University of Victoria
Department of Economics

Econometrics Working Paper EWP0801

ISSN 1485-6441

FINITE-SAMPLE MOMENTS OF THE MLE FOR THE BINARY LOGIT MODEL

Qian Chen

School of Public Finance and Public Policy,
Central University of Finance and Economics,
Beijing, People's Republic of China

&

David E. Giles

Department of Economics, University of Victoria,
Victoria, B.C., Canada

January 2008

Abstract

We examine the finite sample properties of the MLE for the Logit model with random covariates. We derive the second order bias and MSE function for the MLE in this model, and undertake some numerical evaluations to illustrate the analytic results. From these numerical results we find, for example, that the bias correction that we provide is effective, and that the bias-corrected estimator is more efficient than the uncorrected MLE.

Keywords: Logit model; bias; mean squared error; bias correction

JEL Classifications: C01 C13, C25

Author Contact:
David E., Dept. of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, B.C., Canada V8W 2Y2; e-mail: dgiles@uvic.ca; Voice: (250) 721-8540; FAX: (250) 721-6214

FINITE-SAMPLE MOMENTS OF THE MLE FOR THE BINARY LOGIT MODEL

1. Introduction

Qualitative response (QR) models, which are very widely used in empirical economics and elsewhere, have the characteristic that the dependent variable is qualitative, rather than quantitative. To make the model estimable, these qualitative attributes are "coded" numerically so as to partition the sample data appropriately. There is a vast and readily accessible literature relating to inference in the context of these models - for example, see Maddala (1983), Wooldridge (2002), Hensher *et al*. (2005) and Cameron and Trivedi (2005).

The binary choice model is the most widely used of the QR models. In this model the dependent variable is coded as unity or zero according to whether some event occurs or not. In this case, it is well known that conventional (linear) regression methods are inappropriate: the predicted probabilities can be negative, or exceed unity; the error must be heteroskedastic; and the error term clearly cannot follow a normal distribution. These problems can be overcome by making the probability of the event of interest a non-linear, rather than a linear function of the covariates. In particular, if this function is taken to be a cumulative distribution function, it will be monotonically non-decreasing, and bounded between zero and unity. Usually, the distribution that is chosen for this purpose is the normal distribution (giving rise to the Probit model), the logistic distribution (which gives us the Logit model), the Weibull distribution, or the extreme value distribution. Different distributions lead to different non-linear models with somewhat different features. The Logit and Probit models are the two that are encountered most frequently in practice, and they generally yield similar (scaled) estimates. In each of these two cases the likelihood function is strictly concave, so it has a unique maximum.

The likelihood functions for QR models satisfy the usual regularity conditions, so the maximum likelihood estimators (MLE's) are weakly consistent and best asymptotically normal. The strong consistency of the estimator for the Logit model has been established by Gourieroux and Montfort (1981). Surprisingly, there have been very few studies of the finite sample properties of the MLE for QR models. In this paper we derive analytic expressions for some of the finite sample properties of the MLE of the parameters in the Logit model. The approach that we use could also be used to extend our results to other QR models.

The next section discusses some of the related literature, and section 3 introduces the Logit model. In section 4 we present some results due to Rilstone *et al*. (1996) and use them to derive analytic expressions for the bias and MSE of the MLE in the Logit model. Some numerical results follow in section 5; and the final section provides our conclusions.

2. Related Literature

Amemiya (1980) derived the approximate n^{-2}-order mean squared error (MSE) of the MLE and the minimum chi-square estimator (MCSE) of the binary Logit model and provided some numerical results for the relative quality of these two estimators, based on their MSE's. The MCSE was first introduced by Berkson (1944) for the binary Logit model, and Taylor (1953) showed that the MCSE estimator and the MLE have the same asymptotic normal distribution. Berkson (1955) approximated the finite-sample bias and MSE of the MLE and the MCSE estimator for some particular samples of data, and showed that the MCSE is preferred to the MLE in terms of MSE in all of the cases he considered. Following Amemiya's work, several studies advanced Berkson's and Amemiya's results. Ghosh and Sinha (1981) provided the theory to give necessary and sufficient conditions for improving the MSE of the MLE, and applied these to Berkson's models and data. They also showed the relative MSE ranking of the MLE and the MCSE is model-specific. Davis (1984) found some examples in which the MLE has smaller MSE than the MCSE estimator, and Hughes and Savin (1994) provided further results indicating that the choice between these two estimators is not straightforward. Another somewhat related study is that of Mackinnon and Smith (1998). They discussed methods for reducing the bias of consistent estimators that are biased in finite samples, and applied their methods to the parameter estimator in the AR(1) model and the Logit model. Finally, Li (2005) used a Monte Carlo experiment to examine some of the small sample properties of the MLE for three different models - the Probit model, the Logit model and the binary choice model where the underlying distribution is the Extreme Value distribution. She also considered the case where the underlying distributional process is mis-specified, and found that this increases the MSE for each of the estimators.

In this paper, we will apply Rilstone *et al.*'s (1996) results to derive analytic expressions for the bias and MSE functions for the MLE in the Logit model with stochastic covariates. This approach was also used by Rilstone and Ullah (2002) in the context of Heckman's sample selection estimator. Based on the analytic bias and MSE expressions, we can derive a bias-corrected estimator and the standard error associated with the bias-corrected estimator. We also provide some numerical evaluations based on these analytic results. These numerical evaluations show that the bias correction works very well. In order to apply Rilstone *et al.*'s results, we need to assume that both the dependent and independent variables are random and that the observations are i.i.d.. This makes our results incomparable with Amemiya's and Davis's results. Because all of the observations are i.i.d random, the expectations of the random variables or any function of the random variables are the same for all of the observations, which simplifies the application of Rilstone *et al*.'s results.

3. The Logit Model and the Maximum Likelihood Estimator

In a binary choice model we can use a latent dependent variable to incorporate the effects of covariates. The latent regression is:

$$y_i^* = X_i'\beta + \varepsilon \quad , \tag{1}$$

where y^* is the latent dependent variable, and the row vector, X_i', represents the i^{th} observation on all of the covariates.

