Multiplicative Errors-in-Variables Models With Applications to Recent Data Released by the U.S. Department of Energy

JIUNN T. HWANG*

In an errors-in-variables model, the predicting variables are observed with errors. Traditionally, the errors are assumed to be additive. In this article, I consider the case in which the error is multiplicative, a situation that arises when analyzing some recent data released by the U.S. Department of Energy. A consistent estimator $\hat{\beta}$ is provided for regression coefficients β by correcting the asymptotic bias of the least squares estimate. It is shown that $\hat{\beta} - \beta$, after being normalized by an estimate of its covariance, is asymptotically normally distributed. This can, therefore, be used to construct approximate tests and confidence sets for β. The results are essentially nonparametric.

KEY WORDS: Asymptotically consistent estimator; Asymptotic normal distribution.

1. INTRODUCTION

In many statistical practices, one assumes a linear model:

$$Y = X\beta + \varepsilon, \qquad (1.1)$$

where $Y = (Y_1, \ldots, Y_n)'$ is the n-dimensional observation, $\beta = (\beta_1, \ldots, \beta_p)'$ is a p-dimensional unknown vector, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ is an unobserved n-dimensional random error. One also assumes that the $n \times p$ matrix X can be observed exactly. There are many situations, however, in which X can only be observed with errors; that is, instead of X, one can only observe X^*, where

$$X^* = X + U. \qquad (1.2)$$

Model (1.1) and (1.2) is an errors-in-variables model, since observational errors are present in both variables X and Y. The least squares estimator is not asymptotically consistent. This problem has a long history and seems to be well studied. See Anderson (1984) for a recent survey.

The purpose of this article is to consider a different model. Instead of (1.2), we assume that we can only observe X^*, where

$$X^* = X \boxdot U. \qquad (1.3)$$

For two matrices A and B of the same dimension, $A \boxdot B$ denotes the Hadamard product. That is, $A \boxdot B$ is a matrix, also of the same dimension, whose (i, j)th element is the product of the (i, J)th elements of A and B. Similarly, $A \boxed{\div} B$ denotes the elementwise division $a_{ij} \div b_{ij}$. This model is necessary for analyzing the recent data collected by the Department of Energy concerning energy consumption and housing characteristics of households in the United States. To preserve confidentiality, some of the predicting variables that might identify household owners have been multiplied by random numbers, U_{ij}, the (i, j)th element of U. Hence model (1.1) and (1.3) is appropriate for these data and will be referred to as a multiplicative errors-in-variables model. In contrast, model (1.1) and (1.2) will be called an additive errors-in-variables model. Although the exact values of U_{ij} are not given, their distributions are. The U_{ij}'s are truncated normal random variables with known variances and means all equal to 1 (see Sec. 4). Consequently, the distribution of U_{ij} is neither continuous nor discrete, but mixed.

Looking at model (1.1) and (1.3), the statistician's first reaction might be to apply a logarithm transformation so that, say,

$$\log X_{ij}^* = \log X_{ij} + \log U_{ij}$$

will transform the multiplicative error into an additive error. The statistician then hopes to use some existing statistical procedures designed for errors-in-variables models. This approach fails, because the logarithm transformation destroys the linearity of the model. Furthermore, not much is known about nonlinear errors-in-variables models. The only discussions for such models of which I am aware were given in Dolby and Lipton (1972), Wolter and Fuller (1982a, b), and the references cited therein, which do not apply to this case. (The first and third paper just cited deal with the case in which replications are available. The second paper considers only the quadratic model.) Stefanski (1985) suggested some new methods applicable to this problem.

The goal of this article is to develop a consistent estimator and an associated asymptotic distribution theory for estimating β under model (1.1) and (1.3). Asymptotic properties seem to be important, since in the energy problem $n = 5{,}979$. In addition to assumptions (2.2), (2.3), and (2.4), which are reasonable for the energy problem, I assume the *structural model* as opposed to the functional model (see Kendall and Stuart 1967, chap. 29). That is, the rows of X are assumed to be iid random vectors from an unknown distribution. This assumption is especially ap-

* Jiunn T. Hwang is Associate Professor, Department of Mathematics, Cornell University, Ithaca, NY 14853. This research was supported by National Science Foundation Grant MCS-8301139 to Cornell University and Chinese National Science Council Grant VE 84-060 to Academia Sinica in Taiwan. The author wishes to thank Li-Min Hsueh for informing him of this problem, pointing out many references, and commenting on his first draft. He also thanks Roger Farrell, Wayne Fuller, John Preston, Shayle Searle, and Leonard Stefanski for many valuable discussions and comments. In particular, Leonard Stefanski read through the second draft in detail and made many valuable suggestions. The comments of the associate editor and the referee also helped to improve the earlier draft.

Journal of the American Statistical Association
September 1986, Vol. 81, No. 395, Theory and Methods

propriate for the energy problem; the rows of the relevant X are the housing characteristics of the households randomly drawn from the superpopulation of households in the United States. These assumptions are essentially nonparametric.

By correcting the asymptotic bias of the least squares estimator, in Section 2, a strongly consistent estimator $\hat{\beta}$ is derived for β. In Section 3 the asymptotic normality of $\hat{\beta}$ under some extra moment assumptions is derived. If normalized by a (strongly) consistent estimate of its covariance matrix (which is provided here), $\hat{\beta}$ has an asymptotically standard normal distribution. Hence approximate testing procedures and confidence sets can be constructed. Section 4 discusses the energy example in detail. This section is self-contained. Readers interested in the applications and not the theoretical development of our procedures need only read Section 4 and its cited materials.

After the first draft of this article was written, Wayne Fuller kindly pointed out that Fuller and Hidiroglou (1978) and Fuller (1984) considered closely related models (see also the program package SUPER CARP in Hidiroglou, Fuller, and Hickman 1980). In particular, Fuller's (1984) general results apply to models (1.1) and (1.2). His point estimator for β is the same as (2.11), whereas his estimate for the asymptotic covariance matrix is different from (3.12). The covariance matrix estimate (or corresponding tests) of (3.12) reduces to the usual standard maximum likelihood estimator (or the standard t test or F test) under normality assumption and the homoscedasticity assumption of ε when all of the predicting variables are observed exactly. Fuller's covariance matrix estimate does not reduce to a familiar form or a function of the sufficient statistics. Therefore, the present covariance matrix estimate seems to be more appropriate under the homoscedasticity assumption, which may be a reasonable one in the energy problem. Fuller's estimate, however, may work better in a heteroscedastic situation. Both covariance matrix estimators are shown to be strongly consistent for the structural model. Theorem 4.1 provides, under a simpler situation, a simpler formula for calculating the covariance matrix estimator.

