The Annals of Statistics
1988, Vol. 16, No. 1, 147–160

ESTIMATION FOR THE NONLINEAR FUNCTIONAL RELATIONSHIP

BY YASUO AMEMIYA AND WAYNE A. FULLER

Iowa State University

Estimation of the parameters of the nonlinear functional model with known error covariance matrix is discussed. Asymptotic properties of the maximum likelihood estimator for the implicit functional model are presented. The approximate bias in the maximum likelihood estimator due to the nonlinearity of the relationship is given and a bias-adjusted estimator is suggested. Numerical and theoretical results support the superiority of the bias-adjusted estimator relative to the maximum likelihood estimator.

1. Introduction. In the functional relationship model, the true values are assumed to satisfy a given mathematical relationship. The observations are the sums of true values and measurement errors. The model is also called the functional errors-in-variables model. An extensive literature exists for estimation of parameters in linear functional relationship models. Reviews of the literature are contained in Madansky (1959), Moran (1971), Kendall and Stuart (1979), Fuller (1980), Gleser (1981) and Anderson (1984).

To define the nonlinear functional relationship, let $\{b_n\}_{n=1}^{\infty}$ and $\{a_n\}_{n=1}^{\infty}$ be sequences of positive real numbers such that $n = b_n a_n$ for all n and assume that a sequence of experiments indexed by n exists. The explicit functional relationship is

$$y_t^0 = g(\mathbf{x}_t^0; \boldsymbol{\beta}^0), \qquad t = 1, 2, \dots, b_n, \tag{1.1}$$

where $\mathbf{x}_t^0$ are $1 \times (p-1)$ unobservable fixed vectors belonging to a subset of $(p-1)$-dimensional Euclidean space and $\boldsymbol{\beta}^0$ is a $1 \times k$ vector of unknown parameters belonging to a subset of k-dimensional Euclidean space. Throughout the paper all vectors are row vectors. If $g(\mathbf{x}_t^0; \boldsymbol{\beta}^0)$ is nonlinear in $\mathbf{x}_t^0$ for fixed $\boldsymbol{\beta}^0$ or nonlinear in $\boldsymbol{\beta}^0$ for fixed $\mathbf{x}_t^0$, we say the model is nonlinear. A more general model is the implicit functional relationship

$$f(\mathbf{z}_t^0; \boldsymbol{\beta}^0) = 0, \qquad t = 1, 2, \dots, b_n, \tag{1.2}$$

where $\mathbf{z}_t^0$ are unobservable fixed vectors belonging to a parameter space Γ, $\boldsymbol{\beta}^0$ is a $1 \times k$ vector of unknown parameters belonging to a parameter space Ω, $f(\mathbf{z}; \boldsymbol{\beta})$ is a function defined on $\Gamma \times \Omega$ and Γ and Ω are subsets of p-dimensional and k-dimensional Euclidean spaces, respectively. The model is nonlinear if $f(\mathbf{z}_t^0; \boldsymbol{\beta}^0)$ is nonlinear in the sense defined for $g(\mathbf{x}_t^0; \boldsymbol{\beta}^0)$. Clearly, (1.1) is a special case of (1.2) with $f(\mathbf{z}_t^0; \boldsymbol{\beta}^0) = y_t^0 - g(\mathbf{x}_t^0; \boldsymbol{\beta}^0)$, where $\mathbf{z}_t^0 = (y_t^0; \mathbf{x}_t^0)$. Hence, we concentrate our discussion on the implicit functional relationship (1.2).

Received March 1985; revised March 1987.
AMS 1980 *subject classifications.* Primary 62J02; secondary 62F12, 62H12.
Key words and phrases. Nonlinear implicit relationship, measurement errors, asymptotic bias, maximum likelihood estimator, bias-adjusted estimator.

For each $\boldsymbol{\beta}$ in Ω, let Γ_β denote the set of $\mathbf{z}$ in Γ satisfying $f(\mathbf{z}; \boldsymbol{\beta}) = 0$. We assume that for each $\boldsymbol{\beta}$ in Ω, the set Γ_β is nonempty. The observations are the p-dimensional row vectors

$$(1.3) \qquad \mathbf{Z}_{nt} = \mathbf{z}_t^0 + \boldsymbol{\varepsilon}_{nt}, \qquad t = 1, 2, \ldots, b_n,$$

where $\boldsymbol{\varepsilon}_{nt}$ are the vectors of measurement errors. It is assumed that the $\boldsymbol{\varepsilon}_{nt}$ are independently distributed with mean zero and covariance matrix $\boldsymbol{\Sigma}_n = a_n^{-1}\boldsymbol{\Phi}$, where $\boldsymbol{\Phi}$ is a fixed positive definite matrix. The covariance matrix $\boldsymbol{\Sigma}_n$ and the b_n observations are assumed to be known. Often $\boldsymbol{\Sigma}_n$ will have been estimated from replicate observations. An extension to the case where $\boldsymbol{\Sigma}_n = \sigma_n^2\boldsymbol{\Phi}$ with known $\boldsymbol{\Phi}$ and unknown σ_n^2 is discussed briefly at the end of Section 3.

We will derive the limiting properties of the estimators under the assumption that $n \to \infty$. If $b_n \to \infty$ as $n \to \infty$, then the number of data points tends to infinity. If $a_n \to \infty$ as $n \to \infty$, then the measurement error variances approach zero as $n \to \infty$. To avoid uninteresting cases, we assume that a_n and b_n are nondecreasing in n. For example, if a_n replicate observations are made at each of b_n points, then the total number of observations is $n = b_n a_n$ and the vector $\mathbf{Z}_{nt}$ used in the analysis is the mean of the a_n replicates. In practice, the limiting result obtained as $n \to \infty$ with both a_n and b_n increasing can be used as an approximation when the error variances are small and the number of data points is large.

Most research on estimation for the nonlinear functional relationship model has concentrated on the explicit model (1.1). Estimation procedures for $\boldsymbol{\beta}^0$ in (1.1) based on the least squares principle or on the maximum likelihood principle have been suggested by Deming (1931, 1943), Cook (1931), Clutton-Brock (1967), Dolby (1972), Dolby and Lipton (1972) and Egerton and Laycock (1979). Estimation of specific nonlinear models has been discussed by Hey and Hey (1960), McDonald (1962), O'Neill, Sinclair and Smith (1969), Griliches and Ringstad (1970), Wolter and Fuller (1982a) and Amemiya (1985). A Bayesian treatment has been given by Reilly and Patino-Leal (1981). For the logistic regression where the covariate is measured with error, Stefanski and Carroll (1985) derived properties of the ordinary logistic regression estimator, the maximum likelihood estimator and two types of bias-adjusted estimators.

