The Annals of Statistics
1987, Vol. 15, No. 2, 513–540

EFFICIENT ESTIMATION IN THE ERRORS IN VARIABLES MODEL[1]

BY P. J. BICKEL AND Y. RITOV

University of California, Berkeley and The Hebrew University of Jerusalem

We consider efficient estimation of the slope in the errors in variables model with normal error when either the ratio of error variances is known and the distribution of the independent is arbitrary and unknown or the distribution of the independent variable is not Gaussian or degenerate. We calculate information bounds and exhibit estimates achieving these bounds using an initial minimum distance estimate and suitable estimates of the efficient score function.

1. Introduction. Errors in variables models have been the subject of an enormous amount of literature. A fairly recent reference with a good bibliography is Anderson (1984).

In its simplest form the model assumes n independent observations $\mathbf{X}_i = (X_i, Y_i)$, which are written as

$$X_i = X_i' + \varepsilon_{i1},$$
$$Y_i = \alpha + \beta X_i' + \varepsilon_{i2}. \tag{1.1}$$

The X_i' are viewed either as

(i) unknown constants;
(ii) independent identically distributed random variables.

Model (i) is called functional and (ii) structural by Kendall and Stuart (1979), Chapter 29.

The $(\varepsilon_{i1}, \varepsilon_{i2})$ are considered random vectors, which are identically distributed with mean 0, as well as independent of the X_i' in model (ii). In this paper we will deal exclusively with large sample theory in the structural model, although we believe our results generalize to the functional model. Our aim in this paper is the construction of efficient estimates of β under various assumptions in various special cases of (1.1). We also suggest how our results may be extended to instrumental variable models through the special case of repeated observations at the same X_i'.

Write $\mathbf{X}$, X', ε_1, ε_2 for "generic" observations. If we do not make any assumptions on the distributions of X and $(\varepsilon_1, \varepsilon_2)$, then β is clearly unidentifiable. In fact, β is unidentifiable even if we assume $\varepsilon_1, \varepsilon_2$ to be independent Gaussian variables with unknown variances and suppose X' is also Gaussian.

Received October 1984; revised September 1986.

[1]This research was supported in part by ONR Contract N00014-80-C-0163.

AMS 1980 *subject classifications*. Primary 62G20; secondary 62P20.

Key words and phrases. Reiersøl models, semiparametric, minimum distance, structural.

However, β has been shown to be identifiable under various sets of assumptions. These fall into two broad classes:

(A) *Gaussian errors.* $(\varepsilon_1, \varepsilon_2)$ have a bivariate Gaussian distribution with variance–covariance matrix Σ. The usual way to make β identifiable in the literature is to assume $\varepsilon_1, \varepsilon_2$ independent and either

$$\mathrm{Var}(\varepsilon_1) = c_0 \mathrm{Var}(\varepsilon_2) \tag{1.2}$$

or

$$\mathrm{Var}(\varepsilon_1) = c_0, \tag{1.3}$$

with c_0 assumed known. Both (1.2) and (1.3) are plausible under special circumstances [see Kendall and Stuart (1979), Chapter 29, for a discussion]. We shall explore a generalization of (1.2),

$$\Sigma = \sigma^2 \Sigma_0, \tag{1.4}$$

where Σ_0 is known. Model (1.3) can be analyzed in the same way. We shall call (1.4) the *restricted Gaussian error* model. This model and its generalizations to more complicated situations have been extensively studied; see Anderson (1984), for example. A second model in which the identifiability of β was established by Reiersøl (1950) puts no restriction on Σ but requires X' to be non-Gaussian (where constants are viewed as Gaussian). We shall call this the *general Gaussian error* model.

(B) *General independent errors.* Assume $\varepsilon_1, \varepsilon_2$ independent. If (1.2) holds, β is identifiable. This *restricted independent error* has also been extensively studied. If (1.2) is not present but either X' is non-Gaussian or $\varepsilon_1, \varepsilon_2$ have no Gaussian component, then, again according to Reiersøl (1950), β is identifiable. This *arbitrary independent error* model is probably most satisfactory but our results do not bear on it.

We review briefly some results on these models.

The restricted Gaussian model can be reduced to case (1.2) with $c_0 = 1$. The maximum likelihood estimate for β in this case is $\hat{\beta}_P$, which minimizes the sum of squared perpendicular distances of observed points from the fitted line

$$\sum_{i=1}^{n} \frac{(Y_i - \alpha - \beta X_i)^2}{1 + \beta^2}. \tag{1.5}$$

This estimate is well known to be $n^{1/2}$-consistent and asymptotically normal not only under the restricted Gaussian model but also under the restricted independent error model, see Gleser (1981) who considers multivariate generalizations. In the presence of fourth moments, it is not hard to show that $n^{1/2}$-consistency and asymptotic normality persist under the restricted independent error model when Σ_0 is the identity. Estimates of β in the general Gaussian error model, with Σ_0 diagonal, have been proposed by a variety of authors including Neyman and Scott (1948) and Rubin (1956). In the arbitrary independent error model, Wolfowitz in a series of papers ending in 1957, Kiefer and Wolfowitz (1956) and Spiegelman (1979) by a variety of methods gave estimates, which are consistent and in Spiegelman's case $n^{1/2}$-consistent and asymptotically normal.

Little seems to be known about the efficiency of these procedures other than that in the restricted Gaussian model the estimate $\hat{\beta}_P$ is efficient if X' is Gaussian by the classical results for M.L.E.'s in parametric models. Our main aims in this paper are:

In the general Gaussian error model:

(i) To give the structure that efficient estimates in the sense of Stein (1956), Koshevnik and Levit (1976) and Pfanzagl (1982) must have (Theorem 2.1).

(ii) To exhibit a reasonable efficient estimate (Theorem 2.2). In addition, we extend Theorem 2.1 to the simplest instrumental variable model, m repeated measurements with Gaussian errors,

$$X_{ij} = X_i' + \varepsilon_{ij1},$$
$$Y_{ij} = \alpha + \beta X_i + \varepsilon_{ij2}, \qquad j = 1, \dots, m,\ i = 1, \dots, r,\ n = mr,$$

and

$$\mathbf{X}_i = \{(X_{ij}, Y_{ij}),\ j = 1, \dots, m\},$$

where $m \geq 2$.

The ε_{ij2} are independent and identically distributed Gaussian and independent of ε_{ij1} which are also Gaussian. We refer this as the multiple *Gaussian measurements model.* Note that in this model if $m \geq 2$, the assumption of non-Gaussianity of the distribution of X' is unnecessary.

We speak of efficient estimation in the sense of Stein (1956) as developed by Koshevnik and Levit (1976), Pfanzagl (1982), Begun, Hall, Huang and Wellner (1983) and in a forthcoming monograph by Klaassen, Wellner and ourselves. Let $\mathbf{P}$ be the set of possible joint distributions of $\mathbf{X}$. We call $\mathbf{P}_0$ a parametric submodel of $\mathbf{P}$ if $\mathbf{P}_0 \subset \mathbf{P}$ and $\mathbf{P}_0$ can be represented as $\{P_{(\beta, \eta)};\ \beta \in \mathbf{R},\ \eta \in E \text{ open } \subset \mathbf{R}^k\}$. A parametric submodel is regular if at every (β_0, η_0) the mapping $(\beta, \eta) \to P_{(\beta, \eta)}$ is continuously Hellinger differentiable. Suppose that P belongs to $\mathbf{P}_0$—a regular parametric submodel of $\mathbf{P}$. Then the notion of information bound and efficient estimation of β are well defined [e.g., Ibragimov and Has'minskii (1981), pages 158–169]. Let $n^{-1}I^{-1}(P; \beta, \mathbf{P}_0)$ denote the asymptotic variance of an efficient estimate of β when P ranges over $\mathbf{P}_0$. Clearly, if we only assume that $P \in \mathbf{P}$ we can estimate no better than if we assumed that $P \in \mathbf{P}_0$. Accordingly, let $I(P; \beta, \mathbf{P}) = \inf\{I(P; \beta, \mathbf{P}_0)\colon \mathbf{P}_0 \text{ a regular parametric submodel}, P \in \mathbf{P}_0\}$, be the information bound for estimating β under $\mathbf{P}$.

Loosely speaking, $\hat{\beta}_n$ is regular and efficient in $\mathbf{P}$ if

$$\mathbf{L}_P\big(\sqrt{n}\,(\hat{\beta}_n - \beta(P))\big) \to \mathbf{N}(0, I^{-1}(P; \beta, \mathbf{P})),$$

in some sense uniformly in $P \in \mathbf{P}$. Here $\mathbf{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. The weakest kind of uniformity acceptable is that

$$\mathbf{L}_{P_n}\big(\sqrt{n}\,(\hat{\beta}_N - \beta(P_n))\big) \to \mathbf{N}(0, I^{-1}(P; \beta, \mathbf{P})), \tag{1.6}$$

for sequences $P_n \in \mathbf{P}_0$, a regular parametric submodel as above, with $P_n = P_{(\beta_n, \eta_n)}$, $|\beta_n - \beta_0| = O(n^{-1/2}) = |\eta_n - \eta_0|$ for some $\beta_0, \eta_0, P = P_{(\beta_0, \eta_0)}$.

If $I^{-1}(P; \beta, \mathbf{P})$ is assumed at some $\mathbf{P}_0$, we obtain from the Hájek–Le Cam convolution theorem, Ibragimov and Has'minskii (1981), that $\hat{\beta}_n$ is asymptotically linear

$$\hat{\beta}_n = \beta(P) + n^{-1} \sum_{i=1}^{n} \tilde{l}(\mathbf{X}_i, P; \beta, \mathbf{P}) + o_p(n^{-1/2}),$$

where $\tilde{l}$ is defined as the efficient influence function, which has the properties

$$E_P \tilde{l}(\mathbf{X}_i, P; \beta, \mathbf{P}) = 0,$$
$$E_P \tilde{l}^2(\mathbf{X}_i, P, \beta, \mathbf{P}) = I^{-1}(P; \beta, \mathbf{P}).$$

Finding $\tilde{l}$ is equivalent to finding a suitable least favorable $\mathbf{P}_0$ (at each P). We discuss the theory which guides us in this search in Section 3.

Note that an estimate is efficient if

(a) it converges in law uniformly [as in (1.6)] on $\mathbf{P}$ and

(b) it is efficient in some parametric submodel $\mathbf{P}_0$ at each P. By the Hájek–Le Cam theorem (b) holds iff the efficient influence function is the influence function of the (local) maximum likelihood estimate of β in $\mathbf{P}_0$.

In Section 2 (Theorem 2.1), we exhibit $\tilde{l}$ and $\mathbf{P}_0$ for the general Gaussian error model and the restricted Gaussian model and discuss the main features of $I(P; \beta, \mathbf{P})$. In Theorem 2.2 we exhibit, for each of the two models, an estimate $\hat{\beta}$, converging in law uniformly [as in (1.6)] on $\mathbf{P}$, which has $\tilde{l}$ as influence function. By (a) and (b), $\hat{\beta}$ is necessarily efficient. The proof of Theorem 2.1 is deferred to Section 3, and the proof of Theorem 2.2 to Section 4.