2. DERIVATION OF THE CONSISTENT ESTIMATOR

The focus here is on developing a consistent point estimator for β for the multiplicative model

$$Y = X\beta + \varepsilon, \qquad X^* = X \boxdot U. \tag{2.1}$$

Let U_i' and X_i' be the ith row of U and X, respectively. It is assumed that

$$\varepsilon_i, \quad 1 \le i \le n, \quad \text{are iid with zero mean;} \tag{2.2}$$

$$U_i', \quad 1 \le i \le n, \quad \text{are iid random vectors;} \tag{2.3}$$

$$X,\ U, \text{ and } \varepsilon \text{ are independent;} \tag{2.4}$$

and

$$X_i \text{ are iid samples from a population with an unknown distribution.} \tag{2.5}$$

As discussed in Section 1, it is reasonable to assume the structural model (2.5). In the energy problem,

$$EU_i = 1_p, \tag{2.6}$$

where 1_k is a k-component column vector of ones. Note that (2.6) can be assumed without loss of generality as long as EU_{1j} is known and is nonzero for all j, since then one can divide the jth column of X^* and U by EU_{1j}. The joint distribution of U_{ij} is assumed to be known. Next I derive a consistent estimate for β by an approach that appears to be quite general (see Section 5 for further discussion).

The main idea is to start with a standard estimator (say the least squares estimator) and calculate its asymptotic bias. The bias typically depends on β and other unknown parameters. By multiplication or subtraction, the bias of the standard estimator is corrected so that this results in an expression depending on unknown parameters but not on β. Estimating these unknown parameters by consistent estimators (moment estimators usually) then leads to a consistent estimator.

Now, consider the least squares (LS) estimator

$$\hat{\beta}_{\text{LS}} = [X^{*\prime}X^*]^{-1}X^{*\prime}Y,$$

where $X^{*\prime}$ denotes, as usual, the transpose of X^*. The calculation will be simplified if one corrects the bias of $X^{*\prime}Y/n$ instead of that of $\hat{\beta}_{\text{LS}}$, even though both approaches lead to the same consistent estimator. Note that

$$X^{*\prime}Y/n = X^{*\prime}X\beta/n + X^{*\prime}\varepsilon/n.$$

Since the rows $X_i^{*\prime}$ of X^* are iid and independent of ε_i's,

$$\frac{X^{*\prime}\varepsilon}{n} = \frac{1}{n}\frac{\sum_{1=1}^{n}\varepsilon_i X_i^*}{n} \xrightarrow{\text{a.s.}} E\varepsilon_1 X_1^* = E\varepsilon_1 EX_1^* = 0,$$

where $\xrightarrow{\text{a.s.}}$ denotes almost sure convergence. The last statement follows from the strong law of large numbers (see, e.g., Loève 1977, p. 251). Similarly,

$$\frac{X^{*\prime}X}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i^* X_i' \xrightarrow{\text{a.s.}} EX_1^* X_1'.$$

The limit obviously equals EX_1X_1' by the assumptions that U_{ij} has expectation 1 and is independent of X. Hence

$$X^{*\prime}Y/n \xrightarrow{\text{a.s.}} A\beta, \tag{2.7}$$

where

$$A = EX_1X_1'. \tag{2.8}$$

The assumption that A is nonsingular implies that

$$A^{-1}(X^{*\prime}Y/n) \xrightarrow{\text{a.s.}} \beta. \tag{2.9}$$

Of course, A is unknown, otherwise a consistent estimate would have been obtained; however, A can be estimated consistently, as follows: Consider

$$\frac{X^{*\prime}X^*}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i^* X_i^{*\prime} \xrightarrow{\text{a.s.}} EX_1^* X_1^{*\prime}.$$

Furthermore,

$$EX_1^* X_1^{*\prime} = M \boxdot EX_1X_1',$$

where $M = EU_1U_1'$. Hence

$$E(X_1^*X_1^{*\prime}) \boxdiv M = EX_1X_1'.$$

Here $\boxdiv$ denotes the Hadamard division—that is, elementwise division. Consequently,

$$\hat{A} \overset{\text{def}}{=} \left(\frac{X^{*\prime}X^*}{n}\right) \boxdiv M \overset{\text{a.s}}{\to} EX_1X_1' = A \qquad (2.10)$$

(where def means defined). [If one is interested in estimating A, it will be better to modify $\hat{A}$ so that it is always positive definite. See Fuller and Hidiroglou (1978) for an appropriate method to do so.] By replacing A in (2.9) by $\hat{A}$, we arrive at the consistent estimator

$$\hat{\beta} = [(X^{*\prime}X^*) \boxdiv M]^{-1}X^{*\prime}Y. \qquad (2.11)$$

Theorem 2.1. Assume that EX_1X_1' exists and is nonsingular and that the covariance matrix of U_1 exists. Then $\hat{\beta}$ is a strongly consistent estimator for β.

There might be an operational problem, however. It is true that as $n \to \infty$, $\hat{A}$ will be nonsingular. Is it nonsingular, however, for finite n? Under additional assumptions, the matrix is nonsingular with probability 1. The following theorem was proved by Roger Farrell at Cornell University and was personally communicated to me.

Theorem 2.2. The matrix $\hat{A}$ in (2.10) is nonsingular with probability 1, provided that X has a probability density function with respect to the Lebesgue measure in R^{np}, the np Euclidean space, and $\Pr(U_{ij} = 0) = 0$, $\forall\, i, j$.

Proof. See Appendix A.

Returning to Theorem 2.1, note that if X can be observed exactly (i.e., $U_{ij} = 1$ for all i, j), then the estimator $\hat{\beta}$ reduces to the least squares estimator $\hat{\beta}_{\text{LS}}$. The well-known result of the consistency of $\hat{\beta}_{\text{LS}}$ is, therefore, a corollary to Theorem 2.1.

If X can only be observed with error, however, then

$$\hat{\beta}_{\text{LS}} \overset{\text{a.s.}}{\to} (A \boxdot M)^{-1}A\beta. \qquad (2.12)$$

To establish the last statement, one has to show that $A \boxdot M$ is nonsingular. This is the case, since

$$A \boxdot M - A = A \boxdot (M - 1_p1_p'). \qquad (2.13)$$

Furthermore, A is positive definite and $M - 1_p1_p'$ is nonnegative definite, being the covariance matrix of U_1. By a theorem of Schur (1911), which is also proved in Styan (1973, theorem 3.1), the matrix in (2.13) is nonnegative definite. Therefore, $A \boxdot M$ is positive definite (and is hence nonsingular), since A is.

Coming back to (2.12), this shows that $\hat{\beta}_{\text{LS}}$ is consistent if and only if

$$(A \boxdot M)^{-1}A = I_p,$$

where I_p is the $p \times p$ identity matrix. The last equation holds iff $M = 1_p1_p'$. This is equivalent to the assertion that U_{ij} are identically 1 for every i and j. The following theorem was already proved.

Theorem 2.3. The least squares estimator $\hat{\beta}_{\text{LS}}$ is asymptotically consistent iff all of the predicting variables X are observed exactly (i.e., U_{ij}'s are identically 1).