For the explicit model (1.1), Villegas (1969) and Wolter and Fuller (1982b) considered a one-step iterative estimator of $\boldsymbol{\beta}^0$ based on a preliminary estimator. Villegas derived the limiting distribution of the estimator for $b_n = O(1)$ and $\boldsymbol{\varepsilon}_{nt}$ normally distributed. Wolter and Fuller showed that the one-step estimator has a limiting normal distribution under the assumption that $a_n^{-1} = o(n^{-1/2})$. They also proposed a modified one-step estimator which has a limiting normal distribution when $a_n^{-1} = o(n^{-1/3})$.

For the implicit nonlinear functional relationship model (1.2), Britt and Luecke (1973) proposed an algorithm for estimating $\boldsymbol{\beta}^0$ and $\mathbf{z}_t^0$. The algorithm has been implemented and expanded by Schnell (1984). Chan (1965) and Anderson (1981) considered the circular errors-in-variables model.

We consider the estimation of $\boldsymbol{\beta}^0$ and $\mathbf{z}_t^0$ for the implicit model (1.2) and give asymptotic properties of the maximum likelihood estimators of $\boldsymbol{\beta}^0$ and $\mathbf{z}_t^0$.

Bias-adjusted estimators of $\boldsymbol{\beta}^0$ and $\mathbf{z}_t^0$ are proposed and their properties are discussed. Estimators for the parameters of the quadratic model and the circle model are compared in a Monte Carlo study. For the circle,

$$\left(y_t^0 - \mu_y^0\right)^2 + \left(x_t^0 - \mu_x^0\right)^2 - \left(\rho^0\right)^2 = 0,$$

where $\boldsymbol{\beta}^0 = (\mu_y^0, \mu_x^0, \rho^0)$ and $\mathbf{Z}_{nt} = (Y_{nt}, X_{nt}) = (y_t^0, x_t^0) + \varepsilon_{nt}$. The adjusted estimator of the radius displayed better properties than the maximum likelihood estimator. The adjusted estimators also performed better than maximum likelihood for the quadratic model.

2. The maximum likelihood estimator. In this section, we consider the maximum likelihood estimators of $\mathbf{z}_t^0$ and $\boldsymbol{\beta}^0$ for the nonlinear implicit model (1.2) with normally distributed ε_{nt}. The maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{z}}_t$ are the values of $\boldsymbol{\beta}$ in Ω and $\mathbf{z}_t$ in Γ that minimize

$$\sum_{t=1}^{b_n} (\mathbf{Z}_{nt} - \mathbf{z}_t)\Sigma_n^{-1}(\mathbf{Z}_{nt} - \mathbf{z}_t)' \tag{2.1}$$

subject to $f(\mathbf{z}_t; \boldsymbol{\beta}) = 0$, $t = 1, 2, \dots, b_n$. We assume that such $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{z}}_t$ exist and are measurable functions of $\mathbf{Z}_{nt}$ for sufficiently large n. Throughout, we denote the Euclidean norm of a vector $\mathbf{a}$ by $|\mathbf{a}|$.

For sequences in which the error variances approach 0, $\hat{\boldsymbol{\beta}}$ is consistent.

LEMMA 1. *Let the model* (1.2) *and* (1.3) *hold and assume*:

(i) $a_n^{-1} = o(1)$.

(ii) *For each* n, *the measurement errors* ε_{nt}, $t = 1, 2, \dots, b_n$, *are independently distributed with mean zero and known covariance matrix* Σ_n, *where* $\Sigma_n = a_n^{-1}\Phi$ *and* Φ *is a fixed positive definite matrix.*

(iii) *For every* $\varepsilon > 0$, *there exist a* $\delta_\varepsilon > 0$ *and an* $N_\varepsilon > 0$ *such that if* $n \geq N_\varepsilon$,

$$Q_n(\boldsymbol{\beta}) = b_n^{-1} \sum_{t=1}^{b_n} \inf_{\mathbf{z} \in \Gamma_\beta} \left(\mathbf{z}_t^0 - \mathbf{z}\right)\Phi^{-1}\left(\mathbf{z}_t^0 - \mathbf{z}\right)' > \delta_\varepsilon$$

for every $\boldsymbol{\beta}$ *in* Ω *satisfying* $|\boldsymbol{\beta} - \boldsymbol{\beta}^0| > \varepsilon$, *where*

$$\Gamma_\beta = \{\mathbf{z} \text{ in } \Gamma;\ f(\mathbf{z}; \boldsymbol{\beta}) = 0\}.$$

Then

$$\operatorname*{plim}_{n\to\infty} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}^0.$$

PROOF. Let

$$P_n(\boldsymbol{\beta}) = b_n^{-1} \sum_{t=1}^{b_n} \inf_{\mathbf{z} \in \Gamma_\beta} (\mathbf{Z}_{nt} - \mathbf{z})\Phi^{-1}(\mathbf{Z}_{nt} - \mathbf{z})'.$$

Since $\hat{\boldsymbol\beta}$ minimizes $P_n(\boldsymbol\beta)$,

$$P_n(\hat{\boldsymbol\beta}) \le P_n(\boldsymbol\beta^0) \le b_n^{-1} \sum_{t=1}^{b_n} \boldsymbol\varepsilon_{nt}\boldsymbol\Phi^{-1}\boldsymbol\varepsilon_{nt}' = R_n,$$

say. Because for any $\mathbf{z}$,

$$(\mathbf{z}_t^0 - \mathbf{z})\boldsymbol\Phi^{-1}(\mathbf{z}_t^0 - \mathbf{z})' \le 2\left[(\mathbf{Z}_{nt} - \mathbf{z})\boldsymbol\Phi^{-1}(\mathbf{Z}_{nt} - \mathbf{z})' + \boldsymbol\varepsilon_{nt}\boldsymbol\Phi^{-1}\boldsymbol\varepsilon_{nt}'\right],$$

it follows that

$$Q_n(\hat{\boldsymbol\beta}) \le 2\left[P_n(\hat{\boldsymbol\beta}) + R_n\right] \le 4R_n.$$

By Markov's inequality, $R_n = O_p(a_n^{-1})$. Hence, the consistency of $\hat{\boldsymbol\beta}$ follows from assumptions (i) and (iii). □

Assumption (iii) is an identification condition for $\boldsymbol\beta^0$. Suppose that $|\boldsymbol\beta - \boldsymbol\beta^0| > \varepsilon$ and consider the average of b_n distances between $\mathbf{z}_t^0$ satisfying $f(\mathbf{z}_t^0; \boldsymbol\beta^0) = 0$ and the projection of $\mathbf{z}_t^0$ onto Γ_β in the metric of $\boldsymbol\Phi^{-1}$. Under assumption (iii), the average distance is greater than a positive constant for large b_n, and the difference between $\boldsymbol\beta$ and $\boldsymbol\beta^0$ can be detected.