2. The main results. Without loss of generality let $(\varepsilon_{i1}, \varepsilon_{i2}) \sim \mathbf{N}(0, \Sigma)$ where $\Sigma = [\sigma_{ij}]_{2\times 2}$ is nonsingular. Let $\theta = (\alpha, \beta, \Sigma)$ and

$$U(\theta) = U(\mathbf{X}, \theta) = \frac{Y - \alpha - \beta X}{\bar{\sigma}(\theta)}, \tag{2.1}$$

$$T(\theta) = T(\mathbf{X}, \theta) = \bar{\sigma}^{-2}(\theta)\big[(\sigma_{22} - \beta\sigma_{12})X + (\beta\sigma_{11} - \sigma_{12})(Y - \alpha)\big], \tag{2.2}$$

where $\bar{\sigma}^2(\theta)$ is the variance of $Y - \alpha - \beta X$ if θ is true,

$$\bar{\sigma}^2(\theta) = \beta^2\sigma_{11} - 2\beta\sigma_{12} + \sigma_{22}. \tag{2.3}$$

Then given θ, $T(\theta)$ is a complete and sufficient statistic for X' treated as a parameter, i.e., for the model $\{\mathbf{L}_\theta(\mathbf{X}|X' = \eta)\colon \eta \in R\}$. This follows since given $X' = \eta$, (X, Y) have an $\mathbf{N}(\eta, \alpha + \beta\eta, \Sigma)$ distribution. Moreover, $U(\theta)$ is ancillary in this problem. It is necessarily independent of $T(\theta)$ in the original model and is distributed $\mathbf{N}(0,1)$. $T(\theta)$ is also the unbiased predictor of X', i.e., given $X' = \eta$, $T(\theta)$ has a $\mathbf{N}(\eta, \tilde{\sigma}^2(\theta))$ distribution, where

$$\tilde{\sigma}^2(\theta) = \bar{\sigma}^{-2}(\theta)\big(\sigma_{11}\sigma_{22} - \sigma_{12}^2\big).$$

We can write the joint density of $\mathbf{X}$ under (θ, G), where G is the distribution of X',

$$p(\mathbf{x}, \theta, G) = \int K(\mathbf{x}, z, \theta) G(dz), \tag{2.4}$$

where

$$
\begin{aligned}
K(\mathbf{x}, z, \theta) = {} & \left[2\pi\left(\sigma_{11}\sigma_{22} - \sigma_{12}^2\right)^{1/2}\right]^{-1} \\
& \times \exp\Big\{-\left[2\left(\sigma_{11}\sigma_{22} - \sigma_{12}^2\right)\right]^{-1} \\
& \qquad \times \Big[\sigma_{22}(x - z)^2 - 2\sigma_{12}(x - z) \\
& \qquad\qquad \times (y - \alpha - \beta z) + \sigma_{11}(y - \alpha - \beta z)^2\Big]\Big\} \\
= {} & \left[2\pi\left(\sigma_{11}\sigma_{22} - \sigma_{12}^2\right)^{1/2}\right]^{-1} \exp\left\{-\frac{1}{2}U^2(\mathbf{x}, \theta)\right\} \\
& \times \exp\left\{-\frac{\tilde{\sigma}^{-2}(\theta)}{2}(T(\mathbf{x}, \theta) - z)^2\right\},
\end{aligned}
$$

is the conditional density of $\mathbf{X}$ given $X' = z$.

Fix $\theta = \theta_0$, $G = G_0$. Drop the argument θ in $U(\theta)$, $T(\theta)$, $\bar{\sigma}^2(\theta)$, and $\tilde{\sigma}^2(\theta)$. Let

$$
\omega(t) = \omega(t, \theta, G) = \tilde{\sigma}^{-1}\int \phi\left(\tilde{\sigma}^{-1}(t - z)\right)G(dz) \tag{2.5}
$$

be the density of T and let

$$
I_0 = \int \frac{[\omega']^2}{\omega}(t)\, dt
$$

be the Fisher information for location of ω. Let $\eta = (\mu, \tau)$, $\mu \in R$, $\tau > 0$, and

$$
G(\cdot, \eta) = G_0\left(\frac{\cdot - \mu}{\tau}\right).
$$

Define

$$
\mathbf{P}_0 = \left\{P_{(\theta, G(\cdot, \eta))}\right\}. \tag{2.6}
$$

That is, in $\mathbf{P}_0$ we assume G known up to location and scale. $\mathbf{P}_0$ is not the same in the general Gaussian error model and the restricted Gaussian error model since Σ varies freely in the former!

Theorem 2.1. *Assume $\int \eta^2 G(d\eta) < \infty$. Then $\mathbf{P}_0$ is the least favorable regular parametric submodel and the information bounds and the efficient influence functions for estimating β at $\theta = \theta_0$, $G = G_0$, are as follows*:

Restricted Gaussian error model. *Define the random variable*

$$
l_a^* = \bar{\sigma}^{-1}U\left(T - E(T) + \tilde{\sigma}^2\frac{\omega'}{\omega}(T)\right). \tag{2.7}
$$

This is the efficient score function defined by Begun, Hall, Huang and Wellner (1983). *The information bound of* (1.5), *which we write as I_a, is given by*

$$
\begin{aligned}
I_a &= E_0(l_a^*)^2 = \bar{\sigma}^{-2}\left(\operatorname{Var}(T) + \tilde{\sigma}^4 I_0 - 2\tilde{\sigma}^2\right) \\
&= \bar{\sigma}^{-2}\left(\operatorname{Var}(X') + \tilde{\sigma}^2\left(\tilde{\sigma}^2 I_0 - 1\right)\right)
\end{aligned} \tag{2.8}
$$

and the efficient influence function is given by

$$\tilde{l}_a = l_a^*/I_a. \tag{2.9}$$

General Gaussian error model. *Define*

$$l_b^* = \bar{\sigma}^{-1}U\left(T - E(T) + I_0^{-1}\frac{\omega'}{\omega}(T)\right). \tag{2.10}$$

The information bound is given by

$$\begin{aligned} I_b = E(l_b^*)^2 &= \bar{\sigma}^{-2}\left(\operatorname{Var}(T) - I_0^{-1}\right) \\ &= \bar{\sigma}^{-2}\left(\operatorname{Var}(X') + \tilde{\sigma}^2 - I_0^{-1}\right) \end{aligned} \tag{2.11}$$

and the efficient influence function by

$$\tilde{l}_b = l_b^*/I_b. \tag{2.12}$$

NOTES.

Restricted Gaussian error model.

(1) If $\sigma_{11} = 0$, then $\tilde{\sigma} = 0$ and we are in the case where $T = X = X'$ is observed without error. In this case,

$$I_a = \operatorname{Var}(X')/\operatorname{Var}(Y - \alpha - \beta X)$$

is the reciprocal of the asymptotic variance of $n^{1/2}$ times the ordinary least-squares estimate as it should be.

(2) If X' is normal, $\operatorname{Var}(T) = I_0^{-1}$ and (2.7) becomes

$$\begin{aligned} \bar{\sigma}^{-2}\left(\operatorname{Var}(X') + \tilde{\sigma}^2\left(\tilde{\sigma}^2 - \operatorname{Var}(T)\right)I_0\right) &= \bar{\sigma}^{-2}\left(\operatorname{Var} X'\right)\left(1 - \tilde{\sigma}^2 I_0\right) \\ &= \bar{\sigma}^{-2}\operatorname{Var}^2(X')/\operatorname{Var}_0(T), \end{aligned}$$

which we shall call I_c.

This is just the asymptotic variance of $\hat{\beta}_P$ if $\Sigma_0 =$ identity [see, e.g., Gleser (1981)], whatever be G. So we conclude that we can do as well not knowing G as knowing it is Gaussian. This is a special instance of the claim that $\mathbf{P}_0$ given by (2.6) is least favorable.

(3) We can study the asymptotic efficiency I_c/I_a of $\hat{\beta}_P$ if G_0 is *not* normal. We show in Section 5 that, $I_c/I_a \geq (1 + \sigma^2/(\beta^2 + 1)(\operatorname{Var}(X') + \sigma^2))^{-1}$. In particular, if the signal-to-noise ratio in X, $\operatorname{Var}(X')/\sigma^2$, is large $\hat{\beta}_P$ is close to efficiency.

(4) The score function l_a^* can be written as

$$l_a^* = \bar{\sigma}^{-1}U(E(X'|T) - E(X')).$$

The least-squares estimate if X' were known is based on the score function

$$\bar{\sigma}^{-1}U(X' - E(X')).$$

Thus the efficient estimate replaces the unobservable X' by its best "estimate" $E(X'|T)$.

(5) Suppose that with $\Sigma = \sigma^2\Sigma_0$ we have m repeated observations at each X_i'. Then by sufficiency l_a^*, evaluated at the mean of each set of observations with Σ_0 replaced by Σ_0/m, is the efficient score function.

General Gaussian error model.

(1) Normality of X', under which β is unidentifiable, corresponds to G = point mass at 0. Appropriately, $I_b \to 0$ as G tends to point mass since then T approaches normality and $\tilde{\sigma}^2 \sim I_0^{-1}$.

(2) Necessarily, $I_a \geq I_b$. The inequality is always strict since

$$\tilde{\sigma}^2(I_a - I_b) = I_0^{-1}\left(\tilde{\sigma}^4 I_0^2 - 2\tilde{\sigma}^2 I_0 + 1\right)$$
$$= I_0^{-1}\left(\tilde{\sigma}^2 I_0 - 1\right)^2 > 0,$$

since I_0, the Fisher information for $X' + \varepsilon_1$, is always smaller than the Fisher information for ε_1 which is just $\tilde{\sigma}^{-2}$.

Multiple Gaussian measurements model. The efficient influence function can be calculated as for the general Gaussian error model, but is much more complicated.

Let $\mathbf{X} = (X_j, Y_j)$, $j = 1, \ldots, m$, where $X_j = X' + \varepsilon_{j1}$, $Y_j = \alpha + \beta X' + \varepsilon_{j2}$ is a generic observation. We assume the ε_{ji} are independent Gaussian with mean 0 and $\mathrm{Var}(\varepsilon_{j1}) = \sigma_{11}$, $\mathrm{Var}(\varepsilon_{j2}) = \sigma_{22}$. Let

$$U = (\overline{Y} - \beta\overline{X} - \alpha)/\sigma_0, \tag{2.13}$$
$$T = \left(\sigma_{22}\overline{X} + \beta\sigma_{11}(\overline{Y} - \alpha)\right)/\left(\sigma_{22} + \beta^2\sigma_{11}\right),$$

where $\overline{Y} = m^{-1}\Sigma_{j=1}^m Y_j$, $\overline{X} = m^{-1}\Sigma_{j=1}^m X_j$. Let

$$\sigma_0^2 = \left(\sigma_{22} + \beta^2\sigma_{11}\right)/m,$$
$$\tilde{\sigma}^2 = \sigma_{11}\sigma_{22}/m^2\sigma_0^2, \tag{2.14}$$
$$I_0 = \int\left(\frac{w'}{w}\right)^2 w(t)\,dt, \quad \text{where } w \text{ is the density of } T \text{ given by (2.13).}$$

The efficient score function is then

$$l^* = \frac{UT}{\sigma_0\tilde{\sigma}^2} + a_2\frac{U}{\sigma_0\tilde{\sigma}^2}\frac{\omega'}{\omega}(T) + a_3(U^2 - 1) + a_4 S_1 + a_5 S_2, \tag{2.15}$$

where

$$S_1 = \sum_{j=1}^{m}\frac{\left(Y_j - \overline{Y}\right)^2}{\sigma_{22}} - (m-1), \qquad S_2 = \sum_{j=1}^{m}\frac{\left(X_j - \overline{X}\right)^2}{\sigma_{11}} - (m-1)$$

and the a's are functions of m, σ^2, σ_0^2 and I_0. For $m = 1$ the form of l^* agrees with l_b^* as it should. As $m \to \infty$,

$$a_2 \sim \tilde{\sigma}^2,$$

which corresponds to l_a^*. This is as expected since m large corresponds to σ_{11}, σ_{22} essentially known. The information I_d for this problem is I_b plus a complicated positive term vanishing for $m = 1$.

We now construct efficient estimates. The idea is to proceed as in the classical estimation of the location problem:

(a) Find a good estimate $\tilde{\beta}_n$ of β.