It is interesting to note that the asymptotic limit of $\hat{\beta}_{\text{LS}}$ satisfies

$$|(A \boxdot M)^{-1}A\beta| \le |\beta|.$$

Hence the least squares estimator tends to underestimate β, in the sense that $\hat{\beta}_{\text{LS}}$ is closer to the origin than β. Therefore, to be asymptotically consistent, one has to "expand" $\hat{\beta}_{\text{LS}}$. This is precisely what was done in constructing $\hat{\beta}$.

3. ESTIMATING THE ASYMPTOTIC DISTRIBUTION OF THE CONSISTENT ESTIMATOR

In this section, I establish the asymptotic normality of $\hat{\beta}$ and provide two strongly consistent estimates for its asymptotic covariance matrix. At the end of the section, some diagnostic comments are made. Also provided is a measure of overall fit for model (2.1) by modifying the standard R^2 statistic.

Theorem 3.1. In addition to the assumptions of Theorem 2.1, suppose that the elements of X and U have finite fourth moments and that ε_i has a finite second moment. Then $\sqrt{n}(\hat{\beta} - \beta)$ is asymptotically normally distributed with mean zero.

Proof. Note that from (2.11),

$$\begin{aligned}\sqrt{n}\,(\hat{\beta} - \beta) &= \sqrt{n}\,\{[(X^{*\prime}X^*) \boxdiv M]^{-1}X^{*\prime}Y - \beta\} \\ &= \{(1/n)[X^{*\prime}X^*] \boxdiv M\}^{-1}S/\sqrt{n}, \qquad (3.1)\end{aligned}$$

where

$$S = \{X^{*\prime}X - [X^{*\prime}X^*] \boxdiv M\}\beta + X^{*\prime}\varepsilon.$$

Writing S in terms of rows of X^*, Y, and ε, we obtain

$$\frac{1}{\sqrt{n}}\,S = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{X_i^*X_i'\beta - [(X_iX_i^{*\prime}) \boxdiv M]\beta + \varepsilon_iX_i^*\}. \qquad (3.2)$$

Observe that the random vectors within the braces form a sequence of iid random vectors for $i = 1, \ldots, n$, having finite covariance matrix and zero mean. By the standard central limit theorem (see Loève 1977, p. 286A), the last expression is asymptotically normally distributed.

Returning to (3.1), note that (2.10) implies

$$\{(1/n)[X^{*\prime}X^*] \boxdiv M\}^{-1} \overset{\text{a.s.}}{\to} A^{-1}. \qquad (3.3)$$

This establishes the theorem.

Now consider estimation of the asymptotic covariance matrix of $\hat{\beta}$. From (3.1), (3.2), and (3.3), the asymptotic covariance matrix of (3.1) is

$$\Sigma = A^{-1}\operatorname{cov}(Z_1 + Z_2)A^{-1}, \qquad (3.4)$$

where $Z_1 = \varepsilon_1X_1^*$ and

$$Z_2 = X_1^*X_1'\beta - [(X_1^*X_1^{*\prime}) \boxdiv M]\beta. \qquad (3.5)$$

Note that $\operatorname{cov}(Z_1 + Z_2) = \operatorname{cov} Z_1 + \operatorname{cov} X_2$, since the covariance of Z_1 and Z_2 is zero by independence of ε_1 and

$\{X_1, X_1^*\}$ and by the assumption that $E\varepsilon_1 = 0$. Hence

$$\Sigma = A^{-1}(\text{cov } Z_1)A^{-1} + A^{-1}(\text{cov } Z_2)A^{-1},$$

which can be thought of as a decomposition of Σ into the sum of the covariance matrix due to the measurement error of Y and the covariance matrix due to the measurement error of X. The first matrix on the right side of the last expression is the asymptotic covariance matrix of $\hat{\beta}$ when X is observed without error and the second matrix is that of $\hat{\beta}$ when Y is observed exactly. From (2.10) $\hat{A}$ estimates A consistently. Therefore, what remain are the problems of estimating cov Z_i ($i = 1, 2$).

To estimate cov Z_1 consistently, note that cov $Z_1 = \sigma_\varepsilon^2 E X_1^* X_1^{*\prime}$, where σ_ε^2 is the variance of ε_1. The expectation can obviously be estimated consistently by $X^{*\prime}X^*/n$. Therefore, it suffices to estimate σ_ε^2 consistently. By the strong law of large numbers,

$$|Y - X\beta|^2/n \xrightarrow{\text{a.s.}} \sigma_\varepsilon^2.$$

We cannot observe X and β, however. Hence one estimates the left side of the last expression. This quantity equals

$$(Y'Y - 2\beta'X'Y + \beta'X'X\beta)/n.$$

By replacing β by its consistent estimate $\hat{\beta}$, $X'Y$ by $X^{*\prime}Y$, and $X'X$ by $[X^{*\prime}X^*] \boxdiv M$, we arrive at an estimate of σ_ε^2,

$$\begin{aligned} \hat{\text{var}} &= (Y'Y - 2\hat{\beta}X^{*\prime}Y + \hat{\beta}'\{[X^{*\prime}X^*] \boxdiv M\}\hat{\beta})/n \\ &= (Y'Y - \hat{\beta}'X^{*\prime}Y)/n. \end{aligned} \tag{3.6}$$

The estimator is obviously consistent, since $|Y - X\beta|^2/n - \sigma_\varepsilon^2 \to 0$ with probability 1, which follows from

$$\hat{\beta} \xrightarrow{\text{a.s.}} \beta, \qquad (X^* - X)'Y/n \xrightarrow{\text{a.s.}} 0,$$

and

$$\{[X^{*\prime}X^*] \boxdiv M - X'X\}/n \xrightarrow{\text{a.s.}} 0.$$

The last two equations follow from the law of large numbers, (2.7), (2.8), and (2.10). The estimator of (3.6), however, can be negative. Hence one should modify it by using

$$\hat{\sigma}_\varepsilon^2 = \max(0, \hat{\text{var}}), \tag{3.7}$$

which is obviously consistent.

Theorem 3.2. Under the assumptions of Theorem 3.1, $\hat{\sigma}_\varepsilon^2$ is strongly consistent for estimating σ_ε^2. Furthermore,

$$\hat{\Sigma}_1 = \hat{\sigma}_\varepsilon^2 X^{*\prime}X^*/n \tag{3.8}$$

is a strongly consistent estimate for Σ_1.

To estimate cov Z_2, I now introduce two matrix notations. For any $l_1 \times l_2$ matrix L, vec L denotes an $l_1 l_1 \times 1$ vector formed by stacking columns of L one under the other. For any two matrices L and K of sizes $l_1 \times l_2$ and $k_1 \times k_2$, respectively, the Kronecker product is defined as the $l_1k_1 \times l_2k_2$ matrix

$$L \otimes K = \{L_{ij}K\},$$

where L_{ij} is the (i, j)th element of L. [For a good review of the vec operation and the Kronecker product, see Henderson and Searle (1981).] Using these notations, I now state Lemma 3.1, the proof of which is in Appendix B.

