The Annals of Statistics
1992, Vol. 20, No. 1, 385–397

GENERALIZED *M*-ESTIMATORS FOR ERRORS-IN-VARIABLES REGRESSION

BY CHI-LUN CHENG AND JOHN W. VAN NESS

Academia Sinica and University of Texas, Dallas

This paper discusses robust estimation for structural errors-in-variables (EV) linear regression models. Such models have important applications in many areas. Under certain assumptions, including normality, the maximum likelihood estimates for the EV model are provided by orthogonal regression (OR) which minimizes the orthogonal distance from the regression line to the data points instead of the vertical distance used in ordinary regression. OR is very sensitive to contamination and thus efficient robust procedures are needed. This paper examines the theoretical properties of bounded influence estimators for univariate Gaussian EV models using a generalized M-estimate approach. The results include Fisher consistency, most B-robust estimators and the OR version of Hampel's optimality problem.

1. Introduction. Consider the usual linear errors-in-variables (EV) model (also called the measurement error model) which assumes that there are linearly related true variables x and y,

$$y = \alpha + \beta x, \tag{1.1}$$

which cannot be observed directly. Instead X and Y are observed, where

$$X = x + u, \qquad Y = y + \varepsilon \tag{1.2}$$

and u, ε and x are independent with $Eu = 0$ and $E\varepsilon = 0$. Thus, x and y are both observed with errors. If x and y are random variables, then (1.1) is called a structural relation; while if x and y are deterministic, then (1.1) is called a functional relation [Kendall and Stuart (1979), Chapter 29]. Only the structural model will be considered here.

The proofs used in this paper require that all random variables be Gaussian. In this case the model (1.1)–(1.2) is underspecified and an additional assumption is needed. A common assumption is that the ratio of the error variances $\lambda = \sigma_\varepsilon^2/\sigma_u^2$ is known. This is assumed throughout this paper and all variables are assumed scaled so that $\lambda = 1$. Under these assumptions the maximum likelihood estimates of the parameters α and β are given by orthogonal regression (OR) which minimizes the orthogonal distance from the regression plane to the data points instead of the vertical distance.

Surveys of the EV model can be found in Kendall and Stuart [(1979), Chapter 29] and Fuller (1987). Two important early papers are Lindley (1947)

Received October 1987; revised March 1991.
AMS 1980 *subject classifications*. Primary 62F35; secondary 62J05.
Key words and phrases. Errors-in-variables regression, measurement error model, structural model, robust statistics, Fisher consistency, generalized M-estimates.

and Madansky (1959). Other early works on the subject are referenced in Anderson (1984). Fuller (1980), Gleser (1981) and others investigate the multivariate EV model.

Simulation studies have shown that OR is nonrobust in the sense that robust algorithms give much better estimates when the data are generated from models having modest deviations from the standard model [Carroll and Gallo (1982), Brown (1982) and Ammann and Van Ness (1988, 1989)]. The present paper examines the theoretical properties of robust estimates for the parameters of model (1.1)–(1.2) in the Gaussian case. In particular we study bounded influence generalized M-estimates.

The results here parallel those of robust ordinary regression [Hampel, Ronchetti, Rousseeuw and Stahel (1986), Chapter 6]. Standard M-estimators [Huber (1973)] are only the first step in the robustification of a regression estimator since they do not allow bounded influence estimators. Generalized M-estimators [Maronna, Bustos and Yohai (1979)], denoted GM-estimators hereafter, provide such estimators.

Early theoretical work on robust errors-in-variables regression includes Carroll and Gallo (1982) and Brown (1982). Zamar (1985, 1989) has extensive coverage of classical M-estimation and Cheng and Van Ness (1990) discusses GM-estimation for the EV model. The difference in bounded influence properties between classical and GM-estimates in EV regression is similar to that mentioned above for ordinary regression.

In Section 2, the M-estimator approach of Huber (1973) is described for the EV problem [Zamar (1985)]. Section 3 introduces GM-estimators and establishes Fisher consistency. In Section 4, sharp lower bounds for both gross-error sensitivity and self-standardized sensitivity are obtained; and GM-estimators which attain these lower bounds are found. In Section 5, Hampel's optimality problem is solved for the EV model.