Then, the dependent variable can be defined as,

$$\begin{aligned} y_i &= 1; \quad \text{if } X_i'\beta + \varepsilon \geq a \\ y_i &= 0; \quad \text{if } X_i'\beta + \varepsilon < a \end{aligned} \tag{2}$$

where a is the threshold. As is well understood, as long as an intercept is included among the regressors, the threshold value for determining the dependent variable is actually irrelevant, and may be set to zero. Then, (1) and (2) can be simplified to

$$y_i^* = X_i'\beta_i + \varepsilon$$

and

$$\begin{aligned} y_i &= 1; \quad \text{if } y_i^* \geq 0 \\ y_i &= 0; \quad \text{if } y_i^* < 0 . \end{aligned} \tag{3}$$

The basic model can be structured as:

$$P_i = \Pr(y_i = 1 | X) = F(X_i'\beta)$$

$$1 - P_i = \Pr(y_i = 0 | X) = 1 - F(X_i'\beta) .$$

The form of the cumulative distribution function, $F(X_i'\beta)$, will determine which particular model is used. In this paper, we focus on the Logit model:

$$P_i = \Pr(y_i = 1 | X) = F(X_i'\beta) = \Lambda_i \quad (4)$$

where

$$\Lambda_i = \frac{\exp(X_i'\beta)}{1 + \exp(X_i'\beta)} \quad (5)$$

is the c.d.f. for the Logistic distribution.

The MLE for the parameter vector in (4) is derived as follows. The (conditional) joint data density function for the sample is:

$$\Pr(Y_1 = y_1, Y_2 = y_2, \cdots, Y_n = y_n | \beta, X) = \prod_{y_i=1} \Lambda_i \prod_{y_i=0} (1 - \Lambda_i),$$

so the (conditional) likelihood function is:

$$L(\beta | X, y) = \prod_{i=1}^{n} \Lambda_i^{y_i} (1 - \Lambda_i)^{(1-y_i)},$$

and the (conditional) log-likelihood function is:

$$\log L = \sum_{i=1}^{n} \left[y_i \log \Lambda_i + (1 - y_i) \log(1 - \Lambda_i) \right].$$

The log-likelihood equations are:

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} \left[(y_i - \Lambda_i) X_i \right] = 0 . \quad (6)$$

The MLE of β is the solution to (6). Since the log-likelihood function is strictly concave, the MLE is unique, but as (6) is highly non-linear in the parameters, it must be solved numerically. That is, the MLE cannot be written as closed-form expression, and this is what substantially complicates the task of evaluating the characteristics of its (finite-sample) sampling distribution.

4. Analytic Results

Before we derive the analytic results for the Bias and MSE of the MLE in the Logit model, we first introduce the results of Rilstone *et al.* (1996). The class of estimators considered by Rilstone *et al*. (RSU) includes those which can only be expressed implicitly as a function of the data. Suppose we have a regression model of the form

$$y_i = f(X_i; \beta) + \varepsilon_i .$$

The regressor vector, X_i, can include any endogenous or exogenous variables. In order to make the derivation simple, RSU assume that all of the variables are random. Let $Z_i = (y_i, X_i)$ and let Z_1, Z_2, $Z_3, \ldots$ be a sequence of m dimensional i.i.d. random vectors. θ_0 represents the true parameter vector, which could include only β, or any other parameters of interest. The estimator $\hat{\theta}$ can be written in the form:

$$\psi_n(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} g_i(\hat{\theta}) = 0 \ , \tag{7}$$

where $g_i(\theta) = g_i(z_i, \theta)$ is a $k \times 1$ vector involving the known variables and the parameters, and $E[g_i(\theta)] = 0$ only for the true value θ_0. Preceding the derivation of Lamma 1 below, RSU made some assumptions about the function $g_i(\theta)$. (See Ullah, 2004, p.31.)

Assumption 1

The s^{th} order derivatives of $g_i(\theta)$ exist in a neighborhood of θ and $E\left\|\nabla^s g_i(\theta)\right\|^2 < \infty$, where $\|A\|$, for a matrix A, denotes the usual norm, trace $[AA']^{1/2}$, and $\nabla^s A(\theta)$ is the matrix of s^{th} order partial derivations of a matrix $A(\theta)$ with respect to θ and obtained recursively.

Assumption 2

For some neighborhood of θ, $(\nabla \psi_n(\theta))^{-1} = O_p(1)$.

Assumption 3

$\left\|\nabla^s g_i(\theta) - \nabla^s g_i(\theta_0)\right\| \le \|\theta - \theta_0\| M_i$ for some neighborhood of θ_0, where M_i satisfies the condition $E|M_i| \le C < \infty$, $i = 1, 2, \ldots$

In the following, we will suppress the argument for any function of θ when there is no confusion. So, $g_i(\theta)$ will be written as g_i. Then, RSU derived the following Lemma.

Lemma 1 (RSU, 1996; Ullah, 2004, p.32)

Let assumptions 1-3 hold for some $s \ge 2$. Then the bias of $\hat{\theta}$ to order $O(n^{-1})$ is

$$B(\hat{\theta}) = \frac{1}{n} Q\left\{\overline{V_1 d_1} - \frac{1}{2}\overline{H}_2 \overline{[d_1 \otimes d_1]}\right\} \quad , \tag{8}$$

where $\overline{H}_j = \overline{\nabla}^j g_i$, $Q = [\overline{\nabla} g_i]^{-1}$, $V_i = [\nabla g_i - \overline{\nabla} g_i]$, and $d_i = Q g_i$. A bar over a function indicates its expectation, so that $\overline{\nabla} g_i = E[\nabla g_i]$. Further, if Assumptions 1-3 hold for some $s \geq 3$, then the MSE of $\hat{\theta}$ to order $O(n^{-2})$ is

$$MSE(\hat{\theta}) = \frac{1}{n}\Pi_1 + \frac{1}{n^2}(\Pi_2 + \Pi_2') + \frac{1}{n^3}(\Pi_3 + \Pi_4 + \Pi_4') \tag{9}$$

where

$$\Pi_1 = \overline{d_1 d_1'}$$

$$\Pi_2 = Q\left\{-\overline{V_1 d_1 d_1'} + \tfrac{1}{2}\overline{H}_2 \overline{[d_1 \otimes d_1] d_1'}\right\}$$

$$\begin{aligned}
\Pi_3 = {} & Q\left\{\overline{V_1 d_1 d_2' V_2'} + \overline{V_1 d_2 d_1' V_2'} + \overline{V_1 d_2 d_2' V_1'}\right\} Q \\
& + \tfrac{1}{4} Q\overline{H}_2 \left\{\overline{[d_1 \otimes d_1][d_2' \otimes d_2']} + \overline{[d_1 \otimes d_2][d_1' \otimes d_2']} + \overline{[d_1 \otimes d_2][d_2' \otimes d_1']}\right\} \overline{H}_2' Q \\
& - \tfrac{1}{2} Q\left\{\overline{V_1 d_1 d_2' \otimes d_2'} + \overline{V_1 d_2 [d_1' \otimes d_2']} + \overline{V_1 d_2 [d_2' \otimes d_1']}\right\} \overline{H}_2' Q \\
& - \tfrac{1}{2} Q\overline{H}_2 \left\{\overline{d_1 \otimes d_1 d_2' V_2'} + \overline{[d_1 \otimes d_2] d_1' V_2'} + \overline{[d_1 \otimes d_2] d_2' V_1'}\right\} Q
\end{aligned}$$

