The Annals of Statistics
1980, Vol. 8, No. 3, 488–505

THE n^{-2}-ORDER MEAN SQUARED ERRORS OF THE MAXIMUM LIKELIHOOD AND THE MINIMUM LOGIT CHI-SQUARE ESTIMATOR[1]

BY TAKESHI AMEMIYA

Stanford University

The n^{-2}order mean squared errors of the maximum likelihood and the minimum chi-square estimator of the logit regression model are derived and the latter is shown to be superior for many parameter values considered. The maximum likelihood is shown to be better if the bias of each estimator is corrected to the order of n^{-1}; however, the difference is shown to be negligibly small in many practical situations.

1. Introduction. Berkson [1944] proposed a noniterative estimator for the dichotomous logit regression model, which has been extensively used by researchers in various fields and commonly referred to as Berkson's minimum logit chi-square estimator. As was first shown by Taylor [1953], the estimator has the same asymptotic normal distribution as the maximum likelihood estimator when the number of observations of the dependent variable for each value of the vector of independent variables goes to infinity. Note that in situations where this number is small, Berkson's estimator tends to break down even though the maximum likelihood estimator can be used. Berkson [1955] evaluated the exact mean squared errors of the minimum chi-square and the maximum likelihood estimator for certain simple models and showed that the mean squared error of the minimum chi-square estimator is smaller in all the cases considered. In this paper we evaluate the mean squared errors of the two estimators to the order of n^{-2} (n being the "average" number of observations) and show that the superiority of the minimum chi-square estimator in terms of the mean squared error generally holds for a much broader class of models than those considered by Berkson. We also evaluate the n^{-2}-order mean squared error matrices of the two estimators after correcting for the bias to the order of n^{-1} and show that in this case the maximum likelihood estimator is superior. A special case of this last result can be found in Ghosh and Subramanyam [1974], who obtained it as an application of their general theorem concerning the second-order efficiency of the maximum likelihood estimator in the exponential family. The numerical study of various examples shows that the difference in the mean squared errors between Berkson's estimator and the maximum likelihood estimator is considerable, whereas the difference between

Received January 1979; revised July 1979.

[1]This research was supported by National Science Foundation Grant SOC77-14944 at the Institute for Mathematical Studies in the Social Sciences, Stanford University.

AMS 1970 *subject classifications.* Primary 62F20, 62F10; secondary 62J05, 62P20.

Key words and phrases. Logit regression model, maximum likelihood estimator, minimum chi-square estimator, second-order efficiency, dichotomous random variable.

Berkson's estimator and the bias-corrected maximum likelihood estimator is always less pronounced and mostly negligible.

The order of the presentation is as follows: In Section 2 we present our basic model, define frequently used symbols, and give the first few moments of the binomial random variables for the later reference. In Section 3 the validity of the asymptotic expansion to be used later in the paper is briefly considered. The derivation of the approximate mean squared error matrices for the maximum likelihood and the minimum chi-square estimator is given in Sections 4 and 5. The approximate biases of the estimators are also obtained. In Section 6 the approximate mean squared errors of the bias-corrected estimators are given. In Section 7 the mean squared error matrices are numerically evaluated for many examples both real and artificial. Finally, brief conclusions are stated in Section 8.

2. Model and notation. Let the $\Sigma_{t=1}^{T} n_t$ observable dichotomous random variables $y_{t\nu}$ ($t = 1, 2, \cdots, T$; $\nu = 1, 2, \cdots, n_t$) take the value 1 or 0 according to the probability distribution

$$P(y_{t\nu} = 1) = \frac{1}{1 + e^{-x_t'\beta_0}} \equiv P_t, \tag{1}$$

where x_t is a K-dimensional vector of known constants and β_0 is a K-dimensional vector of unknown parameters. We define $r_t = n_t^{-1}\Sigma_{\nu=1}^{n_t} y_{t\nu}$. Then the vector $r = (r_1, r_2, \cdots, r_T)'$ constitutes the set of sufficient statistics of the model. The two estimators of β_0 we consider—the maximum likelihood and Berkson's minimum chi-square—are both given as functions of r. In the asymptotic expansion we will use later, we will assume that T is a fixed number greater than or equal to K and each n_t goes to infinity at the same speed. We will designate this common speed as n.

Some frequently used symbols are defined as follows: $u_t = r_t - P_t$, $P = (P_1, P_2, \cdots, P_T)'$, $u = (u_1, u_2, \cdots, u_T)'$, and $X = (x_1, x_2, \cdots, x_T)'$. Subscripts s, t, and τ will denote a particular element of the vector r, P, or u and a particular row of the matrix X. They range from 1 to T. Subscripts i, j, k, and l will denote a particular element of the vector β or x_t. They range from 1 to K. The subscript ν will be used in all the other circumstances. The symbol $D(c_t)$ denotes the $T \times T$ diagonal matrix whose tth diagonal element is c_t. The symbol $\{c_{st}\}$ denotes the $T \times T$ matrix whose s, tth element is c_{st}. Similarly, $\{c_{ij}\}$ denotes the $K \times K$ matrix whose i, jth element is c_{ij}.

We will need the following moments of u_t, which one can find, for example, in Johnson and Kotz ([1969] page 51);

$$Eu_t^2 = \frac{P_t(1 - P_t)}{n_t}, \tag{2}$$

$$Eu_t^3 = \frac{P_t(1 - P_t)(1 - 2P_t)}{n_t^2}, \tag{3}$$

and

$$Eu_t^4 = \frac{3P_t^2(1-P_t)^2}{n_t^2} + \frac{P_t(1-P_t)[1-6P_t(1-P_t)]}{n_t^3}. \tag{4}$$

We will also use the fact that Eu_t^5, $Eu_t^6, \cdots$ are at most of the order of n_t^{-3}.

3. Asymptotic expansion. Before setting out to evaluate the mean squared errors of the estimators to the order of n^{-2}, we will briefly consider the validity of the asymptotic expansion used in our study so that we can meaningfully relate our results to the exact results obtained by Berkson [1955].