2. Robust orthogonal regression via *M*-estimators. Suppose there are n independent and identically distributed observations $(X_1, Y_1), \dots, (X_n, Y_n)$ from the structural Gaussian EV model (1.1)–(1.2) with $\lambda = 1$. Lindley (1947) and Madansky (1959) show that maximum likelihood leads to minimizing a weighted sum of squares,

$$\min_{\alpha,\beta} \sum_{i=1}^{n} w_i(\beta)(Y_i - \alpha - \beta X_i)^2 = \min_{\alpha,\beta} \sum_{i=1}^{n} \left(\frac{Y_i - \alpha - \beta X_i}{K(\beta)} \right)^2, \tag{2.1}$$

where $K(\beta) = (1 + \beta^2)^{1/2}$ is the weight which converts to orthogonal residuals. For weights under more general assumptions, see Fuller [(1987), page 37]. Huber's M-estimation approach applied to OR modifies (2.1) to a general loss function ρ,

$$\min_{\alpha,\beta} \sum_{i=1}^{n} \rho\left(\frac{Y_i - \alpha - \beta X_i}{K(\beta)} \right) = \min_{\alpha,\beta} \int \rho\left(\frac{Y - \alpha - \beta X}{K(\beta)} \right) dF_n(X, Y), \tag{2.2}$$

where F_n is the empirical distribution function of $(X_1, Y_1), \dots, (X_n, Y_n)$. By choosing ρ suitably, various robust estimates α_n and β_n can be obtained. It is convenient to write the estimators in functional form, $\alpha_n = S(F_n)$ and $\beta_n = T(F_n)$, where $S(F)$ and $T(F)$ are solutions of

$$\min_{S,T} \int \rho\left(\frac{Y - S(F) - T(F)X}{K(T(F))}\right) dF(X,Y). \tag{2.3}$$

Zamar (1985, 1989) discusses the multivariate version of the M-estimators obtained from equations of the form (2.3). He investigates existence, uniqueness, consistency, influence functions and asymptotic normality for such estimates. He does not consider the GM-estimates studied in this paper nor the bounded influence criterion for robustness.

A defining relation for β can be obtained from (2.3) by differentiation:

$$\begin{aligned} &\int \psi\left(\frac{Y - S(F) - T(F)X}{K(T(F))}\right) dF(X,Y) = 0, \\ &\int \psi\left(\frac{Y - S(F) - T(F)X}{K(T(F))}\right)(X + T(F)Y)\, dF(X,Y) = 0, \end{aligned} \tag{2.4}$$

where it is assumed that ρ has derivative ψ and that regularity conditions hold which allow interchange of differentiation and integration. The estimates obtained from (2.4) do not have bounded influence [see (2.6)]. To overcome this limitation we employ GM-estimates in the following sections. We will derive the optimal bounded influence estimator for the slope parameter β only. We can treat the slope parameter separately because of the following. If we were to define $\dot{X} = X - \mu_X$ and $\dot{Y} = Y - \mu_Y$, then (2.4) reduces to

$$\int \psi\left(\frac{\dot{Y} - T(G)\dot{X}}{K(T(G))}\right)(\dot{X} + T(G)\dot{Y})\, dG(\dot{X},\dot{Y}) = 0, \tag{2.5}$$

where G is the distribution of $\dot{X}$ and $\dot{Y}$. Thus, $T(F) = T(G)$ and if we are only interested in optimal estimates of β, we can theoretically study estimates obtained from the appropriate GM-estimation modification of (2.5). In practice, μ_X and μ_Y are not known. If we need bounded influence estimators of both α and β, we can, for example, use bounded influence estimators $\mu_X(F)$ and $\mu_Y(F)$ of location found in Hampel, Ronchetti, Rousseeuw and Stahel [(1986), Section 2.4] and apply the following.

PROPOSITION 2.1. *Let $\mu_X(F)$, $\mu_Y(F)$ and $T(F)$ be bounded influence estimators of μ_X, μ_Y and β, respectively, and assume that either $\mu_X(F)$ or $T(F)$ are continuous functionals in the sense that $\mu_X(H) \to \mu_X(F)$ or $T(H) \to T(F)$ whenever $H \to F$ weakly. Then $S(F) = \mu_Y(F) - T(F)\mu_X(F)$ is a bounded influence estimator of α. If $\mu_X(F_n)$, $\mu_Y(F_n)$ and $T(F_n)$ are strongly consistent, then so is $S(F_n)$.*

