The Annals of Statistics
1982, Vol. 10, No. 2, 539–548

ESTIMATION OF NONLINEAR ERRORS-IN-VARIABLES MODELS

BY KIRK M. WOLTER AND WAYNE A. FULLER

U. S. Bureau of the Census and Iowa State University

An estimation procedure is presented for the coefficients of the nonlinear functional relation, where observations are subject to measurement error. The distributional properties of the estimators are derived, and a consistent estimator of the covariance matrix is given. In deriving the results it is assumed that the covariance matrix of the observational errors is known and that this covariance matrix is $o(n^{-1/3})$, where n is the index of the sequence of estimators.

1. Introduction. The term *errors-in-variables* refers to the general class of statistical models in which the "true" values of a set of variables satisfy a given mathematical relationship. The observable variables are the sum of the "true" values and errors of measurement. Given the observations, the parameters of the relationship are to be estimated. Such models occur in both the physical sciences and the social sciences. The models were popular in economics in the 1930's and have recently enjoyed a resurgence of popularity; cf. Griliches (1974). For other applications, see Bohrnstedt and Carter (1971) and Wolter and Fuller (1975). A wide area of application arises from the fact that almost all survey data collected by personal interview are subject to errors of measurement; cf. Dalenius (1977).

To define the model, let $(\Omega, \mathscr{B}, \mathscr{P})$ be a probability space, and 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 $n = 1, 2, \cdots, \infty$. We assume the existence of a sequence of experiments indexed by n. The functional relationship is

$$y_t^0 = g(\mathbf{x}_t^0; \boldsymbol{\beta}^0), \tag{1.1}$$

and $(Y_{nt}, \mathbf{X}_{nt} : t = 1, 2, \cdots, b_n)$ are observed in the nth experiment, where

$$Y_{nt} = y_t^0 + e_{nt}, \quad \mathbf{X}_{nt} = \mathbf{x}_t^0 + \mathbf{u}_{nt}.$$

The $p \times 1$ parameter vector $\boldsymbol{\beta}^0$ is an interior point of $\boldsymbol{\Theta}$, a compact, convex subset of p dimensional Euclidian space $\mathscr{R}^p$; $\{\mathbf{x}_t^0\}$ is a sequence of fixed, $1 \times q$ vectors; the vectors $\mathbf{x}_t^0$ are members of the set $\mathscr{A} \subset \mathscr{R}^q$; and $g : \mathscr{R}^{p+q} \to \mathscr{R}^1$ is a real valued continuous function. The random variables $(e_{nt}, \mathbf{u}_{nt})$, defined on $(\Omega, \mathscr{B}, \mathscr{P})$, denote errors of measurement. The unknown vector $\boldsymbol{\beta}^0$ is to be estimated. Model (1.1) is nonlinear if $g(\mathbf{x}; \boldsymbol{\beta})$ is nonlinear in either $\mathbf{x}$ or $\boldsymbol{\beta}$.

Algorithms for estimating $\boldsymbol{\beta}^0$ for nonlinear g have been given by Deming (1931, 1943), Cook (1931), Dolby and Lipton (1972), Dolby (1972), Britt and Luecke (1973), and Clutton-Brock (1967). Estimation of specific nonlinear models has been considered by O'Neill *et al.* (1969), Hey and Hey (1960), Chan (1965), Kendall and Stuart (1961), and Griliches and Ringstad (1970).

Anderson (1951) obtained the maximum likelihood estimator of $\boldsymbol{\beta}^0$ for the linear model with $b_n \equiv b$ and $a_n^{-1} = O(n^{-1})$ and proved the asymptotic normality of the estimator. Asymptotic normality of the maximum likelihood estimator for $\boldsymbol{\beta}^0$ for the linear model has been established by Fuller (1980) for the univariate model and by Gleser (1981) for the multivariate model under the assumption that $a_n = 1$ and $b_n = n$.

Received July 1977; revised October 1981.

AMS 1970 *subject classifications.* Primary 62J05; secondary 62H25.

Key words and phrases. Errors-in-variables, nonlinear models, asymptotically normally distributed, measurement errors, functional relationship.

Villegas (1969) considered the nonlinear model with $b_n \equiv k$ and $a_n = n$. This model is statistically equivalent to an experiment with n observations at each of k true values of $\mathbf{x}$. For these conditions, Villegas presented an estimator of $\boldsymbol{\beta}^0$ whose error, normalized by $n^{1/2}$, is asymptotically normal with zero mean.

In this article, alternative estimators for the parameters of the nonlinear functional model are constructed and the statistical properties are examined under less restrictive assumptions than those of Villegas. An estimator is presented that is asymptotically normal under the assumption that $a_n^{-1} = o(n^{-1/3})$. The main distributional results are stated in Section 2. A modification that improves the small sample behavior of the estimators is given in Section 3. A Monte Carlo comparison of the various estimators is presented in Section 4. Section 5 contains the proofs of the theorems.

2. An iterative estimator of $\boldsymbol{\beta}^0$. The estimator developed in this section is a modification of the maximum likelihood estimator for the model with normal measurement error. The statistical properties of the estimator will be derived for a somewhat wider class of distributions. In deriving our results, we assume that the covariance matrix of $(e_{nt}, \mathbf{u}_{nt})$, denoted by $\boldsymbol{\Sigma}_n$, is nonsingular and of order a_n^{-1}. One way of interpretating this assumption is to let a_n denote the number of observations made at each point $(y_t^0, \mathbf{x}_t^0)$. Under this interpretation, the total number of observations is $b_n a_n = n$ and each of the vectors $(Y_{nt}, \mathbf{X}_{nt})$ used in the analysis is the mean of a_n replicates.

We assume $g(\mathbf{x}; \boldsymbol{\beta})$ possesses continuous first and second derivatives with respect to both arguments on $\mathscr{A} \times \ddot{\boldsymbol{\Theta}}$. Let $g_x(\mathbf{z}; \boldsymbol{\theta})$ denote the q-dimensional row vector of partial derivatives of $g(\mathbf{x}; \boldsymbol{\beta})$ with respect to the elements of $\mathbf{x}$ evaluated at $(\mathbf{z}; \boldsymbol{\theta})$; let $\mathbf{g}_\beta(\mathbf{z}; \boldsymbol{\theta})$ denote the p-dimensional column vector of partial derivatives with respect to the elements of $\boldsymbol{\beta}$ evaluated at $(\mathbf{z}; \boldsymbol{\theta})$; let $\mathbf{g}_{xx}(\mathbf{z}; \boldsymbol{\theta})$ denote the $q \times q$ matrix of second partial derivatives with respect to the elements of $\mathbf{x}$ evaluated at $(\mathbf{z}; \boldsymbol{\theta})$; and let $\mathbf{g}_{\beta x}(\mathbf{z}; \boldsymbol{\theta})$ denote the $p \times q$ matrix of second partial derivatives with respect to the elements of $\boldsymbol{\beta}$ and $\mathbf{x}$, evaluated at $(\mathbf{z}; \boldsymbol{\theta})$.