Let b be either the maximum likelihood or the minimum chi-square estimator of β_0. Then b can be written as

$$\begin{aligned} b &= h(r) \qquad \text{if} \quad r \in S \\ &= q(r) \qquad \text{if} \quad r \in \bar{S}, \end{aligned} \tag{5}$$

where h is a bounded infinitely differentiable function, q is a bounded function, S is a subset of the Euclidean T-space such that $\lim_{n\to\infty} P(r \in S) = 1$, and $\bar{S}$ is the complement of S. From (5) we have

$$\begin{aligned} \text{MSE}^*(b) &= \Sigma_{r\in S}[h(r)-\beta_0][h(r)-\beta_0]'P(r) \\ &\quad + \Sigma_{r\in \bar{S}}[q(r)-\beta_0][q(r)-\beta_0]'P(r) \\ &\equiv M^* + N^*, \end{aligned} \tag{6}$$

where MSE* denotes the exact mean squared error matrix. Our approximation of MSE*(b) is done in two steps. First, expand $h(r)$ in a Taylor series around P retaining the terms that involve up to the third derivatives of h. Denote the approximation of $h(r) - \beta_0$ thus obtained as $\varepsilon(r)$ and define

$$\begin{aligned} \text{MSE}(b) &= \Sigma_{r\in S}\varepsilon(r)\varepsilon(r)'P(r) + \Sigma_{r\in \bar{S}}\varepsilon(r)\varepsilon(r)'P(r) \\ &\equiv M + N, \end{aligned} \tag{7}$$

where MSE denotes the approximate mean squared error matrix. Then, $M - M^*$ is the order of n^{-3} because of the result stated at the end of the previous section. Second, we approximate MSE(b) by dropping all the terms of order n^{-3} and smaller (that is, retaining the term of order up to n^{-2}). Thus, our approximation is good to the extent that the term of order n^{-3} is small and to the extent that $\Sigma_{r\in\bar{S}}P(r)$ is small. This latter condition ensures the smallness of N and N^*.

4. Maximum likelihood estimator. The logarithmic likelihood function of the model is given by

$$\log L = \Sigma_t n_t r_t \log F_t + \Sigma_t n_t(1-r_t)\log(1-F_t), \tag{8}$$

where $F_t = [1 + \exp(-x_t'\beta)]^{-1}$. Note that F_t is the logistic function evaluated at $x_t'\beta$, whereas P_t is the same function evaluated at $x_t'\beta_0$. Differentiating (8) with respect to β and setting the vector of derivatives equal to the zero vector, we obtain the normal equations

$$\Sigma_t n_t(r_t - F_t)x_t = 0. \tag{9}$$

We rewrite (9) simply as

$$g(\beta, r) = 0, \tag{10}$$

where g is a K-dimensional vector.

There are certain values of r for which no finite value of β satisfies (10). For example, $r_t = 1$ for all t is such a value. Therefore we choose a set S_1 in the Euclidean T-space such that for any r in S_1, (10) yields a bounded solution β, and define the maximum likelihood estimator $\hat{\beta}$ by

$$\begin{aligned} &g(\hat{\beta}, r) = 0, \qquad \text{or equivalently} \quad \hat{\beta}_1 = h_1(r), \qquad \text{if} \quad r \in S_1 \\ &\hat{\beta} = q_1(r) \qquad \text{if} \quad r \in \bar{S}_1, \end{aligned} \tag{11}$$

where q_1 is an unspecified bounded function. We choose the set S_1 in such a way that $\lim_{n\to\infty} P(r \in S_1) = 1$, which is possible under general assumptions on the matrix X.

We will expand $h(r)$ in a Taylor series around P for r in S_1. Noting $g(\beta_0, P) = 0$, we have

$$\begin{aligned} \hat{\beta}_i - \beta_{i0} &\cong \Sigma_t \frac{\partial \beta_i}{\partial r_t} u_t + \tfrac{1}{2}\Sigma_s\Sigma_t \frac{\partial^2 \beta_i}{\partial r_s \partial r_t} u_s u_t + \tfrac{1}{6}\Sigma_\tau\Sigma_s\Sigma_t \frac{\partial^3 \beta_i}{\partial r_\tau \partial r_s \partial r_t} u_\tau u_s u_t \\ &\equiv v_{1i} + v_{2i} + v_{3i}, \end{aligned} \tag{12}$$

where β_i and β_{i0} are the ith element of β and β_0 respectively and all the partial derivatives are evaluated at $r = P$.

We will express the partial derivatives that appear in the right-hand side of (12) as functions of the derivatives of g with respect to β and r. This can be done by solving the equations obtained by differentiating (10) three times with respect to r. Thus we obtain

$$\frac{\partial \beta}{\partial r'} = (X'D_1X)^{-1}X'D(n_t), \tag{13}$$

where $D_1 = D[n_tP_t(1 - P_t)]$,

$$\frac{\partial^2 \beta_i}{\partial r \partial r'} = \Sigma_k \gamma_{ik} Q_k, \tag{14}$$

where $\{\gamma_{ik}\} = (X'D_1X)^{-1}$, and

$$Q_k = D(n_t)X(X'D_1X)^{-1}X'D(x_{kt})D(2P_t - 1)D_1X(X'D_1X)^{-1}X'D(n_t). \tag{15}$$

We also obtain

$$\frac{\partial^3 \beta_i}{\partial r_\tau \partial r_s \partial r_t} = \left[\Sigma_k Q_k(s, t) \frac{\partial \gamma_{ik}}{\partial \beta'} + \Sigma_k \gamma_{ik} \frac{\partial Q_k(s, t)}{\partial \beta'} \right] \frac{\partial \beta}{\partial r_\tau}, \tag{16}$$

where $Q_k(s, t)$ is the s, tth element of Q_k.

From (12), the mean squared error matrix of $\hat{\beta}$ to the order of n^{-2}, denoted as MSE_1, is given by

$$\begin{aligned} \mathrm{MSE}_1 = {} & \{Ev_{1i}v_{1j}\} + \{Ev_{2i}v_{1j}\} + \{Ev_{2i}v_{1j}\}' + \{Ev_{3i}v_{1j}\} \\ & + \{Ev_{3i}v_{1j}\}' + \{Ev_{2i}v_{2j}\}. \end{aligned} \tag{17}$$

We will evaluate each of the four matrices that appear in the right-hand side of (17) without a transpose. In doing so, we will use the moments of u_t given in equations (2), (3), and (4) and drop any term smaller than the order of n^{-2}. Many lengthy steps are omitted in the following derivations.

From (2) and (13), we readily obtain

$$\{Ev_{1i}v_{1j}\} = (X'D_1X)^{-1}. \tag{18}$$

This is the usual asymptotic mean squared error matrix and the only term which is of order n^{-1}. All other terms are of order n^{-2}.

