The Annals of Statistics
1984, Vol. 12, No. 2, 778–782

COMMENTS ON A PAPER BY T. AMEMIYA ON ESTIMATION IN A DICHOTOMOUS LOGIT REGRESSION MODEL[1]

BY LINDA DAVIS

Stanford University

Amemiya (1980) derived expressions for the n^{-2}-order mean squared errors of the maximum likelihood and the minimum logit chi-squared estimators in the dichotomous logit regression model. He numerically evaluated these expressions for a number of specific examples and in all examples found that the minimum chi-squared estimator has smaller n^{-2}-order mean squared error. In this paper, we demonstrate examples in which the maximum likelihood estimator has smaller n^{-2}-order mean squared error.

1. Introduction. Amemiya (1980) derived expressions for the n^{-2}-order mean squared errors of the maximum likelihood estimator and the minimum logit chi-squared estimator in the dichotomous logit regression model. He evaluated the n^{-2}-order mean squared errors in many examples, both real and artificial, finding in all examples that the n^{-2}-order mean squared error of the minimum logit chi-squared estimator is smaller than that of the corresponding maximum likelihood estimator. He did not show theoretically, however, that the n^{-2}-order mean squared error of the minimum logit chi-squared estimator is smaller than that of the maximum likelihood for all designs and parameter values. The resolution of this question is important in that it can further clarify which estimator should be used in a given situation.

Since Amemiya's paper, little work has been done to determine whether the order relationship observed by Amemiya between the n^{-2}-order mean squared errors is an artifact of the examples he considered or in fact a property of the estimators. The only known work is by Ghosh and Sinha (1981) which suggests that indeed the order relationship is an artifact of the examples. They treat, however, a slightly different estimator than the minimum logit chi-squared estimator and do not give any specific numeric results.

In this paper, we demonstrate specific designs and parameter values for which the n^{-2}-order mean squared error of the maximum likelihood estimator is smaller than that of the corresponding minimum logit chi-squared estimator. The emphasis in these examples is toward determining what aspect of the design and/or parameter value is the main cause of the n^{-2}-order mean squared error being smaller for the maximum likelihood estimator. In particular, we find that the controlling factor in the order relationship between the mean squared errors

Received September 1983; revised November 1983.

[1] This research was supported by a Graduate School Fellowship from Rutgers University and a NSF Postdoctoral Research Fellowship at Stanford University.

AMS 1980 *subject classifications.* Primary 62F12, secondary 62F10.

Key words and phrases. Logit regression model, maximum likelihood estimator, minimum logit chi-squared estimator.

is the number of design points. This result, in fact, explains why Amemiya failed to observe a smaller n^{-2}-order mean squared error for the maximum likelihood estimator in any of his examples.

The paper is organized as follows: In Section 2, we briefly review the pertinent parts of Amemiya's paper, retaining as much as possible of his notation. In Section 3, we report the examples considered and the numerical values of the n^{-2}-order mean squared errors as well as our interpretation of the numerical results. Finally, in Section 4, we summarize the implications of our findings on the original question raised by Amemiya in addition to their broader implications on determining which of these two estimators is preferable in a given situation.

2. Review. The underlying data consists of $\sum_{t=1}^{T} n_t$ dichotomous random variables $y_{t\nu}(t = 1, 2, \cdots, T;\ \nu = 1, 2, \cdots, n_t)$ taking values 0 or 1 with $P\{y_{t\nu} = 1\} = \{1 + \exp(-x_t'\beta_0)\}^{-1} \equiv P_t$ where x_t is a K-dimensional vector of known constants and β_0 is a K-dimensional vector of unknown parameters. We define $r_t = n_t^{-1} \sum_{\nu=1}^{n_t} y_{t\nu}$ and $X = (x_1, x_2, \cdots, x_T)'$.

The two estimators we consider are the maximum likelihood estimator and the minimum logit chi-squared estimator. The maximum likelihood estimator is defined as usual as a solution to the normal equations which in this case are $\sum_t n_t(r_t - F_t)x_t = 0$ where $F_t = \{1 + \exp(-x_t'\beta)\}^{-1}$. The minimum logit chi-squared estimator is defined as

$$\{\textstyle\sum_t n_t r_t(1 - r_t)x_t x_t'\}^{-1} \sum_t n_t r_t(1 - r_t)[\log\{r_t/(1 - r_t)\}]x_t.$$

The asymptotic assumptions are basically that T is a fixed number greater than or equal to K and each n_t goes to infinity at a common rate designated by n. The expression derived by Amemiya for the n^{-2}-order mean squared error is essentially just the Taylor series expansion of the mean squared error truncated after terms of order n^{-2}.

