University of Victoria
Department of Economics

Econometrics Working Paper EWP0906

ISSN 1485-6441

FINITE-SAMPLE PROPERTIES OF THE MAXIMUM LIKELIHOOD ESTIMATOR FOR THE BINARY LOGIT MODEL WITH RANDOM COVARIATES

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

This version: August 2009

Abstract:
We examine the finite sample properties of the maximum likelihood estimator for the binary logit model with random covariates. Analytic expressions for the first-order bias and second-order mean squared error function for the maximum likelihood estimator in this model are derived, and we undertake some numerical evaluations to analyze and illustrate these analytic results for the single covariate case. For various data distributions, the bias of the estimator is signed the same as the covariate's coefficient, and both the absolute bias and the mean squared errors increase symmetrically with the absolute value of that parameter. The behaviour of a bias-adjusted maximum likelihood estimator, constructed by subtracting the (maximum likelihood) estimator of the first-order bias from the original estimator, is examined in a Monte Carlo experiment. This bias-correction is effective in all of the cases considered, and is recommended when the logit model is estimated by maximum likelihood with small samples.

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

MSC Codes: 62F10; 62J12; 62P20

Contact Author:
David E. Giles, 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 PROPERTIESS OF THE MAXIMUM LIKELIHOOD ESTIMATOR FOR THE BINARY LOGIT MODEL WITH RANDOM COVARIATES

1. Introduction

Qualitative response (QR) models are very widely used in various fields, including bioassay, medicine, transportation research, economics, and other social sciences. These models have the characteristic that the dependent variable is qualitative, rather than quantitative. To make the model estimable, these qualitative attributes are "coded" numerically. The binary choice model, with the dependent variable coded as zero or unity (without loss of generality), is the most widely used of the QR models. 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 occurrence for one of the attributes 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. Choosing the standard normal distribution for this function gives rise to the probit model, while the logistic distribution results in the logit (or "logistic regression") model. Of course, other choices are possible, but the logit and probit models are the two that are encountered most frequently in practice, and they generally yield similar (scaled) estimates. The appeal of the logit specification is that the logistic distribution function can be expressed in closed form, and this has certain computational advantages when the model is extended to the multinomial case involving more than two characteristics. In this paper we focus on the logit, rather than probit, model.

The maximum likelihood estimator (MLE) is the usual choice for QR models. For both the logit and probit models the likelihood function is strictly concave, so it has a unique maximum, but the likelihood equations are non-linear in the parameters, and must be solved numerically. If the covariates are non-random, the likelihood functions for QR models satisfy the usual regularity conditions and so the maximum likelihood estimators (MLEs) are weakly consistent and best asymptotically normal, and the strong consistency of the MLE for the logit model has been established by Gourieroux and Montfort (1981). Taylor (1953) showed that the MLE estimator and the minimum chi square estimator (MCSE) proposed by Berkson (1944), and defended vigorously by that author, have the same asymptotic normal distribution for this model. Other estimators that have been suggested for the logit model include the "minimum ϕ-divergence" estimator (Pardo *et al*. 2005, Menéndez *et al*. 2009.), which is a generalization of MLE and is also consistent and asymptotically normal. Turning to the case of random covariates, Fahrmeir and Kaufmann (1986) investigated the asymptotic properties of various discrete and qualitative response models (including the logit model), and provided conditions on the behaviour of the covariates under which the MLE has its usual asymptotic properties. Wilde (2008) discussed the inconsistency of the generalized method of moments estimator for QR models with endogenous (random) regressors, and suggested a suitable modification in the case of the probit, but not logit, model.

A number of results relating to the finite-sample properties of the MLE (and some other estimators) for the logit model have also been established. However, all of these relate to the case of non-stochastic covariates, and it is this last assumption that we relax in this paper. Using the approach of Cox and Snell (1968), Cordeiro and McCullagh (1991) provided analytical expressions for the $O(n^{-1})$ bias of the MLE in the family of generalized linear models. This family includes logistic regression, of course. Several authors have investigated the properties of the MLE for the logit model in the context of a two-stage sampling scheme involving grouped data of a type that arises frequently in the biological sciences. This includes simple random sampling as a special case. Berkson (1955) evaluated the finite-sample bias and MSE of the MLE and the MCSE estimator for some simple examples of this model, and found the MCSE to be preferred to the MLE in terms of MSE in the cases that he considered. Amemiya (1980) derived analytic expressions for the $O(n^{-2})$ MSEs of the MLE and the MCSE for the "dichotomous" (binary) logit model and provided some numerical results for the relative quality of these two estimators. Several other studies have extended Berkson's and Amemiya's results. Ghosh and Sinha (1981) provided 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). Those authors discussed methods for reducing the bias of consistent estimators that are biased in finite samples, and applied these methods to the MLE for the linear AR(1) model and the standard logit model based on simple random sampling with fixed regressors. 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.