PROOF. Let $Z = (X,Y)$ and $F_{Z,\varepsilon} = (1-\varepsilon)F + \varepsilon\Delta_Z$, where Δ_Z represents a unit mass at the Z. If, for example, T is continuous, by the definition of the influence function of S,

$$\begin{aligned}\mathrm{IF}(Z;S,F) &= \lim_{\varepsilon\downarrow 0}\frac{S(F_{Z,\varepsilon})-S(F)}{\varepsilon}\\ &= \lim_{\varepsilon\downarrow 0}\left\{\frac{\mu_Y(F_{Z,\varepsilon})-\mu_Y(F)}{\varepsilon} - T(F_{Z,\varepsilon})\frac{\mu_X(F_{Z,\varepsilon})-\mu_X(F)}{\varepsilon}\right.\\ &\qquad\left. -\mu_X(F)\frac{T(F_{Z,\varepsilon})-T(F)}{\varepsilon},\right\}\end{aligned}$$

and therefore

$$\mathrm{IF}(Z;S,F) = \mathrm{IF}(Z;\mu_Y,F) - T(F)\mathrm{IF}(Z;\mu_X,F) - \mu_X(F)\mathrm{IF}(Z;T,F).$$

Since the influence functions of $\mu_X(F)$, $\mu_Y(F)$ and $T(F)$ are bounded in Z, the influence function of $S(F)$ is bounded in Z. Strong consistency is obvious. □

Of course, the optimality of T (in the sense of Section 5) does not imply that of S. No such optimal estimate for α has been derived even in the ordinary regression case.

Denote the distribution of $\dot{X}$ and $\dot{Y}$ with slope β by G_β. The functional T is said to be Fisher consistent $T(G_\beta) = \beta$. Since all the estimators discussed in this paper are Fisher consistent (see Section 3) the notations $T(G_\beta)$ and β will be used interchangeably as is common practice in the literature.

The influence function and the asymptotic variance of the estimate of β are obtained from (2.4) by appealing to general results on multidimensional M-estimates [see, e.g., Hampel, Ronchetti, Rouseeuw and Stahel (1986), page 230]. The variance results due to Huber (1967) and Clarke (1983) require certain additional regularity conditions not discussed here. The influence function for β is

$$\mathrm{IF}\big(\dot{X},\dot{Y};\beta,G_\beta\big) = M(\psi,G_\beta)^{-1}\psi\left(\frac{\dot{Y}-\beta\dot{X}}{K(\beta)}\right)(\dot{X}+\beta\dot{Y}), \tag{2.6}$$

where

$$M(\psi,G_\beta) = -\int\frac{\partial}{\partial\zeta}\psi\left(\frac{\dot{Y}-\zeta\dot{X}}{K(\zeta)}\right)(\dot{X}+\zeta\dot{Y})\bigg|_{\zeta=\beta} dG_\beta(\dot{X},\dot{Y}).$$

Note that for any nontrivial ψ, the influence function (2.6) remains unbounded. Thus, this form of M-estimation cannot lead to estimates with bounded influence. The same is true in ordinary regression, see Hampel, Ronchetti, Rousseeuw and Stahel (1986). Using the terminology of ordinary regression, call the M-estimator defined by (2.4) a classical M-estimator for the EV model.

3. Generalized *M*-estimators for EV models. Henceforth, we drop the dot notation and write F for G, that is, we assume the data is centered on the mean; (X, Y) has mean zero. GM-estimators for EV models are obtained by replacing $\psi[v(T)]$ in (2.5) by $\eta[w(T), v(T)]$, where we define

$$(3.1) \qquad w(\beta) = \frac{X + \beta Y}{K(\beta)}, \qquad v(\beta) = \frac{Y - \beta X}{K(\beta)}.$$

The second variable added by the η function allows one to remove the term causing the unboundedness of the influence function in (2.6).

Thus a GM-estimate for the slope parameter is defined implicitly by

$$(3.2) \qquad \int \eta[w(T(F)), v(T(F))] w(T(F)) \, dF(X, Y) = 0$$

and the finite sample estimate β_n satisfies $\sum \eta[w_i(\beta_n), v_i(\beta_n)] w_i(\beta_n) = 0$. Here $\eta: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is assumed to satisfy the following conditions [see Maronna and Yohai (1981) and Ronchetti and Rousseeuw (1985)]:

1. $\eta(w, \cdot)$ is continuous on $\mathbb{R} - C(w; \eta)$ for all $w \in \mathbb{R}$, where the set $C(w; \eta)$ has finite cardinality. At each point of $C(w; \eta)$, $\eta(w, \cdot)$ has finite left and right limits.
2. $\eta(w, \cdot)$ is odd and $\eta(w, v) \geq 0$ for all $w \in \mathbb{R}$, $v \in \mathbb{R}^+$.
3. For all w, the set $D(w; \eta)$ of points in which $\eta(w, \cdot)$ is continuous but in which $(\partial/\partial\beta)\eta[w(\beta), v(\beta)]$ is not defined or not continuous, is finite.
4. $M(\eta, F)$ and $Q(\eta, F)$ exist and are nonzero, where

$$(3.3) \qquad M(\eta, F) = -\int \frac{\partial}{\partial t} [\eta[w(t), v(t)] w(t)]_{t=T(F)} \, dF(X, Y)$$

and

$$(3.4) \qquad Q(\eta, F) = \int \eta^2[w(T(F)), v(T(F))] w^2(T(F)) \, dF(X, Y).$$