If $b_n \to \infty$, there are an increasing number of $\mathbf{z}_t^0$ to estimate. However, the $\hat{\mathbf{z}}_t$ are uniformly consistent for $\mathbf{z}_t^0$ provided the error variances decrease faster than the number of data points increases.

LEMMA 2. *Let the model* (1.2) *and* (1.3) *and assumption* (ii) *of Lemma* 1 *hold. Also, assume*

(i′) $b_n a_n^{-1} = o(1)$.

Then, for any $\xi > 0$, *there exists an* N_ξ *such that if* $n > N_\xi$,

$$P\left\{|\hat{\mathbf{z}}_t - \mathbf{z}_t^0| \le \xi \text{ for all } t = 1, 2, \dots, b_n\right\} > 1 - \xi.$$

PROOF. By the proof of Lemma 1,

$$\sum_{t=1}^{b_n} (\mathbf{z}_t^0 - \hat{\mathbf{z}}_t)\boldsymbol\Phi^{-1}(\mathbf{z}_t^0 - \hat{\mathbf{z}}_t)' = b_n Q_n(\hat{\boldsymbol\beta}) = O_p(b_n a_n^{-1})$$

and the result follows from assumption (i′). □

Lemmas 1 and 2 do not require continuity of $f(\mathbf{z}; \boldsymbol\beta)$ and the only restriction on the parameter space is the identification assumption (iii) of Lemma 1.

In deriving the limiting distribution of $\hat{\boldsymbol\beta}$, it is assumed that $f(\mathbf{z}; \boldsymbol\beta)$ possesses continuous first and second derivatives with respect to both arguments on $\Gamma \times \Omega$. Let $\mathbf{f}_z(\mathbf{z}; \boldsymbol\beta)$ denote the $1 \times p$ vector of partial derivatives of $f(\mathbf{z}; \boldsymbol\beta)$ with respect to the elements of $\mathbf{z}$, let $\mathbf{f}_\beta(\mathbf{z}; \boldsymbol\beta)$ denote the $1 \times k$ vector of partial derivatives with respect to the elements of $\boldsymbol{\beta}$ and let $\mathbf{f}_{zz}(\mathbf{z}; \boldsymbol{\beta})$ denote the $p \times p$ matrix of second partial derivatives with respect to the elements of $\mathbf{z}$. We denote the partial derivatives evaluated at $(\mathbf{z}_t^0; \boldsymbol{\beta}^0)$ by a superscript 0 and a subscript t, i.e., $\mathbf{f}_{\beta t}^0 = \mathbf{f}_\beta(\mathbf{z}_t^0; \boldsymbol{\beta}^0)$, $\mathbf{f}_{zt}^0 = \mathbf{f}_z(\mathbf{z}_t^0; \boldsymbol{\beta}^0)$ and $\mathbf{f}_{zzt}^0 = \mathbf{f}_{zz}(\mathbf{z}_t^0; \boldsymbol{\beta}^0)$.

The following theorem shows that $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)$ has a limiting normal distribution with zero mean, provided $a_n \to \infty$ faster than b_n.

THEOREM 1. *Let the model* (1.2) *and* (1.3), *assumptions* (ii) *and* (iii) *of Lemma* 1 *and assumption* (i′) *of Lemma* 2 *hold. Also, assume*:

(iv) *The parameter* $\boldsymbol{\beta}^0$ *is an interior point of* Ω. *There exists a compact subset* Γ_0 *of* Γ *and an* $\eta > 0$ *such that a neighborhood of* $\mathbf{z}_t^0$ *with radius* η *is in* Γ_0 *for all* $t = 1, 2, \ldots, b_n$ *and all* n.

(v) *The partial derivatives through order* 2 *of* $f(\mathbf{z}; \boldsymbol{\beta})$ *exist and are continuous on* $\Gamma \times \Omega$.

(vi) *For all* $\mathbf{z}$ *in* Γ, $\mathbf{f}_z(\mathbf{z}; \boldsymbol{\beta}^0) \neq \mathbf{0}$.

(vii) $\lim_{n \to \infty} \mathbf{m} = \mathbf{M}$, *where*

$$\mathbf{m} = b_n^{-1} \sum_{t=1}^{b_n} \phi_t^{-1} \mathbf{f}_{\beta t}^{0\prime} \mathbf{f}_{\beta t}^0,$$

$\mathbf{M}$ *is positive definite and* $\phi_t = \mathbf{f}_{zt}^0 \boldsymbol{\Phi} \mathbf{f}_{zt}^{0\prime}$.

(viii) *The* $2 + \delta$ *moments of* $a_n^{1/2} \boldsymbol{\varepsilon}_{nt}$ *are bounded for some* $\delta > 0$.

(ix) $b_n^{-1} = o(1)$.

Then, as $n \to \infty$,

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \to_L N(\mathbf{0}, \mathbf{M}^{-1}).$$

PROOF. Because $\boldsymbol{\beta}^0$ is an interior point of Ω, there exists a compact ball Ω_1 about $\boldsymbol{\beta}^0$ in the interior of Ω such that on $\Gamma_0 \times \Omega_1$, the partial derivatives through order 2 of $f(\mathbf{z}; \boldsymbol{\beta})$ are uniformly continuous and bounded and

$$\mathbf{f}_z(\mathbf{z}; \boldsymbol{\beta}) \boldsymbol{\Phi} \mathbf{f}_z'(\mathbf{z}; \boldsymbol{\beta}) > K_0 > 0. \tag{2.2}$$

Thus, by Lemmas 1 and 2, with probability approaching 1, all partial derivatives through order 2 of $f(\mathbf{z}; \boldsymbol{\beta})$ evaluated at $(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}})$ or at any point on the line segment joining $(\mathbf{z}_t^0; \boldsymbol{\beta}^0)$ and $(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}})$ are bounded and satisfy (2.2).