Assume $(e_{nt}, \mathbf{u}_{nt})$ to be independent, normal $(\mathbf{0}, \boldsymbol{\Sigma}_n)$ random variables, where $\boldsymbol{\Sigma}_n$ is known. Then the maximum likelihood estimators (MLE) are those values of $\mathbf{x}_t$ and $\boldsymbol{\beta}$ contained in $\mathscr{A} \times \boldsymbol{\Theta}$ that minimize the sum of squares

$$(2.1)\quad \sum_{t=1}^{b_n} q(\boldsymbol{\beta}, \mathbf{x}_t; Y_{nt}, \mathbf{X}_{nt}) = \sum_{t=1}^{b_n} \{Y_{nt} - g(\mathbf{x}_t; \boldsymbol{\beta}), \mathbf{X}_{nt} - \mathbf{x}_t\} \boldsymbol{\Sigma}_n^{-1} \{Y_{nt} - g(\mathbf{x}_t; \boldsymbol{\beta}), \mathbf{X}_{nt} - \mathbf{x}_t\}'.$$

To condense the notation, we henceforth suppress the subscript n, when no confusion will result.

While an explicit expression for the MLE of $\boldsymbol{\beta}^0$ has not been obtained, one can construct an iterative procedure leading to an estimator of $\boldsymbol{\beta}^0$. To this end, let $\bar{\boldsymbol{\beta}}$ denote a preliminary estimator of $\boldsymbol{\beta}^0$. We suggest the ordinary least squares estimator, and this will be discussed further in Section 5. Let $\hat{\mathbf{x}}_t$ be the value of $\mathbf{x}_t$ contained in $\mathscr{A}$ that minimizes $q(\bar{\boldsymbol{\beta}}, \mathbf{x}_t; Y_t, \mathbf{X}_t)$. Assuming $\hat{\mathbf{x}}_t$ is in the interior of $\mathscr{A}$, $\hat{\mathbf{x}}_t$ satisfies

$$(2.2)\quad \sigma^{ee}\{Y_t - g(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\}\mathbf{g}_x(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}}) + \{Y_t - g(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\} \boldsymbol{\Sigma}^{eu} + \{\mathbf{X}_t - \hat{\mathbf{x}}_t\} \boldsymbol{\Sigma}^{ue} \mathbf{g}_x(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}}) + \{\mathbf{X}_t - \hat{\mathbf{x}}_t\} \boldsymbol{\Sigma}^{uu} = \mathbf{0},$$

where $\boldsymbol{\Sigma}^{-1}$ is partitioned as

$$\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \sigma^{ee} & \boldsymbol{\Sigma}^{eu} \\ \boldsymbol{\Sigma}^{ue} & \boldsymbol{\Sigma}^{uu} \end{bmatrix}.$$

To obtain a geometric interpretation of $\hat{\mathbf{x}}_t$ let $\boldsymbol{\Sigma} = \mathbf{I}$. Then $\hat{\mathbf{x}}_t$ is the $\mathbf{x}$-coordinate of the point on $g(\mathbf{x}; \bar{\boldsymbol{\beta}})$ that is the minimum Euclidian distance from $(Y_t, \mathbf{X}_t)$. If $\boldsymbol{\Sigma}$ is not the identity matrix, then the distance being minimized is a weighted distance.

By expanding $g(\mathbf{x}; \boldsymbol{\beta})$ in a Taylor series about the point $(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})$ and retaining only the linear terms, we obtain

$$(2.3)\quad \Delta y_t \doteq \mathbf{g}_\beta'(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})(\Delta\boldsymbol{\beta}) + \mathbf{g}_x(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})(\Delta\mathbf{x}_t)',$$

where $\hat{y}_t = g(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})$, $\Delta\boldsymbol{\beta} = \boldsymbol{\beta} - \bar{\boldsymbol{\beta}}$, $\Delta\mathbf{x}_t = \mathbf{x}_t - \hat{\mathbf{x}}_t$, and $\Delta y_t = y_t - \hat{y}_t$. Letting $\hat{e}_t = Y_t - \hat{y}_t$ and

$\hat{\mathbf{u}}_t = \mathbf{X}_t - \hat{\mathbf{x}}_t$, the local approximation to the sum of squares (2.1) is

$$\sum_{t=1}^{b} \{\hat{e}_t - \Delta y_t, \hat{\mathbf{u}}_t - \Delta \mathbf{x}_t\} \mathbf{\Sigma}^{-1} \{\hat{e}_t - \Delta y_t, \hat{\mathbf{u}}_t - \Delta \mathbf{x}_t\}'. \tag{2.4}$$

The value of $\Delta\boldsymbol{\beta}$, say $\widehat{\Delta\boldsymbol{\beta}}$, which minimizes (2.4) subject to (2.3) satisfies

$$\hat{\mathbf{M}}_{xx}(\widehat{\Delta\boldsymbol{\beta}}) = n^{-1} \sum_{t=1}^{b} \hat{\sigma}_{vt}^{-2}\, \mathbf{g}_{\beta}(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\{\hat{e}_t - \hat{\mathbf{u}}_t \mathbf{g}_x'(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\}, \tag{2.5}$$

where

$$\hat{\sigma}_{vt}^{2} = \{1, -\mathbf{g}_x(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\} \mathbf{\Sigma} \{1, -\mathbf{g}_x(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\}',$$

$$\hat{\mathbf{M}}_{xx} = n^{-1} \sum_{t=1}^{b} \hat{\sigma}_{vt}^{-2}\, \mathbf{g}_{\beta}(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}}) \mathbf{g}_{\beta}'(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}}).$$

Hence, an improved estimator of $\boldsymbol{\beta}^0$ is given by

$$\hat{\boldsymbol{\beta}} = \bar{\boldsymbol{\beta}} + \Delta\hat{\boldsymbol{\beta}}.$$

The construction of $\hat{\boldsymbol{\beta}}$ is based on a local quadratic approximation to (2.1). Assuming that $a_n^{-1} = o(n^{-1/2})$ and that $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = O_p(\max[a_n^{-1}, n^{-1/2}])$, it is demonstrated in Theorem 1 that $n^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)$ converges in distribution to a normal random variable.

We next develop a modified estimator whose limiting distribution can be established under the weaker assumption that $a_n^{-1} = o(n^{-1/3})$. The weaker assumption is useful because, interpreting a_n as the number of replicates, the modified estimator will be applicable in experiments with fewer replicates. That is, the modified estimator is applicable in situations where the variance of the measurement error is larger.