The assumption that the covariates in the logit model are non-random (or “fixed in repeated samples”) is obviously unsatisfactory in many situations. One example is when survey data are used, as is very common with applications in economics and the other social sciences. So, in this paper, we use results due to Rilstone *et al.* (RSU) (1996), as corrected by Rilstone and Ullah (2005), to derive analytic expressions for the bias and MSE functions for the MLE in the logit model based on simple random sampling with *stochastic* covariates. Based on the analytic bias expression we can derive a bias-corrected MLE and the standard error associated with this bias-corrected estimator. We also provide some numerical evaluations based on these analytic results. The approach that we adopt was also used by Rilstone and Ullah (2002) in the context of Heckman’s sample selection estimator, and could also be used to extend our results to other QR models.

The next section introduces the logit model. In section 3 we summarize the required results of RSU (1996) and use them to derive analytic expressions for the bias and mean square error of the MLE in the binary logit model. Some numerical evaluations and Monte Carlo results follow in section 4; and the final section provides our conclusions.

2. The Logit Model and the Maximum Likelihood Estimator

A binary choice model is structured as follows:

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

where y_i^* is the latent dependent variable to incorporate the effects of covariates; and the row vector, X_i', represents the i^{th} observation on all of the covariates. As is well known, provided that an intercept is included among the covariates, the threshold value, a, may be assigned to zero without affecting the results. We make this assignment in what follows.

The basic model can be structured as:

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

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

The form of the cumulative distribution function, $F(X_i'\beta)$, will determine which particular model is used. As noted above, we focus on the logit model, so:

$$P_i = \Pr(y_i = 1 | X_i) = F(X_i'\beta) = \Lambda_i \tag{2}$$

where

$$\Lambda_i = \frac{\exp(X_i'\beta)}{1 + \exp(X_i'\beta)} \tag{3}$$

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

The MLE for the parameter vector in (2) can be derived as the solution of the following log-likelihood equations:

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^{n} \left[(y_i - \Lambda_i) X_i \right] = 0 . \tag{4}$$

The MLE cannot be written as a closed-form expression, and this is what substantially complicates the task of evaluating the characteristics of its (finite-sample) sampling distribution, whether the covariates are random or not.

3. Analytic Results

Before deriving the analytic results for the bias and MSE of the MLE in the binary logit model, we first introduce the results of RSU (1996). The class of estimators considered by RSU includes those which can only be expressed implicitly as a function of the data. Suppose we have a regression model of the form

$$y_i = f(X_i; \beta_0) + \varepsilon_i . \tag{5}$$

The regressor vector X_i can include any endogenous or exogenous variables. Let $Z_i = (y_i, X_i)$ and assume 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_0, or any other parameters of interest. The estimator $\hat{\theta}$ can be written in the form:

$$\psi_n(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} g_i(\hat{\theta}) = 0 \ , \tag{6}$$

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. Some assumptions about the function $g_i(\theta)$ are needed for the derivation of the Lemmas below. (See Ullah 2004: 31.)

Assumption 1

The s^{th} order derivatives of $g_i(\theta)$ exist in a neighborhood of θ_0 and $E\left\|\nabla^s g_i(\theta_0)\right\|^2 < \infty$, where $\|A\|$ = trace $[AA']^{1/2}$ denotes the usual norm of the matrix *A*; and $\nabla^s A(\theta)$ is the matrix of s^{th} order partial derivations of the matrix $A(\theta)$ with respect to θ, obtained recursively.

Assumption 2

For some neighborhood of θ_0, $(\nabla \psi_n(\theta))^{-1} = O_p(1)$.

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 $E|M_i| \le C < \infty$, $i = 1, 2, \ldots$

As the log-likelihood function of the binary logit model is strictly concave, these three assumptions are satisfied. In the notation that follows, for simplicity we will suppress the argument for any function where this can be done without confusion. So, $g_i(\theta_0)$ will be written as g_i. Then, RSU derived the following lemmas, corrected here according to the corrigendum in Rilstone and Ullah (2005).

Lemma 1 (Proposition 3.2, RSU, 1996; Ullah 2004: 32)

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

$$B(\hat{\theta}) = \frac{1}{n} Q\left\{\overline{V_1 d_1} - \frac{1}{2}\overline{H}_2 \overline{[d_1 \otimes d_1]}\right\} \quad , \tag{7}$$

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 $\overline{\nabla g_i} = E[\nabla g_i]$.

Lemma 2 (Proposition 3.4, RSU, 1996; Ullah 2004: 32)

If Assumptions 1-3 hold for some $s \geq 3$, then the MSE of $\hat{\theta}$ to $O(n^{-2})$ is

$$MSE(\hat{\theta}) = \frac{1}{n}\Pi_1 + \frac{1}{n^2}(\Pi_2 + \Pi_2') + \frac{1}{n^2}(\Pi_3 + \Pi_4 + \Pi_4') \tag{8}$$