Consider the Lagrangian

$$\sum_{t=1}^{b_n} \left\{ \tfrac{1}{2}(\mathbf{Z}_{nt} - \mathbf{z}_t) \boldsymbol{\Phi}^{-1} (\mathbf{Z}_{nt} - \mathbf{z}_t)' + \alpha_t f(\mathbf{z}_t; \boldsymbol{\beta}) \right\}, \tag{2.3}$$

where α_t are Lagrange multipliers. Assuming that $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{z}}_t$ are interior points of Ω and Γ, respectively, the partial derivative equations of (2.3) evaluated at $\hat{\boldsymbol{\beta}}$ and

$\hat{\mathbf{z}}_t$ are

$$-\boldsymbol{\Phi}^{-1}(\mathbf{Z}_{nt}-\hat{\mathbf{z}}_t)' + \alpha_t \mathbf{f}_z'(\hat{\mathbf{z}}_t;\hat{\boldsymbol{\beta}}) = 0, \qquad t = 1,2,\dots,b_n, \tag{2.4}$$

$$\sum_{t=1}^{b_n} \alpha_t \mathbf{f}_\beta(\hat{\mathbf{z}}_t;\hat{\boldsymbol{\beta}}) = 0, \tag{2.5}$$

$$f(\hat{\mathbf{z}}_t;\hat{\boldsymbol{\beta}}) = 0, \qquad t = 1,2,\dots,b_n. \tag{2.6}$$

By assumption (iv) and the consistency of the estimators, the probability that $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{z}}_t$ satisfy (2.4)–(2.6) tends to 1 as $n \to \infty$.

If we expand $f(\hat{\mathbf{z}}_t;\hat{\boldsymbol{\beta}})$ around $(\mathbf{z}_t^0;\boldsymbol{\beta}^0)$, expression (2.6) becomes

$$\mathbf{f}_z(\mathbf{z}_t^*;\boldsymbol{\beta}_t^*)(\hat{\mathbf{z}}_t-\mathbf{z}_t^0)' + \mathbf{f}_\beta(\mathbf{z}_t^*;\boldsymbol{\beta}_t^*)(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0)' = 0, \qquad t = 1,2,\dots,b_n, \tag{2.7}$$

where $(\mathbf{z}_t^*;\boldsymbol{\beta}_t^*)$ is on the line segment joining $(\mathbf{z}_t^0;\boldsymbol{\beta}^0)$ and $(\hat{\mathbf{z}}_t;\hat{\boldsymbol{\beta}})$. Solving (2.4), (2.5) and (2.7), we obtain

$$(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0)\mathbf{m}^* = -b_n^{-1}\sum_{t=1}^{b_n}\mathbf{f}_z(\mathbf{z}_t^*;\boldsymbol{\beta}_t^*)\boldsymbol{\varepsilon}_{nt}'\phi_t^{*-1}\mathbf{f}_\beta(\hat{\mathbf{z}}_t;\hat{\boldsymbol{\beta}}),$$

where $\phi_t^* = \mathbf{f}_z(\mathbf{z}_t^*;\boldsymbol{\beta}_t^*)\boldsymbol{\Phi}\mathbf{f}_z'(\hat{\mathbf{z}}_t;\hat{\boldsymbol{\beta}})$ and

$$\mathbf{m}^* = b_n^{-1}\sum_{t=1}^{b_n}\mathbf{f}_\beta'(\mathbf{z}_t^*;\boldsymbol{\beta}^*)\phi_t^{*-1}\mathbf{f}_\beta(\hat{\mathbf{z}}_t;\hat{\boldsymbol{\beta}}_t).$$

It follows from the consistency of $\hat{\mathbf{z}}_t$ and of $\hat{\boldsymbol{\beta}}$, the boundedness of the derivatives and (2.2) that $\mathbf{m}^* = \mathbf{m} + O_p(a_n^{-1/2})$. By assumption (ii), the boundedness of the derivatives, (2.2) and the consistency of the estimators, we have

$$\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0 = \Delta_1 + O_p(a_n^{-1}), \tag{2.8}$$

where

$$\Delta_1 = -b_n^{-1}\sum_{t=1}^{b_n} v_t\phi_t^{-1}\mathbf{f}_{\beta t}^0\mathbf{m}^{-1}, \qquad v_t = \mathbf{f}_{zt}^0\boldsymbol{\varepsilon}_{nt}'.$$

Because $E\{\Delta_1'\Delta_1\} = O(n^{-1})$, $\Delta_1 = O_p(n^{-1/2})$. By assumption (i′), $a_n^{-1} = o(n^{-1/2})$. The result follows from assumptions (vii) and (viii) and the Liapounov central limit theorem because, by assumption (ix), $b_n \to \infty$ as $n \to \infty$. □

Assumption (viii) can be replaced by the assumption of identically distributed $\boldsymbol{\varepsilon}_{nt}$. Assumption (ix) guarantees that $b_n \to \infty$ so that a central limit theorem can be applied. The limiting normal distribution of $\hat{\boldsymbol{\beta}}$ in Theorem 1 is valid for constant b_n if the $\boldsymbol{\varepsilon}_{nt}$ are normally distributed.

In Theorem 1, the assumption that $a_n^{-1} \to 0$ faster than $n^{-1/2}$ permitted terms of $O_p(a_n^{-1})$ to be ignored. The terms of $O_p(a_n^{-1})$ involve functions of $\mathbf{f}_{zz}(\mathbf{z};\boldsymbol{\beta})$, the curvature of $f(\mathbf{z};\boldsymbol{\beta})$. To study the effect of the curvature on the estimators, we obtain higher order expansions of $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{z}}_t$. Let assumptions (i′), (ii)–(iv), (vi) and (vii) hold. Also, assume:

(v′) The partial derivatives through order 3 of $f(\mathbf{z}; \boldsymbol{\beta})$ exist and are continuous on $\Gamma \times \Omega$.

(viii′) The fourth moments of $a_n^{1/2}\boldsymbol{\varepsilon}_{nt}$ are bounded.