The modification was constructed after investigating the right side of Equation (2.5). If $a_n^{-1} = o(n^{-1/3})$, then

$$\hat{e}_t - \mathbf{g}_x(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\hat{\mathbf{u}}_t' - \tfrac{1}{2}\operatorname{tr}\{\mathbf{g}_{xx}(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})(\hat{\mathbf{u}}_t'\hat{\mathbf{u}}_t - \mathbf{\Sigma}_{uu})\} = \mathbf{g}_{\beta}'(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})(\Delta\boldsymbol{\beta}) + q_t + o_p(n^{-1/2}),$$

where

$$q_t = e_t - \mathbf{g}_x(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\mathbf{u}_t' - \tfrac{1}{2}\operatorname{tr}\{\mathbf{g}_{xx}(\mathbf{x}_t^0; \boldsymbol{\beta}^0)(\mathbf{u}_t'\mathbf{u}_t - \mathbf{\Sigma}_{uu})\}.$$

On the basis of this result, the modified estimator is

$$\tilde{\boldsymbol{\beta}} = \bar{\boldsymbol{\beta}} + \Delta\tilde{\boldsymbol{\beta}},$$

where $\Delta\tilde{\boldsymbol{\beta}}$ satisfies

$$\begin{aligned} \hat{\mathbf{M}}_{xx}(\Delta\tilde{\boldsymbol{\beta}}) &= n^{-1} \sum_{t=1}^{b} \hat{\sigma}_{vt}^{-2}\, \mathbf{g}_{\beta}(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\hat{q}_t, \\ \hat{q}_t &= \hat{e}_t - \hat{\mathbf{u}}_t \mathbf{g}_x'(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}}) - \tfrac{1}{2}\operatorname{tr}\{\mathbf{g}_{xx}(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})(\hat{\mathbf{u}}_t'\hat{\mathbf{u}}_t - \mathbf{\Sigma}_{uu})\}. \end{aligned} \tag{2.6}$$

The limiting distributions of the estimators are given in Theorems 1 and 2. The proofs are presented in Section 5.

THEOREM 1. *Let Model* (1.1) *hold and assume*:

(i) *The partial derivatives through order two of $g(\mathbf{x}; \boldsymbol{\beta})$ are continuous and bounded on $\mathscr{A} \times \mathbf{\Theta}$.*

(ii) *The preliminary estimator $\bar{\boldsymbol{\beta}} \in \mathbf{\Theta}$ satisfies $\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}^0 = O_p(\max[a_n^{-1}, n^{-1/2}])$.*

(iii) *The random variables $(e_{nt}, \mathbf{u}_{nt})$ are mutually independent $(\mathbf{0}, \mathbf{\Sigma}_n)$ random variables.*

(iv) *The error covariance matrices $\mathbf{\Sigma}_n$ are known and positive definite.*

(v) *The $p \times p$ matrix*

$$\mathbf{m}_{xx} = n^{-1} \sum_{t=1}^{b} \sigma_{vt}^{-2}\, \mathbf{g}_{\beta}(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\mathbf{g}_{\beta}'(\mathbf{x}_t^0; \boldsymbol{\beta}^0)$$

is positive definite for all $b \equiv b_n > p$, where

$$\sigma_{vt}^{2} = \{1, -\mathbf{g}_x(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\} \mathbf{\Sigma}_n \{1, -\mathbf{g}_x(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\}'.$$

(vi) $\lim_{n\to\infty} \mathbf{m}_{xx} = \bar{\mathbf{m}}_{xx}$ *exists and is positive definite.*

(vii) *The error covariance matrices satisfy*

$$\lim_{n\to\infty} a_n \mathbf{\Sigma}_n = \mathbf{\Phi},$$

where $\mathbf{\Phi}$ *is nonsingular, and the* $2 + \delta$ *moments of* $a_n^{1/2}(e_{nt}, \mathbf{u}_{nt})$ *are bounded for some* $\delta > 0$.

(viii) *The elements of the sequence* $\{a_n\}_{n=1}^{\infty}$ *satisfy* $a_n^{-1} = o(n^{-1/2})$.

Then,

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \to_{\mathscr{L}} N(\mathbf{0}, \bar{\mathbf{m}}_{xx}^{-1}). \qquad \square$$

THEOREM 2. *Let Model* (1.1) *and Assumptions* (ii) *through* (vi) *of Theorem* 1 *hold. Assume that* (ia) *the partial derivatives through order three of* $g(\mathbf{x}; \boldsymbol{\beta})$ *are continuous and bounded on* $\mathscr{A} \times \mathbf{\Theta}$; (viia) *the error covariance matrices satisfy* $\lim_{n\to\infty} a_n \mathbf{\Sigma}_n = \mathbf{\Phi}$, *where* $\mathbf{\Phi}$ *is nonsingular, and* $E\{|(e_{nt}; \mathbf{u}_{nt})|^4\} = La_n^{-2}$ *for some real L and all t and n*, (viiia) $\lim_{n\to\infty} n^{1/3} a_n^{-1} = 0$.
Then,

$$n^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \to_{\mathscr{L}} N(\mathbf{0}, \bar{\mathbf{m}}_{xx}^{-1}). \qquad \square$$

Assumption (v) is an assumption of convenience because Assumptions (vi) and (vii) guarantee positive definite matrices for sufficiently large n. Also, Assumption (iv) is an assumption of convenience because the data could be transformed to obtain the nonsingular model if $\mathbf{\Sigma}$ were singular.

3. Comments on the estimation procedure. The estimation procedure defined in Section 2 can be summarized as follows:
(a) Compute a preliminary estimator of $\boldsymbol{\beta}^0$, say $\bar{\boldsymbol{\beta}}$.
(b) For $t = 1, \cdots, b$, compute $\hat{\mathbf{x}}_t$ defined by (2.2).
(c) Compute $\tilde{\boldsymbol{\beta}} = \bar{\boldsymbol{\beta}} + \Delta\tilde{\boldsymbol{\beta}}$, where $\Delta\tilde{\boldsymbol{\beta}}$ is defined in (2.6).
This procedure can be iterated using $\tilde{\boldsymbol{\beta}}$ as the preliminary estimator in a second round of calculation. For any finite number of iterations, the asymptotic properties of the final estimator are given in Theorem 2. The maximum likelihood estimator can be obtained as the limit of the iteration by including a modification in the procedure to guarantee convergence.

For many g functions, obtaining a solution to (2.2) will be difficult. Expanding (2.2) in a Taylor series about the point $\mathbf{X}_t$ gives an approximation, $\ddot{\mathbf{x}}_t$, to $\hat{\mathbf{x}}_t$, where $\ddot{\mathbf{x}}_t$ satisfies the linear system

$$\begin{aligned} \mathbf{0} = {} & \sigma^{ee}\{Y_t - g(\mathbf{X}_t; \bar{\boldsymbol{\beta}})\}\{\mathbf{g}_x(\mathbf{X}_t; \bar{\boldsymbol{\beta}}) + \mathbf{\Sigma}^{eu}\} \\ & + (\ddot{\mathbf{x}}_t - \mathbf{X}_t)(\sigma^{ee}[-\mathbf{g}_x'(\mathbf{X}_t; \bar{\boldsymbol{\beta}})\mathbf{g}_x(\mathbf{X}_t; \bar{\boldsymbol{\beta}}) + \{Y_t - g(\mathbf{X}_t; \bar{\boldsymbol{\beta}})\}\mathbf{g}_{xx}(\mathbf{X}_t; \bar{\boldsymbol{\beta}})] \\ & - \mathbf{g}_x'(\mathbf{X}_t; \bar{\boldsymbol{\beta}})\mathbf{\Sigma}^{eu} - \mathbf{\Sigma}^{ue}\mathbf{g}_x(\mathbf{X}_t; \bar{\boldsymbol{\beta}}) - \mathbf{\Sigma}^{uu}). \end{aligned} \tag{3.1}$$

Under the assumptions of Section 2, $\ddot{\mathbf{x}}_t = \hat{\mathbf{x}}_t + O_p(a_n^{-1})$. Thus, $\tilde{\boldsymbol{\beta}}$ may be defined with $\ddot{\mathbf{x}}_t$ replacing $\hat{\mathbf{x}}_t$, and the limiting distribution of $\tilde{\boldsymbol{\beta}}$ will remain as given in Theorem 2. Likewise, in iterating the estimator, the second approximation to $\mathbf{x}_t^0$ can be computed using (3.1) with $\ddot{\mathbf{x}}_t$ replacing $\mathbf{X}_t$ as the initial estimate.