where

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

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

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

$$\begin{aligned}
\Pi_4 = {} & Q\left\{\overline{V_1 Q V_1 d_2 d_2'} + \overline{V_1 Q V_2 d_1 d_2'} + \overline{V_1 Q V_2 d_2 d_1'}\right\} \\
& - \tfrac{1}{2} Q\left\{\overline{V_1 Q\overline{H}_2 [d_1 \otimes d_2] d_2'} + \overline{V_1 Q\overline{H}_2 [d_2 \otimes d_1] d_2'} + \overline{V_1 Q\overline{H}_2 [d_2 \otimes d_2] d_1'}\right\} \\
& + \tfrac{1}{2} Q\left\{\overline{W_1 [d_1 \otimes d_2] d_2'} + \overline{W_1 [d_2 \otimes d_1] d_2'} + \overline{W_1 [d_2 \otimes d_2] d_1'}\right\} \\
& - \tfrac{1}{2} Q\overline{H}_2 \left\{\overline{[d_1 \otimes Q V_1 d_2] d_2'} + \overline{[d_1 \otimes Q V_2 d_1] d_2'} + \overline{[d_1 \otimes Q V_2 d_2] d_1'}\right\} \\
& + \tfrac{1}{4} Q\overline{H}_2 \left\{\overline{d_1 \otimes Q\overline{H}_2 [d_1 \otimes d_2] d_2'} + \overline{d_1 \otimes Q\overline{H}_2 [d_2 \otimes d_1] d_2'} + \overline{d_1 \otimes Q\overline{H}_2 [d_2 \otimes d_2] d_1'}\right\} \\
& - \tfrac{1}{2} Q\overline{H}_2 \left\{\overline{[Q V_1 d_1 \otimes d_2] d_2'} + \overline{[Q V_1 d_2 \otimes d_1] d_2'} + \overline{[Q V_1 d_2 \otimes d_2] d_1'}\right\} \\
& + \tfrac{1}{4} Q\overline{H}_2 \left\{\overline{[Q\overline{H}_2 [d_1 \otimes d_1] \otimes d_2] d_2'} + \overline{[Q\overline{H}_2 [d_1 \otimes d_2] \otimes d_1] d_2'} + \overline{[Q\overline{H}_2 [d_1 \otimes d_2] \otimes d_2] d_1'}\right\} \\
& - \tfrac{1}{6} Q\overline{H}_3 \left\{\overline{[d_1 \otimes d_1 \otimes d_2] d_2'} + \overline{[d_1 \otimes d_2 \otimes d_1] d_2'} + \overline{[d_1 \otimes d_2 \otimes d_2] d_1'}\right\}
\end{aligned} \tag{9}$$

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

To apply the above lemmas to derive the bias and MSE for MLE in the binary logit model, we assume that both the dependent and independent variables are random and i.i.d.. Comparing (4) and (6), we can see that for the logit model we should set $g_i = (y_i - \Lambda_i)X_i$. We know that E($g_i \mid X_i$) = 0, so by the law of iterated expectations, E(g_i) = 0. We then have the following results:

$\nabla g_i = -\Lambda_i^{(1)} X_i X_i'$; $\qquad\qquad \overline{H}_1 = \overline{\nabla g_i} = -E(\Lambda_i^{(1)} X_i X_i')$

$$\nabla^2 g_i = -\Lambda_i^{(2)} X_i (X_i' \otimes X_i'); \qquad \overline{H}_2 = \overline{\nabla^2 g_i} = -E[\Lambda_i^{(2)} X_i (X_i' \otimes X_i')]$$

$$\nabla^3 g_i = -\Lambda_i^{(3)} X_i (X_i' \otimes X_i' \otimes X_i'); \qquad \overline{H}_3 = \overline{\nabla^3 g_i} = -E[\Lambda_i^{(3)} X_i (X_i' \otimes X_i' \otimes X_i')]$$

$$Q = (\overline{\nabla g_i})^{-1} = -[E(\Lambda_i^{(1)} X_i X_i')]^{-1}; \qquad d_i = Q g_i = -[E(\Lambda_i^{(1)} X_i X_i')]^{-1} (y_i - \Lambda_i) X_i$$

$$V_i = \nabla g_i - \overline{\nabla g_i} = -\Lambda_i^{(1)} X_i X_i' + E(\Lambda_i^{(1)} X_i X_i')$$

$$W_i = \nabla^2 g_i - \overline{\nabla^2 g_i} = -\Lambda_i^{(2)} X_i (X_i' \otimes X_i') + E[\Lambda_i^{(2)} X_i (X_i' \otimes X_i')] \quad , \tag{10}$$

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

$$\Lambda_i^{(1)} = \frac{\exp(X_i'\beta)}{[1+\exp(X_i'\beta)]^2}$$

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

$$\Lambda_i^{(3)} = \frac{\exp(X_i'\beta)\{1-4\exp(X_i'\beta)+[\exp(X_i'\beta)]^2\}}{[1+\exp(X_i'\beta)]^4} \quad . \tag{11}$$

Then we can derive the following theorems and corollaries.