(b) (i) Consider $\tilde{l}$ as $\tilde{l}(\mathbf{x}, \beta, \eta, G)$ where $\theta = (\beta, \eta)$, G are now viewed as dummy variables and the argument $\mathbf{x}$ replaces $\mathbf{X}$. For example,

$$\tilde{l}_a(\mathbf{x}, \theta, G) = \bar{\sigma}^{-1}(\theta) U(\mathbf{x}, \theta)\Big(T(\mathbf{x}, \theta) - \int T(\mathbf{x}, \theta) P_{(\theta, G)}(d\mathbf{x}) + \tilde{\sigma}^2(\theta) \frac{\omega'}{\omega}(T(\mathbf{x}, \theta), \theta)\Big) \Big/ I_a(\theta, G),$$

where T is given by (2.2) and $\omega(\cdot, \theta)$ is the marginal density of $T(\mathbf{X}, \theta)$, under $P_{(\theta, G)}$. Construct a suitable estimate $\hat{\tilde{l}}(\mathbf{x}, \beta; \mathbf{X}_1, \dots, \mathbf{X}_n)$ of $\tilde{l}(\mathbf{x}, \beta, \eta, G)$.

(ii) Form

$$\hat{\beta}_n = \tilde{\beta}_n + n^{-1} \sum_{i=1}^{n} \hat{\tilde{l}}(\mathbf{X}_i, \tilde{\beta}_n; \mathbf{X}_1, \dots, \mathbf{X}_n)$$

as the efficient estimate.

Preliminary estimate. We motivate our $\tilde{\beta}_n$ as follows. If we calculate under P_0 and $\beta = \beta_0$, $\mathrm{Var}(Y) \geq \mathrm{Var}(\beta X)$, then

$$\mathbf{L}(Y) = \mathbf{L}(\beta X + \sigma Z + \mu), \tag{2.16}$$

for $Z \sim \mathbf{N}(0, 1)$ independent of X and

$$\mu = E(Y) - \beta E(X),$$
$$\sigma^2 = \mathrm{Var}(Y) - \beta^2 \mathrm{Var}(X).$$

If $\mathrm{Var}(Y) < \mathrm{Var}(\beta X)$, then

$$\mathbf{L}(X) = \mathbf{L}\left(\frac{Y}{\beta} + \sigma Z + \mu\right), \tag{2.17}$$

for $Z \sim \mathbf{N}(0, 1)$ independent of Y, some σ, μ. For $|\beta| \neq |\beta_0|$ neither identity (2.16) nor (2.17) can hold; see Proposition 5.1. Our initial estimate is essentially a minimizing value for the distance between the natural estimates of the laws in (2.16) or (2.17). We believe our estimate may be improved by considering the joint distribution of (X, Y) and not only the marginals. For that note that if (2.16) holds, then

$$\mathbf{L}(\beta X + \sigma Z + \mu, Y) = \mathbf{L}(Y, \beta X + \sigma Z + \mu).$$

Another possible estimate is given by Spiegelman (1979) who does not assume Gaussianity of the errors but does assume ε_1, ε_2 independent. Different estimates $\tilde{\beta}_a$, $\tilde{\beta}_b$ are appropriate for the restricted Gaussian error model and the general Gaussian error model. Essentially, $\tilde{\beta}_b$ works whenever $\tilde{\beta}_a$ does except when G is

Gaussian. We give $\tilde{\beta}_b$ formally and sketch the difference for $\tilde{\beta}_a$. Without loss of generality, we assume $E(\varepsilon_1) = E(\varepsilon_2) = 0$.

Let $\hat{F}_1$ be the empirical distribution function of X_i, $i = 1, \dots, n$, and $F_1(\cdot)$ be the distribution function of X. Let $\hat{F}_2(\cdot)$ and $F_2(\cdot)$ be the empirical distribution function of Y_i and the distribution function of Y, respectively. Let

$$\hat{\mu}(\beta) = \bar{Y} - \beta\bar{X}, \qquad \hat{\sigma}^2(\beta) = |\hat{\sigma}_y^2 - \beta^2\hat{\sigma}_x^2|,$$

$$\hat{\sigma}_y^2 = n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2, \qquad \hat{\sigma}_x^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \qquad \lambda = \hat{\sigma}_y/\hat{\sigma}_x. \tag{2.18}$$

Define, for $\hat{\sigma}_x^2 > 0$, $\hat{\sigma}_y^2 > 0$,

$$\Delta_n(\beta) = \sqrt{n}\int\left|\hat{F}_2(y) - \int\Phi\left(\frac{y - \beta x - \hat{\mu}(\beta)}{\hat{\sigma}(\beta)}\right)d\hat{F}_1(x)\right|^2\phi(y)\,dy, \qquad \text{if } \hat{\sigma}_y^2 > \beta^2\hat{\sigma}_x^2$$

$$= \sqrt{n}\int\left|\hat{F}_1(x) - \int\Phi\left(\frac{\beta x - y + \hat{\mu}(\beta)}{(\operatorname{sgn}\beta)\hat{\sigma}(\beta)}\right)d\hat{F}_2(y)\right|^2\lambda\phi(\lambda y)\,dy, \qquad \text{if } \hat{\sigma}_y^2 < \beta^2\hat{\sigma}_x^2. \tag{2.19}$$

Note that $\Delta_n(\beta)$ can be defined by continuity at $\sigma(\beta) = 0$ since $P[|\beta| + \hat{\sigma}^2(\beta) > 0, \forall\beta] = 1$. For given $a > 0$, let $\Delta_n(\beta, a)$ be the corresponding quantity with Y_i replaced by $Y_i + aX_i$, $i = 1, \dots, n$. Let $\beta_n^*(a)$ minimize $\Delta_n(\beta, a)$. $\beta_0 = 0$ poses difficulties but we can always shift away from this value. Accordingly, let

$$\begin{aligned}\beta_n^* &= \beta_n^*(0), && \text{if } |\beta_n^*(0)| \geq \delta_0\\ &= \beta_n^*(2\delta_0) - 2\delta_0, && \text{if } |\beta_n^*(0)| < \delta_0.\end{aligned}$$

Finally, we need to distinguish between $\pm\beta_n^*$. For that let $\hat{W}_n^+$ be the empirical distribution function of $\hat{\sigma}^{-1}(Y_i - \mu(\beta_n^*) - \beta_n^*X_i)$, where

$$\hat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(Y_i - \mu(\beta_n^*) - \beta_n^*X_i)^2$$

and $\hat{W}_n^-$ the corresponding quantity for $-\hat{\beta}_n^*$. Let

$$\begin{aligned}\tilde{\beta} &= \beta_n^*, \quad \text{if } \int|\hat{W}_n^+(y) - \Phi(y)|^2\phi(y)\,dy \leq \int|W_n^-(y) - \Phi(y)|^2\phi(y)\,dy\\ &= -\beta_n^*, \quad \text{otherwise.}\end{aligned}$$

For the restricted Gaussian error model, $\Sigma_0 =$ identity we proceed as above but change the definition of $\hat{\sigma}^2(\beta)$ to, using the new information,

$$\hat{\sigma}_a^2(\beta) = \frac{|1 - \beta^2|}{1 + \beta^2}n^{-1}\sum_{i=1}^{n}(Y_i - \hat{\mu}(\beta) - \beta X_i)^2$$

and switch the definition of $\Delta_n(\beta)$ as $\beta^2 \leq 1$ or > 1.

EFFICIENT ESTIMATES. Note that

$$\beta\sigma_{11} - \sigma_{12} = \beta\,\mathrm{Var}(X) - \mathrm{cov}(X, Y), \tag{2.20}$$

$$\sigma_{22} - \beta\sigma_{12} = \mathrm{Var}(Y) - \beta\,\mathrm{cov}(X, Y), \tag{2.21}$$

$$\alpha = E(Y) - \beta E(X). \tag{2.22}$$

We can reparametrize the general Gaussian error model using $(\beta, \alpha, \gamma_1, \gamma_2, \sigma_{11}, G)$, where $\gamma_1, \gamma_2, \alpha$ are the expressions in (2.20)–(2.22), respectively. Abusing notation, let $\theta = (\beta, \alpha, \gamma_1, \gamma_2)$ so that

$$U_i(\theta) = (Y_i - \alpha - \beta X_i)/(\beta\gamma_1 + \gamma_2)^{1/2},$$

$$T_i(\theta) = \big(\gamma_2 X_i + \gamma_1(Y_i - \alpha)\big)/(\beta\gamma_1 + \gamma_2).$$

Define $\tilde{\theta}_n = (\tilde{\beta}_n, \tilde{\alpha}_n, \tilde{\gamma}_{1n}, \tilde{\gamma}_{2n})$ by substituting sample moments and $\tilde{\beta}_n$ in the definitions (2.20)–(2.22) for $\beta, \alpha, \gamma_1, \gamma_2$. Let

$$\lambda(t) = e^{-t}(1 + e^{-t})^2,$$

$$\lambda_\nu(t) = \frac{1}{\nu}\lambda\left(\frac{t}{\nu}\right).$$

For sequences c_n, $\nu_n \downarrow 0$, to be characterized later, let $\lambda_n = \lambda_{\nu_n}$ and estimate ω_0 by the kernel estimator,

$$\hat{\omega}_n(t, \theta) = \frac{1}{n}\sum_{i=1}^{n}\lambda_\nu(t - T_i(\theta)) + c_n.$$

Define the efficient estimate for the general Gaussian error model by

$$\tilde{\beta}_{nb} = \tilde{\beta}_n + n^{-1}\hat{I}_b^{-1}\sum_{i=1}^{n}\frac{\tilde{U}_i}{\sigma(\tilde{\theta}_n)}\left(\tilde{T}_i - \tilde{T}. + \hat{I}_0^{-1}\frac{\hat{\omega}_n'}{\hat{\omega}_n}(\tilde{T}_i, \bar{\theta}_n)\right), \tag{2.23}$$

where $\tilde{U}_i$, $\tilde{T}_i$ are used for $U_i(\tilde{\theta}_n)$, $T_i(\tilde{\theta}_n)$, and $\tilde{T}. = n^{-1}\Sigma_{i=1}^n \tilde{T}_i$,

$$\hat{I}_0 = n^{-1}\sum_{i=1}^{n}\left(\frac{\hat{\omega}_n'}{\hat{\omega}_n}\right)^2\big(T_i(\tilde{\theta}_n), \tilde{\theta}_n\big), \tag{2.24}$$

$$\hat{I}_b = \big(\tilde{\beta}_n\tilde{\gamma}_{12} + \tilde{\gamma}_{2n}\big)^{-1}n^{-1}\sum_{i=1}^{n}\left(\tilde{T}_i - \tilde{T}. + \hat{I}_0^{-1}\frac{\omega_n'}{\omega_n}(\tilde{T}_i, \tilde{\theta}_n)\right)^2. \tag{2.25}$$

Similarly, we define the efficient estimate $\hat{\beta}_{na}$ for the restricted Gaussian error model by

$$\hat{\beta}_{na} = \tilde{\beta}_{na} + n^{-1}\hat{I}_a^{-1}\sum_{i=1}^{n}\frac{\tilde{U}_i}{\hat{\sigma}_n}\left(\tilde{T}_{ia} - \tilde{T}_{\cdot a} + \big(1 + \tilde{\beta}_n^2\big)^{-1}\hat{\sigma}_n^2\frac{\hat{\omega}_n'}{\hat{\omega}_n}(\tilde{T}_{ia}, \tilde{\theta}_n)\right),$$

where

$$\hat{\sigma}_n^2 = \left(1 + \tilde{\beta}_n^2\right)^{-1} n^{-1} \sum_{i=1}^{n} \left(Y_i - \mu(\tilde{\beta}_n) - \tilde{\beta}_n X_i\right)^2,$$

$$\tilde{T}_{ia} = \left(\tilde{\beta}_n(Y_i - \tilde{\alpha}_n) + X_i\right)\left(1 + \tilde{\beta}_n^2\right)^{-1},$$

$$\hat{I}_a = \hat{\sigma}_n^{-2} n^{-1} \sum_{i=1}^{n} \left(\tilde{T}_{ia} - \tilde{T}_{\cdot a} + \left(1 + \tilde{\beta}_n^2\right)^{-1} \hat{\sigma}_n^2 \frac{\omega_n'}{\omega_n}\left(\tilde{T}_{ia}, \tilde{\theta}_n\right)\right)^2,$$

in accordance with (2.7) and (2.8).