It is worth noting that our estimation procedure is applicable to an extremely broad class of problems. The procedure is applicable to models where some of the variables are measured without error. If $\mathbf{\Sigma}$ is singular, it is possible to define a linear transformation of $(Y_t, \mathbf{X}_t)$ such that some variables have zero error. The function $g(\cdot)$ is transformed accordingly. In this case, the rows and columns of $\mathbf{\Sigma}$ which correspond to the error-free variables are identically zero and the "estimated values" for such variables are equal to the observed values. Equation (2.2) or (3.1) is applied to that portion of the $\mathbf{x}$-vector measured with error, i.e. to that portion for which the error covariance matrix is nonsingular. If all of the "independent" variables are measured without error, i.e. $\mathbf{\Sigma}_{uu} = \mathbf{0}$, then the errors-in-variables model reduces to the customary nonlinear "regression" model. Implicit models of the form $g(\mathbf{x}_t^0; \boldsymbol{\beta}^0) = 0$ may be estimated by taking $y_t^0 \equiv 0$ and $\sigma_e^2 = 0$.

Preliminary Monte Carlo work demonstrated that the small sample distribution function of the maximum likelihood estimator (the estimator $\hat{\boldsymbol{\beta}}$ iterated to convergence) had "thick tails." This effect can be reduced by using an estimator similar to that studied by Fuller (1980), obtained by replacing $\hat{\mathbf{x}}_t$ with $\tilde{\mathbf{x}}_t$ in the construction of the estimator, where

(3.2) $$\tilde{\mathbf{x}}_t = (1 - b^{-1}\alpha)\hat{\mathbf{x}}_t + b^{-1}\alpha\mathbf{X}_t$$

and α is a fixed number $(1 \leq \alpha \leq 4)$.

Preliminary Monte Carlo work also indicated that $\hat{\mathbf{M}}_{xx}$ was often a poor estimator of the variance of $\tilde{\boldsymbol{\beta}}$. One source of bias in $\hat{\mathbf{M}}_{xx}$ as an estimator of $\bar{\mathbf{m}}_{xx}$ is due to the fact that the expectation of $(\hat{\mathbf{x}}_t'\hat{\mathbf{x}}_t)$ contains a term arising from the variance of $\hat{\mathbf{x}}_t$ as an estimator of $\mathbf{x}_t^0$. An additional bias is associated with the fact that $(Y_t, \mathbf{X}_t)$ is being projected onto a nonlinear surface. Expanding $\mathbf{g}_\beta(\hat{\mathbf{x}}_t; \tilde{\boldsymbol{\beta}})$ about $(\mathbf{x}_t^0; \boldsymbol{\beta}^0)$, we obtain

(3.3) $$\mathbf{g}_\beta(\hat{\mathbf{x}}_t; \tilde{\boldsymbol{\beta}}) = \mathbf{g}_\beta(\mathbf{x}_t^0; \boldsymbol{\beta}^0) + \mathbf{g}_{\beta\beta}(\hat{\mathbf{x}}_t; \boldsymbol{\beta}^\dagger)(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) + \mathbf{g}_{\beta x}(\mathbf{x}_t^0; \boldsymbol{\beta}^0)(\hat{\mathbf{x}}_t - \mathbf{x}_t^0)' + \mathbf{C},$$

where the ith element of $\mathbf{C}$ is

$$C_i = \tfrac{1}{2}(\hat{\mathbf{x}}_t - \mathbf{x}_t^0)\mathscr{R}_i(\mathbf{x}_t^\dagger; \boldsymbol{\beta}^0)(\hat{\mathbf{x}}_t - \mathbf{x}_t^0)',$$

the jkth element of $\mathscr{R}_i(\mathbf{x}_t^\dagger; \boldsymbol{\beta}^0)$ is

$$r_{ijk}(\mathbf{x}_t^\dagger; \boldsymbol{\beta}^0) = \frac{\partial^3 g(\mathbf{x}_t^\dagger; \boldsymbol{\beta}^0)}{\partial\beta_i \partial x_{tj} \partial x_{tk}},$$

$\mathbf{x}_t^\dagger$ is on the line segment joining $\mathbf{x}_t^0$ and $\hat{\mathbf{x}}_t$, and $\boldsymbol{\beta}^\dagger$ is on the line segment joining $\boldsymbol{\beta}^0$ and $\tilde{\boldsymbol{\beta}}$. Because $E\{|\hat{\mathbf{x}}_t - \mathbf{x}_t^0|^2\} = O(a_n^{-1})$, there is a bias in $\hat{\mathbf{M}}_{xx}$ of $O(a_n^{-1})$.

Equation (3.3) suggests that the variance of the limiting distribution of $n^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ be estimated by

(3.4) $$(\hat{\mathbf{M}}_{ww} - \mathbf{A})^{-1},$$

where $$\hat{\mathbf{M}}_{ww} = n^{-1} \textstyle\sum_{t=1}^b \hat{\sigma}_{vt}^{-2}\mathbf{w}_t'\mathbf{w}_t,$$

the ith element of $\mathbf{w}_t$ is

$$w_{ti} = g_{\beta i}(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}}) - \tfrac{1}{2} \operatorname{tr} \hat{\mathbf{\Lambda}}_t^{-1}\mathscr{R}_i(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}}),$$

the ith element of $\mathbf{g}_\beta(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})$ is $g_{\beta i}(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})$,

$$\hat{\mathbf{\Lambda}}_t = \mathbf{\Sigma}^{uu} + \mathbf{\Sigma}^{ue}\mathbf{g}_x(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}}) + \mathbf{g}_x'(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\mathbf{\Sigma}^{eu} + \mathbf{g}_x'(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\mathbf{g}_x(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\sigma^{ee},$$

$$\mathbf{A} = \hat{\lambda}(\alpha)n^{-1} \textstyle\sum_{t=1}^b \hat{\sigma}_{vt}^{-2}\mathbf{g}_{\beta x}(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\hat{\mathbf{\Lambda}}_t^{-1}\mathbf{g}_{\beta x}'(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}}),$$

$$\hat{\lambda}(\alpha) = \begin{cases} 1 - n^{-1}\alpha & \text{if } \hat{\lambda} > 1, \\ \hat{\lambda} - n^{-1}\alpha & \text{if } \hat{\lambda} \leq 1, \end{cases}$$