Let MSE_1 and MSE_2 denote the n^{-2}-order mean squared error matrices of the maximum likelihood estimator and the minimum logit chi-squared estimator respectively. The expressions for MSE_1 and MSE_2 are quite lengthy and thus we simply refer you to Amemiya (1980), equations (34) and (62). There is a slight error, however, in his equation (30) which ultimately affects (34). Equation (30) should read

$$\begin{aligned} Ev_{2i}v_{2j} &= \frac{1}{2} \sum_s \sum_t \frac{\partial^2\beta_i}{\partial r_s \partial r_t} \frac{\partial^2\beta_j}{\partial r_s \partial r_t} \frac{P_t(1 - P_t)}{n_t} \frac{P_s(1 - P_s)}{n_s} \\ &\quad + \frac{1}{4} \left\{\sum_t \frac{\partial^2\beta_i}{\partial r_t^2} \frac{P_t(1 - P_t)}{n_t}\right\} \left\{\sum_t \frac{\partial^2\beta_j}{\partial r_t^2} \frac{P_t(1 - P_t)}{n_t}\right\} \\ &= 2m_{1ij} + m_{3ij}. \end{aligned}$$

The effect of this on (34) is simply to replace $\dot{A}\dot{A}$ by $\dot{A}$. Making this one change in (34) results in the correct expression for MSE_1.

The numerical results in the next section involve comparison of the diagonal elements of MSE_1 and MSE_2 only.

3. Examples. The set of examples we consider is generated by considering all possible pairings of five different designs (specified by the X matrix) and five different parameter values. The five different X matrices which we index by the corresponding values of T are in Table 1 ($n_t \equiv n$ for all $t = 1, 2, \cdots, T$.) The five different values of the paramter $\beta_0' = (\beta_{01}, \beta_{02})$ are in Table 2. These particular values of the parameter were chosen to provide different patterns and ranges of the true probabilities.

Table 3 contains the results of numerically evaluating MSE_1 and MSE_2 for all twenty-five pairings of designs and parameter values. Since for models of the above type $\text{MSE}_i = A/n + B_i/n^2$ for $i = 1, 2$, we simply report the diagonal elements of A, B_1, and B_2 in the table. Note that the comparison of MSE_1 and MSE_2 then reduces to a comparison of B_1 and B_2.

The first thing to note in Table 3 is that there are examples in which the maximum likelihood estimator has smaller n^{-2}-order mean squared error. Thus, these examples resolve the issue raised by Amemiya as to whether in general the n^{-2}-order mean squared error of the minimum logit chi-squared estimator is smaller than that of the maximum likelihood estimator. In general, this statement is not true.

The next point we would like to emphasize by using the table is what aspect of the design and/or parameter value is the main cause of the switch over in the relative sizes of the n^{-2}-order mean squared errors. The general pattern in the table is that for a fixed β_{0i}, the n^{-2}-order mean squared error is smaller for the minimum logit chi-squared estimator when T is small and smaller for the maximum likelihood estimator when T is large. Furthermore, the value of T at which the switch occurs is inversely related to the size of β_{0i} in that the larger β_{0i}, the smaller the value of T at which the switch occurs.

Finally, note that the smallest value of T for which the maximum likelihood estimator has smaller n^{-2}-order mean squared error is always at least nine in

TABLE 1

$$\begin{bmatrix} 1 & -16 \\ 1 & 0 \\ 1 & 16 \end{bmatrix} \quad \begin{bmatrix} 1 & -16 \\ 1 & -8 \\ 1 & 0 \\ 1 & 8 \\ 1 & 16 \end{bmatrix} \quad \begin{bmatrix} 1 & -16 \\ 1 & -12 \\ \vdots & \vdots \\ 1 & 12 \\ 1 & 16 \end{bmatrix} \quad \begin{bmatrix} 1 & -16 \\ 1 & -14 \\ \vdots & \vdots \\ 1 & 14 \\ 1 & 16 \end{bmatrix} \quad \begin{bmatrix} 1 & -16 \\ 1 & -15 \\ \vdots & \vdots \\ 1 & 15 \\ 1 & 16 \end{bmatrix}$$

$$T = 3 \qquad T = 5 \qquad T = 9 \qquad T = 17 \qquad T = 33$$

TABLE 2

Parameter Index	**1**	**2**	**3**	**4**	**5**
β_{01}	0	.67496	1.09861	1.52226	2.19722
β_{02}	.13733	.095141	.068663	.042185	0
Minimum P_t	.1	.3	.5	.7	.9
Maximum P_t	.9	.9	.9	.9	.9