From (3), (13), and (14), we can obtain, after some algebraic manipulation,

$$\{Ev_{2i}v_{1j}\} = -\tfrac{1}{2}(X'D_1X)^{-1}X'D_2\dot{A}D_2X(X'D_1X)^{-1}, \tag{19}$$

where $D_2 = D(2P_t - 1)$ and $\dot{A}$ is the matrix obtained by squaring each element of the matrix A defined by

$$A = D_1^{\frac{1}{2}}X(X'D_1X)^{-1}X'D_1^{\frac{1}{2}}. \tag{20}$$

Defining the Hadamard product $*$ by $\{w_{st}\} * \{z_{st}\} = \{w_{st}z_{st}\}$, we can also write $\dot{A} = A * A$.

The evaluation of the remaining two terms is rather involved. Ignoring the terms of a smaller order than n^{-2}, we have

$$\begin{aligned} Ev_{3i}v_{1j} &= \tfrac{1}{2}E\Sigma_s\Sigma_t n_s\gamma_j'x_s \frac{\partial^3\beta_i}{\partial r_s \partial r_t^2} u_s^2u_t^2 \\ &= \tfrac{1}{2}\Sigma_s\Sigma_t n_s\gamma_j'x_s \frac{\partial^3\beta_i}{\partial r_s \partial r_t^2} \frac{P_s(1-P_s)}{n_s} \frac{P_t(1-P_t)}{n_t}, \end{aligned} \tag{21}$$

where γ_j' is the jth row of $(X'D_1X)^{-1}$. Inserting (16) into (21) and interchanging the summation signs, we obtain

$$\begin{aligned} Ev_{3i}v_{1j} &= \tfrac{1}{2}\Sigma_t\Sigma_k Q_k(t, t)n_t^{-1}P_t(1-P_t)\frac{\partial \gamma_{ik}}{\partial \beta'}\Sigma_s(X'D_1X)^{-1}x_sn_sP_s(1-P_s)\gamma_j'x_s \\ &\quad + \tfrac{1}{2}\Sigma_t\Sigma_k\gamma_{ik}n_t^{-1}P_t(1-P_t)\frac{\partial Q_k(t, t)}{\partial \beta'}\Sigma_s(X'D_1X)^{-1}x_sn_sP_s(1-P_s)\gamma_j'x_s \\ &\equiv d_{ij} + e_{ij}. \end{aligned} \tag{22}$$

We will first evaluate $\{d_{ij}\}$. Using

$$\frac{\partial (X'D_1X)^{-1}}{\partial \beta_l} = (X'D_1X)^{-1}X'D_1D_2D(x_{lt})X(X'D_1X)^{-1}, \tag{23}$$

we obtain after considerable algebra

$$\{d_{ij}\} = \tfrac{1}{2}(X'D_1X)^{-1}X'D_1D_2D_3X(X'D_1X)^{-1}, \tag{24}$$

where $D_3 = D[x_t'(X'D_1X)^{-1}X'D_2D_4\mathbf{1}]$, $\mathbf{1}$ is the vector consisting only of ones, and D_4 is the diagonal matrix whose tth diagonal element is the tth diagonal element of A. (In other words, $D_4 = I * A$, where $*$ is the Hadamard product.) In evaluating e_{ij} next, we need the following expression:

$$\begin{aligned}\frac{\partial Q_k(t,t)}{\partial \beta_l} &= 2n_t^2x_t'\frac{\partial (X'D_1X)^{-1}}{\partial \beta_l}X'D_1D_2D(x_{kt})X(X'D_1X)^{-1}x_t \\ &\quad + 2n_t^2x_t'(X'D_1X)^{-1}X'D(n_t)^{-1}D_1^2D(x_{lt})D(x_{kt})X(X'D_1X)^{-1}x_t \\ &\quad - n_t^2x_t'(X'D_1X)^{-1}X'D_1D_2^2D(x_{lt})D(x_{kt})X(X'D_1X)^{-1}x_t.\end{aligned} \tag{25}$$

When (25) is inserted into e_{ij}, e_{ij} becomes the sum of three terms corresponding to the three terms in the right-hand side of (25). Thus,

$$e_{ij} = e_{1ij} + e_{2ij} + e_{3ij}. \tag{26}$$

Each of the three terms in (26) is evaluated as follows:

$$\{e_{1ij}\} = (X'D_1X)^{-1}X'D_2\dot{A}D_2X(X'D_1X)^{-1}, \tag{27}$$

$$\{e_{2ij}\} = (X'D_1X)^{-1}X'D(n_t)^{-1}D_1D_4X(X'D_1X)^{-1}, \tag{28}$$

$$\{e_{3ij}\} = -\tfrac{1}{2}(X'D_1X)^{-1}X'D_2^2D_4X(X'D_1X)^{-1}. \tag{29}$$

This concludes the evaluation of $\{Ev_{3i}v_{1j}\}$.

Finally we will evaluate $\{Ev_{2i}v_{2j}\}$. We have

$$\begin{aligned}Ev_{2i}v_{2j} &= \tfrac{1}{4}\Sigma_s\Sigma_t\frac{\partial^2\beta_i}{\partial r_s\partial r_t}\frac{\partial^2\beta_j}{\partial r_s\partial r_t}\frac{P_t(1-P_t)}{n_t}\frac{P_s(1-P_s)}{n_s} \\ &\quad + \tfrac{1}{4}\Sigma_t\frac{\partial^2\beta_i}{\partial r_t^2}\frac{\partial^2\beta_j}{\partial r_t^2}\frac{P_t^2(1-P_t)^2}{n_t^2} \\ &\quad + \tfrac{1}{4}\left[\Sigma_t\frac{\partial^2\beta_i}{\partial r_t^2}\frac{P_t(1-P_t)}{n_t}\right]\left[\Sigma_t\frac{\partial^2\beta_j}{\partial r_t^2}\frac{P_t(1-P_t)}{n_t}\right] \\ &\equiv m_{1ij} + m_{2ij} + m_{3ij}.\end{aligned} \tag{30}$$