Let $\{c_n\}$, $\{\nu_n\}$ be such that

$$c_n \to 0, \qquad \nu_n \to 0, \qquad n c_n^2 \nu_n^6 \to \infty.$$

THEOREM 2.2. (i) *Suppose* G_0 *is non-Gaussian,* $\int x^2\, dG_0(x) < \infty$ *and* $\mathbf{P}_0 = \{P_{(\theta, G_0)}: \theta \in \Theta\}$ *is regular. Then, if* $P_0 = P_{(\theta_0, G_0)}$ *satisfies the general Gaussian error model,*

$$\mathbf{L}_{P_0}\left(n^{1/2}\left(\hat{\beta}_{bn} - \beta(P_0)\right)\right) \to \mathbf{N}\left(0, I_b^{-1}(P_0)\right), \tag{2.26}$$

for all $P_0 \in \mathbf{P}_0$.

(ii) *If also* $n\nu_n^{-6}\log n \to 0$, *the convergence in* (2.26) *continues to hold if* P_0 *is replaced by* $P_n = P_{(\theta_n, G_n)}$, *where*

$$\theta_n = \left(\beta_n, \alpha_n, \gamma_{1n}, \gamma_{2n}, \sigma_{11n}\right) \to \theta = \left(\beta, \alpha, \gamma_1, \gamma_2, \sigma_{11}\right)$$

and $G_n \to G$ *weakly and* $\int z^2 G_n(dz) \to \int z^2 G(dz) < \infty$.

(iii) *Write* (2.21)–(2.23) *as* $\hat{\beta}_n = \hat{\beta}_n(\tilde{\beta}_n)$ *and let* $\hat{\beta}_{0n} = \tilde{\beta}_n$, $\hat{\beta}_{in} = \hat{\beta}_n(\hat{\beta}_{i-1,n})$, $i = 1, 2, 3, \ldots$. *Then, for* $i \geq 1$, *all* $\hat{\beta}_{in}$ *are efficient and* $|\hat{\beta}_{in} - \hat{\beta}_{i-1,n}| = o_p(n^{-1/2})$ *for all* $i \geq 2$.

(iv) *If* $\hat{\beta}_n$ *is replaced by* $\hat{\beta}_{an}$ *and the restricted Gaussian error model is considered then claims* (i)–(iii) *continue to hold with* I_b *replaced by* I_a.

NOTES.

(1) Let $K \subset \mathbf{P}$ be compact in the total variation norm topology. Part (ii) of the theorem shows that the convergence in (2.24) is uniform over K if $P \to I_b(P)$ is continuous on K. These are the largest sets over which we may expect uniform convergence.
(2) Part (iii) of the theorem may be interpreted in terms of running the iteration $\hat{\beta}_{in}$ to convergence. Suppose the stopping rule is of the form: Stop as soon as $|\hat{\beta}_{in} - \beta_{i-1,n}| \leq \varepsilon_n$, where $\varepsilon_n \downarrow 0$, $n^{1/2}\varepsilon_n > c > 0$. This is reasonable since the random fluctuations in the estimate are of order $n^{-1/2}$. Then, by part (iii), with probability tending to 1 the iteration stops with $\hat{\beta}_{2n}$.

Under more stringent conditions on ν_n, c_n we conjecture that tedious calculations will show that, in fact, $\lim_i \hat{\beta}_{in}$ exists with probability tending to 1 and is efficient.

3. Information bounds and proof of Theorem 2.1. Let $\mathbf{P}_0$ be a regular parametric submodel of a model $\mathbf{P}$ written in the form $\{P_{(\beta,\gamma)}: \beta \in R, \gamma \in E \subset R^k\}$. Let $l(X, \beta, \gamma)$ denote the log likelihood of an observation from $P_{(\beta,\gamma)}$ and let $\dot{l}_0(X) = \partial l/\partial\beta|_{(\beta_0,\gamma_0)}$, $\dot{l}_j(X) = \partial l/\partial\gamma_j|_{(\beta_0,\gamma_0)}$, $1 \le j \le k$, where $\gamma = (\gamma_1, \dots, \gamma_k)$. Begun, Hall, Huang and Wellner (1983) [see also Efron (1977) and Neyman (1957)] show (in slightly different terms) that, if $P_0 = P_{(\beta_0,\gamma_0)}$

$$I(P_0; \beta, \mathbf{P}_0) = \min\left\{E\left(\dot{l}_0(X) - \sum_{j=1}^{k} c_j\dot{l}_j(X)\right)^2 : (c_1, \dots, c_k) \in R^k\right\}$$
$$= E\{[l^*]^2(X)\},$$

where

$$l^* = \dot{l}_0 - \sum_{j=1}^{k} c_j^*\dot{l}_j, \tag{3.1}$$

and the c_j^* are uniquely determined by the orthogonality condition

$$El^*\dot{l}_j(X) = 0, \qquad j = 1, \dots, k. \tag{3.2}$$

Moreover, the efficient influence function for $\mathbf{P}_0$ is given by

$$\tilde{l}(X, P_0|\beta, \mathbf{P}_0) = l^*(X)/I(P_0; \beta, \mathbf{P}_0). \tag{3.3}$$

Therefore, to calculate $\tilde{l}$ for $\mathbf{P}_0$ we need only calculate the projection $\sum_{j=1}^{k} c_j^*\dot{l}_j(X)$, in $L_2(P_0)$, of $\dot{l}_0$ into $[\dot{l}_j: 1 \le j \le k]$, the linear span of $\dot{l}_1, \dots, \dot{l}_k$. Let $\Pi(h|L)$ denote the projection of $h \in L_2(P_0)$ into a closed linear space $L \subset L_2(P_0)$.

To prove Theorem 2.1 we go through the following steps for the restricted Gaussian error model and an analogous series for the general Gaussian error model.

(i) Identify $(\gamma_1, \gamma_2) = (\alpha, \sigma^2)$, where σ^2 is given by (1.4) and let $\eta = (\eta_1, \dots, \eta_{k-2})$ index G, i.e.,

$$\mathbf{P}_0 = \left\{P_{(\theta, G_\eta)}: \eta \in E, \theta = (\alpha, \beta, \sigma^2), \alpha, \beta \in R\right\}.$$

Calculate formally $\dot{l}_j$, $0 \le j \le k$, at $P_0 = P_{(\theta_0, G_{\eta_0})}$, where $j = 0 \leftrightarrow \beta$, $j = 1,2 \leftrightarrow \alpha, \sigma^2$, $j \ge 3 \leftrightarrow \eta$.

We project $\dot{l}_0$ into $[\dot{l}_j: j \ge 1]$ in two steps. First, calculate, for $0 \le j \le 2$, $\Pi(\dot{l}_j|V)$, where

$$\begin{aligned} V &= \left[\dot{l}_j: j \ge 3\right], \\ l^* &= \dot{l}_0 - \Pi\left(\dot{l}_0|V\right) - \Pi\left(\dot{l}_0 - \Pi\left(\dot{l}_0|V\right)|W\right), \end{aligned} \tag{3.4}$$

where

$$W = \left[\dot{l}_j - \Pi\left(\dot{l}_j|V\right): 1 \le j \le 2\right].$$

Claim (3.4) is well known and can be verified by checking (3.2). We establish that:

(ii) For any regular parametric submodel $\mathbf{P}_0$

$$[\dot{l}_j\colon j \geq 3] \subset \{a(T)\colon a(T) \in L_2(P_0),\ Ea(t) = 0\}$$

and then prove:

(iii) If $\mathbf{P}_0$ is given by (2.6), then $\mathbf{P}_0$ is regular and

$$[\dot{l}_j\colon j \geq 3] \supset [E(\dot{l}_0(X)|T)]. \tag{3.5}$$

The existence of a model $\mathbf{P}_0$ having property (3.5), but not the specific choice (2.6), follows from Theorem 14.3.12 of Pfanzagl (1982). Note that

$$E(h(X) - E(h(X)|T))a(T) = 0, \quad \text{for all } a(T),\ h \in L_2(P_0). \tag{3.6}$$

Now (ii) and (iii) imply that, for $\mathbf{P}_0$ given by (2.6),

$$\Pi(\dot{l}_i|V) = E(\dot{l}_i(X)|T), \qquad 0 \leq i \leq 2,$$

and hence by (3.4) if l_0^* is the l^* of $\mathbf{P}_0$ given by (2.6),

$$l_0^*(X) = \dot{l}_0(X) - E(\dot{l}_0(X)|T) - \sum_{j=1}^{2} d_j(\dot{l}_j(X)) - E(\dot{l}_j(X)|T), \tag{3.7}$$

with $\{d_j\colon 1 \leq j \leq 2\}$ determined by (3.2) for $j = 1, 2$. Take $\mathbf{P}_0$ to be any regular parametric submodel. By (ii) and (3.6)

$$El_0^*(X)\dot{l}_j(X) = 0, \qquad j \geq 3.$$

By (3.2)

$$El_0^*(X)\dot{l}_j(X) = 0, \qquad j = 1, 2.$$

Therefore,

$$\begin{aligned} &E(l^*(X))^2 - E(l_0^*(X))^2 \\ &\quad = E(l^*(X) - l_0^*(X))^2 + 2E(l_0^*(X)(l^* - l_0^*)(X)) \\ &\quad = E(l^*(X) - l_0^*(X))^2 \geq 0, \end{aligned} \tag{3.8}$$

since $l^* - l_0^* \in [\dot{l}_j\colon j \geq 1]$. We conclude that $\mathbf{P}_0$ given by (2.6) is least favorable.

Proof of Theorem 2.1. For mnemonic convenience we write $\dot{l}_0 = l_\beta$ and $\dot{l}_j = l_\alpha,\ l_{\sigma^2},\ l_{\sigma_{11}}$, etc., as appropriate.

Restricted Gaussian error model. (i) Differentiating (2.4) we get, for $\theta = \theta_0$, $G = G_0$,

$$\begin{aligned} l_\beta(\mathbf{X}) &= p^{-1}(\mathbf{X}, \theta, G)\int(\sigma_{11}\sigma_{22} - \sigma_{12}^2)^{-1}(\sigma_{11}(Y - \alpha - \beta z) - \sigma_{12}(X - z)) \\ &\qquad \times zK(\mathbf{X}, z, \theta)G(dz) \\ &= \tilde{\sigma}^{-1}(\theta)\int\left(\frac{U}{\sigma(\theta)} + \frac{\beta\sigma_{11} - \sigma_{12}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}(T - z)\right) \\ &\qquad \times z\phi(\tilde{\sigma}^{-1}(\theta)(T - z))G(dz)/\omega(T), \end{aligned} \tag{3.9}$$

since

$$X = T - \tilde{\sigma}^{-1}(\theta)(\beta\sigma_{11} - \sigma_{12})U,$$
$$Y - \alpha = \beta T + \tilde{\sigma}^{-1}(\theta)(\sigma_{22} - \beta\sigma_{12})U.$$

Similarly,

$$\begin{aligned} l_\alpha &= [\omega(T)\tilde{\sigma}(\theta)]^{-1} \\ &\quad \times \int \left(\frac{U}{\sigma(\theta)} + \frac{\beta\sigma_{11} - \sigma_{22}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}(T - z) \right) \phi(\tilde{\sigma}^{-1}(\theta)(T - z))G(dz), \end{aligned} \tag{3.10}$$

$$\begin{aligned} l_{\sigma^2} &= \frac{1}{2\sigma^2}\Big((U^2 - 1) + \tilde{\sigma}^{-1}(\theta) \\ &\quad \times \int \left(\tilde{\sigma}^{-2}(\theta)(T - z)^2 - 1 \right) \phi(\tilde{\sigma}^{-1}(\theta)(T - z))G(dz)/\omega(T)\Big). \end{aligned} \tag{3.11}$$

(ii) Suppose $\mathbf{P}_0 = \{P_{(\theta, G_\eta)}\}$ is a regular submodel with $G_\eta \ll G_0 = G$. If $g_\eta = dG_\eta/dG$, $g_0 = 1$, and, formally,

$$\dot{l}_{j+2}(X) = \int \exp\left\{ -\frac{\tilde{\sigma}^{-2}}{2}(T - z)^2 \right\} \frac{\partial g_\eta}{\partial \eta_j}(z)G(dz)/\omega(T), \tag{3.12}$$

a function of T only. If $\dot{l}_{j+2}$ exists only in the Hellinger sense it is easy to check that $\dot{l}_{j+2}$ is an L_2 limit of functions of T and hence T measurable.