Lemma 3.1.

$$\begin{aligned} \text{cov } Z_2 &\overset{\text{def}}{=} \Sigma_2 = (I_p \otimes \beta') \\ &\times \{Q \boxdot E\{[\text{vec}(X_1^* X_1^{*\prime})][\text{vec}(X_1^* X_1^{*\prime})]'\}\}(I_p \otimes \beta), \end{aligned} \tag{3.9}$$

where I_p is the $p \times p$ identity matrix,

$$\begin{aligned} Q = \{&E[\text{vec } V'(\text{vec } V')]'\} \\ &\boxdiv \{\text{E}\{(\text{vec } U_1U_1')[\text{vec}(U_1U_1')]'\}\}, \end{aligned} \tag{3.10}$$

and $V = U_1 1_p' - \{(U_1U_1') \boxdiv E(U_1U_1')\}$.

To obtain a consistent estimator for Σ_2, we only have to deal with the estimation of β and the expectation in (3.9). Of course, β is estimated by $\hat{\beta}$. Furthermore, the law of large numbers shows that the expectation can be consistently estimated by

$$\frac{1}{n}\sum_{i=1}^{n} \text{vec}(X_i^* X_i^{*\prime})[\text{vec}(X_i^* X_i^{*\prime})]'.$$

Hence we arrive at a strongly consistent estimate $\hat{\Sigma}_2$ of Σ_2:

$$\begin{aligned} \hat{\Sigma}_2 = (I_p \otimes \hat{\beta}')\{Q \boxdot \frac{1}{n}\sum_{i=1}^{n} [\text{vec}(X_i^* X_i^{*\prime})][\text{vec}(X_i^* X_i^{*\prime})]'\} \\ \times (I_p \otimes \hat{\beta}). \end{aligned} \tag{3.11}$$

This, together with Theorem 3.1, (3.4), and Theorem 3.2, implies the following theorem.

Theorem 3.3. Let $\hat{A}$, $\hat{\Sigma}_1$, and $\hat{\Sigma}_2$ be defined in (2.10), (3.7), (3.8), and (3.11). Define

$$\hat{\Sigma} = \hat{A}^{-1}(\hat{\Sigma}_1 + \hat{\Sigma}_2)\hat{A}^{-1}. \tag{3.12}$$

Under the assumptions of Theorem 3.1, $\hat{\Sigma}$ is a strongly consistent estimate for $\hat{\Sigma}$. Furthermore the distribution of $\sqrt{n}(\hat{\Sigma})^{-1/2}(\hat{\beta} - \beta)$ approaches a p-variate standard normal distribution as $n \to \infty$.

By using Theorem 3.3, one can construct approximate confidence sets for β (or its function) and conduct approximate tests of hypotheses involving functions of β. These procedures will reduce to the usual F or t procedures (F test or t test) when the predicting variables X are observed exactly and ε has a normal distribution.

In practice, it should be better to use $(\hat{\Sigma}_2)_+$ instead of $\hat{\Sigma}_2$ in (3.12). In general, for a symmetric matrix S, S_+ denotes a nonnegative definite matrix with the same eigenvalues and corresponding eigenvectors as S except that the negative eigenvalues of S (if S has any) are replaced by zero. In any situation (even in the situation when S has repeated eigenvalues), S_+ is uniquely defined. Schwarz (1967, p. 698) indicated that eigenvalues are continuous functions of symmetric matrices (with respect to the euclidean norm in $R^{p(p+1)/2}$, where p is the number of rows or columns of S). Therefore, since $\hat{\Sigma}_2$ approaches Σ_2 with probability 1, the same statement holds with $\hat{\Sigma}_2$ replaced by $(\hat{\Sigma}_2)_+$. Consequently Theorem 3.3 holds with $\hat{\Sigma}_2$ replaced by $(\hat{\Sigma}_2)_+$.

Although asymptotically $\hat{\Sigma}$ in (3.12) is nonsingular, it can be singular in finite samples. In the energy problem, where the sample size is very large, this does not occur. It may occur, however, in other applications. In fact, if $\hat{\Sigma}$ is singular (or nearly singular), it serves as a warning to possible mistakes. For example, it may indicate that the sample size is not large enough and hence the implication of the asymptotic theory is doubtful.

Another covariance estimate that is positive definite with probability 1 can be derived. Using (3.1) and (3.2), we obtain

$$\sqrt{n}(\hat{\beta} - \beta) = (\hat{A})^{-1} \sum_i \{X_i^* Y_i - (X_i^* X_i^{*\prime} \boxed{\div} M)\beta\}/\sqrt{n}$$

$$\equiv \hat{A}^{-1} \sum_i e_i/\sqrt{n}, \text{ say.}$$

By the law of large numbers, $\Sigma_i e_i e_i'/n \to Ee_1 e_1'$. Hence

$$\hat{A}^{-1} \sum_i e_i e_i' \hat{A}^{-1}/n \xrightarrow{\text{a.s.}} \Sigma.$$

Since e_i involves β, β is replaced by $\hat{\beta}$. It is easy to show that the modification does not change the limit. Hence another strongly consistent estimator is obtained:

$$\frac{1}{n} \hat{A}^{-1} \sum_{i=1}^{n} \{X_i^* Y_i - (X_i^* X_i^{*\prime} \boxdot M)\hat{\beta}\}$$

$$\times \{X_i^* Y_i - (X_i^* X_i^{*\prime} \boxdot M)\hat{\beta}\}' \hat{A}^{-1}. \quad (3.13)$$

When Leonard Stefanski read an earlier draft of this article, he also derived a strongly consistent estimate similar to (3.13). When specialized to model (2.1), the estimator (3.13) is Fuller's (1984) estimator, except Fuller used $n - p$ instead of n as a divisor.

One can establish a theorem similar to Theorem 3.3 by replacing the covariance matrix in (3.12) by (3.13). Therefore, asymptotic confidence sets and asymptotic tests for β can be constructed using Fuller's estimate. An advantage of Fuller's covariance estimate is that it is easy to compute and appears to be more nonparametric than (3.12). Fuller's estimate, however, does not reduce to the standard estimate when all of the predicting variables are observed exactly. In fact, when all of the X's are observed exactly (3.13) is

$$(X'X/n)^{-1} \sum_i (\hat{\varepsilon}_i)^2 (X_i X_i'/n)(X'X/n)^{-1}, \quad (3.14)$$

where $\hat{\varepsilon}_i$'s are the residuals. In contrast, the covariance estimate (3.12), in this case, reduces to

$$\sum_i [(\hat{\varepsilon}_i)^2/n](X'X/n)^{-1}, \quad (3.15)$$

the maximum likelihood estimate under normal assumption of ε, which is more efficient than (3.14). This seems to indicate that (3.12) may be more appropriate than (3.13) under the homoscedasticity assumption, which may be reasonable in the energy problem. Note, however, that if ε is heteroscedastic, the reverse may be true, since then (3.14) is more appropriate than (3.15) in the standard linear model.