Using (14) we can derive

$$\{m_{1ij}\} = \tfrac{1}{4}(X'D_1X)^{-1}X'D_2\dot{A}D_2X(X'D_1X)^{-1}, \tag{31}$$

$$\{m_{2ij}\} = \tfrac{1}{4}(X'D_1X)^{-1}X'D_2\dot{A}\dot{A}D_2X(X'D_1X)^{-1}, \tag{32}$$

and

$$\{m_{3ij}\} = \tfrac{1}{4}(X'D_1X)^{-1}X'D_2D_4\mathbf{1}\mathbf{1}'D_2D_4X(X'D_1X)^{-1}. \tag{33}$$

Combining (17)–(33), we finally obtain

$$\begin{aligned} \mathrm{MSE}_1 = {} & (X'D_1X)^{-1} \\ & + \tfrac{5}{4}(X'D_1X)^{-1}X'D_2\dot{A}D_2X(X'D_1X)^{-1} \\ & + (X'D_1X)^{-1}X'D_1D_2D_3X(X'D_1X)^{-1} \\ & + 2(X'D_1X)^{-1}X'D(n_t)^{-1}D_1D_4X(X'D_1X)^{-1} \\ & - (X'D_1X)^{-1}X'D_2^2D_4X(X'D_1X)^{-1} \\ & + \tfrac{1}{4}(X'D_1X)^{-1}X'D_2\dot{A}\dot{A}D_2X(X'D_1X)^{-1} \\ & + \tfrac{1}{4}(X'D_1X)^{-1}X'D_2D_4\mathbf{1}\mathbf{1}'D_2D_4X(X'D_1X)^{-1}. \end{aligned} \tag{34}$$

The approximate bias of the maximum likelihood estimator to the order of n^{-1}, denoted by BS_1, can be easily obtained by taking the expectation of the first two terms in the Taylor expansion (12) as

$$\mathrm{BS}_1 = \tfrac{1}{2}(X'D_1X)^{-1}X'D_2D_4\mathbf{1}. \tag{35}$$

5. Minimum chi-square estimator. Berkson's minimum logit chi-square estimator is defined by

$$\begin{aligned} \tilde{\beta} & = [\Sigma_t n_t r_t(1-r_t)x_t x_t']^{-1}\Sigma_t n_t r_t(1-r_t)\left[\log\frac{r_t}{1-r_t}\right]x_t \\ & \equiv h_2(r) \quad \text{if } r \in S_2 \\ & = q_2(r) \quad \text{if } r \in \bar{S}_2, \end{aligned} \tag{36}$$

where q_2 is an unspecified bounded function and S_2 is a subset of the Euclidean T-space such that $\tilde{\beta}$ is bounded for all r in S_2. Like S_1, we can choose S_2 so that $\lim_{n\to\infty}P(r \in S_2) = 1$. Subtracting β_0 from both sides of (36), we obtain

$$\tilde{\beta} - \beta_0 = [\Sigma_t n_t r_t(1-r_t)x_t x_t']^{-1}\Sigma_t n_t r_t(1-r_t)\left[\log\frac{r_t}{1-r_t} - \log\frac{P_t}{1-P_t}\right]x_t \quad \text{for } r \in S_2. \tag{37}$$

The right-hand side of (37) involves the product of three different nonlinear functions of r_t. We will approximate each of the three functions by the first few terms in its Taylor series expansion around P and then calculate the mean squared error matrix of the product of the three terms.

Retaining the terms of the order up to u_t^2, we have

$$
\begin{aligned}
[\Sigma_t n_t r_t(1 - r_t)x_t x_t']^{-1} \cong{} & (X'D_1X)^{-1} + (X'D_1X)^{-1}X'D(n_t)D_2D(u_t)X(X'D_1X)^{-1} \\
& + (X'D_1X)^{-1}X'D(n_t)D(u_t^2)X(X'D_1X)^{-1} \\
& + (X'D_1X)^{-1}X'D(n_t)D_2D(u_t)X(X'D_1X)^{-1} \\
& \quad X'D(n_t)D_2D(u_t)X(X'D_1X)^{-1} \\
\equiv{} & B_1 + B_2 + B_3 + B_4.
\end{aligned} \tag{38}
$$

Similarly we have

$$
\begin{aligned}
n_t r_t(1 - r_t) &\cong n_t P_t(1 - P_t) - n_t(2P_t - 1)u_t - n_t u_t^2 \\
&\equiv w_{1t} + w_{2t} + w_{3t}.
\end{aligned} \tag{39}
$$

Retaining the terms of the order up to u_t^3, we have

$$
\begin{aligned}
& \log\frac{r_t}{1 - r_t} - \log\frac{P_t}{1 - P_t} \\
& \cong \frac{1}{P_t(1 - P_t)}u_t + \frac{2P_t - 1}{2P_t^2(1 - P_t)^2}u_t^2 + \frac{3P_t^2 - 3P_t + 1}{3P_t^3(1 - P_t)^3}u_t^3 \\
& \equiv z_{1t} + z_{2t} + z_{3t}.
\end{aligned} \tag{40}
$$

Inserting (38), (39), and (40) into the right-hand side of (37) produces $4 \times 3 \times 3 = 36$ terms. However we need to retain only 11 terms involving a cubic or smaller power of u_t for our purpose. Denoting these eleven terms by ξ_ν, $\nu = 1,2, \cdots, 11$, we have

$$
\tilde{\beta} - \beta_0 \cong \Sigma_{\nu=1}^{11}\xi_\nu. \tag{41}
$$

We now evaluate ξ's as follows:

$$
\xi_1 = B_1\Sigma_t w_{1t}z_{1t}x_t = (X'D_1X)^{-1}X'D(n_t)u. \tag{42}
$$

$$
\begin{aligned}
\xi_2 + \xi_3 &= B_1\Sigma_t w_{1t}z_{2t}x_t + B_1\Sigma_t w_{2t}z_{1t}x_t \\
&= -\tfrac{1}{2}(X'D_1X)^{-1}X'D(n_t)^2D_1^{-1}D_2u^2,
\end{aligned} \tag{43}
$$

where u^2 is the T-vector whose tth element is u_t^2.

$$
\xi_4 = B_2\Sigma_t w_{1t}z_{1t}x_t = (X'D_1X)^{-1}X'D(n_t)D(u_t)D_2X(X'D_1X)^{-1}X'D(n_t)u. \tag{44}
$$

$$
\xi_5 = B_1\Sigma_t w_{1t}z_{3t}x_t = \tfrac{1}{3}(X'D_1X)^{-1}X'D(n_t)^3D_1^{-2}D(3P_t^2 - 3P_t + 1)u^3, \tag{45}
$$

where u^3 is the T-vector whose tth element is u_t^3.

