

**Finite-Sample Properties of the Maximum Likelihood Estimator for
the Poisson Regression 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*

September 2009

Abstract: We examine the small-sample behaviour of the maximum likelihood estimator for the Poisson regression model with random covariates. Analytic expressions for the first-order bias and second-order mean squared error for this estimator are derived, and we undertake some numerical evaluations to illustrate these results for the single covariate case. The properties of the bias-adjusted maximum likelihood estimator, constructed by subtracting the estimated first-order bias from the original estimator, are investigated in a Monte Carlo experiment. Correcting the estimator for its first-order bias is found to be effective in the cases considered, and we recommend its use when the Poisson regression model is estimated by maximum likelihood with small samples.

Keywords: Poisson regression 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

1. Introduction

The Poisson regression model is widely used to study count data in many disciplines, and is generally fitted using the maximum likelihood estimator (MLE). The likelihood function for the Poisson model is strictly concave, and satisfies the usual regularity conditions. So, the MLE for the Poisson regression model has all of the usual desirable asymptotic properties (Gourieroux *et al.*, 1984). Surprisingly, however, the finite sample properties of the MLE for this model have been studied only for a limited number of particular models, only for the case of non-random covariates, and only using simulation methods. King (1988) and Brännäs (1991) used Monte Carlo experiments to examine the finite sample properties of MLE of the Poisson regression model. The evidence provided by the first of these authors must be treated cautiously, as his Monte Carlo experiment involved only 100 replications. However, for a model with two covariates, King (1988: 850) reports biases as large as -13.4%, 16.4% and -3.7% for sample sizes of $n = 10, 50$ and 100 respectively. From more reliable experiments, Breslow (1990: 568) reports biases in the range 1.2% to 1.9% when $n = 36, 72$; and Brännäs (1991: 234-235) reports biases in the range -2% to 1% when $n = 50$. These last two studies involve a Poisson regression model with a single covariate. The present paper is the first to derive analytic expressions for the first-order bias and second-order mean squared error (MSE) of the MLE for the Poisson regression model, and the first to consider these properties in the context of random covariates. The latter are likely to arise when the model is estimated from survey data, and there is no reason to presume that the relatively small biases noted above will be applicable in this case.

Now, suppose Y is a series of count data, and follows a Poisson distribution with mean (and variance) λ :

$$\Pr.[Y = y_i] = \lambda_i^{y_i} \exp(-\lambda_i) / y_i!; \quad y_i = 0, 1, 2, \dots; \quad i = 1, 2, \dots, n. \quad (1)$$

The Poisson regression model arises when we make the mean a (non-negative) function of certain covariates:

$$\lambda_i = \exp(x_i' \beta); \quad i = 1, 2, 3, \dots, n. \quad (2)$$

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

$$\partial \log L(\beta) / \partial \beta = \sum_{i=1}^n (y_i - \lambda_i) x_i = 0 \quad (3)$$

and the Hessian matrix is:

$$\partial^2 \log L(\beta) / \partial \beta \partial \beta' = - \sum_{i=1}^n \lambda_i x_i x_i' . \quad (4)$$

This Hessian matrix is negative-definite for all x and β , so there is a unique solution to the likelihood equations. However, as (3) is non-linear in the parameters, the MLE has to be solved numerically. Evaluation of the finite-sample properties of the MLE is also complicated by the fact that the estimator cannot be expressed in closed form. In this paper, we apply results from Rilstone *et al.* (RSU) (1996), as corrected by Rilstone and Ullah (2005), to derive analytic expressions for the first-order bias and second-order MSE functions for the MLE in the Poisson regression model with *stochastic* covariates that follow quite general distributions. We also present some numerical evaluations of the analytic bias and MSE expressions, and we explore the effectiveness of bias-adjusting the MLE.

In the next section, we apply the methods of RSU to derive the finite-sample properties of the MLE for the Poisson regression model. Section 3 presents some numerical evaluations of the rather complex analytic expressions derived in section 2, and in section 4 we discuss the results of a Monte Carlo experiment that focuses on bias-adjusting this MLE when the model has a single covariate. Section 5 provides our conclusions.