Results in Huber [(1981), page 45] imply that the influence function of T at F is

$$(3.5) \quad \mathrm{IF}(X, Y; T, F) = M^{-1}(\eta, F) \eta[w(T(F)), v(T(F))] w(T(F)).$$

Under certain regularity conditions [Huber (1967)] including uniqueness of solutions of (3.2), the estimator is strongly consistent and asymptotically normal with asymptotic variance

$$(3.6) \quad V(T, F) = \int \mathrm{IF}^2(X, Y; T, F) \, dF(X, Y) = M^{-2}(\eta, F) Q(\eta, F).$$

The Huber conditions can be weakened for the special case of EV model GM-estimation. The specialization is done for ordinary regression in Maronna and Yohai (1981) and for EV model M-estimators defined by (2.3) in Zamar (1985), but we will not pursue this matter in the present paper.

It is necessary in what follows that we have a unique solution to (3.2) at least in a neighborhood of F_β. Note that even for classical least squares estimation (2.4) has two solutions. Frequently side conditions are imposed to force uniqueness [see, e.g., Zamar (1985) and Clarke (1983)]. Henceforth, it is assumed that a unique solution T has been identified.

Fisher consistency, $T(F_\beta) = \beta$ for all β, is important for several reasons. It says that the defining equation (3.2) is unbiased at the model. Furthermore, if T is continuous with respect to the weak-star topology [see, e.g., Huber (1981), Chapter 2], then since F_n converges almost surely, then $\beta_n = T(F_n)$ will converge almost surely to β. Finally, Fisher consistency will be necessary for the most B-robust and optimal estimates discussed below.

In the EV model with L_2 loss, it is easy to check that the defining equation (2.4) is satisfied at the model. This means that for L_2 loss, the corresponding M-estimator is Fisher consistent regardless of the distribution of (X, Y). Fisher consistency for EV model M-estimators and GM-estimators can be proved under the Gaussian assumption as follows.

THEOREM 3.1. *In the Gaussian EV model, if a unique GM-estimator is defined by* (3.2), *it is Fisher consistent if, at the model* F_β, $E_v\eta(w, v) = 0$ *for all* $w \in \mathbb{R}$.

PROOF. We first rotate to the principal components, w and v, of X and Y. At the model $F = F_\beta$, let θ be such that

$$\tan\theta = \beta = \frac{E_F Y^2 - E_F X^2 + \left[\left(E_F Y^2 - E_F X^2\right)^2 + 4\left(E_F XY\right)^2\right]^{1/2}}{2E_F XY}.$$

Note that $(w(\beta), v(\beta))^t = A(X, Y)^t$, where

$$A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

It is easy to check that

$$\tan 2\theta = \frac{2EXY}{EX^2 - EY^2} = \frac{2\beta}{1 - \beta^2}. \tag{3.7}$$

By elementary calculations, $E(w, v) = (0, 0)$ and

$$\begin{aligned} Ewv &= \tfrac{1}{2}(EY^2 - EX^2)\sin 2\theta + EXY\cos 2\theta \\ &= \tfrac{1}{2}(\sigma_Y^2 - \sigma_X^2)\sin 2\theta + \sigma_{XY}\cos 2\theta. \end{aligned} \tag{3.8}$$

By (3.7) and (3.8), $Ewv = 0$ and since $Ev = 0$, $\sigma_{wv} = 0$; thus w and v are independent. We wish to show that

$$\int \eta[w(\beta), v(\beta)]w(\beta)\, dF_\beta(X, Y) = 0,$$

that is, that

$$\int \eta[w(\beta), v(\beta)]w(\beta)\, dH_\beta(w, v) = 0, \tag{3.9}$$

where H_β is the joint distribution of w and v at the model. Since w and v are independent, the left-hand side of (3.9) becomes

$$\int \left[\int \eta[w(\beta), v(\beta)] \, dH_2(v) \right] w(\beta) \, dH_1(w),$$

where H_1 and H_2 are the marginal distributions of w and v, respectively. The result follows from the assumptions. □

Unfortunately the proof of Theorem 3.1 requires that w and v uncorrelated implies w and v independent. This and the calculations in Section 4 are the reasons for the Gaussian assumption in this paper. It may be possible to add side conditions in order to obtain the uniqueness of the estimates required for this theorem. The theorem can easily be generalized to higher dimensions. Note that the above requirement that $\eta(w, \cdot\,)$ be odd guarantees $E_v \eta(w, v) = 0$.