$$\begin{aligned}
\Pi_4 = {} & Q\left\{\overline{V_1 Q V_1 d_2 d_2'} + \overline{V_1 Q V_2 d_1 d_2'} + \overline{V_1 Q V_2 d_2 d_1'}\right\} \\
& - \tfrac{1}{2} Q\left\{\overline{V_1 Q\overline{H}_2 [d_1 \otimes d_2] d_2'} + \overline{V_1 Q\overline{H}_2 [d_2 \otimes d_1] d_2'} + \overline{V_1 Q\overline{H}_2 [d_2 \otimes d_2] d_1'}\right\} \\
& + \tfrac{1}{2} Q\left\{\overline{W_1 [d_1 \otimes d_2] d_2'} + \overline{W_1 [d_2 \otimes d_1] d_2'} + \overline{W_1 [d_2 \otimes d_2] d_1'}\right\} \\
& - \tfrac{1}{2} Q\overline{H}_2 \left\{\overline{[d_1 \otimes Q V_1 d_2] d_2'} + \overline{[d_1 \otimes Q V_2 d_1] d_2'} + \overline{[d_1 \otimes Q V_2 d_2] d_1'}\right\} \\
& + \tfrac{1}{4} Q\overline{H}_2 \left\{\overline{d_1 \otimes Q\overline{H}_2 [d_1 \otimes d_2] d_2'} + \overline{d_1 \otimes Q\overline{H}_2 [d_2 \otimes d_1] d_2'} + \overline{d_1 \otimes Q\overline{H}_2 [d_2 \otimes d_2] d_1'}\right\} \\
& - \tfrac{1}{2} Q\overline{H}_2 \left\{\overline{[Q V_1 d_1 \otimes d_2] d_2'} + \overline{[Q V_1 d_2 \otimes d_1] d_2'} + \overline{[Q V_1 d_2 \otimes d_2] d_1'}\right\} \\
& + \tfrac{1}{4} Q\overline{H}_2 \left\{\overline{[Q\overline{H}_2 [d_1 \otimes d_1] \otimes d_2] d_2'} + \overline{[Q\overline{H}_2 [d_1 \otimes d_2] \otimes d_1] d_2'} + \overline{[Q\overline{H}_2 [d_1 \otimes d_2] \otimes d_2] d_1'}\right\} \\
& - \tfrac{1}{6} Q\overline{H}_3 \left\{\overline{[d_1 \otimes d_1 \otimes d_2] d_2'} + \overline{[d_1 \otimes d_2 \otimes d_1] d_2'} + \overline{[d_1 \otimes d_2 \otimes d_2] d_1'}\right\}
\end{aligned} \tag{10}$$

where $W_i = [\nabla^2 g_i - \overline{\nabla}^2 g_i]$.

Now we apply the above Lemma to derive the bias and MSE of the MLE in the Logit model. First, we assume that both the dependent and independent variables in the Logit model are random, and the observations are i.i.d. Comparing (6) and (7), we can see that for the Logit model, we should set $g_i = (y_i - \Lambda_i) X_i$, and we know that $E(g_i \mid X_i) = 0$, then according to the law of iterated expectations, $E(g_i) = 0$.

Now we have the following results:

$$\nabla g_i = -\Lambda_i^{(1)} X_i X_i' ; \qquad \overline{H}_1 = \overline{\nabla} g_i = -E(\Lambda_i^{(1)} X_i X_i')$$
$$\nabla^2 g_i = -\Lambda_i^{(2)} X_i (X_i' \otimes X_i') ; \qquad \overline{H}_2 = \overline{\nabla}^2 g_i = -E[\Lambda_i^{(2)} X_i (X_i' \otimes X_i')]$$
$$\nabla^3 g_i = -\Lambda_i^{(3)} X_i (X_i' \otimes X_i' \otimes X_i') ; \qquad \overline{H}_3 = \overline{\nabla}^3 g_i = -E[\Lambda_i^{(3)} X_i (X_i' \otimes X_i' \otimes X_i')]$$
$$Q = (\overline{\nabla} g_i)^{-1} = -[E(\Lambda_i^{(1)} X_i X_i')]^{-1} ; \qquad d_i = Q g_i = -[E(\Lambda_i^{(1)} X_i X_i')]^{-1} (y_i - \Lambda_i) X_i$$
$$V_i = \nabla g_i - \overline{\nabla} g_i = -\Lambda_i^{(1)} X_i X_i' + E(\Lambda_i^{(1)} X_i X_i')$$
$$W_i = \nabla^2 g_i - \overline{\nabla}^2 g_i = -\Lambda_i^{(2)} X_i (X_i' \otimes X_i') + E[\Lambda_i^{(2)} X_i (X_i' \otimes X_i')] \quad , \qquad (11)$$

where $\Lambda_i^{(s)}$ is the s^{th} order derivative of Λ_i with respect to the argument of $X_i'\beta$ and

$$\Lambda_i^{(1)} = \frac{\exp(X_i'\beta)}{[1+\exp(X_i'\beta)]^2}$$
$$\Lambda_i^{(2)} = \frac{\exp(X_i'\beta)[1-\exp(X_i'\beta)]}{[1+\exp(X_i'\beta)]^3}$$
$$\Lambda_i^{(3)} = \frac{\exp(X_i'\beta)\{1-4\exp(X_i'\beta)+[\exp(X_i'\beta)]^2\}}{[1+\exp(X_i'\beta)]^4} \quad . \qquad (12)$$

Then we can derive the following theorem and corollary.