2. Analytic bias and mean squared error expressions

RSU (1996) provide a general framework that allows us to derive the first-order bias and second-order MSE of a fairly wide class of nonlinear estimators. There are several well-known methods for examining the finite-sample properties of statistics, such as the Edgeworth expansion, the bootstrap and the jackknife methods. Compared with these methods, that discussed by RSU has some particular strengths. First, it provides us with

analytic, rather than numerical, results. In addition, it is much simpler to derive than the Edgeworth expansion, especially for the nonlinear case. RSU's method focuses on statistics which can be expressed as a function of the data in the following way:

$$\psi_n(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n g_i(\hat{\theta}) = 0 \quad , \quad (5)$$

where $\hat{\theta}$ is the estimator of interest; $g_i(\theta) = g_i(z_i, \theta)$ is a $k \times 1$ vector involving the known variables z_i and the parameters θ ; and $E[g_i(\theta)] = 0$ only for the true value of the parameter vector, θ_0 .

This paper applies RSU's method to derive our analytic results. Recently, corresponding results for the binary logit model have been obtained by Chen and Giles (2009). For the Poisson model, we set $g_i = (y_i - \lambda_i)x_i$. So, $E(g_i | x_i) = 0$ and the law of iterated expectations implies that $E(g_i) = 0$. There are certain assumptions about the behavior of $g_i(\theta)$ that are needed in order for the results of RSU to hold (Ullah, 2004: 31):

Assumption 1

The s^{th} order derivatives of $g_i(\theta)$ exist in a neighborhood of θ_0 and $E\|\nabla^s g_i(\theta_0)\|^2 < \infty$, where $\|A\| = \text{trace} [AA']^{1/2}$ 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

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

To simplify the following notation we will suppress the argument for any function where this can be done without confusion. So, $g_i(\theta_0)$ is written as g_i , for example. The structure of g_i for the Poisson regression model can easily be shown to satisfy the above three assumptions. We will use the following two results, corrected here as indicated by Rilstone and Ullah (2005).

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

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

$$B(\hat{\theta}) = \frac{1}{n} Q \left\{ \overline{V_1 d_1} - \frac{1}{2} \overline{H_2 [d_1 \otimes d_1]} \right\}, \quad (6)$$

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

$$MSE(\hat{\theta}) = \frac{1}{n} \Pi_1 + \frac{1}{n^2} (\Pi_2 + \Pi'_2) + \frac{1}{n^2} (\Pi_3 + \Pi_4 + \Pi'_4), \quad (7)$$

where

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

$$\Pi_2 = Q \left\{ -\overline{V_1 d_1 d_1'} + \frac{1}{2} \overline{H_2 [d_1 \otimes d_1] d_1'} \right\}$$

$$\begin{aligned} \Pi_3 = Q \left\{ \overline{V_1 d_1 d_2' V_2'} + \overline{V_1 d_2 d_1' V_2'} + \overline{V_1 d_2 d_2' V_1'} \right\} Q \\ + \frac{1}{4} Q \overline{H_2} \left\{ \overline{[d_1 \otimes d_1] [d_2' \otimes d_2']} + \overline{[d_1 \otimes d_2] [d_1' \otimes d_2']} + \overline{[d_1 \otimes d_2] [d_2' \otimes d_1']} \right\} \overline{H_2'} Q \\ - \frac{1}{2} Q \left\{ \overline{V_1 d_1 d_2' \otimes d_2'} + \overline{V_1 d_2 [d_1' \otimes d_2']} + \overline{V_1 d_2 [d_2' \otimes d_1']} \right\} \overline{H_2'} Q \\ - \frac{1}{2} Q \overline{H_2} \left\{ \overline{d_1 \otimes d_1 d_2' V_2'} + \overline{[d_1 \otimes d_2] d_1' V_2'} + \overline{[d_1 \otimes d_2] d_2' V_1'} \right\} Q \end{aligned}$$