Observe that M and Q and thus V depend on β as well as η. This is a major difference from ordinary linear models where M, Q and V do not depend on β [see Hampel, Ronchetti, Rousseeuw and Stahel (1986), Chapter 6]. For simplicity, write $M(\beta)$, $Q(\beta)$ and $V(\beta)$.

4. Bounded influence robust estimators. We adapt the notation of Hampel, Ronchetti, Rousseeuw and Stahel (1986).

DEFINITION. The unstandardized gross-error sensitivity of an estimator (functional) T at the model distribution F_β is

$$\gamma_u^* = \sup_{X, Y} \left| \mathrm{IF}(X, Y; T, F_\beta) \right| = \sup_{w, v} \left| \eta(w, v) w M^{-1}(\beta) \right|.$$

DEFINITION. If $V(T, F_\beta)$ exists, the self-standardized sensitivity is defined by

$$\gamma_s^* = \sup_{X, Y} \left[\mathrm{IF}^2(X, Y; T, F_\beta) V(T, F_\beta)^{-1} \right]^{1/2} = \sup_{w, v} \left| \eta(w, v) w \right| Q^{-1/2}(\beta)$$

if $V(T, F_\beta)$ is nonzero. Otherwise, $\gamma_s^* = \infty$.

Let $\{\eta\}$ denote all η satisfying the conditions in Section 3. We now derive the most B-robust estimator in $\{\eta\}$, that is, the estimator with the smallest sensitivity γ^*.

DEFINITION. T defined by (3.2) is B_u-robust (B_s-robust) if and only if γ_u^* (γ_s^*) is finite.

LEMMA 4.1. *At the model F_β,*

$$\frac{\partial \sigma_w}{\partial \theta} = 0, \qquad \frac{\partial \sigma_v}{\partial \theta} = 0 \tag{4.1}$$

and

$$\frac{\partial}{\partial\beta}\ln f_\beta(X,Y) = a(\beta)wv, \tag{4.2}$$

where f_β is the (Gaussian) density of F_β and

$$a(\beta) = \frac{1}{1+\beta^2}\frac{\sigma_w^2-\sigma_v^2}{\sigma_w^2\sigma_v^2} = \frac{1}{1+\beta^2}\frac{\left[\left(\sigma_X^2-\sigma_Y^2\right)^2+4\sigma_{XY}^2\right]^{1/2}}{\sigma_X^2\sigma_Y^2-\sigma_{XY}^2}. \tag{4.3}$$

PROOF. Algebraic manipulation yields

$$\sigma_w^2 = \sigma_X^2\cos^2\theta + \sigma_{XY}\sin 2\theta + \sigma_Y^2\sin^2\theta, \tag{4.4}$$

$$\sigma_v^2 = \sigma_X^2\sin^2\theta + \sigma_{XY}\sin 2\theta + \sigma_Y^2\cos^2\theta. \tag{4.5}$$

Differentiating (4.4) and (4.5) and using (3.8) yields (4.1). Geometrically, we rotate (X,Y) to the principal components (w,v), where σ_w^2 and σ_v^2 are the extreme values.

Next, let h_β be the density of (w,v), then

$$f_\beta(X,Y) = \frac{1}{|\det A^{-1}|}h_\beta\left[\begin{pmatrix}X\\Y\end{pmatrix}\right]. \tag{4.6}$$

Since A is orthogonal, $|\det A^{-1}| = 1$. Because we have centered the data, $EX = 0$ and (w,v) is Gaussian with mean $(0,0)$ and covariance matrix $\begin{bmatrix}\sigma_w^2 & 0\\ 0 & \sigma_v^2\end{bmatrix}$. Thus (4.6) becomes

$$f_\beta(X,Y) = \frac{1}{2\pi\sigma_w\sigma_v}\exp\left\{-\frac{1}{2}\left(\frac{(X\cos\theta+Y\sin\theta)^2}{\sigma_w^2} + \frac{(-X\sin\theta+Y\cos\theta)^2}{\sigma_v^2}\right)\right\}.$$