It is usually helpful in practice to have an R^2 measure of overall fit. In the standard linear model (when X is observable) with intercept, the R^2 statistic is defined as $R^2 = 1 - \text{SSE}_p/\text{SSE}_1$, where

$$\text{SSE}_p = |Y - X\hat{\beta}_{\text{LS}}|^2 = Y'Y - \hat{\beta}_{\text{LS}}' X'Y$$

and

$$\text{SSE}_1 = |Y - \bar{Y}1_n|^2, \qquad \bar{Y} = \sum_i Y_i/n$$

are the sums of squared errors (SSE) depending on whether the predicting variables are all present or all absent. For model (2.1), one can estimate R^2 by similarly replacing $\hat{\beta}_{\text{LS}}$ and $X'Y$ by $\hat{\beta}$ and $X^{*\prime}Y$, respectively. The resultant statistic is then

$$(R^*)^2 \stackrel{\text{def}}{=} 1 - \text{SSE}_p^*/\text{SSE}_1,$$

where $\text{SSE}_p^* \stackrel{\text{def}}{=} Y'Y - \hat{\beta}' X^{*\prime} Y$ and is a consistent estimate of R^2 in the sense that $(R^*)^2 - R^2 \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$.

4. THE MOTIVATING EXAMPLE AND AN ALTERNATIVE FORMULA FOR THE COVARIANCE ESTIMATOR

In this section, I discuss the example that motivates model (2.1) and its assumptions. The details of the application are given in Hsueh (1984). Here I describe briefly some simplified models and justify the assumptions made at the beginning of Section 2.

4.1 The Motivating Example

Economists and other groups of researchers have been interested in analyzing the relationship between energy consumption of a household in the United States and its housing characteristics. In the past, researchers usually took one of two approaches. In the engineering approach, they built (theoretically or physically) a model house and used it to study the heat (or energy) loss as a function of the housing characteristics (see, e.g., O'Neal, Corum, and Jones 1981; Peterson 1984).

The other possible approach that is discussed here is the statistical approach in which a statistical model is fit to data collected from real households, in this case by the U.S. Department of Energy. These data, based on four surveys of residential energy use starting in 1978, are stored on computer tapes that are available for public use.

For 1980–1981, the data consist of information of 5,979 households randomly chosen from the United States. The yearly energy (in British thermal units) consumed by a chosen family was reported. With the family's permission, this consumption information was obtained directly from fuel records maintained by the household's fuel suppliers. The suppliers also provided information about the local unit price (average rate or marginal rate) of consumed energy source, such as oil, gas, or electricity. The owner of the household (or the owner's spouse) was interviewed. Interviewers gathered information about household conditions such as the number of windows, the number of storm windows, the enclosed heated area, and the inches of wall insulation, roof insulation, and floor insulation. The interviewees provided information about the total family income and about whether there were persons staying in the house during the day and whether there were certain major appliances (such as a water heating tank or air conditioners). Also reported on the computer tapes are a geographic region index and local weather conditions. Weather conditions are usually measured by a variable called yearly heating degree days (or cooling degree days). The yearly heating (cooling) degree days with base, say, 65°F represent the yearly sum of degrees the daily average local temperature is below (above) 65°F (see Energy Information Administration 1982, p. 249).

Hsueh (1984) treated the problem by setting up a linear model that relates the yearly consumed energy to some particular combinations of housing characteristics. She proposed using predicting variables such as the product of enclosed heated area of the house, the heating degree days, and the inverse of roof insulation thickness, or the product of enclosed heated area, heating degree days, and the inverse of wall insulation thickness. Theoretically, the heat (or energy) loss is proportional to the weighted sum of these multiplicative variables (see American Society of Heating, Refrigerating, and Air-Conditioning Engineers 1981, pp. 25.1–25.2). Therefore, it is more reasonable to consider the linear combination of the multiplicative predicting variables than other models in, for example, Neels (1981).

A simplified linear model that is essentially Hsueh's (1984) is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \varepsilon_i. \quad (4.1)$$

The variables X_{i1} and X_{i2} are 0 or 1, indicating whether the ith household uses oil, gas, or electricity as the main source of energy [say (X_{i1}, X_{i2}) is (0, 0), (0, 1), or (1, 0) if the ith house uses oil, gas, or electricity]; X_{i3} indicates whether there are family members staying at home during the day; X_{i4} is the product of enclosed heated area of the house, yearly heating degree days (h_i), and the inverse of roof insulation thickness; X_{i5} denotes the product of enclosed heated area, h_i, and the inverse of wall insulation thickness; X_{i6} is the fuel price marginal rate [Hsueh (1984) used a related variable, the average rate of the fuel price].

If the heating (and cooling) degree days, h_i, and the marginal rate variables were given exactly, then all of the predicting variables in (4.1) could be assumed to be accurately observed. In particular, in 1980 and 1981 surveys, the dimensions of the heated area in the "interviewed" house were measured precisely (not roughly estimated) by the interviewers (see Energy Information Administration 1982, p. 111; 1983b, p. 105). The analysis is then quite standard and the least squares approach can be adopted.

For the 1980 survey or after, however, the U.S. Department of Energy does not provide, on the public-use computer tape, the exact data about h_i and the marginal rate variables (see Energy Information Administration 1983a). Apparently, it is believed that knowledge of h_i and the marginal rate variables could be used to identify house owners. Therefore, to protect the interviewee's identities, the heating (and cooling) degree days and the marginal rate variables were multiplied by randomly generated numbers. Specifically, for the ith household in the area of, say, heating degree days h_i, a random number, r_i, is generated from a normal distribution with mean 1 and variance σ^2. Let u_i be the truncated value of r_i so that $0 \le B_0 \le |u_i - 1| \le B_1$ for some constants B_0 and B_1 (i.e., u_i is the point satisfying these inequalities that is closest to r_i). Hence instead of informing h_i, $h_i u_i$ is reported on the computer tape for public use. Obviously, r_i is not given. The values of B_0, B_1, and σ, however, are informed to be .01, .6, and .15 for heating and cooling degree days and .1, .4, and .1 for marginal rate variables. Furthermore, the random numbers used to contaminate the marginal rate variables are generated independently of those associated with heating degree days (see Energy Information Administration 1983a, p. 25). Note that the distribution of u_i is completely given. It is neither continuous nor discrete but is mixed due to truncation.

Therefore, to analyze the energy data, model (4.1) should be modified as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \varepsilon_i, \quad (4.2)$$

and $X^*_{i4} = U_{i4} X_{i4}$, $X^*_{i5} = U_{i5} X_{i5}$, $X^*_{i6} = U_{i6} X_{i6}$, and $U_{i4} = U_{i5} = u_i$. The variables, X_{i4}, X_{i5}, and X_{i6} cannot be observed exactly, but X_{i1}, X_{i2}, and X_{i3} can be. Furthermore, $U_{i4} = U_{i5} = u_i$ are statistically independent of U_{i6}. Model (4.2) is a multiplicative errors-in-variables linear model of the form (2.1), with $U_{ij} = 1$, $1 \le j \le 3$, and $\forall\ i$. All of the assumptions (2.2) through (2.5) are intuitively reasonable. In particular, the U_{ij}'s in (4.2) are generated independently from the problem; hence the independence assumption of U and (ε, X) is reasonable.