$\hat{\lambda}$ is the smallest root of the determinantal equation

$$|\hat{\mathbf{M}}_{ww} - \lambda n^{-1} \textstyle\sum_{t=1}^b \hat{\sigma}_{vt}^{-2}\mathbf{g}_{\beta x}(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\hat{\mathbf{\Lambda}}_t^{-1}\mathbf{g}_{\beta x}'(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})| = 0,$$

and α is the constant introduced in (3.2). This modification of $\hat{\mathbf{M}}_{xx}$ should result in several improvements. First, the use of $\hat{\lambda}(\alpha)$ guarantees that the matrix $\hat{\mathbf{M}}_{ww} - \mathbf{A}$ is nonsingular. Second, based on (3.3), $\mathbf{w}_t$ should give a better estimator of $\mathbf{g}_\beta(\mathbf{x}_t^0; \boldsymbol{\beta}^0)$ than $\mathbf{g}_\beta(\hat{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})$, because the trace term in w_{ti} estimates the term C_i. Third, $\mathbf{A}$ is an estimator of the bias in $\hat{\mathbf{M}}_{ww}$ arising from the fact that $\hat{\mathbf{x}}_t$ is an estimator of $\mathbf{x}_t^0$. Finally, replacing $\hat{\mathbf{x}}_t$ by $\tilde{\mathbf{x}}_t$ is analogous to the modification introduced by Fuller for the linear model.

The considerations leading to (3.2) and (3.4) suggest a class of estimators whose asymptotic properties are identical with those of $\tilde{\boldsymbol{\beta}}$, but whose small sample properties may be superior. We denote this class of estimators by $\boldsymbol{\beta}^\dagger(\alpha)$, where the estimator is defined by the expression for $\tilde{\boldsymbol{\beta}}$, with $\tilde{\mathbf{x}}_t$ replacing $\hat{\mathbf{x}}_t$ and $(\hat{\mathbf{M}}_{ww} - \mathbf{A})$ replacing $\hat{\mathbf{M}}_{xx}$. That is

(3.5) $$\boldsymbol{\beta}^\dagger(\alpha) = \bar{\boldsymbol{\beta}} + (\hat{\mathbf{M}}_{ww} - \mathbf{A})^{-1}\{n^{-1} \textstyle\sum_{t=1}^b \tilde{\sigma}_{vt}^{-2}\mathbf{g}_\beta(\tilde{\mathbf{x}}_t; \bar{\boldsymbol{\beta}})\tilde{q}_t\},$$

where $\tilde{\sigma}_{vt}^2$ and $\tilde{q}_t$ are defined analogous to $\hat{\sigma}_{vt}^2$ and $\hat{q}_t$ with $\tilde{\mathbf{x}}_t$ replacing $\hat{\mathbf{x}}_t$. In the next section, we investigate the sampling behavior of $\boldsymbol{\beta}^\dagger(1)$, $\boldsymbol{\beta}^\dagger(4)$, $\hat{\boldsymbol{\beta}}$, and $\tilde{\boldsymbol{\beta}}$.

TABLE 1
Monte Carlo Properties of Estimators of β_2 in Model (4.1)
(500 *samples*)

Parameter Set	Estimator	Mean	Var.	MSE	Percentiles 25%	75%
(i)	$\hat{\beta}_2$	1.374	1.808	1.948	0.644	1.861
	$\tilde{\beta}_2$	1.233	1.583	1.638	0.644	1.652
	$\beta_2^\dagger(1)$	1.199	1.213	1.252	0.629	1.633
	$\beta_2^\dagger(4)$	1.046	0.761	0.763	0.592	1.486
(ii)	$\hat{\beta}_2$	1.055	0.131	0.134	0.804	1.266
	$\tilde{\beta}_2$	1.041	0.115	0.117	0.810	1.241
	$\beta_2^\dagger(1)$	1.039	0.113	0.114	0.810	1.234
	$\beta_2^\dagger(4)$	1.014	0.103	0.104	0.800	1.205
(iii)	$\hat{\beta}_2$	1.009	0.0187	0.0188	0.911	1.095
	$\tilde{\beta}_2$	1.004	0.0182	0.0182	0.906	1.090
	$\beta_2^\dagger(1)$	1.001	0.0181	0.0181	0.904	1.086
	$\beta_2^\dagger(4)$	0.995	0.0178	0.0178	0.899	1.078

4. Monte Carlo Results. Estimators of the parameters of the model

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 \tag{4.1}$$

were studied using the Monte Carlo method. In this study, measurement errors (e_t, u_t) were generated as a random sample from a bivariate normal $(\mathbf{0}, \mathbf{\Sigma})$ distribution, the true parameter vector was $\boldsymbol{\beta}^0 = (.25, 1, 1)'$, and the remaining parameters were selected from one of the following sets:

(i) Sample size $b = 33$, the x-values $(x_1, x_2, \cdots, x_{33}) = (1, 1, 1) \otimes (-0.5, -0.4, -0.3, \cdots, 0.4, 0.5)$, and $\mathbf{\Sigma} = 0.0324\mathbf{I}$.

(ii) Sample size $b = 66$, the x-values $(x_1, x_2, \cdots, x_{66}) = (1, 1, 1, 1, 1, 1) \otimes (-0.5, -0.4, -0.3, \cdots, 0.4, 0.5)$, and $\mathbf{\Sigma} = 0.0144\mathbf{I}$.

(iii) Sample size and x-values as in parameter set (ii) but

$$\mathbf{\Sigma} = \begin{pmatrix} 0.0081 & 0.0027 \\ 0.0027 & 0.0036 \end{pmatrix}.$$

The ratio of the standard error σ_u to the root mean square for x is 0.5692, 0.3795, and 0.1897 for parameter sets (i), (ii), and (iii), respectively.

For each parameter set, 500 samples were generated and the four estimators $\hat{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\beta}}$, $\boldsymbol{\beta}^\dagger(1)$, and $\boldsymbol{\beta}^\dagger(4)$ computed for each sample. The estimators are defined in equations (2.4), (2.5) and (3.5) respectively. The ordinary least squares estimator was used as the preliminary estimator and the four estimators were iterated to convergence. While certain x-values were repeated, this fact was not used in the estimation.

Table 1 contains the Monte Carlo properties of the four estimators of β_2^0. The estimator biases are generally in the order

$$|\operatorname{Bias}\{\boldsymbol{\beta}^\dagger(4)\}| \le |\operatorname{Bias}\{\boldsymbol{\beta}^\dagger(1)\}| \le |\operatorname{Bias}\{\tilde{\boldsymbol{\beta}}\}| \le |\operatorname{Bias}\{\hat{\boldsymbol{\beta}}\}|.$$

This result is most evident for parameter set (i), where the bias in $\hat{\beta}_2$ is eight times that of $\beta_2^\dagger(4)$. In most cases the quadratic coefficient was overestimated.