Observe that

$$\frac{\partial w}{\partial\theta} = v, \qquad \frac{\partial v}{\partial\theta} = -w, \qquad \frac{\partial}{\partial\beta}\ln f_\beta(X,Y) = \frac{\partial}{\partial\theta}\ln f_\beta(X,Y)\frac{\partial\theta}{\partial\beta}.$$

A direct calculation yields (4.2). Finally, tedious calculations shows that

$$\sigma_w^2\sigma_v^2 = \sigma_X^2\sigma_Y^2 - \sigma_{XY}^2 \quad\text{and}$$

$$\sigma_w^2 - \sigma_v^2 = \left[\left(\sigma_X^2-\sigma_Y^2\right)^2 + 4\sigma_{XY}^2\right]^{1/2} = (1+\beta^2)\sigma_x^2. \qquad \square$$

Specializing (3.3) using results for Fisher consistent estimators [see, e.g., Hampel, Ronchetti, Rousseeuw and Stahel (1986), page 102] we have at the model distribution

$$(4.7)\qquad M(\beta) = \int \eta[w(\beta), v(\beta)]\, w(\beta) \frac{\partial}{\partial \beta} \ln f_\beta(X,Y)\, dF_\beta(X,Y),$$

where f_β is the density of F_β. By the lemma, (4.7) becomes

$$(4.8)\qquad M(\beta) = a(\beta) \int \eta(w,v) w^2 v\, dH(w,v).$$

The following two theorems give lower bounds for the unstandardized gross-error and self-standardized sensitivities and estimators that attain these lower bounds.

THEOREM 4.2. *Always*

$$(4.9)\qquad \gamma_u^* \ge d_u(\beta) \equiv \frac{\pi \sigma_w \sigma_v (1+\beta^2)}{2(\sigma_w^2 - \sigma_v^2)} = \frac{\pi\left(\sigma_X^2 \sigma_Y^2 - \sigma_{XY}^2\right)^{1/2} (1+\beta^2)}{2\left[\left(\sigma_X^2 - \sigma_Y^2\right)^2 + 4\sigma_{XY}^2\right]^{1/2}}$$

and $\eta(w,v) = \operatorname{sign}(v)/|w|$ *attains this lower bound.*

PROOF. By (4.8), $M = a(\beta) E\eta(w,v) w^2 v$, so

$$1 = M^{-1} a(\beta) E\eta(w,v) w^2 v \le a(\beta) E|\eta(w,v)| |M^{-1} w| |wv|$$
$$\le a(\beta) \gamma_u^* E|wv| = a(\beta) \gamma_u^* E|w| E|v|,$$

since w and v are independent. Moreover, $E|w| = 2\sigma_w/\sqrt{2\pi}$ and $E|v| = 2\sigma_v/\sqrt{2\pi}$ because w and v are Gaussian with mean 0. A simple calculation yields (4.9).

When $\eta(w,v) = \operatorname{sign}(v)/|w|$ (observe that this is Fisher consistent since η is odd in v), $M = a(\beta) E w^2 v \operatorname{sign}(v)/|w| = a(\beta) E|wv|$ and

$$\gamma_u^* = \sup_{w,v} |\eta(w,v)| |M^{-1} w| = \sup_{w,v} \frac{1}{|w|} |M^{-1} w| = d_u(\beta). \qquad \square$$

THEOREM 4.3. *Always* $\gamma_s^* \ge 1$ *and* $\eta(w,v) = \operatorname{sign}(v)/|w|$ *attains this lower bound.*

PROOF. By definition $\gamma_s^{*2} \ge \eta^2(w,v) w^2 Q^{-1}$. Integrate this inequality to see that $\gamma_s^{*2} \ge 1$. A simple calculation shows that when $\eta(w,v) = \operatorname{sign}(v)/|w|$, $\gamma_s^* = 1$. $\square$

The estimator η given in Theorems 4.2 and 4.3 is called the most B_u-robust (B_s-robust) estimator and for a given data set, $\{(X_i, Y_i),\ i = 1, \dots, n\}$ is the solution of

$$(4.10)\qquad \sum_{i=1}^{n} \frac{\operatorname{sign}(Y_i - \beta X_i)}{|X_i + \beta Y_i|} (X_i + \beta Y_i) = 0.$$

It is interesting to compare the most B-robust estimator in ordinary linear models, which is

$$\sum_{i=1}^{n} \frac{\operatorname{sign}(Y_i - \beta X_i)}{|X_i|} X_i = 0. \tag{4.11}$$