The least squares estimator is inappropriate, since it is shown in Theorem 2.3 to be asymptotically inconsistent. (Asymptotic properties are important, since $n = 5{,}979$ is large.) The recommended (strongly) consistent estimator $\hat{\beta}$ is given in (2.11). The asymptotic distribution of $\sqrt{n}(\hat{\beta} - \beta)$ is normal with mean zero and a covariance matrix Σ in (3.4), which can be consistently estimated by $\hat{\Sigma}$ of (3.12). For the pathological cases when $\hat{\Sigma}_2$ is not positive definite, see the discussion at the end of Section 3. Also see (3.13) for another consistent estimate for Σ. A modified R^2 statistic for the overall fit is also provided at the end of Section 3.

4.2 A Simpler Formula

In applying the consistent covariance estimate $\hat{\Sigma}$, the matrix computation [for $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ in (3.8) and (3.11)] is not difficult (by using computer programs such as SAS) if the dimensions of matrices are not too big. When the dimensions of the matrices are huge, however, calculation may become impossible. But there are situations in which a simpler formula for $\hat{\Sigma}_2$ is available. This formula was used in Hsueh (1984).

The simpler formula applies to the situation in which

$$U_i' = (1_a', u_i 1_b'), \qquad a + b = p. \quad (4.3)$$

Hence the error-ridden predicting variables are multiplied by the *same* error u_i. Because of (4.3) one can establish the simpler formula in Theorem 4.1. Let m_i denote the ith moment of u_1; that is, $m_i = Eu_1^i$.

Theorem 4.1. Under (4.3),

$$\hat{\Sigma}_2 = (1/n) \sum_i \{[(0_a', \hat{\beta}_2')X_i^*]^2 K \boxdot X_i^* X_i^{*\prime}\}, \quad (4.4)$$

where $(0_a', \hat{\beta}_2')'$ is the estimator $\hat{\beta}'$, but the first a elements are replaced by zero, and

$$K = \begin{pmatrix} k_1 1_a 1_a' & k_2 1_a 1_b' \\ k_2 1_b 1_a' & k_3 1_b 1_b' \end{pmatrix},$$

where $k_1 = (m_2 - 1)/m_2$, $k_2 = -m_2/m_3 + 1/m_2$, and $k_3 = m_2/m_4 - (2m_3/m_2 m_4) + m_2^{-2}$.

Proof. See Appendix B.

The low-order moments $m_1, \ldots, m_4$ of u_1 for the energy problem can be computed using well-known facts concerning the truncated normal distribution.

Theorem 4.1 applies to (4.2) if the marginal fuel price X_6 variable is not included in the regression or if, as in Hsueh (1984), one uses instead a related variable, the average rate of the fuel price, which (unlike X_6) was given exactly by the Department of Energy. Theorem 4.1 can be generalized to situations like (4.2) with X_6 in the equation. The resultant expression for $\hat{\Sigma}_2$, however, is not much simpler than (3.11). Therefore, (3.11) may as well be used.

It may appear that the consistent (point and set) estimators (when tackling the real problems) would be very similar to those derived by using the least squares approach. To the contrary, this is not so. In fact the estimate of β will be fairly different. When applied to the testing hypothesis problems, the modified procedures and the least squares estimator can draw fairly different conclusions. For numerical details, see Hsueh (1984, sec. 5.3.2). She fitted the energy consumption to 28 predicting variables similar to the X's in (4.2). For the situation in which the heating energy source is natural gas, the percentage of change of the consistent estimate over the least squares estimate can range from 480% to .4%. Six of the 28 parameters were estimated by the consistent estimator that is at least 30% different from the least squares estimate. Of course, the latter is not a legitimate procedure for the problem because of inconsistency. It was also observed that the differences are larger for coefficients related to variables observed with error than for those observed exactly.

5. CONCLUSION

In this article, a technique has been developed to construct a consistent estimator by correcting an inconsistent one. This appears to be a fairly general method. It can be applied to the traditional linear errors-in-variables model, when the predicting variables have an additive error, assuming further conditions so that the problem is identifiable. Typically, one assumes that the variances of the errors of the predicting variables are known or that the ratios of the variances of the errors of the predicting and response variables are known. One can also apply this method to obtain a consistent estimator to more complex cases, such as those considered in Fuller (1980).

In regard to multiplicative models considered in this article, note that the present method can be applied easily to polynomial models [instead of the linear model in (2.1)],

$$Y = g(\beta, X) + \varepsilon, \quad (5.1)$$

where the function g is linear in β but is polynomial in X's. When X can be observed exactly, this is essentially a linear model (linear in β). Hence the least squares approach can be applied and no new problem is created. When X can only be observed with error, however, (5.1) is substantially different from the linear model (2.1) (see Wolter and Fuller 1982a, who dealt only with the quadratic model). For the general polynomial model (5.1), the present technique works with only a minor modification so that (2.6) is satisfied. [One can achieve this by rescaling, as discussed in the sentences following (2.6).] After the modification, the present estimator and its associated confidence set are asymptotically consistent and are applicable.

This technique can also be applied to the situation in which some of the predicting variables have multiplicative errors and the rest have additive errors, given that the first and second moments are known. To be specific, assume (1.1) and, instead of observing X, one can only observe

$$X^* = X \boxdot \begin{pmatrix} U_1' & 1_{p-k}' \\ \vdots & \vdots \\ U_n' & 1_{p-k}' \end{pmatrix} + \begin{pmatrix} 0_k' & \lambda_1' \\ \vdots & \vdots \\ 0_k' & \lambda_n' \end{pmatrix},$$

where 0_k is a column vector of k zeros. The U_i's and λ_i's are, respectively, $k \times 1$ and $(p - k) \times 1$ random vectors with $EU_i = 1_k$ and $E\lambda_i = 0_{n-k}$. The rows of the three matrices on the right side are assumed to be iid. Therefore, this is a structural model with the first k predicting variables contaminated by the multiplicative errors, U's, and the rest contaminated by the additive errors, λ's. Similar to the derivation of (2.11), a consistent estimator is

$$\hat{\beta} = \left\{\left[X^{*\prime}X^* - \begin{pmatrix} 0_{k\times k}' & 0_{k\times(n-k)} \\ 0_{(n-k)\times k} & nE\lambda_1\lambda_1' \end{pmatrix}\right] \boxdiv \begin{pmatrix} EU_1U_1' & 1_{k\times(n-k)} \\ 1_{(n-k)\times k} & 1_{(n-k)\times(n-k)} \end{pmatrix}\right\}^{-1} X^{*\prime}Y,$$

where $0_{l_1\times l_2}$ and $1_{l_1\times l_2}$ are $l_1 \times l_2$ matrices of zeros and ones, respectively.