Theorem 1

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

$$Bias(\hat{\beta}) = \frac{1}{2n} Q \overline{H}_2 vecQ \quad . \tag{12}$$

Theorem 2

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

$$MSE(\hat{\beta}) = \frac{1}{n}\Pi_1 + \frac{1}{n^2}(\Pi_2 + \Pi_2') + \frac{1}{n^2}(\Pi_3 + \Pi_4 + \Pi_4'), \tag{13}$$

where

$$\Pi_1 = -Q$$

$$\Pi_2 = -Q\Big\{E(\Lambda_1^{(1)}V_1QX_1X_1'Q) - \tfrac{1}{2}\overline{H}_2E\big\{\Lambda_1^{(2)}[vec(QX_1X_1'Q)]X_1'Q\big\}\Big\}$$

$$\Pi_3 = Q\Big\{E[V_1Q(\Lambda_2^{(1)}X_2X_2')QV_1']\Big\}Q \quad + \tfrac{1}{4}Q\overline{H}_2\big\{(vecQ)(vecQ)' + (Q\otimes Q)$$
$$+ (Q\otimes Q)\Big\{E[(vec\Lambda_1^{(1)}X_2X_1')(vec\Lambda_2^{(1)}X_1X_2')']\Big\}(Q\otimes Q)\Big\}\overline{H}_2'Q$$

$$\Pi_4 = -QE(V_1QV_1Q) + \tfrac{1}{4}Q\overline{H}_2(Q\otimes Q)E\big\{\Lambda_1^{(1)}\Lambda_2^{(1)}(X_1\otimes\overline{H}_2)[vec(QX_2X_1'Q)X_2'$$
$$+ vec(QX_1X_2'Q)X_2' + vec(QX_2X_2'Q)X_1']\big\}Q$$
$$+ \tfrac{1}{4}Q\overline{H}_2E\big\{\Lambda_1^{(1)}\Lambda_2^{(1)}\big[(Q\overline{H}_2vec(QX_1X_1'Q)\otimes QX_2)X_2'$$
$$+ (Q\overline{H}_2vec(QX_2X_1'Q)\otimes QX_1)X_2' + (Q\overline{H}_2vec(QX_2X_1'Q)\otimes QX_2)X_1'\big]\big\}Q$$
$$- \tfrac{1}{6}Q\overline{H}_3E\big\{\Lambda_1^{(1)}\Lambda_2^{(1)}\big[(vec(QX_1X_1'Q)\otimes QX_2)X_2' + (vec(QX_2X_1'Q)\otimes QX_1)X_2'$$
$$+ (vec(QX_2X_1'Q)\otimes QX_2)X_1'\big]\big\}Q \qquad (14)$$

Now we consider the logit model with only one regressor, which implies that the coefficient of the intercept term in the latent regression model equals the true threshold a in (1). For this simple model, we have the following corollaries.

Corollary 1

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

$$Bias(\hat{\beta}) = -\frac{1}{2n}\frac{E(\Lambda_i^{(2)}X_i^3)}{[E(\Lambda_i^{(1)}X_i^2)]^2} \qquad . \qquad (15)$$

Corollary 2

If Assumptions 1-3 hold for some $s \ge 3$, then the $O(n^{-2})$ MSE of the MLE of β in the logit model with only one regressor, is

$$MSE(\hat{\beta}) = \frac{1}{n}\Pi_1 + \frac{1}{n^2}(\Pi_2 + \Pi_2') + \frac{1}{n^2}(\Pi_3 + \Pi_4 + \Pi_4') \qquad , \qquad (16)$$

where

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

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

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

$$\Pi_4 = \frac{1}{[E(\Lambda_1' X_1^2)]^3}\left\{E(\Lambda_1^{(1)} X_1^2)^2 - [E(\Lambda_1^{(1)} X_1^2)]^2 + \frac{3[E(\Lambda_1^{(2)} X_1^3)]^2}{2E(\Lambda_1^{(1)} X_1^2)} - \tfrac{1}{2} E(\Lambda_1^{(3)}) X_1^4]\right\}. \tag{17}$$

The proofs of the Theorems and Corollaries are given in the Appendix.

Based on these results, we can obtain two bias-adjusted estimators, $\hat{\beta}_{BC}$ and $\tilde{\beta}_{BC}$, defined as follows:

$$\hat{\beta}_{BC} = \hat{\beta} - B(\hat{\beta}), \tag{18}$$

$$\tilde{\beta}_{BC} = \hat{\beta} - \hat{B}(\hat{\beta}), \tag{19}$$

where $B(\hat{\beta})$ is the bias based on (15) and the true parameter β, and $\hat{B}(\hat{\beta})$ is the estimated bias based on (15) and the MLE, $\hat{\beta}$. In practice, of course, $\hat{\beta}_{BC}$ is an infeasible estimator as it involves the unknown true parameter. It can be shown that both of these bias-adjusted estimators are unbiased to $O(n^{-2})$.

Finally, in the single covariate case the true standard deviation, s.d.($\hat{\beta}$), and the standard error, s.e.($\hat{\beta}$), can be obtained as:

$$\text{s.d.}(\hat{\beta}) = \sqrt{MSE(\hat{\beta}) - [B(\hat{\beta})]^2}\ , \tag{20}$$

$$\text{s.e.}(\hat{\beta}) = \sqrt{\hat{M}SE(\hat{\beta}) - [\hat{B}(\hat{\beta})]^2}\ . \tag{21}$$

where $MSE(\hat{\beta})$ is based on (16) and the true parameter β, and $\hat{M}SE(\hat{\beta})$ is the estimated MSE based on (16) and the MLE, $\hat{\beta}$.