Theorem 1

If assumptions 1-3 hold for some $s \geq 2$. Then the bias of the MLE in the Logit model, to the order of $O(n^{-1})$ is

$$Bias(\hat{\beta}) = \frac{1}{2n} Q \overline{H}_2 vecQ \qquad (13)$$

Further if Assumptions 1-3 hold for some $s \geq 3$, then the MSE of MLE in the Logit model to order $O(n^{-2})$ is

$$MSE(\hat{\beta}) = \frac{1}{n}\Pi_1 + \frac{1}{n^2}(\Pi_2 + \Pi_2') + \frac{1}{n^3}(\Pi_3 + \Pi_4 + \Pi_4') \qquad (14)$$

where

$$\Pi_1 = -Q$$
$$\Pi_2 = -Q\left\{E(\Lambda_1^{(1)} V_1 Q X_1 X_1' Q) - \tfrac{1}{2}\overline{H}_2 E\left\{\Lambda_1^{(2)}[vec(QX_1X_1'Q)]X_1'Q\right\}\right\}$$

$$\Pi_3 = Q\left\{E[V_1 Q(\Lambda_2^{(1)} X_2 X_2')QV_1']\right\}Q \quad + \tfrac{1}{4} Q\overline{H}_2 \left\{(vecQ)(vecQ)' + (Q \otimes Q) + (Q \otimes Q)\left\{E[(vec\Lambda_1^{(1)} X_2 X_1')(vec\Lambda_2^{(1)} X_1 X_2')']\right\}(Q \otimes Q)\right\}\overline{H}_2' Q$$

$$\Pi_4 = -QE(V_1 Q V_1 Q) + \tfrac{1}{4} Q\overline{H}_2 (Q \otimes Q) E\left\{\Lambda_1^{(1)} \Lambda_2^{(1)} (X_1 \otimes \overline{H}_2)[vec(QX_2 X_1' Q)X_2' + vec(QX_1 X_2' Q)X_2' + vec(QX_2 X_2' Q)X_1']\right\}Q$$
$$+ \tfrac{1}{4} Q\overline{H}_2 E\left\{\Lambda_1^{(1)} \Lambda_2^{(1)} \left[(Q\overline{H}_2 vec(QX_1 X_1' Q) \otimes QX_2)X_2' + (Q\overline{H}_2 vec(QX_2 X_1' Q) \otimes QX_1)X_2' + (Q\overline{H}_2 vec(QX_2 X_1' Q) \otimes QX_2)X_1'\right]\right\}Q$$
$$- \tfrac{1}{6} Q\overline{H}_3 E\left\{\Lambda_1^{(1)} \Lambda_2^{(1)} \left[(vec(QX_1 X_1' Q) \otimes QX_2)X_2' + (vec(QX_2 X_1' Q) \otimes QX_1)X_2' + (vec(QX_2 X_1' Q) \otimes QX_2)X_1'\right]\right\}Q \qquad (15)$$

Now we consider a simple case of (4) with only one regressor, which implies that the constant term α in the latent regression model (1) equals the true threshold α in (2). For this simple model, we derive the following corollary.

Corollary 1

If assumptions 1-3 hold for some $s \geq 2$. Then the bias of the MLE of β in the Logit model with only one regressor, to the order of $O(n^{-1})$, is

$$Bias(\hat{\beta}) = -\frac{1}{2n} \frac{E(\Lambda_i^{(2)} x_i^3)}{[E(\Lambda_i^{(1)} x_i^2)]^2} \quad . \qquad (16)$$

Further if Assumptions 1-3 hold for some $s \geq 3$, then the MSE of the MLE of β in the Logit model with only one regressor, to order $O(n^{-2})$ is

$$MSE(\hat{\beta}) = \frac{1}{n}\Pi_1 + \frac{1}{n^2}(\Pi_2 + \Pi_2') + \frac{1}{n^3}(\Pi_3 + \Pi_4 + \Pi_4') \qquad (17)$$

where

$$\Pi_1 = \frac{1}{E(\Lambda_1^{(1)} X_1^2)}$$

$$\Pi_2 = -\frac{1}{[E(\Lambda_1^{(1)} X_1^2)]^3}\left\{E(\Lambda_1^{(1)} X_1^2)^2 - [E(\Lambda_1^{(1)} X_1^2)]^2 + \frac{[E(\Lambda_1^{(2)} X_1^3)]^2}{2E(\Lambda_1^{(1)} X_1^2)}\right\}$$

$$\Pi_3 = \frac{1}{[E(\Lambda_1^{(1)} X_1^2)]^3}\left\{E(\Lambda_1^{(1)} X_1^2)^2 - [E(\Lambda_1^{(1)} X_1^2)]^2 + \frac{3[E(\Lambda_1^{(2)} X_1^3)]^2}{4E(\Lambda_1^{(1)} X_1^2)}\right\}$$

$$\Pi_4 = \frac{1}{[E(\Lambda_1' X_1^2)]^3}\left\{E(\Lambda_1^{(1)} X_1^2)^2 - [E(\Lambda_1^{(1)} X_1^2)]^2 + \frac{3[E(\Lambda_1^{(2)} X_1^3)]^2}{2E(\Lambda_1^{(1)} X_1^2)} - \tfrac{1}{2} E(\Lambda_1^{(3)}) X_1^4]\right\} \qquad (18)$$

The proofs of Theorem 1 and Corollary 1 are given in the Appendix.

5. Numerical Evaluations

In this section, we present some numerical results based on Corollary 1. These evaluations are undertaken for the one-regressor Logit model with different distributional assumptions for the covariates, and different sample sizes. We choose values for the parameters which ensure a sensible signal/noise ratio for the model. The latter is determined by considering the goodness-of-fit for the model - specifically the measure suggested by Efron (1978):

$$R_{Ef}^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{P}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad . \tag{19}$$

Here, we replace the predicted probability $\hat{P}_i$ with the P_i calculated from y_i^*. Usually, with the cross-section data, only modest goodness-of-fit values are achieved. Therefore, we choose parameter values to ensure that this goodness-of-fit measure is of the order 0.7 to 0.8.

The distributions from which the regressor values are generated include both symmetric and asymmetric examples - the standard normal distribution, the uniform distribution on [-2 , 2], and the Chi-Square distribution with 3 degrees of freedom. In Tables 1 to 3 the first two columns give the true value for the parameter and the sample size. The third column, $\hat{\beta}$, is the average MLE of β based on 2,000 replications of a Monte Carlo experiment.

Specifically, the steps associated with this experiment are:

(i) Set a value for the parameter.

(ii) Generate an (*n*x1) vector of observations for the random regressor *X*,.

(iii) Generate (*n*x1) vectors of observations for *y** and *y* based on (3) with a Logistic-distributed disturbance term.

(iv) Estimate a Logit model based on *y* and *X*, and record the MLE for β and the asymptotic standard error of the MLE of β.

(v) Repeat steps (ii)-(iv) 2,000 times.