$$\begin{aligned} \Pi_4 = Q \left\{ \overline{V_1 Q V_1 d_2 d_2'} + \overline{V_1 Q V_2 d_1 d_2'} + \overline{V_1 Q V_2 d_2 d_1'} \right\} \\ - \frac{1}{2} Q \left\{ \overline{V_1 Q \overline{H_2} [d_1 \otimes d_2] d_2'} + \overline{V_1 Q \overline{H_2} [d_2 \otimes d_1] d_2'} + \overline{V_1 Q \overline{H_2} [d_2 \otimes d_2] d_1'} \right\} \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{2}Q\{\overline{W_1[d_1 \otimes d_2]d'_2} + \overline{W_1[d_2 \otimes d_1]d'_2} + \overline{W_1[d_2 \otimes d_2]d'_1}\} \\
& - \frac{1}{2}Q\overline{H_2}\{\overline{[d_1 \otimes QV_1d_2]d'_2} + \overline{[d_1 \otimes QV_2d_1]d'_2} + \overline{[d_1 \otimes QV_2d_2]d'_1}\} \\
& + \frac{1}{4}Q\overline{H_2}\{\overline{d_1 \otimes Q\overline{H_2}[d_1 \otimes d_2]d'_2} + \overline{d_1 \otimes Q\overline{H_2}[d_2 \otimes d_1]d'_2} + \overline{d_1 \otimes Q\overline{H_2}[d_2 \otimes d_2]d'_1}\} \\
& - \frac{1}{2}Q\overline{H_2}\{\overline{[QV_1d_1 \otimes d_2]d'_2} + \overline{[QV_1d_2 \otimes d_1]d'_2} + \overline{[QV_1d_2 \otimes d_2]d'_1}\} \\
& + \frac{1}{4}Q\overline{H_2}\{\overline{[Q\overline{H_2}[d_1 \otimes d_1] \otimes d_2]d'_2} + \overline{[Q\overline{H_2}[d_1 \otimes d_2] \otimes d_1]d'_2} + \overline{[Q\overline{H_2}[d_1 \otimes d_2] \otimes d_2]d'_1}\} \\
& - \frac{1}{6}Q\overline{H_3}\{\overline{[d_1 \otimes d_1 \otimes d_2]d'_2} + \overline{[d_1 \otimes d_2 \otimes d_1]d'_2} + \overline{[d_1 \otimes d_2 \otimes d_2]d'_1}\} \tag{8}
\end{aligned}$$

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

It is readily shown that for the Poisson regression model we have the following expressions:

$$\Psi_n(\hat{\beta}) = \frac{1}{n} \sum g_i(\hat{\beta})$$

$$\nabla g_i = -\lambda_i x_i x'_i \quad ;$$

$$\overline{H_1} = \overline{\nabla g_i} = -E(\lambda_i x_i x'_i)$$

$$\nabla^2 g_i = -\lambda_i x_i (x'_i \otimes x'_i) \quad ;$$

$$\overline{H_2} = \overline{\nabla^2 g_i} = -E[\lambda_i x_i (x'_i \otimes x'_i)]$$

$$\nabla^3 g_i = -\lambda_i x_i (x'_i \otimes x'_i \otimes x'_i) \quad ;$$

$$\overline{H_3} = \overline{\nabla^3 g_i} = -E[\lambda_i x_i (x'_i \otimes x'_i \otimes x'_i)]$$

$$V_i = \nabla g_i - \overline{\nabla g_i} = -\lambda_i x_i x'_i + E(\lambda_i x_i x'_i) = \nabla g_i - Q^{-1}$$

$$W_i = \nabla^2 g_i - \overline{\nabla^2 g_i} = -\lambda_i x_i (x'_i \otimes x'_i) + E[\lambda_i x_i (x'_i \otimes x'_i)]$$

$$Q = \overline{H_1}^{-1} \quad ;$$

$$d_i = Qg_i \quad .$$

Then, applying RSU's method we can derive the following theorems and corollaries.

Theorem 1

For the Poisson regression model, the bias of the MLE to order $O(n^{-1})$ is

$$Bias(\hat{\beta}) = \frac{1}{2n} Q\overline{H_2}vec(Q) \quad ,$$

and the MSE of MLE to order $O(n^{-2})$ is

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

where

$$\Pi_1 = -Q$$

$$\Pi_2 = Q\left\{-E(\lambda_1 V_1 Q X_1 X_1' Q) + \frac{1}{2}\bar{H}_2 E\left\{\lambda_1 [\text{vec}(Q X_1 X_1' Q)] X_1' Q\right\}\right\}$$

$$\Pi_3 = Q\left\{E[V_1 Q(\lambda_2 X_2 X_2') Q V_1']\right\} Q + \frac{1}{4} Q \bar{H}_2 \left\{(\text{vec} Q)(\text{vec} Q)' + (Q \otimes Q) + (Q \otimes Q) \left\{E[(\text{vec} \lambda_1 X_2 X_2')(\text{vec} \lambda_2 X_1 X_1')]'\right\} (Q \otimes Q)\right\} \bar{H}_2' Q$$