4. Numerical Evaluations

Given their complexity, it is difficult to interpret the above analytical results. Here we provide some numerical evaluations and Monte Carlo simulation results obtained using code for the R (2008) statistical package. For the one-regressor model we first consider how the true value of the coefficient, β_0, and the distribution of the regressor affect the finite sample properties of the MLE. Second, we conduct a small Monte Carlo experiment to evaluate the performance of the feasible bias-corrected estimator of β relative to that of the infeasible bias-corrected estimator and the basic MLE. In that experiment we also evaluate s.e.($\hat{\beta}$) as an estimator of s.d.($\hat{\beta}$).

In both the numerical evaluations and the Monte Carlos simulation, three different distributions are used to generate the random regressor – normal, chi-square and Student-t. This enables us to allow for different degrees of variability, skewness and kurtosis in covariate data by varying the moments of each distribution. The value of β_0 is chosen for each case to control the signal-to-noise ratios to sensible levels. Sample sizes of n = 25, 50 and 100 are considered.

The numerical evaluations of $\hat{\beta}$ are summarized in Tables 1 to 3 where, to conserve space, the results for n = 100 are omitted, but they corroborate the tabulated results. We note the following. The sign of the bias of the MLE is the same as the sign of β_0, and both the absolute bias and the MSE increase (symmetrically) with the absolute value of β_0. For each of the distributions considered for the regressor, as n increases, the absolute bias and MSE both decrease for all β_0 values, reflecting the mean-square consistency of the MLE. Whatever distribution the covariate follows, as the moments of this distribution change, the bias and MSE of the MLE change in a manner that depends on the absolute value of β_0. In percentage terms, the absolute biases of the MLE can be substantial. For example, when n = 25 (50), and for the range of values of β_0 that are tabulated, the percentage biases associated with a normally distributed covariate range up to 48% (24%). The corresponding figures for a chi-square distributed covariate are 40% (20%); and those for a Student-t distributed covariate are 25% (12%). Percentage MSEs range in value up to 27% (14%), 19% (11%) and 14% (8%) in Tables 1, 2 and 3 respectively, again for the values of β_0 that are considered in these evaluations.

In addition, a small Monte Carlo experiment has been undertaken to check the finite-sample performance of the feasible bias-corrected estimator, compared with the MLE and the infeasible bias-corrected estimator. The same sample sizes and distributions for the random regressor are considered, but only a limited selection of parameter values for these distributions, and a small number of values for β_0 are considered. Table 4 reports results averaged over 1,000 Monte Carlo replications. Several features of these are noteworthy. First, and as expected, both the feasible and infeasible bias correction substantially reduces the bias of the MLE. For instance, when the covariate is chi-square distributed, n = 25 and β_0 = -0.2, the absolute percentage biases of the infeasible and feasible bias-adjusted MLEs are 3.5% and 5.0% respectively, as compared with a corresponding bias of 14.4% for $\hat{\beta}$ itself. In addition, it is noteworthy that the feasible bias-corrected estimator ($\tilde{\beta}_{BC}$) often out-performing the infeasible "estimator" ($\hat{\beta}_{BC}$). For example, for the normally distributed covariate, when n = 25 and β_0 = 1.5, the absolute percentage biases of $\hat{\beta}$, $\hat{\beta}_{BC}$ and $\tilde{\beta}_{BC}$ are 17.6%, 5.0% and 2.2% respectively; and the corresponding figures for the Student-t distributed covariate, when n = 25 and β_0 = -0.7, are 22.9%, 8.3% and 3.2%.

Finally, in each part of Table 4 we see that the (average values) of both the true standard deviation and the standard error of $\hat{\beta}$ decrease monotonically as n increases, for a given value of β_0 . The first of these results reflects the mean-square consistency of the estimator noted above. In all but one of the cases considered, s.e.($\hat{\beta}$) is an upward-biased estimator of s.d.($\hat{\beta}$) – the exception is for the chi-square distributed covariate when β_0 = 0.1. For the situations tabulated, the largest bias for this standard error is 11.1%, which occurs for the Student-t distributed covariate when n = 25 and β_0 = -0.7. This (maximum) bias reduces to 5% when n = 100. In each part of Table 4, s.e.($\hat{\beta}$) converges to s.d.($\hat{\beta}$) as n increases, as expected.

5. Concluding remarks

In this paper we apply results from Rilstone *et al*. (1996) and Rilstone and Ullah (2005) to derive analytic expressions for the first two moments of the maximum likelihood estimator for the binary logit regression model with random covariates. Our analysis extends the 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. The bias and mean squared error of this estimator for the logit model are determined by both the value of the true parameter and the data generating process of the regressor. For the one-regressor case, the bias takes the sign of the coefficient of the regressor. The absolute bias and the mean squared error increase with the absolute true value of this coefficient, and (of course) decrease as the sample size increases.