Then, it can be shown that

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = \boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2 + O_p\left(n^{-1/2}a_n^{-1/2}\right), \tag{2.9}$$

$$\hat{\mathbf{z}}_t - \mathbf{z}_t^0 = \mathbf{d}_{1t} + \mathbf{d}_{2t} + O_p\left(n^{-1/2}a_n^{-1/2}\right), \tag{2.10}$$

where

$$(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2) = -b_n^{-1} \sum_{t=1}^{b_n} (v_t, c_t)\phi_t^{-1}\mathbf{f}_{\beta t}^0 \mathbf{m}^{-1},$$

$$\mathbf{d}_{1t} = \boldsymbol{\varepsilon}_{nt}\left[\mathbf{I} - \mathbf{f}_{zt}^{0\prime}\phi_t^{-1}\mathbf{f}_{zt}^0\boldsymbol{\Phi}\right],$$

$$\mathbf{d}_{2t} = -\left[c_t + \mathbf{f}_{\beta t}^0(\boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_2)'\right]\phi_t^{-1}\mathbf{f}_{zt}^0\boldsymbol{\Phi} - a_n v_t \mathbf{d}_{1t}\phi_t^{-1}\mathbf{f}_{zzt}^0\mathbf{V}_t,$$

$$v_t = \mathbf{f}_{zt}^0\boldsymbol{\varepsilon}_{nt}', \qquad c_t = \tfrac{1}{2}\mathbf{d}_{1t}\mathbf{f}_{zzt}^0\mathbf{d}_{1t}',$$

$$\mathbf{V}_t = E\{\mathbf{d}_{1t}'\mathbf{d}_{1t}\} = a_n^{-1}\left[\boldsymbol{\Phi} - \boldsymbol{\Phi}\mathbf{f}_{zt}^{0\prime}\phi_t^{-1}\mathbf{f}_{zt}^0\boldsymbol{\Phi}\right]$$

and ϕ_t and $\mathbf{m}$ are defined in assumption (vii). The leading terms in the expansions, $\boldsymbol{\Delta}_1$ and $\mathbf{d}_{1t}$, are $O_p(n^{-1/2})$ and $O_p(a_n^{-1/2})$, respectively, and both have zero expectation. The second terms in the expansions are both functions of c_t, where c_t is a quadratic form in the original errors $\boldsymbol{\varepsilon}_{nt}$ and depends on the matrix of second derivatives. The expectations of the second terms are

$$E\{\boldsymbol{\Delta}_2\} = -b_n^{-1} \sum_{t=1}^{b_n} B_t\phi_t^{-1}\mathbf{f}_{\beta t}^0\mathbf{m}^{-1} = O\left(a_n^{-1}\right), \tag{2.11}$$

$$E\{\mathbf{d}_{2t}\} = -\left[B_t + \mathbf{f}_{\beta t}^0 E\{\boldsymbol{\Delta}_2'\}\right]\phi_t^{-1}\mathbf{f}_{zt}^0\boldsymbol{\Phi} = O\left(a_n^{-1}\right), \tag{2.12}$$

where $B_t = E\{c_t\} = \frac{1}{2}\operatorname{tr}[\mathbf{f}_{zzt}^0\mathbf{V}_t]$. These results are derived in Amemiya and Fuller (1985).

The expectation in (2.11) is a weighted average of B_t, $t = 1, 2, \ldots, b_n$, while the expectation in (2.12) is a linear function of B_t and $E\{\boldsymbol{\Delta}_2\}$. Each B_t can be considered to be the contribution of the estimation variance of $\mathbf{z}_t^0$ to the bias approximations. The tth bias contribution is small when the elements of $\mathbf{f}_{zzt}^0$ are small or when the elements of the covariance matrix $\mathbf{V}_t$ of the leading term in the expansion of the error made in estimating $\mathbf{z}_t^0$ are small. If the relationship $f(\mathbf{z}; \boldsymbol{\beta})$ is linear in $\mathbf{z}$, the expectations (2.11) and (2.12) are zero.

3. Bias-adjusted estimator. We use expansions (2.9) and (2.10) to develop modifications of the maximum likelihood estimators. To motivate the modification, recall that the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{z}}_t$ were obtained by minimizing (2.1) subject to

$$f(\mathbf{z}_t; \boldsymbol{\beta}) = 0, \qquad t = 1, 2, \ldots, b_n. \tag{3.1}$$

The terms in (2.11) and (2.12) with nonzero expectations are functions of c_t and c_t is a term in the expansion of $f(\hat{\mathbf{z}}_t, \hat{\boldsymbol{\beta}})$,

$$0 = f(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}}) = \mathbf{f}_{zt}^0(\hat{\mathbf{z}}_t - \mathbf{z}_t^0)' + \mathbf{f}_{\beta t}^0(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)' + c_t + O_p(n^{-1/2}a_n^{-1/2}).$$

Because $E\{c_t\} = B_t$, replacing the restriction (3.1) by a restriction adjusted with an estimator of B_t should produce estimators of $\boldsymbol{\beta}^0$ and $\mathbf{z}_t^0$ whose expansions have smaller biases than those of the maximum likelihood estimators. Hence, we suggest that estimators of $\boldsymbol{\beta}$ and $\mathbf{z}_t$ be chosen to minimize (2.1) subject to

$$f(\mathbf{z}_t; \boldsymbol{\beta}) - \tfrac{1}{2}\operatorname{tr}\left[\mathbf{f}_{zz}(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}})\hat{\mathbf{V}}_t\right] = 0, \qquad t = 1, 2, \ldots, b_n, \tag{3.2}$$

where

$$\hat{\mathbf{V}}_t = \Sigma_n - \Sigma_n \mathbf{f}_z'(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}})\left[\mathbf{f}_z(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}})\Sigma_n \mathbf{f}_z'(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}})\right]^{-1}\mathbf{f}_z(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}})\Sigma_n.$$

To simplify our discussion, we consider one-step linearized improved estimators of $\boldsymbol{\beta}^0$ and $\mathbf{z}_t^0$ using the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{z}}_t$ as preliminary estimators. Let $\Delta\hat{\boldsymbol{\beta}}$ and $\Delta\hat{\mathbf{z}}_t$ be the values of $\Delta\boldsymbol{\beta}$ and $\Delta\mathbf{z}_t$ that minimize

$$\sum_{t=1}^{b_n} (\mathbf{Z}_{nt} - \hat{\mathbf{z}}_t - \Delta\mathbf{z}_t)\Sigma_n^{-1}(\mathbf{Z}_{nt} - \hat{\mathbf{z}}_t - \Delta\mathbf{z}_t)'$$

subject to the restrictions

$$\Delta\boldsymbol{\beta}\mathbf{f}_\beta'(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}}) + \Delta\mathbf{z}_t\mathbf{f}_z'(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}}) - \tfrac{1}{2}\operatorname{tr}\left[\mathbf{f}_{zz}(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}})\hat{\mathbf{V}}_t\right] = 0,$$

for $t = 1, 2, \ldots, b_n$, where the restrictions are linear approximations to restrictions (3.2). Then the one-step bias-adjusted estimators $\tilde{\mathbf{z}}_t$ and $\tilde{\boldsymbol{\beta}}$ are

$$\tilde{\mathbf{z}}_t = \hat{\mathbf{z}}_t + \Delta\hat{\mathbf{z}}_t \quad \text{and} \quad \tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + \Delta\hat{\boldsymbol{\beta}}. \tag{3.3}$$

It is understood that $\tilde{\mathbf{z}}_t$ and $\tilde{\boldsymbol{\beta}}$ are replaced by projections onto Γ and Ω, respectively, whenever $\tilde{\mathbf{z}}_t$ is outside Γ or $\tilde{\boldsymbol{\beta}}$ is outside Ω. It can be shown that $(-\Delta\hat{\mathbf{z}}_t)$ and $(-\Delta\hat{\boldsymbol{\beta}})$ are estimators of the expectations (2.11) and (2.12).