(vi) Calculate the averages of the 2,000 MLE's of β and of its asymptotic standard error to get the values referred to as $\hat{\beta}$ in column (1) and ASE($\hat{\beta}$) in column (6) of the tables.

Two bias-adjusted estimators, $\hat{\beta}_{BC}$ and $\tilde{\beta}_{BC}$, are then defined as follows:

$$\hat{\beta}_{BC} = \hat{\beta} - B(\hat{\beta}),$$

and

$$\tilde{\beta}_{BC} = \hat{\beta} - \hat{B}(\hat{\beta}),$$

where $B(\hat{\beta})$ is the bias based on (16) and the true parameter β, and $\hat{B}(\hat{\beta})$ is the bias based on (16) and the MLE $\hat{\beta}$. In practice, $\hat{\beta}_{BC}$ is an infeasible estimator as it involves the true parameter. However, $\tilde{\beta}_{BC}$ is the feasible counterpart to this estimator. Given the complication involved in deriving the properties of $\tilde{\beta}_{BC}$, here we focus on $\hat{\beta}$ and $\hat{\beta}_{BC}$. The sixth column in the tables gives the asymptotic standard error from the maximum likelihood estimation across the 2,000 repetitions. The standard deviation, SD($\hat{\beta}_{BC}$) and the standard error, SE($\hat{\beta}_{BC}$), corresponding to the bias-adjusted estimators $\hat{\beta}_{BC}$, are provided in the following two columns. They are defined as follows:

$$\text{SD}(\hat{\beta}_{BC}) = \sqrt{Var(\hat{\beta}_{BC})} = \sqrt{Var(\hat{\beta})} = \sqrt{MSE(\hat{\beta}) - B(\hat{\beta})B'(\hat{\beta})} \ , \tag{20}$$

and

$$\text{SE}(\hat{\beta}_{BC}) = \sqrt{\hat{M}SE(\hat{\beta}) - \hat{B}(\hat{\beta})\hat{B}'(\hat{\beta})} \, . \tag{21}$$

From (20) and (21), we can also see that SD($\hat{\beta}_{BC}$) and SE($\hat{\beta}_{BC}$) are also the second order approximations to the standard deviation and the standard error of $\hat{\beta}$. In the last two columns, we report the MSE of $\hat{\beta}$ and $\hat{\beta}_{BC}$. MSE($\hat{\beta}$) is based on equation (17) and the true parameter value. MSE($\hat{\beta}_{BC}$) is the square of SD($\hat{\beta}_{BC}$), because $\hat{\beta}_{BC}$ is unbiased. From (20), we can see MSE($\hat{\beta}_{BC}$) is always smaller than MSE($\hat{\beta}$).

The Monte Carlo experiment and the calculation of Bias and MSE in (16) and (17) were conducted with code written for the SHAZAM package (Whistler, *et al*., 2001). Tables 1 to 3 report the results for cases where the regressor follows the standard normal distribution, the uniform distribution on [-2, 2] and the chi-square distribution with three degrees of freedom, respectively. From the information in the tables, we can conclude the following. The bias-corrected estimator yields some efficiency gains over the uncorrected MLE, in terms of MSE. Further, from the columns for ASE($\hat{\beta}$), SE($\hat{\beta}$) and SD($\hat{\beta}$), we can see that both the asymptotic and the finite-sample standard errors tend to overestimate the finite-sample standard deviation of the MLE, and the asymptotic standard error tends to be worse in this respect than the finite-sample standard error of the MLE. Comparing SD($\hat{\beta}_{BC}$) and SE($\hat{\beta}_{BC}$), we can see that the standard error of the bias-corrected estimator tends to overestimate the true standard deviation. Further, we also see that there are gains from bias-correction, regardless of whether this is based on the true analytic bias or the estimated analytic bias. Moreover, these gains increase with the sample size. For all the cases we consider, the bias is positive.

6. Conclusions

In this paper we apply results from Rilstone *et al*. (1996) to derive analytic expressions for the first two moments of the MLE for the standard Logit regression model, and undertake some numerical evaluations to illustrate our analytic results. Our analysis extends the very limited literature on this topic, notably by allowing for random covariates. The analytic expressions that we derive are complex, but some simple numerical evaluations provide some clear messages. First, the bias correction of the MLE also leads to gains in efficiency. Second, the accuracy of the corrected estimator generally increases as the sample size increases. Third, the estimated analytical results are closer to the analytical results as the sample size increases. Finally, the asymptotic standard error overestimates the finite-sample standard deviation of the MLE. These results are consistent with the argument that the results derived from the large-n approximations lie between the true value and the asymptotic approximations, and that their accuracy increases as the sample size increases.

The techniques that are used in this paper can be applied readily to determine the first two moments of other maximum likelihood estimators that are defined only implicitly because the likelihood equations cannot be solved analytically. For example, work in progress deals with such estimators for models of count data.

Table 1: Parameter Estimator and Standard Error Estimates

With Standard Normal Regressor

β	n	$\hat{\beta}$	$\hat{\beta}_{BC}$	$\tilde{\beta}_{BC}$	ASE($\hat{\beta}$)	SD($\hat{\beta}$) (SD($\hat{\beta}_{BC}$))	SE($\hat{\beta}$) (SE($\hat{\beta}_{BC}$))	MSE($\hat{\beta}$)	MSE($\hat{\beta}_{BC}$)
1.4	100	1.4655	1.4222	1.4197	0.3262	0.2974	0.3053	0.0903	0.0884
	200	1.4210	1.3993	1.3989	0.2247	0.2159	0.2178	0.0471	0.0466
1.5	100	1.5770	1.5298	1.5267	0.3425	0.3096	0.3192	0.0981	0.0958
	200	1.5185	1.4948	1.4945	0.2347	0.2252	0.2269	0.0513	0.0507
1.6	100	1.6785	1.6273	1.6240	0.3587	0.3222	0.3323	0.1064	0.1038
	200	1.6208	1.5952	1.5947	0.2455	0.2348	0.2369	0.0558	0.0551
1.7	100	1.7708	1.7154	1.7123	0.3731	0.3352	0.3446	0.1154	0.1123
	200	1.7205	1.6928	1.6923	0.2564	0.2448	0.2469	0.0607	0.0599
1.8	100	1.8633	1.8034	1.8006	0.3888	0.3485	0.3571	0.1250	0.1215
	200	1.8195	1.7895	1.7891	0.2677	0.2551	0.2571	0.0660	0.0651
1.9	100	1.9722	1.9078	1.9044	0.4069	0.3622	0.3722	0.1353	0.1312
	200	1.9304	1.8982	1.8974	0.2807	0.2657	0.2689	0.0716	0.0706
2.0	100	2.0835	2.0143	2.0102	0.4263	0.3761	0.3880	0.1463	0.1415
	200	2.0285	1.9939	1.9932	0.2928	0.2765	0.2797	0.0777	0.0765
2.1	100	2.1928	2.1186	2.1138	0.4457	0.3904	0.4038	0.1579	0.1524
	200	2.1235	2.0864	2.0858	0.3046	0.2877	0.2904	0.0841	0.0828
2.2	100	2.2663	2.1869	2.1834	0.4593	0.4049	0.4147	0.1702	0.1640
	200	2.2282	2.1885	2.1878	0.3179	0.2991	0.3024	0.0910	0.0895
2.3	100	2.3675	2.2828	2.2789	0.4788	0.4197	0.4299	0.1833	0.1761
	200	2.3375	2.2952	2.2941	0.3322	0.3108	0.3152	0.0984	0.0966