$$\Pi_4 = \left\{ \begin{array}{l} \lambda_1 \lambda_2 (X_1 \otimes \bar{H}_2) \{[\text{vec}(Q X_2 X_1' Q)] X_2' \\ + [\text{vec}(Q X_1 X_2' Q)] X_2' + [\text{vec}(Q X_2 X_2' Q)] X_1'\} \end{array} \right\} Q + \frac{1}{4} Q \bar{H}_2 E \left\{ \begin{array}{l} \lambda_1 \lambda_2 \{[Q \bar{H}_2 \text{vec}(Q X_1 X_1' Q) \otimes Q X_2] X_2' + [Q \bar{H}_2 \text{vec}(Q X_2 X_1' Q) \otimes Q X_1] X_2' + \\ [Q \bar{H}_2 \text{vec}(Q X_2 X_2' Q) \otimes Q X_2] X_1'\} \end{array} \right\} Q - \frac{1}{6} Q \bar{H}_3 E \left\{ \begin{array}{l} \lambda_1 \lambda_2 \{[\text{vec}(Q X_1 X_1' Q) \otimes Q X_2] X_2' + [\text{vec}(Q X_2 X_1' Q) \otimes Q X_1] X_2' \\ + [\text{vec}(Q X_2 X_2' Q) \otimes Q X_2] X_1'\} \end{array} \right\} Q$$

When the model contains only one covariate, the expressions in Theorem 1 simplify considerably. Corollary 1 reports the general result for the one-regressor case, no matter what distribution the random covariate follows. The other corollaries report the bias and MSE expressions when the random regressor follows a normal, uniform or chi-square distribution.

Corollary 1

For the Poisson regression model, the bias of the MLE to order $O(n^{-1})$ is

$$\text{Bias}(\hat{\beta}) = -\frac{1}{2n} \frac{E(\lambda_i x_i^3)}{[E(\lambda_i x_i^2)]^2}$$

and the MSE of MLE to order $O(n^{-2})$ is

$$MSE(\hat{\beta}) = \left(\frac{1}{n} - \frac{1}{n^2}\right) \frac{1}{E(\lambda_1 X_1^2)} + \frac{1}{n^2 [E(\lambda_1 X_1^2)]^3} \left\{ E(\lambda_1^2 X_1^4) + \frac{19[E(\lambda_1 X_1^3)]^2}{4E(\lambda_1 X_1^2)} - E(\lambda_1 X_1^4) \right\}$$

Corollary 2

In the Poisson regression model, when the single covariate follows a normal distribution with mean u and variance σ^2 , the $O(n^{-1})$ order bias of the MLE is

$$Bias(\hat{\beta}) = -\frac{(\mu + \sigma^2 \beta) [3\sigma^2 + (\mu + \sigma^2 \beta)^2]}{2n [\sigma^2 + (\mu + \sigma^2 \beta)^2]^2 \exp(\mu\beta + 0.5\sigma^2 \beta^2)},$$

and the $O(n^{-2})$ order MSE is

$$\begin{aligned} MSE(\hat{\beta}) &= \left(\frac{1}{n} - \frac{1}{n^2} \right) \frac{1}{[\sigma^2 + (\mu + \sigma^2 \beta)^2] \exp(\mu\beta + 0.5\sigma^2 \beta^2)} \\ &+ \frac{1}{n^2 [[\sigma^2 + (\mu + \sigma^2 \beta)^2] \exp(\mu\beta + 0.5\sigma^2 \beta^2)]^2} \\ &\left\{ \begin{aligned} &[\exp(\mu\beta + 1.5\sigma^2 \beta^2)] [3\sigma^4 + 6\sigma^2(\mu + 2\sigma^2 \beta)^2 + 2(\mu + 2\sigma^2 \beta)^4] \\ &+ \frac{19 \{ (\mu + \sigma^2 \beta) [3\sigma^2 + (\mu + \sigma^2 \beta)^2] \}^2}{4 [\sigma^2 + (\mu + \sigma^2 \beta)^2]} - [3\sigma^4 + 6\sigma^2(\mu + \sigma^2 \beta)^2 + 2(\mu + \sigma^2 \beta)^4] \end{aligned} \right\} \end{aligned}$$

Corollary 3

In the Poisson regression model, when the only covariate follows a uniform distribution on the interval $[a, b]$, the $O(n^{-1})$ order bias of the MLE is

$$Bias(\hat{\beta}) = -\frac{1}{2n} \frac{E(\lambda_i x_i^3)}{[E(\lambda_i x_i^2)]^2},$$

and the $O(n^{-2})$ order MSE is

$$MSE(\hat{\beta}) = \left(\frac{1}{n} - \frac{1}{n^2} \right) \frac{1}{E(\lambda_i X_i^2)} + \frac{1}{n^2 [E(\lambda_i X_i^2)]^3} \left\{ E(\lambda_i^2 X_i^4) + \frac{19 [E(\lambda_i X_i^3)]^2}{4 E(\lambda_i X_i^2)} - E(\lambda_i X_i^4) \right\}$$