(46) $$\xi_6 = B_1\Sigma_t w_{2t}z_{2t}x_t = -\tfrac{1}{2}(X'D_1X)^{-1}X'D(n_t)^3D_1^{-2}D_2^2u^3.$$

(47) $$\xi_7 = B_1\Sigma_t w_{3t}z_{1t}x_t = -(X'D_1X)^{-1}X'D(n_t)^2D_1^{-1}u^3.$$

(48)
$$\begin{aligned}\xi_8 + \xi_9 &= B_2\Sigma_t w_{1t}z_{2t}x_t + B_2\Sigma_t w_{2t}z_{1t}x_t \\ &= -\tfrac{1}{2}(X'D_1X)^{-1}X'D(n_t)D(u_t)D_2X(X'D_1X)^{-1}X'D(n_t)^2D_1^{-1}D_2u^2.\end{aligned}$$

(49) $$\xi_{10} = B_3\Sigma_t w_{1t}z_{1t}x_t = (X'D_1X)^{-1}X'D(n_t)D(u_t^2)X(X'D_1X)^{-1}X'D(n_t)u.$$

(50) $$\begin{aligned}\xi_{11} &= B_4\Sigma_t w_{1t}z_{1t}x_t \\ &= (X'D_1X)^{-1}X'D(n_t)D(u_t)D_2X(X'D_1X)^{-1}X'D(n_t)D(u_t)D_2 \\ &\quad X(X'D_1X)^{-1}X'D(n_t)u.\end{aligned}$$

Note that of the above eleven ξ's, ξ_1 is a linear function of u, ξ_2 through ξ_4 are quadratic functions of u, and the rest are cubic functions of u.

Next we will evaluate $E\xi_\nu\xi_{\nu'}'$ for $\nu, \nu' = 1,2,\cdots,11$ omitting the terms of the order of n^{-3} and smaller. Some of the derivations are rather involved, but we will only state the final expressions. We have

(51) $$E\xi_1\xi_1' = (X'D_1X)^{-1}.$$

(52) $$E\xi_1(\xi_2+\xi_3+\xi_4)' = \tfrac{1}{2}(X'D_1X)^{-1}X'D_2^2(I-2D_4)X(X'D_1X)^{-1}.$$

(53) $$E(\xi_2+\xi_3)(\xi_2+\xi_3)' = \tfrac{1}{4}(X'D_1X)^{-1}X'D_2(\mathbf{11}'+2I)D_2X(X'D_1X)^{-1}.$$

(54) $$E\xi_4\xi_4' = (X'D_1X)^{-1}X'D_2(\dot{A}+D_4\mathbf{11}'D_4+D_4)D_2X(X'D_1X)^{-1}.$$

(55) $$\begin{aligned}E(\xi_2+\xi_3)\xi_4' = &-(X'D_1X)^{-1}X'D_2^2D_4X(X'D_1X)^{-1} \\ &-\tfrac{1}{2}(X'D_1X)^{-1}X'D_2\mathbf{11}'D_2D_4X(X'D_1X)^{-1}.\end{aligned}$$

(56) $$E\xi_1\xi_5' = (X'D_1X)^{-1}X'D(3P_t^2-3P_t+1)X(X'D_1X)^{-1}.$$

(57) $$E\xi_1\xi_6' = -\tfrac{3}{2}(X'D_1X)^{-1}X'D_2^2X(X'D_1X)^{-1}.$$

(58) $$E\xi_1\xi_7' = -3(X'D_1X)^{-1}X'D(n_t)^{-1}D_1X(X'D_1X)^{-1}$$

(59) $$\begin{aligned}E\xi_1(\xi_8+\xi_9)' = &-\tfrac{1}{2}(X'D_1X)^{-1}X'D_1D_2D_5X(X'D_1X)^{-1} \\ &-(X'D_1X)^{-1}X'D_2^2D_4X(X'D_1X)^{-1},\end{aligned}$$

where $D_5 = D[x_t'(X'D_1X)^{-1}X'D_2\mathbf{1}]$.

(60) $$E\xi_1\xi_{10}' = (X'D_1X)^{-1}X'D(n_t)^{-1}D_1X(X'D_1X)^{-1} + 2(X'D_1X)^{-1}X'D(n_t)^{-1}D_1D_4X(X'D_1X)^{-1}.$$

(61) $$\begin{aligned} E\xi_1\xi_{11}' = &(X'D_1X)^{-1}X'D_1D_2D_3X(X'D_1X)^{-1} \\ &+ (X'D_1X)^{-1}X'D_2^2D_4X(X'D_1X)^{-1} \\ &+ (X'D_1X)^{-1}X'D_2\dot{A}D_2X(X'D_1X)^{-1}. \end{aligned}$$

From (41) and (51)–(61), we finally obtain the mean squared error matrix MSE_2 of $\tilde{\beta}$ to the order of n^{-2} as

(62) $$\begin{aligned} \mathrm{MSE}_2 = &(X'D_1X)^{-1} \\ &+\tfrac{1}{2}(X'D_1X)^{-1}X'D_2^2X(X'D_1X)^{-1} \\ &+\tfrac{1}{4}(X'D_1X)^{-1}X'D_2\mathbf{1}\mathbf{1}'D_2X(X'D_1X)^{-1} \\ &-3(X'D_1X)^{-1}X'D_2^2D_4X(X'D_1X)^{-1} \\ &+3(X'D_1X)^{-1}X'D_2\dot{A}D_2X(X'D_1X)^{-1} \\ &+ (X'D_1X)^{-1}X'D_2D_4\mathbf{1}\mathbf{1}'D_2D_4X(X'D_1X)^{-1} \\ &-2(X'D_1X)^{-1}X'D(n_t)^{-1}D_1X(X'D_1X)^{-1} \\ &- (X'D_1X)^{-1}X'D_1D_2D_5X(X'D_1X)^{-1} \\ &+4(X'D_1X)^{-1}X'D(n_t)^{-1}D_1D_4X(X'D_1X)^{-1} \\ &+2(X'D_1X)^{-1}X'D_1D_2D_3X(X'D_1X)^{-1} \\ &-\tfrac{1}{2}(X'D_1X)^{-1}X'D_2D_4\mathbf{1}\mathbf{1}'D_2X(X'D_1X)^{-1} \\ &-\tfrac{1}{2}(X'D_1X)^{-1}X'D_2\mathbf{1}\mathbf{1}'D_2D_4X(X'D_1X)^{-1}. \end{aligned}$$