Table 2: Parameter Estimator and Standard Error Estimates

With Uniform Distribution (-2, 2) Regressor

β	n	$\hat{\beta}$	$\hat{\beta}_{BC}$	$\tilde{\beta}_{BC}$	ASE($\hat{\beta}$)	SD($\hat{\beta}$) (SD($\hat{\beta}_{BC}$))	SE($\hat{\beta}$) (SE($\hat{\beta}_{BC}$))	MSE($\hat{\beta}$)	MSE($\hat{\beta}_{BC}$)
1.4	100	1.4807	1.4437	1.4401	0.2874	0.2584	0.2678	0.0681	0.0668
	200	1.4126	1.3941	1.3939	0.1946	0.1872	0.1883	0.0354	0.0350
1.5	100	1.5843	1.5428	1.5388	0.3032	0.2701	0.2804	0.0747	0.0730
	200	1.5194	1.4986	1.4982	0.2055	0.1963	0.1981	0.0390	0.0385
1.6	100	1.6582	1.6119	1.6089	0.3145	0.2824	0.2898	0.0819	0.0797
	200	1.6267	1.6035	1.6028	0.2172	0.2059	0.2086	0.0429	0.0424
1.7	100	1.7619	1.7104	1.7070	0.3317	0.2952	0.3033	0.0898	0.0871
	200	1.7286	1.7028	1.7020	0.2288	0.2161	0.2190	0.0473	0.0467
1.8	100	1.8639	1.8069	1.8031	0.3504	0.3084	0.3171	0.0984	0.0951
	200	1.8265	1.7980	1.7973	0.2406	0.2266	0.2295	0.0522	0.0514
1.9	100	1.9751	1.9122	1.9074	0.3716	0.3221	0.3326	0.1077	0.1038
	200	1.9343	1.9028	1.9017	0.2543	0.2377	0.2416	0.0575	0.0565
2.0	100	2.0665	1.9972	1.9927	0.3888	0.3361	0.3457	0.1178	0.1130
	200	2.0313	1.9966	1.9956	0.2672	0.2492	0.2528	0.0633	0.0621
2.1	100	2.1790	2.1028	2.0972	0.4113	0.3505	0.3621	0.1287	0.1229
	200	2.1396	2.1015	2.1001	0.2823	0.2610	0.2659	0.0696	0.0681
2.2	100	2.3119	2.2286	2.2200	0.4393	0.3652	0.3819	0.1403	0.1334
	200	2.2355	2.1938	2.1925	0.2963	0.2733	0.2778	0.0764	0.0747
2.3	100	2.4312	2.3402	2.3295	0.4668	0.3801	0.3999	0.1527	0.1445
	200	2.3412	2.2957	2.2940	0.3122	0.2860	0.2913	0.0839	0.0818

Table 3: Parameter Estimator and Standard Error Estimates

With Chi-Square (3) Regressor

β	n	$\hat{\beta}$	$\hat{\beta}_{BC}$	$\tilde{\beta}_{BC}$	ASE($\hat{\beta}$) (SD($\hat{\beta}_{BC}$))	SD($\hat{\beta}$) (SE($\hat{\beta}_{BC}$))	SE($\hat{\beta}$)	MSE($\hat{\beta}$)	MSE($\hat{\beta}_{BC}$)
1.7	200	1.8024	1.7204	1.7086	0.3651	0.2421	0.2574	0.0653	0.0586
	500	1.7455	1.7127	1.7107	0.2206	0.1772	0.1838	0.0325	0.0314
1.8	200	1.9066	1.8130	1.7996	0.3898	0.2570	0.2720	0.0748	0.0661
	500	1.8502	1.8127	1.8103	0.2361	0.1918	0.1992	0.0382	0.0368
1.9	200	1.9962	1.8901	1.8770	0.4113	0.2711	0.2836	0.0848	0.0735
	500	1.9521	1.9097	1.9069	0.2512	0.2067	0.2146	0.0445	0.0427
2.0	200	2.0917	1.9720	1.9585	0.4346	0.2841	0.2947	0.0950	0.0807
	500	2.0557	2.0078	2.0046	0.2667	0.2219	0.2305	0.0515	0.0492
2.1	200	2.2114	2.0769	2.0591	0.4641	0.2956	0.3063	0.1055	0.0874
	500	2.1622	2.1085	2.1046	0.2832	0.2374	0.2471	0.0592	0.0563
2.2	200	2.3383	2.1879	2.1640	0.4962	0.3052	0.3148	0.1158	0.0932
	500	2.2645	2.2044	2.2000	0.2989	0.2531	0.2633	0.0677	0.0640
2.3	200	2.4645	2.2970	2.2661	0.5281	0.3127	0.3184	0.1258	0.0978
	500	2.3656	2.2987	2.2939	0.3148	0.2689	0.2794	0.0768	0.0723
2.4	200	2.5582	2.3723	2.3406	0.5519	0.3173	0.3169	0.1352	0.1007
	500	2.4628	2.3885	2.3836	0.3302	0.2849	0.2949	0.0867	0.0812
2.5	200	2.6745	2.4690	2.4314	0.5822	0.3183	0.3087	0.1436	0.1013
	500	2.5643	2.4821	2.4767	0.3465	0.3009	0.3112	0.0973	0.0905
2.6	200	2.7732	2.5467	2.5070	0.6082	0.3149	0.2943	0.1504	0.0991
	500	2.6750	2.5844	2.5777	0.3644	0.3169	0.3288	0.1086	0.1004