It is possible to use $\tilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{z}}_t$ as new preliminary estimators and to iterate the procedure. The iterative procedure leads to estimators of $\boldsymbol{\beta}^0$ and $\mathbf{z}_t^0$ that minimize (2.1) subject to (3.2).

The second order expansions of the one-step estimators $\tilde{\mathbf{z}}_t$ and $\tilde{\boldsymbol{\beta}}$ are given in Theorem 2. The theorem also holds for estimators constructed with a finite number of steps of the iterative procedure. The proof of Theorem 2 contains rather tedious Taylor expansions and is omitted. The proof is available in Amemiya and Fuller (1985).

THEOREM 2. *Let the model* (1.2) *and* (1.3) *with assumptions* (i′), (ii)–(iv), (v′), (vi), (vii) *and* (viii′) *hold. Then*

$$\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = \Delta_1 + O_p(n^{-1/2}a_n^{-1/2}),$$

$$\tilde{\mathbf{z}}_t - \mathbf{z}_t^0 = \mathbf{d}_{1t} + \mathbf{d}_{3t} + O_p(n^{-1/2}a_n^{-1/2}),$$

where

$$\mathbf{d}_{3t} = -\left[c_t - \mathbf{d}_{1t}\mathbf{f}^0_{zzt}\boldsymbol{\varepsilon}'_{nt} + B_t + \mathbf{f}^0_{\beta t}\Delta'_1\right]\phi_t^{-1}\mathbf{f}^0_{zt}\boldsymbol{\Phi} - a_n v_t \mathbf{d}_{1t}\phi_t^{-1}\mathbf{f}^0_{zztt}\mathbf{V}_t,$$

Δ_1, $\mathbf{d}_{1t}$, v_t, c_t *and* B_t *are defined in* (2.9)–(2.12) *and* $E\{\mathbf{d}_{3t}\} = \mathbf{0}$. *If, in addition, assumption* (ix) *of Theorem* 1 *holds, then*

$$n^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \to_L N(\mathbf{0}, \mathbf{M}^{-1}).$$

Theorem 2 shows that the term Δ_2 in the expansion of $\hat{\boldsymbol{\beta}}$ disappears in the expansion of $\tilde{\boldsymbol{\beta}}$. The expectations of the terms in the expansions of $\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0$ and $\tilde{\mathbf{z}}_t - \mathbf{z}^0_t$ up to $O_p(n^{-1/2}a_n^{-1/2})$ are zero.

In our discussion, we have assumed Σ_n to be known. The results can be extended to the case where $\Sigma_n = \sigma_n^2\Phi$, Φ is a known positive definite matrix, σ_n^2 is unknown and $\sigma_n^2 = a_n^{-1}\psi$ for some $\psi > 0$. For this model with the error covariance matrix known up to a multiple, the maximum likelihood estimators of $\boldsymbol{\beta}^0$ and $\mathbf{z}^0_t$ are equal to $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{z}}_t$ because replacing Σ_n with Φ in (2.1) leaves those estimators unchanged. The results in Lemmas 1 and 2, Theorem 1 and the expansions (2.9) and (2.10) remain valid for $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{z}}_t$ with an obvious modification in assumption (ii). An estimator of σ_n^2 is

$$\hat{\sigma}_n^2 = (b_n - k)^{-1}\sum_{t=1}^{b_n}(\mathbf{Z}_{nt} - \hat{\mathbf{z}}_t)\Phi^{-1}(\mathbf{Z}_{nt} - \hat{\mathbf{z}}_t)'.$$

Under assumption (i′), the distribution of $(b_n - k)\sigma_n^{-2}\hat{\sigma}_n^2$ can be approximated by that of a chi-square random variable with $b_n - k$ degrees of freedom. Bias-adjusted estimators of $\boldsymbol{\beta}^0$ and $\mathbf{z}^0_t$ of the form (3.3) can be obtained by replacing Σ_n in the definition of $\hat{\mathbf{V}}_n$ in (3.2) with $\hat{\sigma}_n^2\Phi$. Rough calculations suggest that the expansions of the bias-adjusted estimators with estimated σ_n^2 are those of Theorem 2. Monte Carlo results support this conjecture but also show that the estimation of σ_n^2 reduces the small sample effectiveness of the bias adjustment.

4. Monte Carlo studies. In this section, we present two Monte Carlo studies of the modification of the maximum likelihood estimator suggested in Section 3. For these examples, the bias-adjusted estimator has small sample properties that are superior to those of maximum likelihood.

In the first study, the model is the quadratic

$$y_t - \beta_0 - \beta_1 x_t^2 = 0, \qquad \mathbf{Z}_t = (Y_t, X_t) = (y_t, x_t) + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}'_t \sim NI(0, \Sigma)$ and $\Sigma = 0.0625\mathbf{I}$. In this section, we omit the superscript 0 and the subscript n because no confusion will result.

We generated 200 samples of size 50. Each sample of size 50 was generated by creating five $\mathbf{Z}_t = (Y_t, X_t)$ vectors for each of the ten x-values $\{-1.35, -1.05, -0.75, \ldots, 1.05, 1.35\}$. The parameters (β_0, β_1) were set equal to $(0, 1)$ in the data generation. For each sample of 50 observations, the maximum likelihood estimate $(\hat{\beta}_0, \hat{\beta}_1)$ and the bias-adjusted estimate $(\tilde{\beta}_0, \tilde{\beta}_1)$ were computed by minimizing (2.1) subject to (3.1) and (3.2), respectively. In constructing the estimates, the covariance matrix $\Sigma = 0.0625\mathbf{I}$ was treated as known.