TABLE 3
n^{-2}-order mean squared errors

Parameter Index	T	Estimation of β_{01}			Estimation of β_{02}		
		A	B_1	B_2	A	B_1	B_2
1	3	2.33	7.30	1.58	.0217	.2056	.1171
	5	1.24	2.17	−.76	.0143	.0559	.0005
	9	.65	.58	−1.06	.0086	.0166	−.0018
	17	.33	.15	−.73	.0048	.0046	.0127
	33	.17	.04	−.42	.0025	.0012	.0253
2	3	2.10	7.57	3.02	.0143	.0737	.0379
	5	1.16	2.25	−.37	.0105	.0289	.0006
	9	.62	.63	−.54	.0066	.0094	−.0039
	17	.32	.17	−.04	.0038	.0027	.0026
	33	.16	.04	.40	.0020	.0007	.0094
3	3	2.21	10.35	5.02	.0134	.0616	.0287
	5	1.26	3.17	−.06	.0102	.0258	−.0005
	9	.68	.91	−.13	.0066	.0088	−.0052
	17	.36	.25	.77	.0038	.0026	−.0000
	33	.18	.06	1.53	.0021	.0007	.0058
4	3	2.51	15.44	8.81	.0146	.0731	.0353
	5	1.47	5.07	.68	.0114	.0315	−.0022
	9	.81	1.51	.59	.0075	.0112	−.0097
	17	.42	.42	2.23	.0044	.0034	−.0051
	33	.22	.11	3.63	.0024	.0010	.0009
5	3	3.70	40.05	26.61	.0217	.1648	.0804
	5	2.22	14.42	4.64	.0174	.0755	−.0120
	9	1.23	4.45	4.24	.0116	.0275	−.0346
	17	.65	1.25	9.32	.0068	.0085	−.0287
	33	.34	.33	13.76	.0037	.0024	−.0181

these examples. In all of the examples reported by Amemiya (1980), T is at most six, which explains why he did not observe a smaller n^{-2}-order mean squared error for the maximum likelihood estimator in any of his examples.

There might be some skepticism as to whether such large T values arise in practice as well as whether the asymptotic expansions are valid for such large values of T. Amemiya (1979), however, computed the n^{-2}-order mean squared errors of both estimators for a real example adopted from Amemiya and Nold (1975) in which $T = 16$. (In this example, he found the n^{-2}-order mean squared error smaller for the minimum logit chi-squared estimator. If the range of $(x_t)_2$ in this example is converted to $[-16, 16]$, however, the converted parameter value is $\beta_{01} = .184$ and $\beta_{02} = .0408$. Thus, the results in Table 1 indicate that the value of β_{0i} is too small for the maximum likelihood estimator to have smaller n^{-2}-order mean squared error even for $T = 16$.)

4. Conclusions. In this paper, we have shown that the minimum logit chi-squared estimator has smaller n^{-2}-order mean squared error than the maximum likelihood estimator only for certain designs and parameter values. Thus, we have resolved the question raised by Amemiya (1980). Furthermore, we have shown that the minimum logit chi-squared estimator has smaller n^{-2}-order mean squared error when the true parameter value is small or the number of design points is small. Thus, it appears that consideration of the size of T should enter into the choice of which estimator to use in a given situation.

There are of course other estimators of β_0 that have been suggested. One such estimator is the weighted least squares estimator, i.e., the value of β minimizing $\sum_t n_t(r_t - F_t)^2/\{F_t(1 - F_t)\}$. This is easily seen to be equivalent to the minimum Pearson's chi-squared estimator. We derived an expression for the n^{-2}-order mean squared error of this estimator and found it to be equivalent to that derived by Amemiya (1980) for the minimum logit chi-squared estimator. Thus, all of the above discussion applies directly to the comparison of the weighted least squares estimator and the maximum likelihood estimator.

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 of Mathematical Studies in the Social Sciences, Stanford University.

AMEMIYA, T. (1980). The n^{-2}-order mean squared errors of the maximum likelihood and the minimum logit chi-square estimator. *Ann. Statist.* **8** 488–505. Correction: *Ann. Statist.* **12** 783.

AMEMIYA, T. and NOLD, F. (1975). A modified logit model. *Rev. Econom. Statist.* **57** 255–257.

GHOSH, J. K. and SINHA, B. K. (1981). A necessary and sufficient condition for second order admissibility with applications to Berkson's bioassay problem. *Ann. Statist.* **9** 1334–1338.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
SEQUOIA HALL
STANFORD, CALIFORNIA 94305