where

$$E(\lambda_i x_i^2) = \frac{[b^2 \exp(b\beta) - a^2 \exp(a\beta)]}{\beta(b-a)} - \frac{2[b \exp(b\beta) - a \exp(a\beta)]}{\beta^2(b-a)} + \frac{2[\exp(b\beta) - \exp(a\beta)]}{\beta^3(b-a)}$$

$$E(\lambda_i x_i^3) = \frac{[b^3 \exp(b\beta) - a^3 \exp(a\beta)]}{\beta(b-a)} - \frac{3E(\lambda_i x_i^2)}{\beta}$$

$$E(\lambda_i x_i^4) = \frac{[b^4 \exp(b\beta) - a^4 \exp(a\beta)]}{\beta(b-a)} - \frac{4E(\lambda_i x_i^3)}{\beta}$$

$$E(\lambda_i^2 x_i^4) = \frac{[b^4 \exp(2b\beta) - a^4 \exp(2a\beta)]}{2\beta(b-a)} - \frac{[b^3 \exp(2b\beta) - a^3 \exp(2a\beta)]}{\beta^2(b-a)} \\ + \frac{3[b^2 \exp(2b\beta) - a^2 \exp(2a\beta)]}{2\beta^3(b-a)} - \frac{3[b \exp(2b\beta) - a \exp(2a\beta)]}{2\beta^4(b-a)} + \frac{3[\exp(2b\beta) - \exp(2a\beta)]}{4\beta^5(b-a)}$$

Corollary 4

In the Poisson regression model, when the only covariate follows a χ^2 distribution with r degrees of freedom, the $O(n^{-1})$ order bias of the MLE is

$$Bias(\hat{\beta}) = -\frac{(r+4)(1-2\beta)^{r/2+1}}{2nr(r+2)},$$

and the $O(n^{-2})$ order MSE is

$$MSE(\hat{\beta}) = \left(\frac{1}{n} - \frac{1}{n^2}\right) \frac{(1-2\beta)^{r/2+2}}{r(r+2)} + \frac{(1-2\beta)^{r+2}(r+4)}{n^2 r^2 (r+2)^2} \left\{ (r+6) \left[\frac{(1-2\beta)}{(1-4\beta)} \right]^{r/2+4} + \frac{15}{4} r + 13 \right\}$$

The proofs of Theorem 1 and the various corollaries follow directly from the results provided by RSU (1996), Rilstone and Ullah (2005), and Chen and Giles (2009).

Although the direction of the bias of the MLE when the covariate is uniformly distributed is difficult to discern from Corollary 3, we see immediately from Corollary 2 that if the regressor is normally distributed then the bias will be positive if and only if $\beta < -\mu/\sigma^2$. From Corollary 4, if the covariate follows a chi-square distribution with r degrees of freedom then the bias will be positive if and only if $\beta > 1/2$ and r is a multiple of 4.

Using the results from Corollaries 1 to 4, two bias-adjusted estimators, $\hat{\beta}_{BC}$ and $\tilde{\beta}_{BC}$, can be defined as follows:

$$\hat{\beta}_{BC} = \hat{\beta} - Bias(\hat{\beta}), \quad (9)$$

$$\tilde{\beta}_{BC} = \hat{\beta} - \hat{Bias}(\hat{\beta}), \quad (10)$$

where $Bias(\hat{\beta})$ is the bias we derived in the Theorem 1 and the corollaries, and $\hat{Bias}(\hat{\beta})$ is the corresponding estimator of $Bias(\hat{\beta})$, which is obtained by substituting $\hat{\beta}$ for β in the bias expression. Of course, $\hat{\beta}_{BC}$ is an infeasible estimator, but $\tilde{\beta}_{BC}$ is a feasible estimator, and both of these bias-adjusted estimators are unbiased to $O(n^{-2})$.

Corresponding to the bias-corrected estimators, we can also obtain the true standard deviation, $s.d.(\hat{\beta})$, and the standard error, $s.e.(\hat{\beta})$, to order $O(n^{-1})$ for the single covariate case as follows:

$$s.d.(\hat{\beta}) = \sqrt{MSE(\hat{\beta}) - [Bias(\hat{\beta})]^2}, \quad (11)$$

$$s.e.(\hat{\beta}) = \sqrt{MSE(\hat{\beta}) - [\hat{Bias}(\hat{\beta})]^2}, \quad (12)$$

where $MSE(\hat{\beta})$ and $\hat{MSE}(\hat{\beta})$ are constructed in a manner similar to $Bias(\hat{\beta})$ and $\hat{Bias}(\hat{\beta})$.