(iii) If $\mathbf{P}_0$ is given by (2.6),

$$\begin{aligned} \frac{\partial l}{\partial \mu}(X,\theta,G_\eta)\Big|_{\mu=0,\,\tau=1} &= \omega^{-1}(T)\frac{\partial}{\partial \mu}\int \exp\left\{ -\frac{\tilde{\sigma}^{-2}}{2}\left(T - \frac{(z - \mu)}{\tau} \right)^2 \right\} G(dz) \\ &= \omega^{-1}(T)\int (T - z)\exp\left\{ -\frac{\tilde{\sigma}^{-2}}{2}(T - z)^2 \right\} G(dz), \end{aligned} \tag{3.13}$$

$$\frac{\partial l}{\partial \tau}(X, \theta, G_\eta)\Big|_{\mu=0,\,\tau=1} = \omega^{-1}(T)\int z(T - z)\exp\{ -\tilde{\sigma}^{-2}(T - z)^2 \} G(dz). \tag{3.14}$$

The independence of U and T and $EU = 0$ yield from (3.9)

$$E(l_\beta | T) = \tilde{\sigma}^{-1}\frac{\beta\sigma_{11} - \sigma_{12}}{\sigma_{11}\sigma_{22} - \sigma_{12}^2}\int z(T - z)\phi(\tilde{\sigma}^{-1}(T - z))G(dz)/\omega_0(T),$$

which is proportional to $\partial l/\partial \tau$ as required. Therefore,

$$l_\beta - E(l_\beta | T) = \tilde{\sigma}^{-1}U\left[\int \phi(\tilde{\sigma}^{-1}(T - z))G(dz) \right]^{-1} \int z\phi(\tilde{\sigma}^{-1}(T - z))G(dz).$$

From (2.5)

$$l_\beta - E(l_\beta | T) = \tilde{\sigma}^{-1}U\left(T + \tilde{\sigma}^2\frac{\omega'}{\omega}(T) \right). \tag{3.15}$$

Similarly,

$$l_\alpha - E(L_\alpha|T) = \bar{\sigma}^{-1}U,$$

$$l_{\sigma^2} - E(l_{\sigma^2}|T) = \frac{1}{2\bar{\sigma}^2}(U^2 - 1).$$

Now, from (3.10) and (3.13)

$$l_\alpha - \Pi(l_\alpha|V) = l_\alpha - E(l_\alpha|T) \tag{3.16}$$

and necessarily by (ii)

$$l_{\sigma^2} - \Pi(l_{\sigma^2}|V) = l_{\sigma^2} - E(l_{\sigma^2}|T) + b(T). \tag{3.17}$$

Therefore, (3.17) is orthogonal to both (3.16) and (3.15) so that $d_2 = 0$. On the other hand, it is easy to see that $d_1 = E(T)$. From (3.7), (3.15) and (3.16) we obtain Theorem 2.1 for a restricted Gaussian model.

General Gaussian error model. We find after some computation

$$\begin{aligned} l_{\sigma_{11}} &= \alpha_{11}(U^2 - 1) + \beta_{11}U\frac{\omega'}{\omega}(T) + \gamma_{11}b(T), \\ l_{\sigma_{22}} &= \alpha_{22}(U^2 - 1) + \beta_{22}U\frac{\omega'}{\omega}(T) + \gamma_{22}b(T), \\ l_{\sigma_{12}} &= \alpha_{12}(U^2 - 1) + \beta_{12}U\frac{\omega'}{\omega}(T) + \gamma_{12}b(T), \end{aligned} \tag{3.18}$$

where

$$b(T) = \tilde{\sigma}^{-1}\int z^2\phi(\tilde{\sigma}^{-1}(T - z))G(dz)/\omega(T),$$

and the matrix $\begin{pmatrix} \alpha_{11} & \beta_{11} \\ \alpha_{22} & \beta_{22} \\ \alpha_{12} & \beta_{12} \end{pmatrix}$ has dimension 2. Let $V = [l_\mu(\mathbf{x}), l_\tau(\mathbf{x})]$.

From (3.18) the linear span of $l_\alpha - E(l_\alpha|T)$, $l_{\sigma_{ij}} - \Pi(l_{\sigma_{ij}}|V)$, $i, j = 1, 2$, is

$$\left[U, U^2 - 1, U\frac{\omega'}{\omega}(T), c(T)\right], \tag{3.19}$$

where $c(T) = \Pi(b(T)|V)$. We find the projection of $l_\beta - E(l_\beta|T)$ on (3.19) by using the independence of U and T, $EU = 0$, $EU^2 = 1$. We obtain

$$\begin{aligned} &\tilde{\sigma}^{-1}(\Pi(UT|[U])) + \Pi\left(UT\middle|\left[U\frac{\omega'}{\omega}(T)\right]\right) + \tilde{\sigma}^2U\frac{\omega'}{\omega}(T) \\ &= \tilde{\sigma}^{-1}UE(T) + \left(\tilde{\sigma}^2 - \frac{1}{I_0}\right)U\frac{\omega'}{\omega}(T), \end{aligned}$$

since $E(T(\omega'/\omega)(T) = -1$. We conclude that under the submodel (2.6), with Σ varying freely, l_b^* is the efficient score function. But clearly, $El_b^*(\mathbf{X})a(T) = 0$ for all $a(T) \in L_2(P_0)$ and, in view of (ii), the argument leading to (3.8) applies to l_b^* also and (2.6) is least favorable. □

4. Proof of Theorem 2.2 and miscellaneous results. We begin by studying β_n^*.

PROPOSITION 4.1. *If either*

$$\mathbf{L}_{P_0}(Y) = \mathbf{L}_{P_0}(\beta X) * \mathbf{N} \tag{4.1}$$

or

$$\mathbf{L}_{P_0}(X) = \mathbf{L}_{P_0}(Y/\beta) * \mathbf{N} \tag{4.2}$$

(where $\mathbf{N}$ *is a Gaussian law and* $*$ *denotes convolution), then* $|\beta| = |\beta_0|$ *or* G_0 *is Gaussian. If* $\beta = \beta_0$ *one of these relations holds.*

PROOF. Let ψ be the characteristic function of X'. The case $\beta_0 = 0$ is simple. Assume $\beta_0 \neq 0$. Without loss of generality, take $E_0(X) = E_0(Y) = 0$ and $\beta_0 = 1$. Suppose $|\beta| \neq 1$ and without loss of generality, take $|\beta| > 1$. Then (4.1) becomes

$$\psi(t) = \psi(\beta t) e^{at^2}, \tag{4.3}$$

for some a. Iterating (4.3) we get for all k, t

$$\psi(\beta^k t) = \exp\left(-at^2 \frac{(\beta^{2k} - 1)}{(\beta^2 - 1)}\right) \psi(t).$$

Putting $u = \beta^k t$ and letting $k \to \infty$,

$$\psi(u) = \exp\left(-au^2(\beta^2 - 1)^{-1}(1 + o(1))\right)(1 + o(1))$$

and we get G_0 Gaussian. The same argument works for (4.2). □

PROPOSITION 4.2. *Suppose that* $\mathbf{P}$ *consists of all probabilities satisfying the general Gaussian error model with* $\int x^2\, dG(x) < \infty$. *Then for every* $P_0 \in \mathbf{P}$

$$\lim_{M \to \infty} \lim_{n \to \infty} P_0\left[\sqrt{n}\, \tilde{\beta}_n - \beta(P_0)| \geq M\right] = 0.$$

PROOF. Let

$$Z_n(y, \beta) = \sqrt{n}\left\{(\hat{F}_2(y) - F_2(y)) - \int \Phi\left(\frac{y - \beta x - \mu(\beta)}{\sigma(\beta)}\right) d(\hat{F}_1(x) - F_1(x))\right\}$$

$$= \sqrt{n}\left\{(\hat{F}_2(y) - F_2(y)) + \operatorname{sgn}\beta \int (\hat{F}_1((y - \mu(\beta) - z\sigma(\beta))/\beta) - F_1((y - \mu(\beta) - z\sigma(\beta))/\beta))\phi(z)\, dz\right\}, \tag{4.4}$$

where F_1, F_2 are the marginal distribution functions of X and Y under P_0 and $\mu(\beta)$, $\sigma(\beta)$ are obtained by substituting population for sample moments in (2.18) and (2.19). By strong approximation, e.g., Csörgő (1981), we can construct $Z(\cdot, \cdot)$, a mean 0 Gaussian process in $C([-\infty, \infty] \times [-\infty, \infty])$ such that

$$\sup_{y, \beta} |Z_n(y, \beta) - Z(y, \beta)| = o_p(1). \tag{4.5}$$

Let $\hat{Z}_n(\cdot,\cdot)$ be defined by replacing $\mu(\beta), \sigma(\beta)$ by $\hat{\mu}(\beta), \hat{\sigma}(\beta)$ in (4.4). For $\sigma(\beta) \geq \varepsilon$, the family of functions $x \to \Phi((y - \beta x - \mu(\beta))/\sigma(\beta))$ is uniformly bounded and equicontinuous. Moreover,

$$\sup\{\sigma^{-1}(\beta)\beta\phi((y - \beta x - \mu(\beta)/\sigma(\beta)) - \hat{\sigma}(\beta)\beta\phi(y - \beta x - \mu(\hat{\beta}))/\sigma(\hat{\beta})) \colon \sigma(\beta) \geq \varepsilon\} \to_P 0.$$

From (4.4) we then conclude that

$$\sup_y \{|\hat{Z}_n(y,\beta) - Z_n(y,\beta)| \colon \sigma(\beta) \geq \varepsilon\} \to_P 0.$$

Now there exist $\varepsilon, \delta > 0$ such that $\inf\{\sigma(\beta)\colon |\beta| \leq \delta\} \geq \varepsilon$ and so

$$\sup_y \{|\hat{Z}_n(y,\beta) - Z_n(y,\beta)| \colon |\beta| \leq \delta\} \to_P 0.$$

On the other hand, from (4.5)

$$\sup_y \{|\hat{Z}_n(y,\beta) - Z_n(y,\beta)| \colon \delta \leq |\beta|\} \to_P 0$$

and so

$$\sup\{|\hat{Z}_n(y,\beta) - Z(y,\beta)|\} \to_P 0. \tag{4.6}$$

Similarly,

$$\sup\{|\hat{Z}_n^*(x,\beta) - Z^*(x,\beta)|\} \to_P 0, \tag{4.7}$$

where

$$\hat{Z}_n^*(x,\beta) = \sqrt{n}\left(\hat{F}_1(x) - F_1(x) - \int\Phi\left(\frac{\beta x - y + \hat{\mu}(\beta)}{\hat{\sigma}(\beta)}\right) d(\hat{F}_2(y) - F_2(y))\right)$$

and Z^* is an appropriately defined Gaussian process. A weak consequence of (4.6) and (4.7) is that for all $\varepsilon > 0$,

$$\inf\{\Delta_n(\beta)\colon \varepsilon \leq |\beta^2 - \beta_0^2|\} \to_P \infty$$

and

$$\Delta_n(\beta_0) = O_p(1).$$

Therefore, by Proposition 4.1

$$\min\{|\beta_n^*(0) - \beta_0|, |\beta_n^*(0) + \beta_0|\} \to_P 0.$$

Since $Y - \mu(\beta) - \beta X$ is normal if and only if $\beta = \beta_0$, we conclude that $\tilde{\beta}_n$ is consistent.