Appendix: Proof of Theorem 1 and Corollary 1

Proof of Theorem 1

First, for the Logit model in (4), we know the following propositions.

$$E(y_i^j | X) = \Lambda_i \tag{22}$$

By applying (11) and the law of iterated expectations, we can derive the following results.

$$\overline{V_1 d_1} = 0$$

$$\overline{V_1 d_1 d_2' V_2'} = 0$$

$$\overline{V_1 d_2 d_1' V_2'} = 0$$

$$\overline{V_1 d_1 d_2' \otimes d_2'} = 0$$

$$\overline{V_1 d_2 [d_1' \otimes d_2']} = 0$$

$$\overline{V_1 d_2 [d_2' \otimes d_1']} = 0$$

$$\overline{d_1 \otimes d_1 d_2' V_2'} = 0$$

$$\overline{[d_1 \otimes d_2] d_1' V_2'} = 0$$

$$\overline{[d_1 \otimes d_2] d_2' V_1'} = 0$$

$$\overline{V_1 Q V_2 d_1 d_2'} = 0$$

$$\overline{V_1 Q V_2 d_2 d_1'} = 0$$

$$\overline{V_1 Q \overline{H}_2 [d_1 \otimes d_2] d_2'} = 0$$

$$\overline{V_1 Q \overline{H}_2 [d_2 \otimes d_1] d_2'} = 0$$

$$\overline{V_1 Q \overline{H}_2 [d_2 \otimes d_2] d_1'} = 0$$

$$\overline{W_1 [d_1 \otimes d_2] d_2'} = 0$$

$$\overline{W_1 [d_2 \otimes d_1] d_2'} = 0$$

$$\overline{W_1 [d_2 \otimes d_2] d_1'} = 0$$

$$\overline{[d_1 \otimes Q V_1 d_2] d_2'} = 0$$

$$\overline{[d_1 \otimes Q V_2 d_1] d_2'} = 0$$

$$\overline{[d_1 \otimes Q V_2 d_2] d_1'} = 0$$

$$\overline{[QV_1 d_1 \otimes d_2]d_2'} = 0$$

$$\overline{[QV_1 d_2 \otimes d_1]d_2'} = 0$$

$$\overline{[QV_1 d_2 \otimes d_2]d_1'} = 0$$

$$\overline{d_1 \otimes d_1} = -vecQ$$

$$\overline{V_1 d_1 d_1'} = E(\Lambda_1^{(1)} V_1 Q X_1 X_1' Q)$$

$$\overline{[d_1 \otimes d_1]d_1'} = E\left\{\Lambda_1^{(2)}[vec(QX_1 X_1' Q)]X_1' Q\right\}$$

$$\overline{V_1 d_2 d_2' V_1'} = E[V_1 Q(\Lambda_2^{(1)} X_2 X_2')QV_1']$$

$$\overline{[d_1 \otimes d_1]}\overline{[d_2' \otimes d_2']} = (vecQ)(vecQ)'$$

$$\overline{[d_1 \otimes d_2][d_1' \otimes d_2']} = Q \otimes Q$$

$$\overline{V_1 Q V_1 d_2 d_2'} = -EV_1 Q V_1 Q$$

$$\overline{[d_1 \otimes d_2][d_2' \otimes d_1']} = (Q \otimes Q)\left\{E[(vec\Lambda_1^{(1)} X_2 X_1')(vec\Lambda_2^{(1)} X_1 X_2')']\right\}(Q \otimes Q)$$

$$\overline{d_1 \otimes Q\overline{H}_2[d_1 \otimes d_2]d_2'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}(Q \otimes Q)(X_1 \otimes \overline{H}_2)[vec(QX_2 X_1' Q)]X_2' Q$$

$$\overline{d_1 \otimes Q\overline{H}_2[d_2 \otimes d_1]d_2'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}(Q \otimes Q)(X_1 \otimes \overline{H}_2)[vec(QX_1 X_2' Q)]X_2' Q$$

$$\overline{d_1 \otimes Q\overline{H}_2[d_2 \otimes d_2]d_1'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}(Q \otimes Q)(X_1 \otimes \overline{H}_2)[vec(QX_2 X_2' Q)]X_1' Q$$

$$\overline{[Q\overline{H}_2[d_1 \otimes d_1] \otimes d_2]d_2'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}[Q\overline{H}_2 vec(QX_1 X_1' Q) \otimes QX_2]X_2' Q$$

$$\overline{[Q\overline{H}_2[d_1 \otimes d_2] \otimes d_1]d_2'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}[Q\overline{H}_2 vec(QX_2 X_1' Q) \otimes QX_1]X_2' Q$$

$$\overline{[Q\overline{H}_2[d_1 \otimes d_2] \otimes d_2]d_1'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}[Q\overline{H}_2 vec(QX_2 X_1' Q) \otimes QX_2]X_1' Q$$

$$\overline{[d_1 \otimes d_1 \otimes d_2]d_2'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}[vec(QX_1 X_1' Q) \otimes QX_2]X_2' Q$$

$$\overline{[d_1 \otimes d_2 \otimes d_1]d_2'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}[vec(QX_2 X_1' Q) \otimes QX_1]X_2' Q$$

$$\overline{[d_1 \otimes d_2 \otimes d_2]d_1'} = E\Lambda_1^{(1)}\Lambda_2^{(1)}[vec(QX_2 X_1' Q) \otimes QX_2]X_1' Q \qquad (23)$$

Therefore, based on Lemma 1 and (23), Theorem 1 is proved.