3. Numerical evaluations

In this section we present some numerical evaluations using the analytic results in Corollaries 2 to 4 for the one-regressor case. In all of the experiments, the value of β_0 is chosen to control the signal-to-noise ratios to sensible levels, and sample sizes of $n = 25, 50$ and 100 are considered. These numerical evaluations enable us to check how the finite sample properties of the estimator change as the characteristics of the model change, and they appear in Tables 1 to 3. Consistent with Corollaries 2 to 4, we consider a random regressor that is either normally, uniformly, or chi-square distributed. We vary the moments of each of these distributions by changing the parameter values to check the implications for the finite-sample properties of the MLE. In all of the cases we have considered, the results show that the bias and MSE depend on the values of the parameters of the regressor distributions – a result that could be anticipated by differentiating the expressions in the corollaries with respect to these

parameters.

We note the following results. No matter which distribution the regressor follows and what value β_0 is, the absolute bias and the MSE decrease as the sample size increases. This finding is further reinforced by the results of sample size of 100, which are not shown here in order to conserve space. This finding reflects the mean square consistency of MLE, which is consistent with the previous literature. With respect to the percentage bias, the magnitude is quite substantial for certain cases. Among the cases we considered, when $n = 25$ and the regressor is normally distributed, the absolute percentage bias can be as large as 120.8%, depending on the mean and variance of the data. Further, for this sample size the reported absolute percentage biases are as large as 57% for a uniformly distributed covariate, and 72.9% for a chi-square distributed covariate, again depending on the moments of the regressor's distribution. It will be recalled that these (absolute) biases are considerably greater than those found by Breslow (1990) and Brännäs (1991) for models with non-random covariates.

4. Bias-corrected estimation

We have also conducted a Monte Carlo experiment to evaluate the relative performances of the feasible bias-corrected MLE, infeasible bias-corrected MLE and the MLE itself, and also the performance of $\text{s.e.}(\hat{\beta})$ as an estimator of $\text{s.d.}(\hat{\beta})$. Table 4 shows the results of our experiment using 1000 replications. Of course, the feasible bias-corrected estimator is the one that we are most interested in. In all the cases we have considered, the feasible bias-corrected MLE improves the performance of the original MLE substantially. For example, when the regressor follows a chi-square distribution, $n = 25$ and $\beta_0 = -0.3$, the absolute percentage bias of the feasible bias-corrected MLE is 0.91%, compared with 10.15% for the MLE. Similarly, when the regressor follows a uniform distribution, $n = 25$ and $\beta_0 = -2.7$, the absolute percentage bias of the feasible bias-corrected MLE is 1.10%, compared with 11.81% for the MLE. Interestingly, there are negligible differences between the performances of the feasible and infeasible bias-corrected MLEs.

Our Monte Carlo experiment also provides insights into the performance of the $s.e.(\hat{\beta})$ as an estimator of the $s.d.(\hat{\beta})$. The $s.e.(\hat{\beta})$ tends to overestimate the $s.d.(\hat{\beta})$ in all but three of the cases considered. The largest bias of $s.e.(\hat{\beta})$ as an estimator of $s.d.(\hat{\beta})$ among all of the cases tabulated is 34.77%, which occurs when the regressor follows a chi-square distribution, $n = 25$ and $\beta_0 = -0.5$. This bias decreases to 2.53%, when the sample size increases to 100. For all the cases we considered, the $s.e.(\hat{\beta})$ converges to the $s.d.(\hat{\beta})$ and both decrease monotonically in value as the sample size increases, as expected from the consistency of the MLE.

5. Concluding remarks

In this paper, we apply techniques developed by Rilstone *et al.* (1996) to derive analytic expressions for the first-order bias and second-order mean squared error of the maximum likelihood estimator for the Poisson regression model with random regressors. Our study is the first to investigate the finite-sample properties of this estimator when this model has random covariates, and the first to obtain analytic (rather than simulation-based) results for the bias and mean squared error. Not surprisingly, we find that the bias of the maximum likelihood estimator when the covariates are random is greater than that obtained from Monte Carlo experiments by other authors for the non-random regressor case. The magnitude of the bias and mean squared error can be substantial, and this motivates us to consider bias-adjusting the maximum likelihood estimator. Using a Monte Carlo experiment we are able to confirm that our bias-corrected estimator can substantially reduce the bias of the maximum likelihood estimator, and we recommend using this bias-adjusted estimator when the sample size is less than 100.