We also find that a feasible bias-corrected estimator, constructed by subtracting the estimated bias from the maximum likelihood estimator, substantially reduces the bias in all of the situations examined in a limited Monte Carlo experiment. We recommend the use of this bias correction when the logit model is estimated from samples of 100 observations or less. The Monte Carlo experiment also indicates that the standard error associated with the maximum likelihood estimator is quite a reliable estimator of the true standard deviation of that estimator. Although the standard error is generally upward-biased for the cases considered, it converges rapidly to the true standard deviation as the sample size increases.

The techniques that are used in this paper can be applied readily to determine analytic expressions for 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 for count data.

Table 1: $O(n^{-1})$ Bias and $O(n^{-2})$ MSE of MLE. Normal (μ, σ^2) Regressors

	Normal (0, 1)		Normal (3, 1)		Normal (5, 1)		Normal (0, 2)		Normal (0, 4)	
					***n* = 25**					
$\|\beta_0\|$	|Bias|	MSE	|Bias|	MSE	|Bias|	MSE	|Bias|	MSE	|Bias|	MSE
0.1	0.0120	0.1644	0.0056	0.0223	0.0048	0.0116	0.0119	0.0540	0.0118	0.0267
0.2	0.0239	0.1678	0.0116	0.0251	0.0107	0.0160	0.0236	0.0577	0.0234	0.0325
0.3	0.0356	0.1735	0.0185	0.0300	0.0190	0.0241	0.0350	0.0640	0.0362	0.0422
0.4	0.0471	0.1815	0.0267	0.0372	0.0316	0.0375	0.0467	0.0730	0.0512	0.0552
0.5	0.0586	0.1917	0.0370	0.0470	0.0512	0.0583	0.0590	0.0845	0.0692	0.0713
0.6	0.0700	0.2041	0.0500	0.0599	0.0817	0.0894	0.0723	0.0984	0.0903	0.0904
0.7	0.0816	0.2186	0.0665	0.0763	0.1289	0.1338	0.0867	0.1147	0.1147	0.1126
0.8	0.0934	0.2352	0.0874	0.0968	0.2015	0.1930	0.1025	0.1335	0.1426	0.1378
0.9	0.1056	0.2539	0.1137	0.1219	0.3123	0.2621	0.1197	0.1548	0.1740	0.1662
1.0	0.1181	0.2748	0.1468	0.1518	0.4798	0.3186	0.1384	0.1788	0.2090	0.1976
					***n* = 50**					
$\|\beta_0\|$	|Bias|	MSE	|Bias|	MSE	|Bias|	MSE	|Bias|	MSE	|Bias|	MSE
0.1	0.0060	0.0814	0.0028	0.0097	0.0024	0.0046	0.0060	0.0238	0.0059	0.0095
0.2	0.0119	0.0831	0.0058	0.0108	0.0053	0.0060	0.0118	0.0256	0.0117	0.0118
0.3	0.0178	0.0861	0.0092	0.0128	0.0095	0.0088	0.0175	0.0287	0.0181	0.0157
0.4	0.0236	0.0901	0.0134	0.0157	0.0158	0.0134	0.0234	0.0329	0.0256	0.0211
0.5	0.0293	0.0953	0.0185	0.0198	0.0256	0.0206	0.0295	0.0385	0.0346	0.0281
0.6	0.0350	0.1016	0.0250	0.0251	0.0408	0.0317	0.0362	0.0452	0.0452	0.0369
0.7	0.0408	0.1090	0.0333	0.0322	0.0645	0.0479	0.0434	0.0533	0.0574	0.0476
0.8	0.0467	0.1176	0.0437	0.0411	0.1008	0.0709	0.0512	0.0627	0.0713	0.0603
0.9	0.0528	0.1272	0.0569	0.0524	0.1562	0.1008	0.0598	0.0736	0.0870	0.0753
1.0	0.0590	0.1380	0.0734	0.0663	0.2399	0.1345	0.0692	0.0860	0.1045	0.0927

Table 2: $O(n^{-1})$ Bias and $O(n^{-2})$ MSE of MLE. Chi-square (d.f.) regressors

	Chi-square (3)		Chi-square (4)		Chi-square (5)	
			***n* = 25**			
$\lvert\beta_0\rvert$	**|Bias|**	**MSE**	**|Bias|**	**MSE**	**|Bias|**	**MSE**
0.3	0.0391	0.0440	0.0343	0.0383	0.0326	0.0370
0.4	0.0524	0.0556	0.0485	0.0510	0.0486	0.0518
0.5	0.0677	0.0699	0.0657	0.0668	0.0693	0.0708
0.7	0.1050	0.1068	0.1114	0.1085	0.1285	0.1215
0.9	0.1523	0.1553	0.1743	0.1640	0.2172	0.1883
1.0	0.1799	0.1842	0.2131	0.1967	0.2750	0.2255
1.1	0.2104	0.2161	0.2574	0.2322	0.3431	0.2627
1.3	0.2802	0.2891	0.3635	0.3086	0.5142	0.3228
			***n* = 50**			
$\lvert\beta_0\rvert$	**|Bias|**	**MSE**	**|Bias|**	**MSE**	**|Bias|**	**MSE**
0.3	0.0195	0.0168	0.0171	0.0138	0.0163	0.0127
0.4	0.0262	0.0220	0.0242	0.0191	0.0243	0.0185
0.5	0.0339	0.0285	0.0329	0.0259	0.0347	0.0261
0.7	0.0525	0.0463	0.0557	0.0450	0.0642	0.0485
0.9	0.0761	0.0712	0.0871	0.0730	0.1086	0.0822
1.0	0.0900	0.0867	0.1066	0.0908	0.1375	0.1037
1.1	0.1052	0.1045	0.1287	0.1115	0.1715	0.1284
1.3	0.1401	0.1471	0.1818	0.1614	0.2571	0.1859