This has solution, $\beta = \operatorname{median}\{Y_i/X_i\}$ [see Krasker (1980)]. The solution of (4.10) is similar. Rotate coordinates so that the regression line is the new X-axis, then the solution to (4.10) is $\beta = \operatorname{median}\{Y_i^*/X_i^*\}$, where (X_i^*, Y_i^*) are the rotated observations. The estimator (4.11) has the property that it is optimal in the sense that it minimizes the maximum bias among all regression equivariant estimators over an ε-contaminated neighborhood [Martin, Yohai and Zamar (1989), Theorem 5.2]. We conjecture that if one uses this minimax approach in the EV model, a similar result is true for (4.10). If one is only interested in the smallest sensitivity for both the unstandardized and the self-standardized cases (or is working in an ε-contaminated neighborhood), (4.10) should be used. However, if optimality in Hampel's sense is required, then use the estimator of the next section.

5. Hampel's optimality problem. In this section the so-called Hampel's optimality problem is solved for the EV model. An estimator is constructed that has the following property: it minimizes the asymptotic variance at the model, subject to a bound on the supremum of the influence function (also at the model). It was first proposed and solved by Hampel (1968) for one-dimensional M-estimators. Huber [(1981), Chapter 11] has a discussion of this and obtains some further results. In Hampel, Ronchetti, Rousseeuw and Stahel [(1986), Chapter 4], this idea is extended to multidimensional M-estimators where it is shown that if such estimators exist, they must take a certain form but the existence remains unresolved in general. In ordinary linear models, the problem was first studied by Krasker (1980). Further results were obtained by Ronchetti and Rousseeuw (1985) with some restrictions on the distribution of the independent variables.

For the EV model, Hampel's optimality problem has a satisfactory solution for the univariate Gaussian case. The resulting estimator bears strong resemblance to that of ordinary linear models. It might be called the Hampel–Krasker estimator for EV models.

THEOREM 5.1. *The estimator within* $\{\eta\}$ *(defined in Section* 4*) which minimizes the asymptotic variance subject to a bound* $c(\beta)$ *on the unstandardized gross-error sensitivity exists and is unique if and only if* $c(\beta) > d_u(\beta)$. *This estimator is defined by*

$$\eta_{c(\beta)}(w, v) = \frac{\psi_{c(\beta)}(a(\beta)b(\beta)|w|v)}{b(\beta)|w|},$$

where $\psi_c(t) = \operatorname{sign}(t)\min[|t|, c]$ *is Huber's function,* $a(\beta)$ *is given in* (4.3) *and*

$b(\beta) > 0$ *is defined implicitly by*

$$a(\beta)E\psi_{c(\beta)}(a(\beta)b(\beta)|w|v)|w|v = 1.$$

PROOF. Fix β and write $a(\beta)$, $b(\beta)$ and $c(\beta)$ as a, b and c, respectively. Define, for all $b > 0$, $f(b) = E\psi_c(ab|w|v)a|w|v$, then

$$f'(b) = E\psi_c'(ab|w|v)a^2w^2v^2 > 0.$$

Observe that

$$\psi_c(ab|w|v) = \begin{cases} ab|w|v, & \text{if } -c/b \leq a|w|v \leq c/b, \\ c, & \text{if } a|w|v > c/b, \\ -c, & \text{if } a|w|v < -c/b. \end{cases}$$

Hence $\lim_{b \to 0^+} f(b) = 0$. Furthermore,

$$\lim_{b \to \infty} \psi_c(ab|w|v) = \begin{cases} c, & \text{if } v > 0, \\ -c, & \text{if } v < 0, \end{cases}$$

which implies that $\lim_{b \to \infty} f(b) = Eca|wv|$. Now consider $f(b) = 1$. The above argument shows that if $c > 1/Ea|wv|$, there is a solution and if $c < 1/Ea|wv|$, there is no solution. Moreover, $f'(b) > 0$ means that f is strictly increasing and the solution is unique if it exists.

The conditions required by Theorem 1, Hampel, Ronchetti, Rousseeuw and Stahel [(1986), page 241] are satisfied in this case. This gives the optimality of $\eta_{c(\beta)}$. $\square$

Note that $b(\beta)$ is simply $1/M(\beta)$. An immediate corollary to this theorem is:

COROLLARY 5.2. *If* β *is fixed, then* $\gamma_u^*(\eta_c) = c$.

The major difference between the Hampel–Krasker estimators for ordinary linear models and for EV models is that the former is defined independently of β while the latter is defined pointwise by β. Compare Theorem 5.1 with Proposition 2 of Hampel, Ronchetti, Rousseeuw and Stahel [(1986), page 319]. This dependence on the unknown parameters is not uncommon in the general M-estimation problem [see, e.g., Section 2.4 of Hampel, Ronchetti, Rousseeuw and Stahel (1986)]. The existence and uniqueness of the η-functions which define the optimal estimators do not guarantee the existence and uniqueness of a value β of such an estimator for a given sample or distribution.