We need to distinguish several cases for $n^{1/2}$-consistency:

(a) $|\beta_0| \geq \frac{3}{2}\delta_0$, $\sigma^2(\beta_0) > 0$;
(b) $|\beta_0| \geq \frac{3}{2}\delta_0$, $\sigma^2(\beta_0) = 0$;
(c) $\frac{1}{2}\delta_0 \leq |\beta_0| < \frac{3}{2}\delta_0$;
(d) $|\beta_0| \leq \frac{1}{2}\delta_0$.

(a) Suppose also that $\mathrm{Var}(Y) > \beta_0^2\mathrm{Var}(X)$. Then, by (4.4) and (4.5)

$$\Delta_n(\beta) = \int\left|\sqrt{n}\left(F_2(y) - \int\Phi\left(\frac{y-\beta x-\mu(\beta)}{\sigma(\beta)}\right)dF_1(x)\right) + Z(y,\beta)\right|^2\phi(y)\,dy$$
$$+Q_n(\beta),$$

where

$$\sup\{|Q_n(\beta)|: |\beta-\beta_0| \le \varepsilon_n\} = o_p(1).$$

Now, under these conditions,

$$\frac{\partial}{\partial\beta}\int\Phi\left(\frac{y-\beta x-\mu(\beta)}{\sigma(\beta)}\right)dF_1(x)$$
$$= -\sigma^{-1}(\beta)\int\phi\left(\frac{y-\beta x-\mu(\beta)}{\sigma(\beta)}\right)(x - E(X) - \beta\sigma^{-2}(\beta)$$
$$\times(y-\beta x-\mu(\beta))\mathrm{Var}\,X)\,dF_1(x),$$

$$\frac{\partial}{\partial\beta}\int\Phi\left(\frac{y-\beta x-\mu(\beta)}{\sigma(\beta)}\right)dF_1(x)\bigg|_{\beta_0} \tag{4.8}$$
$$= -\beta_0^{-1}((y-EY)f_2(y) + \mathrm{Var}\,Yf_2'(y)),$$

which cannot vanish identically as a function of y unless Y is normal (i.e., $\beta_0 = 0$ or G_0 is normal). Moreover, the derivative in (4.8) is bounded as a function of y and continuous in β. We can conclude that $\tilde{\beta}_n$ is $n^{1/2}$-consistent in this case. This follows since $\Delta_n(\beta_0) = O_p(1)$ and

$$\Delta_n(\tilde{\beta}_n) \ge \int\left(Z(y,\beta_0) + n^{1/2}(\beta_n - \beta_0)c(y)\right)^2\phi(y)\,dy + o_p(1),$$

where $c(y)$ is the derivative in (4.6). Unboundedness of $n^{1/2}(\tilde{\beta}_n - \beta_0)$ leads to a contradiction since $c(y)$ does not vanish identically.

Case (a) with $\mathrm{Var}(Y) < \beta_0^2\mathrm{Var}(X)$ is dealt with similarly using Z^*.

(b) If $\sigma(\beta_0) = 0$, calculate (taking $\beta_0 > 0$)

$$\lim_{\beta\to\beta_0}(\beta-\beta_0)^{-1}\left[\int\phi\left(\frac{y-\beta x-\mu(\beta)}{\sigma(\beta)}\right)dF_1(x) - F_1\left(\frac{y-\mu(\beta_0)}{\beta_0}\right)\right]$$
$$= \lim_{\beta\to\beta_0}(\beta-\beta_0)^{-1}\int(F_1((y-\mu(\beta)$$
$$-z\sigma(\beta))/\beta) - F_1((y-\mu(\beta_0))/\beta_0))\,d\Phi(z)$$
$$= -\frac{y-EY}{\beta_0^2}f_1((y-\mu(\beta_0))/\beta_0) - \mathrm{Var}\,Xf_1'((y-\mu(\beta_0))/\beta_0).$$

Again this expression cannot vanish identically in y unless F_1 and hence G_0 is normal. Boundedness in y and continuity in β again hold. (i) and case (b) follow.

(c) In this range since $\tilde{\beta}_n$ is consistent, we are driven to minimizing either $\Delta_n(\beta, 0)$ or $\Delta_n(\beta, 2\delta)$. In the first case, we are minimizing at $|\beta_0| > \delta/2$ and get $n^{1/2}$-consistency. In the second case, after reparametrization, we again minimize at $\delta/2 \le \beta_0 \le 7\delta/2$ and again get $n^{1/2}$-consistency.

(d) In this range since $\tilde{\beta}_n$ is consistent, we minimize $\Delta_n(\beta, 2\delta)$ with probability tending to 1. But after reparametrizing this corresponds to minimizing at $\beta_0 \ge 3\delta/2$ and we again get $n^{1/2}$-consistency. □

NOTES.

(1) For cases (ii) and (iii) of Theorem 2.2 we need to check that convergence in our arguments holds uniformly for sequences with $\|P_n - P_0\| \to 0$, $\int x^2\, dG_n(x) \to \int x^2\, dG_0(x)$, where $\|\cdot\|$ is total variation. A careful examination of the argument shows that for consistency, we need only check that
$$\bar{Y} \to_{P_n} E_0(Y), \qquad \hat{\sigma}_x^2 \to_{P_n} \operatorname{Var}_{P_0}(X),$$
$$\bar{X} \to_{P_n} E_0(X), \qquad \hat{\sigma}_y^2 \to_{P_n} \operatorname{Var}_{P_0}(Y).$$
For $n^{1/2}$-consistency, the derivatives in (4.8) and (4.9) are now evaluated at $\beta_{0n} \leftrightarrow P_n$ and depend on the marginals of T, $F_{1n} \leftrightarrow P_n$ with $\|F_{1n} - F_{10}\| \to 0$ and $F_{10} \leftrightarrow P_0$ non-Gaussian. The derivatives still converge to that for F_{10} uniformly for β bounded and are bounded uniformly in y, since $\sup_n \int |x|\, dF_{1n} < \infty$. The argument can now be made at the limit F_{10} as before.

(2) Under the restricted Gaussian error model the same argument yields that $\tilde{\beta}_{na}$ is $n^{1/2}$-consistent.

We now proceed to study the correction term which gives efficiency.

PROPOSITION 4.3. *Whatever be* G_0
$$\left|\frac{\omega_0'}{\omega_0}(t)\right| \le \tilde{\sigma}^{-2}(\theta_0)\left(|t| + \int |\eta| G_0(d\eta)\right). \tag{4.9}$$

PROOF. By a standard Laplace transform theorem, writing $\tilde{\sigma}$ for $\tilde{\sigma}(\theta_0)$,
$$\frac{\omega_0'}{\omega_0}(t) = \tilde{\sigma}^{-2}\,\frac{\int(\eta - t)\phi(\tilde{\sigma}^{-1}(t-\eta))G_0(d\eta)}{\int \phi(\tilde{\sigma}^{-1}(t-\eta))G_0(d\eta)},$$
$$\left|\int(\eta - t)\phi((\tilde{\sigma}^{-1}(t-\eta)))G_0(d\eta)\right| \le \int|\eta - t|\phi((\tilde{\sigma}^{-1}(t-\eta)))G_0(d\eta)$$
$$\le \int|\eta - t| G_0(d\eta)\int \phi((\tilde{\sigma}^{-1}(t-\eta)))G_0(d\eta),$$
by an inequality of Chebyshev [Hardy, Littlewood and Pólya (1952), page 43] since $\phi(t)$ is decreasing for $t \ge 0$. □

PROPOSITION 4.4. *Suppose* $H_n \to H$ *weakly and* $\int x^2\, dH_n(x) \to \int x^2\, dH(x)$. *Then*
$$I(H_n * \Phi) \to I(H * \Phi),$$
where I *denotes Fisher information for location.*

PROOF. By dominated convergence for all t

$$H_n * \phi(t) \to H * \phi(t),$$
$$[H_n * \phi]'(t) \to [H * \phi]'(t).$$

By Proposition 4.3

$$\frac{|[H_n * \phi]'(t)|^2}{[H_n * \phi]} \le V(t, H_n), \tag{4.10}$$

where

$$V(t,H) = 4[H * \phi](t)\left(t^2 + \int \eta^2 H(d\eta)\right).$$

But

$$V(t, H_n) \to V(t, H) \quad \text{for all } t$$

and

$$\int V(t, H_n)\, dt = 8\int \eta^2 H_n(d\eta) + 4 \to 8\int \eta^2 H(d\eta) + 4 = \int V(t,H)\, dt.$$

The sequence in (4.10) is uniformly integrable and the result follows. □

PROPOSITION 4.5. *Let*

$$\omega_{0n}(t) = \int \omega_0(t - \sigma_n s)\lambda(s)\, ds + c_n. \tag{4.11}$$

Then if we write T_i for $T_i(\theta_0)$,

$$E\left(\frac{\hat{\omega}_n'}{\hat{\omega}_n}(T_1) - \frac{\omega_{0n}'}{\omega_{0n}}(T_1)\right)^2 \to 0, \tag{4.12}$$

$$E\left(\frac{\omega_{0n}'}{\omega_{0n}}(T_1) - \frac{\omega_0'}{\omega_0}(T_1)\right)^2 \to 0. \tag{4.13}$$

PROOF. We repeatedly use the inequalities

$$|\omega_{0n}^{(i)}| \le \sigma_n^{-i}\omega_{0n}, \qquad \omega_{0n} \le \sigma_{0n}^{-1}.$$

Write

$$\frac{\hat{\omega}_n'}{\hat{\omega}_n}(T_1) \frac{\omega_{0n}'}{\omega_{0n}}(T_1) = \frac{n^{-1}\Sigma_{j=1}^n\left[\lambda_n'(T_1 - T_j) - \omega_{0n}'(T_1)\right]}{\hat{\omega}_n}$$
$$- \frac{\omega_{0n}'(T_1)}{\omega_{0n}\hat{\omega}_n}\left(\frac{1}{n}\sum_{j=1}^n \lambda_n(T_1 - T_j) - \omega_{0n}(T_1)\right).$$

The first term has L_2 norm bounded by

$$c_n n^{-1/2} E^{1/2}\left([\lambda_n']^2(T_1 - T_2)\right) = O\left(c_n^{-1}\sigma_n^{-2} n^{-1/2}\right).$$

The second term is similarly norm bounded by

$$O\left(c_n^{-1}\sigma_n^{-2}n^{-1/2}\right)$$

and (4.12) follows.