Table 3: $O(n^{-1})$ Bias and $O(n^{-2})$ MSE of MLE. Student-t (d.f.) regressors

	t (5)		*t* (10)		*t* (15)	
	***n* = 25**					
$\|\beta_0\|$	**|Bias|**	**MSE**	**|Bias|**	**MSE**	**|Bias|**	**MSE**
0.5	0.0793	0.1485	0.0682	0.1679	0.0647	0.1754
1.0	0.1389	0.2494	0.1287	0.2598	0.1251	0.2643
1.5	0.2138	0.4038	0.2016	0.4059	0.1974	0.4075
2.0	0.3079	0.6190	0.2925	0.6123	0.2873	0.6107
2.5	0.4229	0.9017	0.4038	0.8854	0.3973	0.8805
3.0	0.5597	1.2565	0.5365	1.2302	0.5286	1.2219
3.5	0.7191	1.6847	0.6912	1.6488	0.6819	1.6370
4.0	0.9014	2.1847	0.8686	2.1400	0.8577	2.1252
4.5	1.1071	2.7509	1.0688	2.6996	1.0561	2.6822
	***n* = 50**					
$\|\beta_0\|$	**|Bias|**	**MSE**	**|Bias|**	**MSE**	**|Bias|**	**MSE**
0.5	0.0397	0.0729	0.0341	0.0830	0.0323	0.0869
1.0	0.0695	0.1252	0.0643	0.1305	0.0626	0.1327
1.5	0.1069	0.2075	0.1008	0.2079	0.0987	0.2085
2.0	0.1540	0.3261	0.1463	0.3209	0.1437	0.3195
2.5	0.2115	0.4877	0.2019	0.4758	0.1987	0.4721
3.0	0.2799	0.6988	0.2682	0.6790	0.2643	0.6727
3.5	0.3596	0.9652	0.3456	0.9367	0.3410	0.9274
4.0	0.4507	1.2923	0.4343	1.2540	0.4288	1.2414
4.5	0.5535	1.6844	1.0688	2.6996	0.5281	1.6194

Table 4: Average of 1,000 Monte Carlo simulation replications

		Normal (0, 1)							**Chi-square (3)**				
$\boldsymbol{\beta_0}$	$\boldsymbol{n}$	$\hat{\beta}$	$\tilde{\beta}_{BC}$	$\hat{\beta}_{BC}$	s.e.($\hat{\beta}$)	s.d.($\hat{\beta}$)	$\boldsymbol{\beta_0}$	$\boldsymbol{n}$	$\hat{\beta}$	$\tilde{\beta}_{BC}$	$\hat{\beta}_{BC}$	s.e.($\hat{\beta}$)	s.d.($\hat{\beta}$)
1.5	25	1.7643	1.4664	1.5754	0.6756	0.6135	0.4	25	0.4912	0.4574	0.4388	0.2601	0.2299
	50	1.6302	1.5412	1.5357	0.4806	0.4489		50	0.4262	0.3858	0.4000	0.1595	0.1458
	100	1.5412	1.4877	1.4940	0.3311	0.3226		100	0.4131	0.4011	0.4000	0.0986	0.0967
1.1	25	1.2893	1.1750	1.1582	0.5801	0.5297	0.1	25	0.1133	0.1102	0.0982	0.1612	0.1764
	50	1.1913	1.1409	1.1257	0.4035	0.3817		50	0.1082	0.1051	0.1007	0.0970	0.1041
	100	1.1304	1.0981	1.0977	0.2784	0.2724		100	0.1020	0.1000	0.0982	0.0619	0.0651
-0.7	25	-0.7844	-0.7166	-0.7028	0.4926	0.4604	-0.2	25	-0.2288	-0.1900	-0.2017	0.1980	0.1861
	50	-0.7319	-0.6677	-0.6911	0.3391	0.3277		50	-0.2066	-0.1940	-0.1930	0.1165	0.1131
	100	-0.7191	-0.7015	-0.6987	0.2365	0.2325		100	-0.2096	-0.2015	-0.2028	0.0742	0.0725
-1.3	25	-1.5161	-1.4173	-1.3574	0.6209	0.5702	-0.3	25	-0.3520	-0.2209	-0.3130	0.2220	0.2061
	50	-1.3627	-1.2890	-1.2833	0.4316	0.4139		50	-0.3212	-0.3045	-0.3017	0.1334	0.1280
	100	-1.3429	-1.3031	-1.3032	0.3047	0.2964		100	-0.3089	-0.2972	-0.2992	0.0852	0.0836

Student-t (5)