Thus far, the derivation depends heavily on the explicit expression of the least squares estimator. This is not necessary, however. To demonstrate the alternative way, write the sum of squares (normalized by n) as

$$\frac{1}{n}(Y - X\beta)'(Y - X\beta) = \frac{Y'Y}{n} - \frac{2\beta'X'Y}{n} + \beta'\frac{X'X\beta}{n}.$$

Replace $X'Y/n$ and $X'X/n$ by the consistent estimate $(X^*)'Y/n$ and $[(X^{*\prime}X^*) \boxdiv M]/n$ and then minimize by varying β. The resultant estimator is exactly $\hat{\beta}$. Using this approach, one might be able to correct the asymptotic bias of the least squares estimator under nonlinear models even when its explicit formula is not available. Such situations can occur when the model is neither linear nor polynomial. One might also be able to apply this method to other minimum distance estimators.

APPENDIX A: PROOF OF THEOREM 2.2

With probability 1, all of the U_{ij}'s are nonzero simultaneously. Condition on $U = u$, where u is a matrix without any zero element. Then the determinant of $\hat{A}$ is a polynomial in X_{ij}'s. This is a nonzero polynomial as long as we can find an X so that $\det \hat{A} \neq 0$. Let $X = \binom{I}{0}$, where I is the $p \times p$ identity matrix and 0 is the zero matrix of size $(n - p) \times p$. Then

$$\det[(X^{*\prime}X^*) \boxed{\div} M] = u_{11}^2 \cdots u_{pp}^2/M_{11} \cdots M_{pp} \neq 0,$$

where u_{ij} and M_{ij} are the (i, j)th elements of u and M, respectively. Returning to a general X, by the lemma of Okamoto (1973, p. 763), the set of X_{ij}'s for which $\det \hat{A} = 0$ has Lebesgue measure zero. Consequently,

$$\Pr(\det \hat{A} = 0 \mid U = u) = 0,$$

which obviously implies the theorem.

APPENDIX B: PROOFS OF LEMMA 3.1 AND THEOREM 4.1

In this Appendix, I will use repeatedly the following equations. Let A, B, C, and D be matrices and let V_1, V_2, V_3, and V_4 be column vectors so that all of the operations in the following equations are meaningful. One can easily establish that

$$(A \otimes B)(C \otimes D) = AC \otimes BD, \tag{B.1}$$

$$(A \otimes B) \boxdot (C \otimes D) = (A \boxdot C) \otimes (B \boxdot D), \tag{B.2}$$

$$(V_1 \boxdot V_2)(V_3 \boxdot V_4)' = (V_1V_3') \boxdot (V_2V_4'), \tag{B.3}$$

$$\operatorname{vec}(V_1V_2') = V_2 \otimes V_1, \tag{B.4}$$

$$\operatorname{vec}(A \boxdot B) = \operatorname{vec} A \boxdot \operatorname{vec} B, \tag{B.5}$$

$$(A \otimes B)' = A' \otimes B'. \tag{B.6}$$

Furthermore, these equations hold if $\boxdot$ is replaced by $\boxed{\div}$ whenever expressions are well defined.

Proof of Lemma 3.1. From (3.5), $Z_2 = C\beta$, where

$$C = X_1^*X_1' - (X_1^*X_1^{*\prime}) \boxed{\div} M = (X_1X_1') \boxdot V.$$

(For a definition of V, see the statement of the lemma.) The last equation follows from (B.3). Let C_i' denote the ith row of C, and let B denote $I_p \otimes \beta'$. Then

$$C\beta = \begin{pmatrix} C_1'\beta \\ \vdots \\ C_p'\beta \end{pmatrix} = \begin{pmatrix} \beta'C_1 \\ \vdots \\ \beta'C_p \end{pmatrix} = B \operatorname{vec} C',$$

and hence

$$\operatorname{cov} Z_2 = BE \operatorname{vec} C'(\operatorname{vec} C')'B'. \tag{B.7}$$

Since, by (B.5), $\operatorname{vec} C' = \operatorname{vec}(X_1X_1') \boxdot \operatorname{vec} V'$, it follows from (B.3) and the independence of X_1 and U_1 that

$$E \operatorname{vec} C'(\operatorname{vec} C')'$$
$$= \{E[\operatorname{vec}(X_1X_1')][\operatorname{vec}(X_1X_1')]'\} \boxdot \{E \operatorname{vec} V'(\operatorname{vec} V')'\}. \tag{B.8}$$

Using independence of X_1 and U_1, (B.3), and (B.5), we establish that

$$E \operatorname{vec}(X_1X_1')[\operatorname{vec}(X_1X_1')]'$$
$$= E \operatorname{vec}(X_1^*X_1^{*\prime})[\operatorname{vec}(X_1^*X_1^{*\prime})]' \boxed{\div} E \operatorname{vec}(U_1U_1')[\operatorname{vec}(U_1U_1')]'.$$

Hence (B.8) equals $\{E[\operatorname{vec}(X_1^*X_1^{*\prime})(\operatorname{vec} X_1^*X_1^{*\prime})']\} \boxdot Q$, which, together with (B.6) and (B.7), imply (3.9).

Proof of Theorem 4.1. We must show that (3.11) reduces to (4.4) under the assumption (4.3). To calculate Q [in (3.10)], first note that by the fact that $Eu_1 = 1$,

$$V = \begin{pmatrix} 0 & (1 - u_1)1_a1_b' \\ 0 & (u_1 - u_1^2/Eu_1^2)1_b1_b' \end{pmatrix} = \begin{pmatrix} (1 - u_1)1_a \\ (u_1 - u_1^2/m_2)1_b \end{pmatrix}(0 \ 1_b').$$

In the last equation and the remainder of the proof, 0 can represent a zero matrix, column vector, or row vector. The dimensions are obvious and therefore are not specified.

Applying (B.4) and (B.1), one can then obtain

$$E(\operatorname{vec} V')(\operatorname{vec} V')' = \begin{pmatrix} (m_2 - 1)1_a1_a' & d_21_a1_b' \\ d_21_b1_a' & d_31_b1_b' \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1_b1_b' \end{pmatrix},$$

where the last matrix is a $p \times p$ matrix with all elements zero or one, $d_2 = E(1 - u_1)(u_1 - u_1^2/m_2)$, and $d_3 = E(u_1 - u_1^2/m_2)^2$. Hence

$$Q = \begin{pmatrix} [(m_2 - 1)/m_2]1_a1_a' & (d_2/m_3)1_a1_b' \\ (d_2/m_3)1_b1_a' & (d_3/m_4)1_b1_b' \end{pmatrix} \otimes \begin{pmatrix} 0 & 0 \\ 0 & 1_b1_b' \end{pmatrix}. \tag{B.9}$$

Direct calculations show that the first matrix on the right side is exactly K, where K is defined in the theorem. Using (B.4), (B.6), and (B.1), we have

$$(\operatorname{vec} X_i^*X_i^{*\prime})(\operatorname{vec} X_i^*X_i^{*\prime})' = (X_i^* \otimes X_i^*)(X_i^{*\prime} \otimes X_i^{*\prime})$$
$$= (X_i^*X_i^{*\prime}) \otimes (X_i^*X_i^{*\prime}).$$