The estimator mean squared errors are in the same order as the bias

$$\operatorname{MSE}\{\boldsymbol{\beta}^\dagger(4)\} \le \operatorname{MSE}\{\boldsymbol{\beta}^\dagger(1)\} \le \operatorname{MSE}\{\tilde{\boldsymbol{\beta}}\} \le \operatorname{MSE}\{\hat{\boldsymbol{\beta}}\}.$$

The 25th and 75th percentiles are also given for each estimator and parameter set. The order of the differences between these percentiles generally agrees with the order of the estimator variances. The distributions of $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^\dagger(1)$ tend to be more symmetric than the distribution of $\boldsymbol{\beta}^\dagger(4)$.

For the small error variances of parameter set (iii), the four estimators are very similar.

TABLE 2
Monte Carlo Percentiles for Studentized Statistics
(500 samples)

Parameter Set	Statistic	Percentiles				
		5%	10%	50%	90%	95%
(1)	$\hat{t}_2$	−3.550	−2.446	0.331	2.122	2.355
	$\tilde{t}_2$	−3.538	−2.437	0.250	1.933	2.211
	$t_2^\dagger(1)$	−2.355	−1.676	0.063	1.158	1.483
	$t_2^\dagger(4)$	−2.554	−1.745	−0.083	1.059	1.307
(ii)	$\hat{t}_2$	−2.491	−2.015	0.089	2.132	2.539
	$\tilde{t}_2$	−2.437	−1.987	0.088	1.982	2.354
	$t_2^\dagger(1)$	−2.010	−1.575	0.055	1.458	1.766
	$t_2^\dagger(4)$	−2.080	−1.686	−0.062	1.377	1.692
(iii)	$\hat{t}_2$	−1.763	−1.399	0.001	1.450	1.919
	$\tilde{t}_2$	−1.786	−1.421	−0.043	1.385	1.841
	$t_2^\dagger(1)$	−1.690	−1.366	−0.056	1.295	1.724
	$t_2^\dagger(4)$	−1.737	−1.412	−0.115	1.240	1.673

TABLE 3
Ratio of Asymptotic Variance to Monte Carlo Variance

Parameter Set	Estimator			
	$\hat{\beta}_2$	$\tilde{\beta}_2$	$\beta_2^\dagger(1)$	$\beta_2^\dagger(4)$
(i)	0.15	0.17	0.22	0.35
(ii)	0.46	0.52	0.53	0.58
(iii)	0.85	0.88	0.88	0.90

For large error variances and (or) small sample sizes, $\boldsymbol{\beta}^\dagger(1)$ and $\boldsymbol{\beta}^\dagger(4)$ are the preferred estimators. It is somewhat surprising that the Monte Carlo moments of $\hat{\boldsymbol{\beta}}$ are of the same order as those of the other estimators, because it is known that the maximum likelihood estimator of $\boldsymbol{\beta}^0$ does not possess finite moments for the linear errors-in-variables model.

Statistics analogous to Student's t were also computed for each parameter set, and the sample percentiles are given in Table 2. These statistics used the estimated variance from the inverse of the M-matrices on the left side of the expressions defining $\hat{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\beta}}$, $\boldsymbol{\beta}^\dagger(1)$, and $\boldsymbol{\beta}^\dagger(4)$, respectively. There is reasonable agreement between the sample percentiles for $\mathbf{t}^\dagger(1)$ and $\mathbf{t}^\dagger(4)$ and the theoretical percentiles for the $N(0, 1)$ distribution. The agreement improves as the sample size increases and (or) the error variance decreases. The statistics $t_2^\dagger(1)$ and $t_2^\dagger(4)$ have negatively skewed distributions. The tails of the distributions of $\hat{\mathbf{t}}$ and $\tilde{\mathbf{t}}$ are heavy in comparison to the $N(0, 1)$ distribution for parameter sets (i) and (ii). The modification to the M-matrix introduced in (3.4) substantially improves the estimated variances and the corresponding studentized statistics. That is, the tail percentiles of $t_2^\dagger(1)$ and $t_2^\dagger(4)$ are in much better agreement with those of the normal distribution than are the tail percentiles of $\hat{t}_2$ and $\tilde{t}_2$.

The variance of the asymptotic distribution is generally smaller than the Monte Carlo variance for all estimators and parameter sets, though the discrepancy declines as the error variance declines relative to the variability in x, i.e. as one moves from parameter set (i) to (iii). See Table 3.

In summary, the Monte Carlo results suggest that the estimators $\boldsymbol{\beta}^\dagger(1)$ and $\boldsymbol{\beta}^\dagger(4)$ can be recommended in samples as small as 30 with the ratio of error variance to variability in x as large as 0.3. In such samples, the variance of the estimator is much larger than theory suggests, but the "t-statistics" have a distribution reasonably close to $N(0, 1)$. Therefore, error rates for inferences based upon the sample statistics will not deviate greatly from the nominal rates.

5. Proofs of primary results. The classical results on the consistency and asymptotic normality of the maximum likelihood estimates are not immediately applicable for the errors-in-variables model of Section 2, because the number of unknown incidental parameters increases with increasing sample size.

The proof of Theorem 1 is not presented because it is nearly identical to that of Theorem 2. As a first step in obtaining the limiting distribution of $\tilde{\boldsymbol{\beta}}$, the order of the error in $\hat{x}_t$ is established.

LEMMA 1. *Let Model* (1.1) *and Assumptions* (i) *through* (iv), (viia), *and* (viiia) *hold. Then*

$$\hat{\mathbf{x}}_t = \mathbf{x}_t^0 + \boldsymbol{\delta}_t + O_p(\max[a_n^{-1}, n^{-1/2}])$$

for $t = 1, \cdots, b$, *where*

$$\boldsymbol{\delta}_t = [\mathbf{u}_t\{\boldsymbol{\Sigma}^{uu} + \boldsymbol{\Sigma}^{ue}\mathbf{g}_x(x_t^0; \boldsymbol{\beta}^0)\} + e_t\{\boldsymbol{\Sigma}^{eu} + \sigma^{ee}\mathbf{g}_x(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\}]\Lambda_t^{-1} \tag{5.1}$$

$$\Lambda_t = \boldsymbol{\Sigma}^{uu} + \boldsymbol{\Sigma}^{ue}\mathbf{g}_x(\mathbf{x}_t^0; \boldsymbol{\beta}^0) + \mathbf{g}_x'(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\boldsymbol{\Sigma}^{eu} + \mathbf{g}_x'(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\mathbf{g}_x(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\sigma^{ee},$$

and $\hat{\mathbf{x}}_t$ is the root of (2.1) maximizing the likelihood. Furthermore,

$$E(|\hat{\mathbf{x}}_t - \mathbf{x}_t^0|^4) = O(a_n^{-2}). \tag{5.2}$$

PROOF. By definition, $\hat{\mathbf{x}}_t$ is the value of $\mathbf{x}_t$ in $\mathscr{A}$ that minimizes $D^2(\mathbf{x}_t) = q(\tilde{\boldsymbol{\beta}}, \mathbf{x}_t; Y_t, \mathbf{X}_t)$. Since $D^2(\mathbf{x}_t^0) = O_p(1)$ it follows that $\hat{\mathbf{x}}_t - \mathbf{x}_t^0 = O_p(a_n^{-1/2})$. Similarly $E(|\hat{\mathbf{x}}_t - \mathbf{x}_t^0|^4) = O(a_n^{-2})$ because $E\{D^4(\mathbf{x}_t^0)\} = O(1)$.