$\boldsymbol{\beta_0}$	$\boldsymbol{n}$	$\hat{\beta}$	$\tilde{\beta}_{BC}$	$\hat{\beta}_{BC}$	s.e.($\hat{\beta}$)	s.d.($\hat{\beta}$)	$\boldsymbol{\beta_0}$	$\boldsymbol{n}$	$\hat{\beta}$	$\tilde{\beta}_{BC}$	$\hat{\beta}_{BC}$	s.e.($\hat{\beta}$)	s.d.($\hat{\beta}$)
1.5	25	1.7667	1.5593	1.5529	0.6459	0.5984	-0.7	25	-0.8600	-0.7221	-0.7581	0.4609	0.4152
	50	1.5888	1.4851	1.4819	0.4668	0.4428		50	-0.7521	-0.7024	-0.7011	0.3096	0.2966
	100	1.5437	1.4847	1.4902	0.3290	0.3199		100	-0.7368	-0.7094	-0.7113	0.2166	0.2108
0.9	25	1.0386	0.8967	0.9126	0.4983	0.4574	-1.7	25	-2.0947	-1.8176	-1.8456	0.7200	0.6482
	50	0.9444	0.8691	0.8814	0.3414	0.3295		50	-1.8295	-1.7031	-1.7050	0.5186	0.4845
	100	0.9385	0.8941	0.9070	0.2416	0.2351		100	-1.7899	-1.7392	-1.7277	0.3691	0.3515

Appendix: Proofs of Theorems and Corollaries

Proofs of Theorems 1 and 2

For the logit model in (3),

$$E(y_i^j \mid X) = \Lambda_i \quad . \tag{22}$$

By applying (10) and the law of iterated expectations, we can derive the following results. The terms, which are in Lemmas to derive the bias and MSE, but not specified below, are all equal to zero.

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

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

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

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

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

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

$$\overline{V_1 Q V_1 d_2 d_2'} = -E[V_1 Q V_1] Q$$

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

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

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

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

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

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

$$\overline{[Q\overline{H}_2[d_1 \otimes d_2] \otimes d_2]d_1'} = E\left[\Lambda_1^{(1)}\Lambda_2^{(1)}[Q\overline{H}_2 vec(QX_2X_1'Q) \otimes QX_2]X_1'\right]Q$$

$$\overline{[d_1 \otimes d_1 \otimes d_2]d_2'} = E\left[\Lambda_1^{(1)}\Lambda_2^{(1)}[vec(QX_1X_1'Q) \otimes QX_2]X_2'\right]Q$$

$$\overline{[d_1 \otimes d_2 \otimes d_1]d_2'} = E\left[\Lambda_1^{(1)}\Lambda_2^{(1)}[vec(QX_2X_1'Q) \otimes QX_1]X_2'\right]Q$$

$$\overline{[d_1 \otimes d_2 \otimes d_2]d_1'} = E\left[\Lambda_1^{(1)}\Lambda_2^{(1)}[vec(QX_2X_1'Q) \otimes QX_2]X_1'\right]Q \qquad (23)$$

Therefore, based on Lemmas 1, 2 and (23), Theorems 1 and 2 are proved.

Proofs of Corollaries 1 and 2

When the logit model only includes just one regressor, (23) reduces to

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Based on Lemmas 1, 2 and (24), Corollaries 1 and 2 are proved.

References

Amemiya T (1980) The n^{-2}-order mean squared errors of the maximum likelihood and the 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.

Cordeiro GM, McCullagh P (1991) Bias correction in generalized linear models. Journal of the Royal Statistical Society, B 53: 629-643.

Cox DR, Snell EJ (1968) A general definition of residuals. Journal of the Royal Statistical Society, B 30: 248-275.

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.

Fahrmeir L, Kaufmann H (1986) Asymptotic inference in discrete response models. Statistical Papers 27: 197-205.

Ghosh JK, Sinha BK (1981) A necessary and sufficient condition for second order admissibility with applications to Berkson's bioassay problem. Annals of Statistics 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.

Hughes GA, Savin NE (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 JG, Smith AA (1998) Approximate bias correction in econometrics. Journal of Econometrics 85: 205-230.

Menéndez ML, Pardo, L, Pardo MC (2009) Preliminary phi-divergence test estimators for linear restrictions in a logistic regression model. Statistical Papers 50: 277-300.

Pardo JA, Pardo L, Pardo MC (2005) Minimum ϕ-divergence estimator in logistic regression models. Statistical Papers 47: 91-108.

R (2008) The R Project for Statistical Computing, http://www.r-project.org

Rilstone P, Srivatsava VK, Ullah A (1996) The second order bias and MSE of nonlinear estimators. Journal of Econometrics 75: 369-395.

Rilstone P, Ullah A (2002) Sampling bias in Heckman's sample selection estimator. In: Chaubey YP (ed) Recent advances in statistical methods. World Scientific, Hackensack NJ.

Rilstone P, Ullah A (2005) Corrigendum to: The second order bias and mean squared error of non-linear estimators. Journal of Econometrics 124: 203-204.

Taylor WF (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.

Wilde J (2008) A note on GMM estimation of probit models with endogenous regressors. Statistical Papers 49: 471-484.