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

		Normal (0, 1)		Normal (5, 1)		Normal (10, 1)		Normal (0, 2)		Normal (0, 4)	
$n = 25$											
β_0	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	
-0.6	0.0182	0.0395	-0.0796	0.1081	-0.7248	6.0283	0.0044	0.0081	0.0001	0.0070	
-0.4	0.0173	0.0413	-0.0309	0.0238	-0.1061	0.1482	0.0079	0.0090	0.0010	0.0024	
-0.2	0.0110	0.0399	-0.0115	0.0061	-0.0149	0.0056	0.0087	0.0103	0.0039	0.0022	
-0.1	0.0059	0.0388	-0.0069	0.0033	-0.0055	0.0015	0.0055	0.0100	0.0043	0.0026	
0.3	-0.0149	0.0409	-0.0008	0.0003	-0.0001	0.0000	-0.0091	0.0099	-0.0022	0.0020	
0.5	-0.0184	0.0408	-0.0003	0.0001	0.0000	0.0000	-0.0061	0.0083	-0.0004	0.0037	
0.7	-0.0172	0.0378	-0.0001	0.0000	0.0000	0.0000	-0.0030	0.0084	0.0000	0.0158	
0.9	-0.0140	0.0343	0.0000	0.0000	0.0000	0.0000	-0.0012	0.0115	0.0000	0.1363	
1.1	-0.0104	0.0324	0.0000	0.0000	0.0000	0.0000	-0.0004	0.0198	0.0000	2.3616	
$n = 50$											
β_0	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	
-0.6	0.0091	0.0160	-0.0398	0.0353	-0.3624	1.5448	0.0022	0.0025	0.0001	0.0017	
-0.4	0.0087	0.0183	-0.0154	0.0090	-0.0531	0.0425	0.0039	0.0034	0.0005	0.0006	
-0.2	0.0055	0.0194	-0.0058	0.0026	-0.0075	0.0021	0.0043	0.0046	0.0020	0.0008	
-0.1	0.0029	0.0196	-0.0035	0.0015	-0.0028	0.0006	0.0028	0.0048	0.0022	0.0011	
0.3	-0.0075	0.0190	-0.0004	0.0002	0.0000	0.0000	-0.0046	0.0040	-0.0011	0.0006	
0.5	-0.0092	0.0173	-0.0001	0.0000	0.0000	0.0000	-0.0030	0.0028	-0.0002	0.0009	
0.7	-0.0086	0.0147	0.0000	0.0000	0.0000	0.0000	-0.0015	0.0024	0.0000	0.0039	
0.9	-0.0070	0.0123	0.0000	0.0000	0.0000	0.0000	-0.0006	0.0030	0.0000	0.0341	
1.1	-0.0052	0.0106	0.0000	0.0000	0.0000	0.0000	-0.0002	0.0050	0.0000	0.5904	

Table 2: $O(n^{-1})$ Bias and $O(n^{-2})$ MSE of MLE. Uniform (a, b) Regressors

	U (0, 1)		U (0, 5)		U (0, 10)		U (-5, 1)		U (-10, 1)	
$n = 25$										
β_0	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
-1.9	-0.1546	0.8175	-0.5379	4.6888	-1.0830	17.5419	0.0000	0.0000	0.0000	0.0000
-1.4	-0.1140	0.5252	-0.2867	1.4317	-0.5878	5.3088	0.0000	0.0000	0.0000	0.0000
-1.0	-0.0883	0.3687	-0.1439	0.4083	-0.2986	1.4313	0.0002	0.0001	0.0000	0.0000
-0.6	-0.0679	0.2593	-0.0573	0.0851	-0.1042	0.1989	0.0011	0.0006	0.0000	0.0000
-0.2	-0.0517	0.1828	-0.0177	0.0147	-0.0164	0.0089	0.0052	0.0032	0.0011	0.0003
0.7	-0.0273	0.0842	-0.0006	0.0003	0.0000	0.0000	0.0652	0.1097	0.1277	0.2868
1.1	-0.0203	0.0600	-0.0001	0.0001	0.0000	0.0000	0.0692	0.1500	0.1470	0.3761
1.3	-0.0174	0.0506	0.0000	0.0000	0.0000	0.0000	0.0432	0.1285	0.0914	0.2306
1.7	-0.0128	0.0361	0.0000	0.0000	0.0000	0.0000	-0.0136	0.1602	-0.0232	0.3179
2.0	-0.0102	0.0281	0.0000	0.0000	0.0000	0.0000	-0.0326	0.1868	-0.0595	0.4358
$n = 50$										
β_0	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
-1.9	-0.0773	0.3201	-0.2690	1.3444	-0.5415	4.7284	0.0000	0.0000	0.0000	0.0000
-1.4	-0.0570	0.2137	-0.1434	0.4286	-0.2939	1.4644	0.0000	0.0000	0.0000	0.0000
-1.0	-0.0442	0.1544	-0.0719	0.1306	-0.1493	0.4080	0.0001	0.0001	0.0000	0.0000
-0.6	-0.0339	0.1115	-0.0286	0.0306	-0.0521	0.0612	0.0006	0.0003	0.0000	0.0000
-0.2	-0.0259	0.0805	-0.0088	0.0062	-0.0082	0.0035	0.0026	0.0015	0.0006	0.0001
0.7	-0.0136	0.0386	-0.0003	0.0002	0.0000	0.0000	0.0326	0.0407	0.0639	0.0894
1.1	-0.0101	0.0279	-0.0001	0.0000	0.0000	0.0000	0.0346	0.0655	0.0735	0.1423
1.3	-0.0087	0.0236	0.0000	0.0000	0.0000	0.0000	0.0216	0.0658	0.0457	0.1181
1.7	-0.0064	0.0170	0.0000	0.0000	0.0000	0.0000	-0.0068	0.0763	-0.0116	0.1457
2.0	-0.0051	0.0133	0.0000	0.0000	0.0000	0.0000	-0.0163	0.0792	-0.0298	0.1685