Now expanding $g(\hat{\mathbf{x}}_t; \tilde{\boldsymbol{\beta}})$ and $\mathbf{g}_x(\hat{\mathbf{x}}_t; \tilde{\boldsymbol{\beta}})$ about $(\mathbf{x}_t^0; \boldsymbol{\beta}^0)$ gives

$$Y_t - g(\hat{\mathbf{x}}_t; \tilde{\boldsymbol{\beta}}) = e_t - \mathbf{g}_x'(\mathbf{x}_t^0; \boldsymbol{\beta}^0)(\hat{\mathbf{x}}_t - \mathbf{x}_t^0)' + O_p(\max[a_n^{-1}, n^{-1/2}])$$

and

$$\mathbf{g}_x(\hat{\mathbf{x}}_t; \tilde{\boldsymbol{\beta}}) = \mathbf{g}_x(\mathbf{x}_t^0; \boldsymbol{\beta}^0) + O_p(a_n^{-1/2}),$$

respectively. The lemma follows by substituting these results into (2.2). □

LEMMA 2. *Let Model* (1.1) *and Assumptions* (i), (iii), *and* (iv) *hold. Then*

$$\mathrm{Cov}\{\boldsymbol{\delta}_t, e_t - \mathbf{u}_t\mathbf{g}_x'(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\} = \mathbf{0}.$$

PROOF. The result is obtained by evaluating the covariance using (5.1) and the relation between a partitioned matrix and its partitioned inverse. □

PROOF OF THEOREM 2. Expanding $\mathbf{g}_\beta(\hat{\mathbf{x}}_t; \tilde{\boldsymbol{\beta}})$, $\mathbf{g}_x(\hat{\mathbf{x}}_t, \tilde{\boldsymbol{\beta}})$, and $\hat{\sigma}_{vt}^2$ about $(\mathbf{x}_t^0; \boldsymbol{\beta}^0)$, we obtain

$$\mathbf{g}_\beta(\hat{\mathbf{x}}_t; \tilde{\boldsymbol{\beta}}) = \mathbf{g}(\mathbf{x}_t^0; \boldsymbol{\beta}^0) + \mathbf{g}_{\beta x}(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\boldsymbol{\delta}_t' + O_p(\max[a_n^{-1}, n^{-1/2}]),$$

$$\mathbf{g}_x(\hat{\mathbf{x}}_t, \tilde{\boldsymbol{\beta}}) = \mathbf{g}_x(\mathbf{x}_t^0; \boldsymbol{\beta}^0) + \boldsymbol{\delta}_t\mathbf{g}_{xx}(\mathbf{x}_t^0; \boldsymbol{\beta}^0) + O_p(\max[a_n^{-1}, n^{-1/2}]),$$

$$a_n\hat{\sigma}_{vt}^2 = a_n\sigma_{vt}^2 + a_n(\Delta\sigma_{vt}^2) + O_p(\max[a_n^{-1}, n^{-1/2}]),$$

where

$$\Delta\sigma_{vt}^2 = 2\boldsymbol{\delta}_t\mathbf{g}_{xx}(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\{\boldsymbol{\Sigma}_{uu}\mathbf{g}_x'(\mathbf{x}_t^0; \boldsymbol{\beta}^0) - \boldsymbol{\Sigma}_{ue}\}.$$

In all cases the part of the remainder associated with $(\hat{\mathbf{x}}_t - \mathbf{x}_t^0)$ has expectation that is of the same order as the order in probability of the term.

By these results and Lemmas 1 and 2, it follows that

$$\hat{\mathbf{M}}_{xx} = \mathbf{m}_{xx} + O_p(\max[a_n^{-1}, n^{-1/2}]) \tag{5.3}$$

and

$$n^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) = \mathbf{m}_{xx}^{-1}[n^{1/2}b^{-1}\textstyle\sum_{t=1}^{b}(a_n^{-1}\sigma_{vt}^{-2})\mathbf{g}_\beta(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\{e_t - \mathbf{u}_t g_x'(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\}] + o_p(1). \tag{5.4}$$

Hence, the limiting distribution of $n^{1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)$ is the same as that of the leading term of (5.4). The theorem then follows by application of the Liapounov central limit theorem. □

Theorems 1 and 2 assume the existence of a preliminary estimator $\bar{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}^0$ whose error is $O_p(\max[a_n^{-1}, n^{-1/2}])$. We now demonstrate that the ordinary least squares estimator satisfies this condition. Let

$$T_n(\boldsymbol{\beta}) = b^{-1} \textstyle\sum_{t=1}^{b} \{Y_t - g(\mathbf{X}_t; \boldsymbol{\beta})\}^2, \tag{5.5}$$

where it is understood that in (5.5) and in the ensuing development, $\mathbf{X}_t$ is replaced by its projection onto the space $\mathscr{A}$, whenever $\mathbf{X}_t$ is outside of $\mathscr{A}$. Then the ordinary least squares estimator, denoted by $\hat{\boldsymbol{\beta}}_\ell$, is the $\boldsymbol{\beta}$ that minimizes $T_n(\boldsymbol{\beta})$. We assume:

(ix) The matrix

$$\mathbf{H}_n(\boldsymbol{\theta}, \boldsymbol{\beta}) = b^{-1} \textstyle\sum_{t=1}^{b} \mathbf{L}'(\mathbf{x}_t^0; \boldsymbol{\theta})\mathbf{L}(\mathbf{x}_t^0; \boldsymbol{\beta})$$

converges to the matrix $\mathbf{H}(\boldsymbol{\theta}, \boldsymbol{\beta})$ uniformly for all $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ in $\dot{\boldsymbol{\Theta}}$, where

$$\mathbf{L}(\mathbf{x}_t^0; \boldsymbol{\theta}) = [g(\mathbf{x}_t^0; \boldsymbol{\theta}), \mathbf{g}'_\beta(\mathbf{x}_t^0; \boldsymbol{\theta}), \operatorname{Vec}\{\mathbf{g}_{\beta\beta}(\mathbf{x}_t^0; \boldsymbol{\theta})\}]$$

and $\operatorname{Vec}\{\mathbf{g}_{\beta\beta}(\mathbf{x}_t^0; \boldsymbol{\theta})\}$ denotes the row vector created by listing the rows of $\mathbf{g}_{\beta\beta}(\mathbf{x}_t^0; \boldsymbol{\theta})$ in order.

(x) The function

$$\lim_{n\to\infty} b^{-1} \textstyle\sum_{t=1}^{b} \{g(\mathbf{x}_t^0; \boldsymbol{\beta}) - g(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\}^2$$

defined on $\boldsymbol{\Theta}$ has a unique minimum at $\boldsymbol{\beta} = \boldsymbol{\beta}^0$.