TABLE 1
Empirical properties of alternative estimators for the quadratic model

	β_0		β_1	
Percentile	**ML**	**Adj.**	**ML**	**Adj.**
5	−0.209	−0.152	0.849	0.826
10	−0.165	−0.110	0.873	0.849
25	−0.121	−0.066	0.952	0.920
50	−0.069	−0.014	1.033	0.998
75	−0.010	0.040	1.140	1.096
90	0.036	0.084	1.232	1.190
95	0.057	0.108	1.313	1.244
Mean	−0.0685	−0.0152	1.055	1.019
Variance	0.0065	0.0060	0.022	0.019
MSE	0.0112	0.0062	0.025	0.020

By Theorem 1, (2.9) and (2.11), the theoretical asymptotic mean and covariance matrix of the approximate distribution of the maximum likelihood estimator $(\hat{\beta}_0, \hat{\beta}_1)$ are

$$E\{(\hat{\beta}_0, \hat{\beta}_1)\} = (0,1) + E\{\Delta_2\} = (-0.0500, 1.0323),$$

$$V\{(\hat{\beta}_0, \hat{\beta}_1)\} = n^{-1}\mathbf{m}^{-1} = \sum_{t=1}^{b_n} \mathbf{f}_{\beta t}^{0\prime}(\mathbf{f}_{zt}^{0}\Sigma_n \mathbf{f}_{zt}^{0\prime})^{-1}\mathbf{f}_{\beta t}^{0} \tag{4.1}$$

$$= \begin{pmatrix} 0.0046 & -0.0045 \\ -0.0045 & 0.0127 \end{pmatrix}.$$

Using the results for the one-step version of the bias-adjusted estimator given in Theorem 2, the theoretical asymptotic mean and covariance matrix of $(\tilde{\beta}_0, \tilde{\beta}_1)$ are (0, 1) and the matrix in (4.1), respectively.

Characteristics of the empirical distribution of the estimators for 200 samples are given in Table 1. The Monte Carlo variances are larger than the theoretical approximations for both parameters. We have not shown that the estimators have finite variances and there are theoretical arguments to the contrary. However, there were no outlying estimates in our 200 samples and the ratios of the empirical interquartile range to the standard error for the adjusted estimators $\tilde{\beta}_0$ and $\tilde{\beta}_1$ are 1.39 and 1.27. The observed ratios are close to the value of 1.35 that holds for the normal distribution.

For both estimators, the Monte Carlo biases are slightly larger in absolute value than the approximate theoretical biases. The empirical bias reduction due to the adjustment is (0.053, 0.036) and is roughly equal to the theoretical bias reduction (0.050, 0.032). The bias of the adjusted estimator $\tilde{\beta}_0$ is about three times its standard error and the bias of $\tilde{\beta}_1$ is about twice its standard error.

By almost any criterion, the adjusted estimators display properties superior to those of the maximum likelihood estimator. It is somewhat surprising that the adjustment reduced both the variance and the absolute value of the bias relative to maximum likelihood.

For each sample, we computed Studentized statistics $\hat{t}_i$ and $\tilde{t}_i$, where $\hat{t}_i$ and $\tilde{t}_i$ are the differences $\hat{\beta}_i - \beta_i$ and $\tilde{\beta}_i - \beta_i$ divided by the square roots of the corresponding estimated variances for $i = 0, 1$. By analogy to variance estimators for the linear model [see, for example, Fuller (1980)] and based on arguments presented in Wolter and Fuller (1982b), we estimated the covariance matrix of the approximate distribution of the maximum likelihood estimator $(\hat{\beta}_0, \hat{\beta}_1)$ by the formula

$$\hat{V}\{(\hat{\beta}_0, \hat{\beta}_1)\} = \hat{\mathbf{m}}^{-1}\hat{\mathbf{G}}\hat{\mathbf{m}}^{-1}, \tag{4.2}$$

where

$$\hat{\mathbf{G}} = b_n^{-1} \sum_{t=1}^{b_n} \mathbf{f}_\beta'(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}})\hat{\gamma}_t^{-1}\mathbf{f}_\beta(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}}),$$

$$\hat{\mathbf{m}} = \hat{\mathbf{G}} - b_n^{-1} \sum_{t=1}^{b_n} \mathbf{f}_{\beta z}(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}})\hat{\mathbf{V}}_t\mathbf{f}_{\beta z}'(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}})\hat{\gamma}_t^{-1},$$

$$\hat{\gamma}_t = \mathbf{f}_z(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}})\boldsymbol{\Sigma}_n\mathbf{f}_z'(\hat{\mathbf{z}}_t; \hat{\boldsymbol{\beta}}),$$

$\hat{\mathbf{V}}_t$ is defined in (3.2) and $\mathbf{f}_{\beta z}(\mathbf{z}; \boldsymbol{\beta})$ is the $k \times p$ matrix of second partial derivatives of $f(\mathbf{z}; \boldsymbol{\beta})$ with respect to the elements of $\boldsymbol{\beta}$ and $\mathbf{z}$. For the adjusted estimator $(\tilde{\beta}_0, \tilde{\beta}_1)$, the covariance matrix of the approximate distribution was estimated by the formula (4.2) with the adjusted estimators $\tilde{\mathbf{z}}_t$ and $\tilde{\boldsymbol{\beta}}$ replacing $\hat{\mathbf{z}}_t$ and $\hat{\boldsymbol{\beta}}$. The empirical percentiles of $\hat{t}_i$ and $\tilde{t}_i$ are given in Table 2. Because the maximum likelihood estimator is seriously biased, the use of $\hat{t}_i$ to construct confidence intervals would lead to seriously biased results. On the other hand, the distributions of $\tilde{t}_i$ for the adjusted estimators, while slightly skewed, are in reasonable agreement with the percentiles of the standard normal distribution.

The second model for which the estimators were studied is that of a circle

$$(y_t - \mu_y)^2 + (x_t - \mu_x)^2 - \rho^2 = 0,$$

$$\mathbf{Z}_t = (Y_t, X_t) = (y_t, x_t) + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t \sim NI(\mathbf{0}, \mathbf{I})$. Samples of size 50 were created by generating 5 observations for the 10 points on the circle associated with angles of $0, 25, 50, \ldots, 175, 200, 250°$.