One way to check the accuracy of the mean squared error formulas (34) and (62) is to see if both formulas are reduced to the same expression when we assume $T = K$, the case where the maximum likelihood and minimum chi-square estimators are identical. Assuming X is invertible, both formulas become

(63) $$\begin{aligned} \mathrm{MSE} = &X^{-1}D_1^{-1}(X')^{-1} + \tfrac{3}{2}X^{-1}D_1^{-2}D_2^{-2}(X')^{-1} \\ &+ 2X^{-1}D(n_t)^{-1}D_1^{-1}(X')^{-1} + \tfrac{1}{4}X^{-1}D_1^{-1}D_2\mathbf{1}\mathbf{1}'D_2D_1^{-1}(X')^{-1}. \end{aligned}$$

The approximate bias of the minimum chi-square estimator to the order of n^{-1}, denoted by BS_2, can be easily obtained by taking the expectation of $\Sigma_{\nu=1}^4\xi_\nu$ as

(64) $$\mathrm{BS}_2 = (X'D_1X)^{-1}X'D_2D_4\mathbf{1} - \tfrac{1}{2}(X'D_1X)^{-1}X'D_2\mathbf{1}.$$

6. Correction for bias. Ghosh and Subramanyam [1974] proved a general theorem about the second-order efficiency of the maximum likelihood estimator in the exponential family and applied it to a special case ($n_t = n$, $K = 2$, and only the intercept is unknown) of our model to show that the n^{-2}-order mean squared error of the maximum likelihood estimator after the bias to the order of n^{-1} has been corrected is smaller than the corresponding expression for Berkson's estimator. We will prove that this is also true for our model. The theorem of Ghosh and Subramanyam cannot be directly applied to our model since they assume $n_t = n$ for all t.

We define the bias-corrected maximum likelihood and minimum chi-square estimators respectively as follows:

$$\hat{\beta}_c = \hat{\beta} - \mathrm{BS}_1(\hat{\beta}) \tag{65}$$

and

$$\tilde{\beta}_c = \tilde{\beta} - \mathrm{BS}_2(\tilde{\beta}), \tag{66}$$

where BS_1 and BS_2 have been shown in equations (35) and (64) and the term within the parenthesis indicates the value at which the bias is evaluated.

We will first derive the n^{-2}-order mean squared error matrix of $\hat{\beta}_c$. Write (12) in vector notation as

$$\hat{\beta} - \beta_0 \cong v_1 + v_2 + v_3. \tag{67}$$

Then we have

$$\hat{\beta}_c - \beta_0 \cong v_1 + (v_2 - Ev_2) + v_3 - \left[\mathrm{BS}_1(\hat{\beta}) - \mathrm{BS}_1(\beta_0)\right] \tag{68}$$

since $Ev_2 = \mathrm{BS}_1(\beta_0)$. Applying a Taylor expansion to the last bracketed term above and omitting the terms that do not contribute to the n^{-2}-order mean squared error, we get

$$\hat{\beta}_c - \beta_0 \cong v_1 + (v_2 - Ev_2) + v_3 - \frac{\partial \mathrm{BS}_1}{\partial \beta'} v_1. \tag{69}$$

Therefore, the n^{-2}-order mean squared error matrix of $\hat{\beta}_c$, denoted by CMSE_1, is given by

$$\begin{aligned} \mathrm{CMSE}_1 &= \mathrm{MSE}_1 - \mathrm{BS}_1 \cdot \mathrm{BS}_1' - E\frac{\partial \mathrm{BS}_1}{\partial \beta'} v_1 v_1' - Ev_1 v_1' \frac{\partial \mathrm{BS}_1'}{\partial \beta} \\ &= (X'D_1X)^{-1} + \tfrac{1}{4}(X'D_1X)^{-1}X'D_2(\dot{A} + \dot{A}\dot{A})D_2X(X'D_1X)^{-1}. \end{aligned} \tag{70}$$

For the derivation of the second equality above, the reader is referred to Amemiya [1979].

The n^{-2}-order mean squared error matrix of $\tilde{\beta}_c$, denoted by CMSE_2, can be similarly obtained as

$$\begin{aligned} \mathrm{CMSE}_2 &= \mathrm{MSE}_2 - \mathrm{BS}_2 \cdot \mathrm{BS}_2' - E\frac{\partial \mathrm{BS}_2}{\partial \beta'} v_1 v_1' - Ev_1 v_1' \frac{\partial \mathrm{BS}_1'}{\partial \beta} \\ &= (X'D_1X)^{-1} + (X'D_1X)^{-1}X'D_2\left(\tfrac{1}{2}I - D_4 + \dot{A}\right)D_2X(X'D_1X)^{-1}. \end{aligned} \tag{71}$$

From (70) and (71), we have

(72)

$$\text{CMSE}_2 - \text{CMSE}_1 = (X'D_1X)^{-1}X'D_2\left(\tfrac{3}{4}\dot{A} + \tfrac{1}{2}I - D_4 - \tfrac{1}{4}\dot{A}\dot{A}\right)D_2X(X'D_1X)^{-1}.$$

The above is semi-positive definite because

$$\tfrac{3}{4}\dot{A} + \tfrac{1}{2}I - D_4 - \tfrac{1}{4}\dot{A}\dot{A} = \tfrac{1}{4}(\dot{A} - \dot{A}\dot{A}) + \tfrac{1}{2}(A - I) * (A - I) \tag{73}$$

and matrices $\dot{A} - \dot{A}\dot{A}$ and $(A - I) * (A - I)$ are semipositive definite. The last statement follows immediately once one observes that the Hadamard product $A * B$ is a principal submatrix of the Kronecher product $A \otimes B$. (See Minc and Marcus [1964] page 120.)