For (4.13) note that, for all t, by dominated convergence,

$$\frac{\omega_{0n}'}{\omega_{0n}}(t) \to \frac{\omega_0'}{\omega_0}(t). \tag{4.14}$$

Without loss of generality, take $\tilde{\sigma}(\theta_0) = 1$. Then

$$\omega_{0n}(t) = \int \phi(t-\eta)\, d(G_0 * \lambda_n)(\eta) + c_n,$$

$$\omega_{0n}'(t) = \int \omega_0'(t - \sigma_n s)\lambda(s)\, ds.$$

By Proposition 4.3 we get

$$\frac{[\omega_{0n}']^2}{\omega_{0n}^2}(t) \le 2\left(t^2 + \int \eta^2\, dG_{s0} * \lambda_n(\eta)\right).$$

But

$$\int t^2 \omega_0(t)\, dt < \infty,$$

so that by dominated convergence and (4.14)

$$\int \left(\frac{\omega_{0n}'}{\omega_{0n}}\right)^2 (t)\omega_0(t)\, dt \to \int \frac{[\omega_0']^2}{\omega_0}(t)\, dt.$$

L_2 convergence of ω_{0n}'/ω_{0n} to ω_0'/ω_0 follows. □

PROPOSITION 4.6. *For sequences $\{P_n\}, \{c_n\}, \{\nu_n\}$ as in Theorem* 2.2(ii), *and all M finite*,

$$\sup\left\{\left|n^{-1/2}\sum_{i=1}^n U_i(\theta)\left(\frac{\hat{\omega}_n'}{\hat{\omega}_n}(T_i(\theta),\theta) - \frac{\omega_{0n}'}{\omega_{0n}}(T_i(\theta))\right)\right| : n^{1/2}|\theta - \theta_{0n}| \le M\right\} \to_{P_n} 0, \tag{4.15}$$

where $\theta_{0n} \leftrightarrow P_{0n}$,

$$\begin{aligned}\sup\Bigg\{ n^{1/2}\Bigg|\frac{1}{n}\sum_{i=1}^n \Bigg\{ & U_i(\theta)\left(T_i(\theta) - E_{P_n}(T_i(\theta)) + I_{0n}^{-1}\frac{\omega_{0n}'}{\omega_{0n}}(T_i(\theta))\right) \\ & - U_i(\theta_{0n})\left(T_i(\theta_{0n}) - E_{P_n}(T_i(\theta_{0n})) - I_{0n}^{-1}\frac{\omega_{0n}'}{\omega_{0n}}(T_i(\theta_{0n}))\right)\Bigg\} \\ & + I_{bn}(\theta - \beta_{0n})\Bigg| : n^{1/2}|\theta - \theta_{0n}| \le M\Bigg\} \to_{P_n} 0.\end{aligned} \tag{4.16}$$

This proposition reduces the proof of case (ii) to establishing that if $U_i \triangleq U_i(\theta_{0n})$, $T_i \triangleq T_i(\theta_{0n})$

$$\mathbf{L}_{P_0}\left(n^{-1/2}\sum_{i=1}^{n}\left\{U_i\big(T_i - E_{P_0}(T_i)\big) + I_{0n}^{-1}\frac{\omega_{0n}'}{\omega_{0n}}(T_i)\right\}\right) \to \mathbf{N}\big(0, I_b^{-1}(P_0)\big) \tag{4.17}$$

and

$$n^{-1}\sum_{i=1}^{n}\left(\frac{\omega_{0n}'}{\omega_{0n}}\right)^2(T_i) \to_{P_n} I_0(P_0), \tag{4.18}$$

$$n^{-1}\sum_{i=1}^{n}U_i^2\left(T_i + I_0^{-1}(P_0)\frac{\omega_{0n}'}{\omega_{0n}}(T_i)\right)^2 \to_{P_n} I_b(P_0). \tag{4.19}$$

All three claims follow since

$$\mathbf{L}_{P_n}(U_1, T_1) \to \mathbf{L}_{P_0}(U_1, T_1),$$
$$\frac{\omega_{0n}'}{\omega_{0n}}(t) \to \frac{\omega_0'}{\omega_0}(t), \quad \text{for all } t,$$

and $E_{P_n}(U_1^2)$, $E_{P_n}(T_1^2)$, $\int([\omega_{0n}']^2/\omega_{0n})(t)\,dt$ all converge to the appropriate limits under P_0. The last claim is a consequence of Proposition 4.4.

Proof of Proposition 4.6. Denote the (random) functions in absolute values in (4.15) by

$$Q_n(\Delta), \quad \text{where } \Delta = (\theta - \theta_{0n})n^{1/2}.$$

Now

$$Q_n(0) \to_{P_n} 0 \tag{4.20}$$

by Proposition 4.5.

Write

$$Q_{1n}(\Delta) = n^{-1}\sum_{i=1}^{n}T_i\left(\frac{\hat\omega_n'}{\hat\omega_n}(T_i(\theta),\theta) - \frac{\omega_{0n}'}{\omega_{0n}}(T_i(\theta))\right),$$
$$Q_{2n}(\Delta) = n^{-1/2}\sum_{i=1}^{n}U_{in}\left(\frac{\hat\omega_n'}{\hat\omega_n}(T_i(\theta),\theta) - \frac{\omega_{0n}'}{\omega_{0n}}(T_i(\theta))\right).$$

It is easy to see that for (4.15) we need only check that

$$\sup\{|Q_{in}(\Delta)|: |\Delta| \le M\} \to_{P_n} 0, \qquad i = 1,2. \tag{4.21}$$

Throughout this calculation we write $\lambda_n = \lambda_{\nu_n}$ and repeatedly use

$$\hat\omega_n \ge c_n, \qquad \omega_{0n} \ge c_n, \qquad |\lambda_n^{(i)}| \le \nu_n^{-i}\lambda_n.$$

We begin with $i = 1$. Let

$$V_{1n}(\Delta) = \frac{\hat\omega_n'}{\hat\omega_n}(T_1(\theta),\theta) - \frac{\omega_{0n}'}{\omega_{0n}}(T_1(\theta),\theta).$$

By Cauchy-Schwarz and uniform integrability of T_1^2 (as P_n varies), it is enough to check that

$$E \sup_{\Delta} (V_{1n}(\Delta))^2 = O\big(n^{-1}\nu_n^{-4}(c_n^{-2} + \log n)\big). \tag{4.22}$$

Note first that

$$|V_{1n}(0)| \le c_n^{-1}|\hat{\omega}_n'(T_1, \theta_{0n}) - \omega_{0n}'(T_1)| + c_n^{-1}\nu_n^{-1}|\hat{\omega}_n(T_1, \theta_{0n}) - \omega_{0n}(T_1)|. \tag{4.23}$$

Let $\hat{F}_n$ be the empirical distribution function of $T_1, \dots, T_n$ and F its expectation. Then

$$\begin{aligned} |\hat{\omega}_n'(t, \theta_{0n}) - \omega_{0n}'(t)| &= O(n^{-1}\sigma_n^{-2}) + n^{-1} \sum_{i=2}^{n} [\lambda_n'(t - T_i) - E\lambda_n'(t - T_i)] \\ &= O(n^{-1}\sigma_n^{-2}) + O(1)\int (\hat{F}_n(s) - F(s))\lambda_n''(t - s)\, ds \\ &\le O(n^{-1}\sigma_n^{-2}) + O(1)\sup_s |\hat{F}_n(s) - F(s)| \int |\lambda''(s)|\, ds, \end{aligned}$$

where the 0 terms are nonstochastic and independent of t. A similar bound holds for the second term in (4.23) and hence

$$EV_{1n}^2(0) = O(n^{-1}\nu_n^{-4}c_n^{-2}). \tag{4.24}$$

Next we write

$$T_i(\theta) = T_i + \frac{a}{\sqrt{n}}U_i + \frac{b}{\sqrt{n}}T_i,$$

so that a, b are well defined functions of Δ and note that

$$\begin{aligned} \frac{\partial}{\partial a} V_{1n}(\Delta) = n^{-1/2}\Bigg\{ &\frac{\Sigma(U_1 - U_j)\lambda_n''(T_1(\theta) - T_j(\theta))}{nc_n + \Sigma\lambda_n(T_1(\theta) - T_j(\theta))} - \frac{\hat{\omega}_n'}{\hat{\omega}_n}(T_1(\theta), \theta) \\ &\times \frac{\Sigma(U_1 - U_j)\lambda_n'(T_1(\theta) - T_j(\theta))}{nc_n + \Sigma\lambda_n(T_1(\theta) - T_j(\theta))} - U_1\left(\frac{\omega_{0n}'}{\omega_{0n}}\right)'(T_1(\theta))\Bigg\}. \end{aligned}$$

Therefore,

$$\begin{aligned} &E \sup\left\{\left|\frac{\partial}{\partial a} V_{1n}(\Delta)\right| : (\Delta) \le M\right\}^2 \\ &\quad \le C(M)n^{-1}\nu_n^{-4}E\left(\max_j (U_1 - U_j)^2 + U_1^2\right) = O\left(\frac{\log n}{n}\sigma_n^{-4}\right). \end{aligned} \tag{4.25}$$

Similarly, we can bound

$$\begin{aligned} \left|\frac{\partial}{\partial b} V_{1n}(\Delta)\right| \le n^{-1/2}\Bigg| &\frac{\Sigma(T_1 - T_j)\lambda_n''(T_1(\theta) - T_j(\theta))}{\hat{\omega}_n(T_1(\theta), \theta)} \\ &- \frac{\hat{\omega}_n'}{\hat{\omega}_n^2}(T_1(\theta), \theta)\sum|T_1 - T_j|\lambda_n'(T_1(\theta) - T_j(\theta)) \\ &- T_1\left(\frac{\omega_{0n}'}{\omega_{0n}}\right)'(T_1(\theta))\Bigg|. \end{aligned} \tag{4.26}$$

Representing $T_i = kT_i(\theta) + (c/\sqrt{n})U_i$, $k \to 1$, we can bound (4.26) by

$$An^{-1/2}\left\{\nu_n^{-2}\frac{\Sigma|T_1(\theta) - T_j(\theta)|\lambda_n(T_1(\theta) - T_j(\theta))}{nc_n + \Sigma\lambda_n(T_1(\theta) - T_j(\theta))}\right.$$
$$\left. + n^{-1/2}\left|\frac{\partial V_{1n}}{\partial a}(\Delta)\right| + \nu_n^{-2}(|U_1| + |T_1|)\right\}.$$

Representing $T_i = kT_i(\theta) + (c/\sqrt{n})U_i$, $k \to 1$, we can bound (4.26) by

$$An^{-1/2}\left\{\nu_n^{-2}\frac{\Sigma|T_1(\theta) - T_j(\theta)|\lambda_n(T_1(\theta) - T_j(\theta))}{nc_n + \Sigma\lambda_n(T_1(\theta) - T_j(\theta))}\right.$$
$$\left. + n^{-1/2}\left|\frac{\partial V_{1n}}{\partial a}(\Delta)\right| + \nu_n^{-2}(|U_1| + |T_1|)\right\},$$

for a constant A depending on M only. Since $\lambda_\nu(|t|)$ is decreasing, the first term in curly brackets is bounded using the Chebyshev inequality by

$$n^{-1}\sum|T_1(\theta) - T_j(\theta)|. \tag{4.27}$$

Since (4.27) is bounded by

$$B\left\{n^{-1}\sum\left(|T_j| + |T_1| + n^{-1/2}(|U_j| + |U_1|)\right)\right\},$$

for B depending on M only, we obtain

$$E\sup\left\{\left|\frac{\partial V_{1n}}{\partial b}(\Delta)\right|^2 : |\Delta| \le M\right\} = O\left(n^{-1}\nu_n^{-4}\log n\right). \tag{4.28}$$

Combining (4.24), (4.25) and (4.28), we get (4.21) for $i = 1$.

The proof of (4.21) for $i = 2$ is similar, but more complicated using the almost independence of $U_i(\theta)$, $T_i(\theta)$.