(xi) The matrix

$$\lim_{n\to\infty} b^{-1} \textstyle\sum_{t=1}^{b} \mathbf{g}_\beta(\mathbf{x}_t^0; \boldsymbol{\beta}^0)\mathbf{g}'_\beta(\mathbf{x}_t^0; \boldsymbol{\beta}^0)$$

is nonsingular.

LEMMA 4. *Let Model* (1.1) *and Assumptions* (i), (iii), (vii), *and* (viiia) *through* (xi) *hold. Then*

$$\hat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}^0 = O_p(\max[a_n^{-1}, n^{-1/2}]).$$

PROOF. Let $\hat{\boldsymbol{\beta}}(x)$ be the estimator of $\boldsymbol{\beta}$ constructed using the true $\mathbf{x}$-values. That is, $\hat{\boldsymbol{\beta}}(\mathbf{x})$ is the $\boldsymbol{\beta}$ for which

$$b^{-1} \textstyle\sum_{t=1}^{b} \{Y_t - g(\mathbf{x}_t^0; \boldsymbol{\beta})\}^2 \tag{5.6}$$

is a minimum, where the subscript n on b is suppressed. By the results of Jennrich (1969),

$$\hat{\boldsymbol{\beta}}(\mathbf{x}) - \boldsymbol{\beta}^0 = O_p(n^{-1/2}).$$

Now

$$\begin{aligned} & b^{-1} \textstyle\sum_{t=1}^{b} \{Y_t - g(\mathbf{X}_t; \boldsymbol{\beta})\}^2 \\ &= b^{-1} \textstyle\sum_{t=1}^{b} \{Y_t - g(\mathbf{x}_t^0; \boldsymbol{\beta}) - \mathbf{g}_x(\mathbf{x}_t^\dagger; \boldsymbol{\beta})\mathbf{u}'_t\}^2 \\ &= b^{-1} \textstyle\sum_{t=1}^{b} [\{Y_t - g(\mathbf{x}_t^0; \boldsymbol{\beta})\}^2 - 2\{Y_t - g(\mathbf{x}_t^0; \boldsymbol{\beta})\}\mathbf{g}_x(\mathbf{x}_t^\dagger; \boldsymbol{\beta})\mathbf{u}'_t + \{\mathbf{g}_x(\mathbf{x}_t^\dagger; \boldsymbol{\beta})\mathbf{u}'_t\}^2] \\ &= b^{-1} \textstyle\sum_{t=1}^{b} \{Y_t - g(\mathbf{x}_t^0; \boldsymbol{\beta})\}^2 + O_p(a_n^{-1}), \end{aligned} \tag{5.7}$$

where $\mathbf{x}_t^\dagger$ is on the line segment joining $\mathbf{X}_t$ and $\mathbf{x}_t^0$. The order of the remainder in (5.7) follows from assumptions (i) and (vii). Because of the uniqueness of the minimum, it follows that

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

By the mean value theorem (see Jennrich, 1969, equation (3)), there exists a $\ddot{\boldsymbol{\beta}}$ such that

$$b^{-1} \sum_{t=1}^{b} \mathbf{g}_\beta(\mathbf{X}_t; \boldsymbol{\beta}^0)\{Y_t - g(\mathbf{X}_t; \boldsymbol{\beta}^0)\} = b^{-1}\sum_{t=1}^{b} \mathbf{g}_\beta(\mathbf{X}_t; \hat{\boldsymbol{\beta}}_\ell)\{Y_t - g(\mathbf{X}_t; \hat{\boldsymbol{\beta}}_\ell)\}$$

(5.8)

$$+ [b^{-1} \sum_{t=1}^{b} \mathbf{g}_\beta(\mathbf{X}_t; \ddot{\boldsymbol{\beta}})\mathbf{g}'_\beta(\mathbf{X}_t; \ddot{\boldsymbol{\beta}}) - b^{-1} \sum_{t=1}^{b} \mathbf{g}_{\beta\beta}(\mathbf{X}_t; \ddot{\boldsymbol{\beta}})\{Y_t - g(\mathbf{X}_t; \ddot{\boldsymbol{\beta}})\}](\hat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}^0)$$

where $|\ddot{\boldsymbol{\beta}} - \boldsymbol{\beta}^0| \leq |\hat{\boldsymbol{\beta}}_\ell - \boldsymbol{\beta}^0|$. The inner products on the right side of the equality are zero by the least squares property; $O_p(1)$ by assumption (i); and $o_p(1)$ by assumptions (vii) and (viiia), respectively. The conclusion follows from assumption (xi) and the fact that the quantity on the left side of (5.8) is $O_p(\max[a_n^{-1}, n^{-1/2}])$. □

Acknowledgments. This research was partly supported by joint statistical research agreements J.S.A. 74-1 and J.S.A. 81-14 with the U.S. Bureau of the Census. We thank Dr. Hans-Peter Höschel and the referees for useful comments on earlier drafts of the paper.

REFERENCES

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for normal distributions. *Ann. Math. Statist.* **22** 327–351.

Bohrnstedt, G. W. and Carter, T. M. (1971). Robustness in regression analysis. In *Sociological Methodology* (H. L. Costner, ed.) 118–146. Jossey-Bass, San Francisco.

Britt, H. I. 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** 261–269.

Cook, W. R. (1931). On curve fitting by means of least squares. *Philosophical Magazine, Series 7* **12** 1025–1039.

Dalenius, T. (1977a). Bibliography on nonsampling errors in surveys-I. *Internat. Statist. Rev.* **45** 71–89.

Dalenius, T. (1977b). Bibliography on nonsampling errors in surveys-II. *Internat. Statist. Rev.* **45** 181–197.

Dalenius, T. (1977c). Bibliography on nonsampling errors in surveys-III. *Internat. Statist. Rev.* **45** 303–317.

Deming, W. E. (1931). The application of least squares. *Philosophical Magazine Series 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 relationship with replicated observations and correlated errors. *Biometrika* **59** 121–129.

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

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

Griliches, Z. (1974). Errors in variables and other unobservables. *Econometrica* **42** 971–998.

Griliches, Z. and Ringstad, V. (1970). Errors-in-the-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.

Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.* **40** 633–643.

Kendall, M. G. and Stuart, A. (1961). *The Advanced Theory of Statistics.* Chapter 27. Griffin, London.

O'Neill, M., Sinclair, L. G., and Smith, F. J. (1969). Polynomial curve fitting when abscissas and ordinate are both subject to error. *Computer J.* **12** 52–56.

Villegas, C. (1969). On the least squares estimation of nonlinear relations. *Ann. Math. Statist.* **40** 462–466.

Wolter, K. M. (1974). Estimators for a nonlinear functional relationship. Unpublished Ph.D. dissertation, Iowa State University.

Wolter, K. M. and Fuller, W. A. (1975). Estimating a nonlinear errors-in-variables model with singular error covariance matrix. 1975 Proceedings of the Business and Economic Statistics Section, Amer. Statist. Assoc.

Bureau of the Census
U.S. Department of Commerce
Washington, D.C. 20233

Department of Statistics
Iowa State University
Snedecor Hall
Ames, Iowa 50010