Consider another estimator obtained by correcting the maximum likelihood estimator in such a way that it has the same bias to the order of n^{-1} as the minimum chi-square estimator. That is to say, define

$$\hat{\beta}_d = \hat{\beta} - \text{BS}_1(\hat{\beta}) + \text{BS}_2(\hat{\beta}). \tag{74}$$

If we denote its n^{-2}-order mean squared error matrix by DMSE_1, we can easily show

$$\text{DMSE}_1 = \text{CMSE}_1 + \text{BS}_2 \cdot \text{BS}_2' + E\frac{\partial \text{BS}_2}{\partial \beta'}v_1v_1' + Ev_1v_1'\frac{\partial \text{BS}_2'}{\partial \beta}. \tag{75}$$

Therefore, from (71), (72), and (75), we get

(76)

$$\text{DMSE}_1 = \text{MSE}_2 - (X'D_1X)^{-1}X'D_2\left(\tfrac{3}{4}\dot{A} + \tfrac{1}{2}I - D_4 - \tfrac{1}{4}\dot{A}\dot{A}\right)D_2X(X'D_1X)^{-1},$$

which shows that $\hat{\beta}_d$ has a smaller n^{-2}-order mean squared error matrix than Berkson's minimum chi-square estimator.

7. Numerical evaluation. We have evaluated MSE_1 given in (34) and MSE_2 given in (62) numerically by choosing various values for the independent variables and the parameters. In all the examples we have computed, both reported and unreported below, each diagonal element of MSE_1 turned out to be larger than the corresponding diagonal element of MSE_2. However, we have been unable to prove or disprove that $\text{MSE}_1 - \text{MSE}_2$ is semipositive definite. For the same examples, we have also computed DMSE_1 given in (76), which we have theoretically proved to be smaller than MSE_2. In the following tables, we report the values of the diagonal elements of MSE_1, MSE_2, and DMSE_1. The last value always appears within a parenthesis. We have also computed CMSE_1 but will not report its values. It has turned out that in a majority of cases CMSE_1 lies between MSE_1 and MSE_2 but in a few cases it is larger than MSE_1 or smaller than DMSE_1.

We will first evaluate these matrices in the same examples for which Berkson [1955] calculated the exact mean squared errors. The characteristics of Berkson's four examples are as follows: Throughout the four examples we have $\beta' = (0, 0.08473)$ and $n_t = 10$ for all t.

EXAMPLE 1.

$$X' = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

$$P' = (0.3, \quad 0.5, \quad 0.7).$$

EXAMPLE 2.

$$X' = \begin{bmatrix} 1 & 1 & 1 \\ -0.52297 & 0.47854 & 1.48006 \end{bmatrix}$$

$$P' = (0.391, \quad 0.6, \quad 0.778).$$

EXAMPLE 3.

$$X' = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2.00155 \end{bmatrix}$$

$$P' = (0.5, \quad 0.7, \quad 0.845).$$

EXAMPLE 4.

$$X' = \begin{bmatrix} 1 & 1 & 1 \\ 0.63827 & 1.63614 & 2.63309 \end{bmatrix}$$

$$P' = (0.632, \quad 0.8, \quad 0.903).$$

Table 1 gives the exact mean squared errors calculated by Berkson and the approximate mean squared errors calculated according to the formulas (34) and (62) in each of the four examples described above. The mean squared errors of the bias-corrected maximum likelihood estimator defined in (74) are given within parentheses.

Table 1

Exact and approximate mean squared errors in Berkson's four examples

		Estimation of β_1		Estimation of β_2	
		MLE	MIN χ^2	MLE	MIN χ^2
Example 1	Exact	.187	.154	.322	.271
	Approx.	.175	.162	.301	.286
		(.162)		(.285)	
Example 2	Exact	.230	.206	.341	.272
	Approx.	.220	.208	.315	.295
		(.208)		(.293)	
Example 3	Exact	.430	.394	.404	.274
	Approx.	.408	.391	.371	.336
		(.391)		(.330)	
Example 4	Exact	1.103	.689	.466	.208
	Approx.	1.078	.983	.532	.447
		(.974)		(.428)	

An agreement between the exact and the approximate mean squared errors in Table 1 is satisfactory. From (6) and (7) we know that the difference between the exact and the approximate mean squared errors is equal to the sum of $M - M^*$ and $N - N^*$, both of which decrease as n_t increases. $M - M^*$ decreases at the rate approximately proportional to $[n_t P_t(1 - P_t)]^{-3}$, whereas $N - N^*$ decreases at a less predictable way but eventually at a faster rate. The results in Table 1 are encouraging when we consider the fact that $n_t P_t(1 - P_t)$ in these examples is at most 2.5.

The advantage of the approximate formulas (34) and (62) is that they can be readily computed in many models for which the evaluation of the exact mean squared errors would be extremely difficult. We have computed them for many examples, of which we will report here four artificially created examples and one example taken from econometric applications. The results of the other examples can be found in Amemiya [1979].

In the next four examples we fix

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix}$$

and vary β as follows:

EXAMPLE 5.

$$\beta' = (-0.5677, 0.1622)$$
$$P' = (.04, 0.4395, 0.4797, 0.5203, 0.5605, 0.6).$$

EXAMPLE 6.

$$\beta' = (-1.9408, 0.5545)$$
$$P' = (0.2, 0.3033, 0.4311, 0.5689, 0.6967, 0.8).$$

EXAMPLE 7.

$$\beta' = (0.2093, 0.1962)$$
$$P' = (0.6, 0.6460, 0.6895, 0.7299, 0.7668, 0.8).$$

EXAMPLE 8.

$$\beta' = (-1.5825, 0.1962)$$
$$P' = (0.2, 0.2332, 0.2701, 0.3105, 0.3540, 0.4).$$

In these examples, β's are chosen to produce different patterns of the spread of probabilities. We fix $n_t = 1$ for all t and give the diagonal elements of $(X'D_1X)^{-1}$ and MSE $- (X'D_1X)^{-1}$ separately in Table 2. The mean squared errors of the two estimators differ only in the second term and the sign of the difference is independent of n_t as long as n_t is the same for all t. If $n_t = n$ for all t, one should compute the mean squared error matrix by the formula $n^{-1}(X'D_1X)^{-1} + n^{-2}[\text{MSE} - (X'D_1X)^{-1}]$. For example, if $n = 10$, one computes the mean squared error of the maximum likelihood estimator of β_1 to be $0.3573 + 0.03329 = 0.39059$. In the results of Table 2, the minimum chi-square estimator comes out the winner in all the eight cases. In this and the subsequent tables, the diagonal elements of DMSE_1 are given within the parentheses.