First, since $\hat{\omega}_n(\cdot, \theta_0)$ does not depend on the U_i,

$$\begin{aligned} EQ_{2n}^2(0) &= EU_1^2E\left(V_{1n}^2(0)\right) \\ &= O\left(n^{-2}\nu_n^{-4}c_n^{-2}\right). \end{aligned} \tag{4.29}$$

Next,

$$\begin{aligned} \frac{\partial Q_{2n}}{\partial a}(\Delta) &= \frac{1}{n}\sum_j U_j^2\left(\left(\frac{\hat{\omega}_n'}{\hat{\omega}_n}\right)'(T_j(\theta), \theta_0) - \left(\frac{\omega_{0n}'}{\omega_{0n}}\right)'(T_j(\theta))\right) \\ &\quad + n^{-1}\sum_i U_i\frac{\Sigma_j U_j\lambda_n''(T_i(\theta) - T_j(\theta))}{nc_n + \Sigma\lambda_n(T_i(\theta) - T_j(\theta))} \\ &\quad - n^{-1}\sum_i U_i\frac{\hat{\omega}_n'}{\hat{\omega}_n}(T_i(\theta), \theta)\frac{\Sigma_j U_j\lambda_n'(T_i(\theta) - T_j(\theta))}{nc_n + \Sigma\lambda_n(T_i(\theta) - T_j(\theta))} \\ &= R_{1n}(\Delta) + R_{2n}(\Delta) + R_{3n}(\Delta), \quad \text{say.} \end{aligned}$$

By arguing as for (4.22)

$$\sup\{ER_{1n}^2(\Delta): |\Delta| \le M\} = O\big(n^{-1}\nu_n^{-6}(c_n^{-2} + \log n)\big).$$

The additional ν_n^{-2} comes from the third derivatives in λ_n we have to deal with.

To deal with R_{2n} and R_{3n}, note that we can define $c(\theta)$ such that the Gaussian random variable

$$\tilde{U}_i(\theta) = U_i + \frac{c(\theta)}{\sqrt{n}}(T_i - X_i') \tag{4.30}$$

is independent of $T_i(\theta)$. This follows since $T_i(\theta)$ is a linear combination of X_i' and the Gaussian variables U_i and $T_i - X_i'$, both of which are independent of X_i'. Using (4.30)

$$\begin{aligned} ER_{2n}^2(\Delta) &\le 4E\left(n^{-2}\sum_{i,j}\tilde{U}_i\tilde{U}_j(\theta)\frac{\lambda_n''(T_i(\theta) - T_j(\theta))}{\hat{\omega}_n(T_i(\theta),\theta)}\right)^2 \\ &\quad + 4E\left(n^{-2}\sum_{i,j}\big(U_iU_j - \tilde{U}_i\tilde{U}_j(\theta)\big)\frac{\lambda_n''(T_i(\theta) - T_j(\theta))}{\hat{\omega}_n(T_i(\theta),\theta)}\right)^2 \\ &= O\big(n^{-1}\nu_n^{-4}\big) + O\big(n^{-1}\log n\nu_n^{-4}\big), \end{aligned}$$

since

$$E\tilde{U}_i^2(\theta) = O(1),$$

$$E\max\big(\tilde{U}_i\tilde{U}_j(\theta) - U_iU_j\big)^2 = O\big(n^{-1}\log n\big).$$

We can bound $ER_{3n}^2(\Delta)$ similarly to get

$$\sup\left\{E\left(\frac{\partial}{\partial a}Q_{2n}(\Delta)\right)^2 : |\Delta| \le M\right\} = O\big(n^{-1}\nu_n^{-6}(c_n^{-2} + \log n)\big). \tag{4.31}$$

Finally, we need to study $(\partial/\partial b)Q_{2n}(\Delta)$. It is possible to pass from the bound on $E((\partial/\partial a)Q_{2n}(\Delta))^2$ to the bound on $E((\partial/\partial b)Q_{2n}(\Delta))^2$ as was done in the passing from the bound on $(\partial/\partial a)V_{1n}(\Delta)$ to the bound on $(\partial/\partial b)V_{1n}(\Delta)$. We conclude

$$\sup\left\{E\left(\frac{\partial}{\partial b}Q_{2n}(\Delta)\right)^2 : |\Delta| \le M\right\} = O\big(n^{-1}\sigma_n^{-6}(c_n^{-2} + \log n)\big). \tag{4.32}$$

If we combine (4.31) and (4.32) with (4.29), we get by the standard Billingsley–Chentsov fluctuation inequalities [Billingsley (1968)],

$$\sup\{|V_{2n}(\Delta)|: |\Delta| \le M\} = O_{P_n}\big(n^{-1}\sigma_n^{-6}(c_n^{-2} + \log n)\big).$$

The proof of (4.15) is complete.

We now prove (4.16). Let

$$W_n(\Delta) = n^{-1/2}\bar{\sigma}(\theta)\sum_{i=1}^{n}U_i(\theta)\left(T_i(\theta) - E_{P_n}(T_i(\theta)) + I_{0n}^{-1}\frac{\omega_{0n}'}{\omega_{0n}}(T_i(\theta))\right),$$

where $\theta = \theta_0 + \Delta n^{-1/2}$, $\Delta = (\Delta_1, \dots, \Delta_4)$, $\Delta_1 = \beta$, etc. Claim (4.16) is equivalent to

$$(4.33)\qquad \sup\left\{\left|W_n(\Delta) - W_n(0) - \sum_{j=1}^{4} \frac{\partial W_n(0)}{\partial \Delta_j}\Delta_j\right| : |\Delta| \le M\right\} \to_{P_n} 0$$

and

$$(4.34)\qquad \left|\frac{\partial W_n(0)}{\partial \Delta_j} - I_{bn}\bar{\sigma}(\theta_0)\delta_{1j}\right| \to_{P_n} 0, \qquad j = 1, \dots, 4.$$

Now,

$$\frac{\partial W_n(0)}{\partial \Delta_1} = n^{-1} \sum_{i=1}^{n} \left[X_i\left(T_i + I_{0n}^{-1}\frac{\omega_{0n}'}{\omega_{0n}}(T_i)\right) + U_i(\gamma_1 U_i + \gamma_2(T_i - ET_i)) \times\left(1 + I_{0n}^{-1}\left(\frac{\omega_{0n}'}{\omega_{0n}}\right)'(T_i)\right)\right],$$

for suitable γ_1, γ_2, the laws of the summands converge to $\mathbf{L}_0(A)$, where

$$A = X\left(T + I_0^{-1}\frac{\omega_0'}{\omega_0}(T)\right) + U(\gamma_1 U + \gamma_2(T - E_0 T))\left(1 + I_0^{-1}\left(\frac{\omega_0'}{\omega_0}\right)'(T)\right),$$

and the summands are uniformly integrable (P_n) by Proposition 4.4. Therefore,

$$\frac{\partial W_n}{\partial \Delta_1}(0) \to_{P_n} E_0(A) = I_b\bar{\sigma}(\theta_0),$$

after some computation. A similar argument establishes (4.34) for $j > 1$. For (4.33) we check that for $1 \le j \le k \le 4$,

$$(4.35)\qquad \sup\left\{\left|\frac{\partial^2 W_n}{\partial \Delta_j\, \partial \Delta_k}(\Delta)\right| : |\Delta| \le M\right\} \to 0.$$

We give the argument for a typical term, $\Delta_3 \leftrightarrow \nu_1$,

$$(4.36)\qquad \frac{\partial^2 W_n}{\partial \Delta_3^2} = n^{-3/2} \sum_{i=1}^{n} \sigma(\theta) U_i(\theta) I_{0n}^{-1} X_{in}^2 \left(\frac{\omega_{0n}'}{\omega_{0n}}\right)''(T_i(\theta)).$$

Since $|\omega_{0n}^{(i)}/\omega_{0n}| \le \sigma_n^{-i}$, we bound (4.35) uniformly in $|\Delta| \le M$ by

$$(4.37)\qquad n^{-1/2}\sigma_n^{-2}O(1)\left\{n^{-1}\sum_{i-1}^{n} |U_i|\left(T_i^2 + U_i^2\right) + n^{-3/2}\sum_{i=1}^{n} |T_i|^3\right\}.$$

Since T_i^2 are uniformly integrable under P_n,

$$(4.38)\qquad n^{-1/2} \max_i |T_i| \to_{P_n} 0.$$

Claim (4.35) for $j = k = 3$ follows from (4.37) and (4.38). The other terms are dealt with similarly and the result follows.

Proposition 4.6 establishes claim (ii) of the theorem. For part (iii) note that Proposition 4.6 shows that if β_n^* is $n^{1/2}$-consistent so is $\hat{\beta}_n(\beta_n^*)$ and, in fact,

$$\hat{\beta}_n(\beta_n^*) = \beta_{0n} + n^{-1} \sum_{i=1}^{n} \tilde{I}_b(X_i, P_n) + o_{P_n}(n^{-1/2}).$$

Therefore, taking β_n^* successively as $\hat{\beta}_{0n}, \hat{\beta}_{1n}, \dots$, we get

$$\hat{\beta}_{in} - \hat{\beta}_{1n} = o_{P_n}(n^{-1/2})$$

and claim (iii) follows. Claim (iv) is established in exactly the same way as claims (i)–(iii). □

Proposition 4.7. *The efficiency of* $\hat{\beta}_P$ *under model (Identity,* Φ*),* I_c/I_a*, satisfies*

$$I_c/I_a \geq \left(1 + \sigma^2/(\beta^2 + 1)(\mathrm{Var}(X') + \sigma^2)\right)^{-1}.$$

Proof.

$$I_a/I_c = [\mathrm{Var}(X')]^{-2}\mathrm{Var}(T)\left[\mathrm{Var}(T) - 2\tilde{\sigma}^2 + \tilde{\sigma}^4 I_0\right]$$
$$= 1 + \tilde{\sigma}^4(I_0\mathrm{Var}(T) - 1)/(\mathrm{Var}(X'))^2,$$

since $\mathrm{Var}(T) = \mathrm{Var}(X') + \tilde{\sigma}^2$. Since T is, in general, an inefficient estimate of η in the location model $T = \eta + \varepsilon$ we must have $\tilde{\sigma}^2 \geq I_0^{-1}$ so that

$$I_a/I_c - 1 \leq \tilde{\sigma}^4(\mathrm{Var}(T)/\tilde{\sigma}^2 - 1)/(\mathrm{Var}(X'))^2$$
$$= \tilde{\sigma}^2/\mathrm{Var}(X') = \sigma^2/(\beta^2 + 1)\mathrm{Var}(X')$$

and the result follows. □

Acknowledgment. We thank Cliff Spiegelman for helpful discussions.

REFERENCES

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

Begun, J. M., Hall, W. J., Huang, W.-M. and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11** 432–452.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Csörgő, M. (1981). *Strong Approximations in Probability and Statistics*. Academic, New York.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.* **72** 557–565.

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

Hardy, G. H., Littlewood, J. E. and Pólya, G. (1952). *Inequalities*, 2nd ed. Cambridge Univ. Press.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, New York.

Kendall, M. G. and Stuart, A. (1979). *The Advanced Theory of Statistics* **2**, 4th ed. Hafner, New York.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

KOSHEVNIK, YU. A. and LEVIT, B. YA. (1976). On a non-parametric analogue of the information matrix. *Theory Probab. Appl.* **21** 738–753.

NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics* (U. Grenander, ed.) 213–234. Wiley, New York.

NEYMAN, J. and SCOTT, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16** 1–32.

PFANZAGL, J. (1982). *Contributions to a General Asymptotic Statistical Theory. Lectures Notes in Statist.* **13**. Springer, New York.

REIERSØL, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica* **18** 375–389.

RUBIN, H. (1956). Uniform convergence of random functions with applications to statistics. *Ann. Math. Statist.* **27** 200–203.

SPIEGELMAN, C. (1979). On estimating the slope of a straight line when both variables are subject to error. *Ann. Statist.* **7** 201–206.

STEIN, C. (1956). Efficient nonparametric testing and estimation. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 187–195. Univ. of California Press.

COURANT INSTITUTE OF MATHEMATICAL SCIENCES
NEW YORK UNIVERSITY
251 MERCER STREET
NEW YORK, NEW YORK 10012

DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY
JERUSALEM 91905
ISRAEL