B* **9**, 359–369.

Brown, M. (1982). Robust line estimation with errors in both variables. *J. Amer. Statist. Assoc.* **77**, 71–79.

Carroll, R. and Gallo, P. (1982). Some aspects of robustness in the functional errors in variables model. *Commun. Statist. A* **11**, 2573–2585.

Carroll, R. J. and Eltinge, J. L. and Ruppert, D. (1993). Robust linear regression in replicated measurement error models. *Statist. Probab. Lett.* **16**, 169–175.

Cheng, C. and J. W. Van Ness (1990). Bounded influence errors in variables regression, (P. Brown and W. A. Fuller, eds.), Contemporary Mathematics: Statistical Analysis of Measurement Error Models and Applications (Amer. Math. Soc. Providence, RI).

Cheng, C. and J. W. Van Ness (1992). Generalized M-estimators for errors-in-variables regression. *Ann. Statist.* **20**, 385–397.

Croos, J. and Fuller, W. A. (1991). Robust estimation in measurement error models. American Statistical Association 1991 Proceedings of the Business and Statistics Section, 283–288.

Fernholz, L.T. (1983). *Von Mises Calculus for statistical Functionals*. New York: Springer Verlag.

Fuller, W. A. (1987). *Measurement error models*. New York: Wiley

Holland, P. W. and Welsch, R. E. (1977). Robust regression using iteratively reweighted least squares. *Commun. Statist. A* **6**, 813–827.

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

Kelly, G. (1984). The influence function in the errors in variables problem. *Ann. Statist.* **12**, 87–100.

Lindsay, B. G. (1994). Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Ann. Statist.* **22**, 1081–1114.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* **61**, 439–447.

Zamar, R. H. (1989). Robust estimation in the errors-in-variables model. *Biometrika* **76**, 149–160.