6. Conclusion. This paper derives bounded influence robust estimates for the slope and intercept parameters in the univariate, Gaussian, structural EV model with the ratio of the error variances known. For the slope parameter, Fisher consistency, sharp lower bounds on the gross-error and self-standardized sensitivities and Hampel's optimal estimator are obtained. The Gaussian assumption is crucial to these derivations. Outstanding problems in this area include the removal of the Gaussian assumption; the establishment of conditions for the uniqueness and weak continuity of the implicitly defined functional T; the study of bounded influence estimation in the case where one error variance is known instead of the ratio λ; and the extension of the results to the multivariate case.

Acknowledgments. We wish to thank the referees and an Associate Editor for their detailed comments and many helpful suggestions. We also thank Dr. Paul Whitney for his valuable comments. The first author gratefully acknowledges a research award from the National Science Council of Taiwan, Republic of China.

REFERENCES

AMMANN, L. and VAN NESS, J. (1988). A routine for converting regression algorithms into corresponding orthogonal regression algorithms. *ACM Trans. Math. Software* **14** 76–87.

AMMANN, L. and VAN NESS, J. (1989). Standard and robust orthogonal regression. *Comm. Statist. Simulation Comput.* **18** 145–162.

ANDERSON, T. W. (1984). Estimating linear statistical relationships. *Ann. Statist.* **12** 1–45.

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

CARROLL, R. J. and GALLO, P. P. (1982). Some aspects of robustness in the functional errors-in-variables model. *Comm. Statist. Theory Methods* **11** 2573–2585.

CHENG, C. L. and VAN NESS, J. W. (1990). Bounded influence errors-in-variables regression. In *Statistical Analysis of Measurement Error Models and Applications. Contemp. Math.* **112** 227–242 Amer. Math. Soc., Providence, R.I.

CLARKE, B. R. (1983). Uniqueness and Frechet differentiability of functional solutions to maximum likelihood type equations. *Ann. Statist.* **11** 1196–1205.

FULLER, W. A. (1980). Properties of some estimators for the errors-in-variables model. *Ann. Statist.* **8** 407–422.

FULLER, W. A. (1987). *Measurement Error Models.* Wiley, New York.

GLESER, L. J. (1981). Estimation in a multivariate "errors in variables" regression model: Large sample results. *Ann. Statist.* **9** 24–44.

HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Ph.D dissertation, Univ. California, Berkeley.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics.* Wiley, New York.

HUBER. P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.

HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

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

KENDALL, M. and STUART, A. (1979). *The Advanced Theory of Statistics* **2**, 4th ed. Charles Griffin, London.

KRASKER, W. S. (1980). Estimation in linear regression models with disparate data points. *Econometrica* **48** 1333–1346.

LINDLEY, D. B. (1947). Regression lines and the functional relationship. *J. Roy. Statist. Soc. Suppl.* **9** 219–244.

MADANSKY, A. (1959). The fitting of straight lines when both variables are subject to error. *J. Amer. Statist. Assoc.* **54** 173–205.

MARONNA, R. A., BUSTOS, O. H. and YOHAI, V. J. (1979). Bias- and efficiency-robustness of general M-estimators for regression with random carriers. *Smoothing Techniques for Curve Estimation. Lecture Notes in Math.* **757** 91–116. Springer, Berlin.

MARONNA, R. A. and YOHAI, V. J. (1981). Asymptotic behavior of general M-estimates for regression and scale with random carriers. *Z. Wahrsch. Verw. Gebiete* **58** 7–20.

MARTIN, R. D., YOHAI, V. J. and ZAMAR, R. H. (1989). Min-max bias robust regression. *Ann. Statist.* **17** 1608–1630.

RONCHETTI, E. and ROUSSEEUW, P. J. (1985). Change-of-variance sensitivities in regression analysis. *Z. Wahrsch. Verw. Gebiete* **68** 503–519.

ZAMAR, R. H. (1985). Orthogonal regression M-estimators. Ph.D dissertation, Dept. Statistics, Univ. Washington, Seattle.

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

INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
TAIPEI, TAIWAN
REPUBLIC OF CHINA

PROGRAM IN MATHEMATICAL SCIENCES
UNIVERSITY OF TEXAS AT DALLAS
P.O. BOX 830688
RICHARDSON, TEXAS 75083-0688