TABLE 2
Empirical percentiles of Studentized statistics

	t_0		t_1		
Percentile	**ML**	**Adj.**	**ML**	**Adj.**	**Normal**
5	−2.44	−1.76	−1.54	−1.83	−1.64
10	−2.07	−1.38	−1.21	−1.48	−1.28
25	−1.49	−0.84	−0.41	−0.69	−0.68
50	−0.88	−0.19	0.26	−0.01	0
75	−0.13	0.56	0.96	0.68	0.68
90	0.55	1.22	1.46	1.20	1.28
95	0.84	1.58	1.77	1.51	1.64

TABLE 3
Empirical properties of alternative estimators for the circle model

	μ_x		ρ	
Percentile	**ML**	**Adj.**	**ML**	**Adj.**
5	−0.365	−0.366	3.360	3.195
10	−0.318	−0.319	3.418	3.256
25	−0.149	−0.152	3.515	3.358
50	−0.015	−0.013	3.626	3.474
75	0.101	0.099	3.735	3.589
90	0.217	0.214	3.842	3.698
95	0.317	0.315	3.903	3.762
Mean	−0.0241	−0.0260	3.625	3.472
Variance	0.0421	0.0419	0.0274	0.0297
MSE	0.0427	0.0426	0.0430	0.0305

The center of the circle is $(\mu_y, \mu_x) = (0, 0)$ and the radius is $\rho = 3.5$. A total of 200 samples were generated and the maximum likelihood and the bias-adjusted estimates of (μ_y, μ_x, ρ) were constructed for each sample. The covariance matrix of ε_t was treated as known in the estimation.

Empirical properties of the distributions of estimators of μ_x and ρ are given in Table 3. The properties of the estimators of μ_y were very similar to those of the estimators of μ_x. The approximate theoretical bias of the maximum likelihood estimator of (μ_y, μ_x, ρ) is $(0, 0, 0.143)$. Thus, the bias adjustment does not alter the theoretical properties of the maximum likelihood estimator of μ_x and this is reflected in the similar empirical properties of $\hat{\mu}_x$ and $\tilde{\mu}_x$. The absolute value of the empirical bias is only slightly greater than its standard error for both $\hat{\mu}_x$ and $\tilde{\mu}_x$.

The observed bias in the maximum likelihood estimator of ρ is about 1.5 standard errors smaller than the theoretical bias. The reduction in bias in the estimator of ρ associated with the adjustment is about equal to the approximate

TABLE 4
Empirical percentiles of Studentized statistics

	t: $\mu_x = 0$		t: $\rho = 3.5$		
Percentile	**ML**	**Adj.**	**ML**	**Adj.**	**Normal**
5	−1.74	−1.75	−0.89	−1.84	−1.65
10	−1.43	−1.44	−0.51	−1.42	−1.28
25	−0.67	−0.69	0.09	−0.82	−0.68
50	−0.06	−0.06	0.73	−0.15	0
75	0.48	0.48	1.37	0.51	0.68
90	1.03	1.02	1.96	1.06	1.28
95	1.47	1.46	2.29	1.44	1.65

theoretical bias of $\hat{\rho}$. The adjustment provides a definite improvement in the mean square error properties of the estimator of ρ.

The use of the bias-adjusted estimator and the modified covariance matrix estimator of the form (4.2) gives a Studentized statistic whose empirical distribution can be well approximated by the standard normal distribution. See Table 4.

Acknowledgments. This research was partly supported by Joint Statistical Agreements JSA 84-2 and JSA 85-1 with the U.S. Bureau of the Census. We thank Dan Schnell for the computations of the Monte Carlo study and anonymous referees for useful comments.

REFERENCES

AMEMIYA, Y. (1985). Instrumental variable estimator for the nonlinear errors-in-variables model. *J. Econometrics* **28** 273–289.

AMEMIYA, Y. and FULLER, W. A. (1985). Estimation for the nonlinear functional relationship. Report to the U.S. Bureau of the Census on JSA 85-1. Iowa State Univ., Ames, Ia.

ANDERSON, D. A. (1981). The circular structural model. *J. Roy. Statist. Soc. Ser. B* **43** 131–141.

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

BRITT, H. L. and LUECKE, R. H. (1973). The estimation of parameters in nonlinear implicit models. *Technometrics* **15** 233–247.

CHAN, N. N. (1965). On circular functional relationships. *J. Roy. Statist. Soc. Ser. B* **27** 45–56.

CLUTTON-BROCK, M. (1967). Likelihood distributions for estimating functions when both variables are subject to error. *Technometrics* **9** 261–299.

COOK, W. R. (1931). On curve fitting by means of least squares. *Philos. Mag. Ser. 7* **12** 1025–1039.

DEMING, W. E. (1931). The application of least squares. *Philos. Mag. Ser. 7* **11** 146–158.

DEMING, W. E. (1943). *Statistical Adjustment of Data.* Wiley, New York.

DOLBY, G. R. (1972). Generalized least squares and maximum likelihood estimation of nonlinear functional relationships. *J. Roy. Statist. Soc. Ser. B* **34** 393–400.

DOLBY, G. R. and LIPTON, S. (1972). Maximum likelihood estimation of the general nonlinear functional relationship with replicated observation and correlated errors. *Biometrika* **59** 121–129.

EGERTON, M. F. and LAYCOCK, P. J. (1979). Maximum likelihood estimation of multivariate non-linear functional relationships. *Math. Operationsforsch. Statist. Ser. Statist.* **10** 273–280.

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

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

GRILICHES, Z. and RINGSTAD, V. (1970). Errors in variables bias in nonlinear contexts. *Econometrica* **38** 368–370.

HEY, E. N. and HEY, M. H. (1960). The statistical estimation of a rectangular hyperbola. *Biometrics* **16** 606–617.

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

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

MCDONALD, R. P. (1962). A general approach to nonlinear factor analysis. *Psychometrika* **27** 397–415.

MORAN, P. A. P. (1971). Estimating structural and functional relationships. *J. Multivariate Anal.* **1** 232–255.

O'NEILL, M., SINCLAIR, L. G. and SMITH, F. J. (1969). Polynomial curve fitting when abscissas and ordinates are both subject to error. *Comput. J.* **12** 52–56.

REILLY, P. M. and PATINO-LEAL, H. (1981). A Bayesian study of the error-in-variables model. *Technometrics* **23** 221–231.

SCHNELL, D. (1984). Maximum likelihood estimation of parameters in the implicit nonlinear errors-in-variables model. Creative component for the M.S. degree, Iowa State Univ., Ames, Ia.

STEFANSKI, L. A. and CARROLL, R. J. (1985). Covariate measurement error in logistic regression. *Ann. Statist.* **13** 1335–1351.

VILLEGAS, C. (1969). On the least squares estimation of non-linear relations. *Ann. Math. Statist.* **40** 462–466.

WOLTER, K. M. and FULLER, W. A. (1982a). Estimation of the quadratic errors-in-variables model. *Biometrika* **69** 175–182.

WOLTER, K. M. and FULLER, W. A. (1982b). Estimation of nonlinear errors-in-variables models. *Ann. Statist.* **10** 539–548.

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50011