Hence the ith term of (3.11) is

$$(1/n)(I_p \otimes \hat{\beta}')\{[(X_i^*X_i^{*\prime}) \otimes (X_i^*X_i^{*\prime})] \boxdot Q\}(I_p \otimes \hat{\beta}),$$

which, by (B.2) and (B.8), equals

$$(1/n)(I_p \otimes \hat{\beta}')\Big\{((X_i^*X_i^{*\prime}) \boxdot K) \otimes \Big[(X_i^*X_i^{*\prime}) \boxdot \begin{pmatrix} 0 & 0 \\ 0 & 1_b1_b' \end{pmatrix}\Big]\Big\}(I_p \otimes \hat{\beta}).$$

Using (B.1), one shows that the last expression equals

$$\frac{1}{n}\{(X_i^*X_i^{*\prime}) \boxdot K\} \otimes \Big\{\hat{\beta}'\Big[(X_i^*X_i^{*\prime}) \boxdot \begin{pmatrix} 0 & 0 \\ 0 & 1_b1_b' \end{pmatrix}\Big]\hat{\beta}\Big\}. \tag{B.10}$$

Now,

$$(X_i^*X_i^{*\prime}) \boxdot \begin{pmatrix} 0 & 0 \\ 0 & 1_b1_b' \end{pmatrix} = \Big[(X_i^*X_i^{*\prime}) \boxdot \begin{pmatrix} 0 \\ 1_b \end{pmatrix}(0 \ \ 1_b')\Big]$$
$$= \Big[X_i^* \boxdot \begin{pmatrix} 0 \\ 1_b \end{pmatrix}\Big]\Big[X_i^* \boxdot \begin{pmatrix} 0 \\ 1_b \end{pmatrix}\Big]'$$

by (B.3). Hence

$$\hat{\beta}'\Big[(X_i^*X_i^{*\prime}) \boxdot \begin{pmatrix} 0 & 0 \\ 0 & 1_b1_b' \end{pmatrix}\Big]\hat{\beta} = \Big\{\hat{\beta}'\Big[X_i^* \boxdot \begin{pmatrix} 0 \\ 1_b \end{pmatrix}\Big]\Big\}^2. \tag{B.11}$$

Direct calculation shows that

$$\hat{\beta}'\Big[X_i^* \boxdot \begin{pmatrix} 0 \\ 1_b \end{pmatrix}\Big] = (0, \hat{\beta}_2)X_i^*. \tag{B.12}$$

Plugging (B.12) and (B.11) into (B.10) establishes (4.4).

[*Received July 1984. Revised December 1985.*]

REFERENCES

American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc. (1981), *Ashrae Handbook 1981, Fundamentals*, New York: Author.

Anderson, T. W. (1984), "Estimating Linear Statistical Relationships," *The Annals of Statistics*, 12, 1–45.

Dolby, G. R., and Lipton, S. (1972), "Maximum Likelihood Estimation of the General Nonlinear Functional Relationship With Replicated Observations and Correlated Errors," *Biometrika*, 59, 121–129.

Energy Information Administration (1982), "Residential Energy Consumption Survey: Consumption and Expenditures, April 1980 Through March 1981, Part 1: National Data," technical report, Washington, DC: U.S. Department of Energy.

——— (1983a), "Residential Energy Consumption Survey, Housing Characteristics 1980, and Consumption and Expenditure, April 1980 Through March 1981, Public Use Tape Technical Documents," technical report, Washington, DC: National Technical Information Service.

——— (1983b), "Residential Energy Consumption Survey: Consumption and Expenditures, April 1981 Through March 1982, Part 1: National Data," technical report, Washington, DC: U.S. Department of Energy.

Fuller, W. A. (1980), "Properties of Some Estimators for the Errors-in-Variables Model," *The Annals of Statistics*, 8, 407–422.

——— (1984), "Measurement Error Models With Heterogeneous Error Variances," in *Topics in Applied Statistics*, eds. Y. P. Chaubey and T. D. Dwivedi, Montreal: Concordia University, pp. 257–289.

Fuller, W. A., and Hidiroglou, M. A. (1978), "Regression Estimation After Correcting for Attenuation," *Journal of the American Statistical Association*, 73, 99–104.

Henderson, H. V., and Searle, S. R. (1981), "The Vec-Permutation Matrix, the Vec Operator and Kronecker Products: A Review," *Linear and Multilinear Algebra*, 9, 271–288.

Hidiroglou, M. A., Fuller, W. A., and Hickman, R. D. (1980), "Super Carp," technical report, Iowa State University, Dept. of Statistics.

Hsueh, L. M. (1984), "A Model of Home Heating and Calculation of Rates of Return to Household Energy Conservation Investment," unpublished Ph.D. thesis submitted to Cornell University, Dept. of Consumer Economics and Housing.

Kendall, M. G., and Stuart, A. (1967), *The Advanced Theory of Statistics* (2nd ed.), London: Charles W. Griffin.

Loève, M. (1977), *Probability Theory I* (4th ed.), New York: Springer-Verlag.

Neels, K. (1981), "Families, Houses and the Demand for Energy: Housing Assistance Supply Experiment," Rand Note N-1542-HUD, Santa Monica, CA: Rand Corporation.

Okamoto, M. (1973), "Distinctness of the Eigenvalues of a Quadratic Form in a Multivariate Sample," *The Annals of Statistics*, 1, 763–765.

O'Neal, D. L., Corum, K. R., and Jones, J. L. (1981), "Estimate of Consumer Discount Rate Implicit in Single-Family-Housing Construction Practice," Technical Report ORNL/CON-62, Oak Ridge, TN: Oak Ridge National Laboratory.

Peterson, S. R. (1984), "Economics and Energy Conservation of New Single-Family Housing," technical report, Washington, DC: U.S. Department of Commerce, National Bureau of Standards.

Schur, J. (1911), "Bemerkungen zur Theorie der Beschränkten Bilinearformen mit Unendlich Vielen Veränderlichen," *Journal für die Reine und Angewandte Mathematik*, 140, 1–28.

Schwarz, R. E. (1967), "Admissible Tests in Multivariate Analysis of Variance," *Annals of Mathematical Statistics*, 38, 698–710.

Stefanski, L. (1985), "The Effects of Measurement Error on Parameter Estimation," *Biometrika*, 72, 583–592.

Styan, G. (1973), "Hadamard Products and Multivariate Statistical Analysis," *Linear Algebra and Its Applications*, 6, 217–240.

Wolter, K. M., and Fuller, W. A. (1982a), "Estimation of the Quadratic Errors-in-Variables Model," *Biometrika*, 69, 175–182.

——— (1982b), "Estimation of Nonlinear Errors-in-Variables Models," *The Annals of Statistics*, 10, 539–548.