Proof of Corollary 1

When the Logit model only include one regressor, (23) reduces to

$$\overline{V_1 d_1} = 0$$

$$\overline{V_1 d_1 d_2' V_2'} = 0$$

$$\overline{V_1 d_2 d_1' V_2'} = 0$$

$$\overline{V_1 d_1 d_2' \otimes d_2'} = 0$$

$$\overline{V_1 d_2 [d_1' \otimes d_2']} = 0$$

$$\overline{V_1 d_2 [d_2' \otimes d_1']} = 0$$

$$\overline{d_1 \otimes d_1 d_2' V_2'} = 0$$

$$\overline{[d_1 \otimes d_2] d_1' V_2'} = 0$$

$$\overline{[d_1 \otimes d_2] d_2' V_1'} = 0$$

$$\overline{V_1 Q V_2 d_1 d_2'} = 0$$

$$\overline{V_1 Q V_2 d_2 d_1'} = 0$$

$$\overline{V_1 Q \overline{H}_2 [d_1 \otimes d_2] d_2'} = 0$$

$$\overline{V_1 Q \overline{H}_2 [d_2 \otimes d_1] d_2'} = 0$$

$$\overline{V_1 Q \overline{H}_2 [d_2 \otimes d_2] d_1'} = 0$$

$$\overline{W_1 [d_1 \otimes d_2] d_2'} = 0$$

$$\overline{W_1 [d_2 \otimes d_1] d_2'} = 0$$

$$\overline{W_1 [d_2 \otimes d_2] d_1'} = 0$$

$$\overline{[d_1 \otimes Q V_1 d_2] d_2'} = 0$$

$$\overline{[d_1 \otimes Q V_2 d_1] d_2'} = 0$$

$$\overline{[d_1 \otimes Q V_2 d_2] d_1'} = 0$$

$$\overline{[Q V_1 d_1 \otimes d_2] d_2'} = 0$$

$$\overline{[Q V_1 d_2 \otimes d_1] d_2'} = 0$$

$$\overline{[Q V_1 d_2 \otimes d_2] d_1'} = 0$$

$$\overline{d_1 \otimes d_1} = -\frac{1}{E(\Lambda_1^{(1)} X_1^2)}$$

$$\overline{V_1 d_1 d_1'} = 1 - \frac{E(\Lambda_1^{(1)} X_1^2)^2}{[E(\Lambda_1^{(1)} X_1^2)]^2}$$

$$\overline{[d_1 \otimes d_1]d_1'} = -\frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{V_1 d_2 d_2' V_1'} = \frac{E(\Lambda_1^{(1)} X_1^2)^2 - [E(\Lambda_1^{(1)} X_1^2)]^2}{E(\Lambda_1^{(1)} X_1^2)}$$

$$\overline{[d_1 \otimes d_1][d_2' \otimes d_2']} = \frac{1}{[E(\Lambda_1^{(1)} X_1^2)]^2}$$

$$\overline{[d_1 \otimes d_2][d_1' \otimes d_2']} = \frac{1}{[E(\Lambda_1^{(1)} X_1^2)]^2}$$

$$\overline{[d_1 \otimes d_2][d_2' \otimes d_1']} = \frac{1}{[E(\Lambda_1^{(1)} X_1^2)]^2}$$

$$\overline{V_1 Q V_1 d_2 d_2'} = 1 - \frac{E(\Lambda_1^{(1)} X_1^2)^2}{[E(\Lambda_1^{(1)} X_1^2)]^2}$$

$$\overline{d_1 \otimes Q\overline{H}_2[d_1 \otimes d_2]d_2'} = \frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{d_1 \otimes Q\overline{H}_2[d_2 \otimes d_1]d_2'} = \frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{d_1 \otimes Q\overline{H}_2[d_2 \otimes d_2]d_1'} = \frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{[Q\overline{H}_2[d_1 \otimes d_1] \otimes d_2]d_2'} = \frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{[Q\overline{H}_2[d_1 \otimes d_2] \otimes d_1]d_2'} = \frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{[Q\overline{H}_2[d_1 \otimes d_2] \otimes d_2]d_1'} = \frac{E(\Lambda_1^{(2)} X_1^3)}{[E(\Lambda_1^{(1)} X_1^2)]^3}$$

$$\overline{[d_1 \otimes d_1 \otimes d_2]d_2'} = \frac{1}{[E(\Lambda_1^{(1)} X_1^2)]^2}$$

$$\overline{[d_1 \otimes d_2 \otimes d_1]d_2'} = \frac{1}{[E(\Lambda_1^{(1)} X_1^2)]^2}$$

$$\overline{[d_1 \otimes d_2 \otimes d_2]d_1'} = \frac{1}{[E(\Lambda_1^{(1)} X_1^2)]^2} \tag{24}$$

Based on (24) and Lemma 1, Corollary 1 is proved.

References

Amemiya, T. (1980). The n^{-2}-order mean squared errors of the maximum likelihood minimum logit chi-square estimator. Annals of Statistics 8, 488-505.

Berkson, J. (1944). Application of the logistic function to bio-assay. Journal of the American Statistical Association 39, 357-365.

Berkson, J. (1955). Maximum likelihood and minimum χ^2 estimates of the logistic function. Journal of the American Statistical Association 50, 130-162.

Cameron, A. C., Trivedi, P. K. (2005) Microeconometrics: Methods and Applications, Cambridge University Press, Cambridge.

Davis, L. (1984). Comments on a paper by T. Amemiya on estimation in a dichotomous logit regression model. Annals of Statistics 12, 778-782.

Efron, B. (1978). Regression and ANOVA with zero-one data: measures of residual variation. Journal of the American Statistical Association 73, 113-212.

Ghosh, J. K., Sinha, B. K. (1981). A necessary and sufficient condition for second order admissibility with applications to Berkson's bioassay problem. Annals of Statistic 9, 1334-1338.

Gourieroux, C., Montfort, A. (1981). Asymptotic properties of the maximum likelihood estimator in dichotomous logit models. Journal of Econometrics 17, 83-97.

Hensher, D. A., J. M. Rose, Greene, W. H. (2005). Applied Choice Analysis. Cambridge University Press, Cambridge.

Hughes, G. A., Savin, N. E. (1994). Is the minimum chi-square estimator the winner in logit regression? Journal of Econometric 61, 345-366.

Li, J. (2005). Small sample properties of discrete choice model estimators based on symmetric and asymmetric cumulative distribution functions. M.A. Extended Essay, Department of Economics, University of Victoria.

Mackinnon, J. G., Smith, A. A. (1998). Approximate bias correction in econometrics. Journal of Econometrics 85, 205-230.

Maddala, G. S. (1983), Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge.

Rilstone, P., Srivastava, V. K., Ullah, A. (1996). The second order bias and MSE of nonlinear estimators. Journal of Econometrics 75, 239-395.

Rilstone, P., Ullah, A. (2002). Sampling bias in Heckman's sample selection estimator. In Y. P. Chaubey (ed.), Recent Advances in Statistical Methods. World Scientific, Hackensack NJ.

Taylor, W. F. (1953). Distance functions and regular best asymptotically normal estimates. Annals of Mathematical Statistics 24, 85-92.

Ullah, A. (2004). Finite Sample Econometrics. Oxford University Press, Oxford.

Whistler, D., White, K. J. Wong, S. D., Bates, D. (2001). SHAZAM User's Reference Manual Version 9. Northwest Econometrics, Vancouver B.C.

Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge MA.