TABLE 2
Approximate mean squared errors in four artificial examples with two regressors

		Estimation of β_1		Estimation of β_2	
		MLE	MIN χ^2	MLE	MIN χ^2
Example 5	Order of n^{-1}	3.573		0.236	
	n^{-2}	3.329		0.233	
		(−1.117)	−1.004	(−0.053)	−0.045
Example 6	n^{-1}	4.797		0.325	
	n^{-2}	7.786		0.571	
		(−3.930)	−1.825	(−0.253)	−0.098
Example 7	n^{-1}	3.912		0.290	
	n^{-2}	4.099		0.379	
		(−2.070)	−1.333	(−0.221)	−0.105
Example 8	n^{-1}	4.854		0.290	
	n^{-2}	7.317		0.379	
		(−4.351)	−1.916	(−0.200)	−0.105

We also did similar calculations for

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 4 & 9 & 16 & 25 & 36 \end{bmatrix},$$

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix},$$

and

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 4 & 9 & 16 & 25 & 36 \end{bmatrix},$$

and the results were similar to those in Table 2, the minimum chi-square estimator having a smaller mean squared error in every case.

The next example is taken from an econometric application. Parameter values are set at actually estimated values.

EXAMPLE 9. Adopted from Theil ([1971] page 635)

$$
\begin{aligned}
y_t &= 1 \text{ positive production plan revision} \\
&= 0 \text{ negative revision} \\
x_{1t} &= 1 \\
x_{2t} &= 1 \text{ negative surprise on orders received} \\
&= 0 \text{ positive surprise} \\
x_{3t} &= 1 \text{ inventories are considered too large} \\
&= 0 \text{ too small} \\
X' &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \\
n_1 &= 89,\ n_2 = 82,\ n_3 = 43,\ n_4 = 164 \\
\beta' &= (1.30,\ -2.23,\ -0.98) \\
P' &= (0.7858,\ 0.5793,\ 0.2829,\ 0.1290).
\end{aligned}
$$

TABLE 3
Approximate mean squared errors in Theil's example

	MLE	MIN χ^2
Estimation of β_1	0.05291 (0.05211)	0.05218
Estimation of β_2	0.06919 (0.06807)	0.06826
Estimation of β_3	0.07113 (0.06996)	0.07003

Thus, in this actual econometric example, the minimum chi-square estimator has a smaller mean squared error in all the cases. We obtained similar results with another econometric example adopted from Amemiya and Nold [1975].

In each of the examples both reported and unreported, we also computed the approximate bias and its square for the two estimators according to the formulas (35) and (63). The minimum chi-square estimator usually had a smaller squared bias than the maximum likelihood estimator but with less regularity than in the comparison of the mean squared errors. At any rate, the squared bias was much smaller than the mean squared error in every case, which implies that the estimator with a smaller mean squared error always had a smaller variance (to the order of n^{-2}) as well. As for the signs of biases, there was no significant pattern to speak of. This fact does not contradict the results of Section 6 since the bias is evaluated at the value of the estimator in defining the bias-corrected estimator.

8. Conclusions. In this paper we have derived the mean squared error matrices to the order of n^{-2} for the maximum likelihood and Berkson's minimum chi-square estimator in the dichotomous logit regression model. We have numerically evaluated the matrices in many examples both real and artificial and found that in all of the cases the minimum chi-square estimator has a smaller mean squared error. Only a few of the results are reported here. See Amemiya [1979] for the other results. This confirms the results obtained by Berkson [1955] in a few examples for which he was able to calculate the exact mean squared errors of the two estimators. The usefulness of our approximate formulas lies in that they can be computed at a minimal computational cost. However, we have not been able to show theoretically that the n^{-2}-order mean squared error of the minimum chi-square estimator is smaller than that of the maximum likelihood estimator in general.

We have also derived the n^{-2}-order mean squared error of the maximum likelihood estimator after correcting its bias in such a way that it has the same n^{-1}-order bias as the minimum chi-square estimator and showed that it is smaller than that of the minimum chi-square estimator. However, the numerical evaluation in many examples, both reported and unreported in this paper, show that the difference in the mean squared error between the minimum chi-square estimator and the bias-corrected maximum likelihood estimator is never significantly large: i.e., not large enough to negate the computational advantage of the minimum chi-square estimator.

Berkson [1957] computed the exact mean squared errors of the maximum likelihood and the minimum probit chi-square estimator in the examples analogous to those considered by Berkson [1955] and showed that the superiority of the minimum chi-square estimator holds in the probit models as well. We have not derived the n^{-2}-order mean squared errors of the probit estimators. The formulas would be more lengthy than the analogous formulas for the logit estimators, though they could be obtained by the same techniques.

Acknowledgments. The author is indebted to Tom Rothenberg for valuable comments and to Jim Powell and Paul Hunt for carrying out the computation in Section 7.

REFERENCES

AMEMIYA, T. [1979]. The n^{-2}-order mean squared errors of the maximum likelihood and the minimum logit chi-square estimator, Technical Report No. 267, Institute for Mathematical Studies in the Social Sciences, Stanford University.

AMEMIYA, T. and NOLD, F. [1975]. A modified logit model. *Rev. Econom. Statist.* **57** 255-257.

BERKSON, J. [1944]. Application of the logistic function to bio-assay. *J. Amer. Statist. Assoc.* **39** 357-365.

BERKSON, J. [1955]. Maximum likelihood and minimum χ^2 estimates of the logistic function. *J. Amer. Statist. Assoc.* **50** 130-162.

BERKSON, J. [1957]. Tables for use in estimating the normal distribution function by normit analysis. *Biometrika* **44** 411-435.

GHOSH, J. K. and SUBRAMANYAM, K. [1974]. Second-order efficiency of maximum likelihood estimators. *Sankhyā* Ser. A **36** Pt. 4, 325-358.

JOHNSON, N. L. and KOTZ, S. [1969]. *Discrete Distributions*, Houghton Mifflin, Boston.
MINC, H. and MARCUS, M. [1964]. *A Survey of Matrix Theory and Matrix Inequalities*, Prindle, Weber and Schmidt, Boston.
TAYLOR, W. F. [1953]. Distance functions and regular best asymptotically normal estimates. *Ann. Math. Statist.* **24** 85-92.
THEIL, H. [1971], *Principles of Econometrics*, Wiley, New York.

DEPARTMENT OF ECONOMICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305