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

	Chi-square (4)		Chi-square (6)		Chi-square (8)	
$n = 25$						
β_0	Bias	MSE	Bias	MSE	Bias	MSE
-1.0	-0.1800	0.5908	-0.3375	1.8272	-0.7290	8.0089
-0.7	-0.0922	0.1747	-0.1382	0.3387	-0.2389	0.9164
-0.5	-0.0533	0.0667	-0.0667	0.0900	-0.0960	0.1645
-0.3	-0.0273	0.0215	-0.0273	0.0194	-0.0315	0.0226
-0.2	-0.0183	0.0112	-0.0160	0.0082	-0.0161	0.0075
0.3	-0.0004	0.0001	-0.0001	0.0000	0.0000	0.0000
1.5	0.0533	0.0655	-0.0667	0.0375	0.0960	0.1628
1.7	0.0922	0.1722	-0.1382	0.2079	0.2389	0.9099
1.9	0.1463	0.3988	-0.2561	0.7945	0.5163	4.0536
2.1	0.2185	0.8374	-0.4369	2.4448	1.0066	15.0453
$n = 50$						
β_0	Bias	MSE	Bias	MSE	Bias	MSE
-1.0	-0.0900	0.1814	-0.1688	0.5074	-0.3645	2.0933
-0.7	-0.0461	0.0575	-0.0691	0.1013	-0.1194	0.2530
-0.5	-0.0267	0.0233	-0.0333	0.0292	-0.0480	0.0491
-0.3	-0.0137	0.0081	-0.0137	0.0070	-0.0157	0.0077
-0.2	-0.0091	0.0044	-0.0080	0.0032	-0.0081	0.0028
0.3	-0.0002	0.0000	-0.0001	0.0000	0.0000	0.0000
1.5	0.0267	0.0230	-0.0333	0.0027	0.0480	0.0487
1.7	0.0461	0.0569	-0.0691	0.0354	0.1194	0.2514
1.9	0.0732	0.1253	-0.1281	0.1628	0.2582	1.0736
2.1	0.1092	0.2531	-0.2185	0.5413	0.5033	3.8955

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

Normal (0, 1)							Chi-square (3)						
β_0	n	$\hat{\beta}$	$\tilde{\beta}_{BC}$	$\hat{\beta}_{BC}$	s.e.($\hat{\beta}$)	s.d.($\hat{\beta}$)	β_0	n	$\hat{\beta}$	$\tilde{\beta}_{BC}$	$\hat{\beta}_{BC}$	s.e.($\hat{\beta}$)	s.d.($\hat{\beta}$)
- 1.7	25	-1.6939	-1.6971	-1.6970	0.2034	0.2040	- 0.5	25	-0.6472	-0.5728	-0.5944	0.3471	0.2576
	50	-1.6975	-1.6991	-1.6991	0.1048	0.1049		50	-0.5284	-0.5001	-0.5020	0.1644	0.1553
	100	-1.6995	-1.7002	-1.7002	0.0553	0.0553		100	-0.5136	-0.4999	-0.5004	0.1016	0.0990
- 1.1	25	-1.0832	-1.0939	-1.0936	0.1798	0.1796	- 0.3	25	-0.3304	-0.2973	-0.3002	0.1737	0.1607
	50	-1.0889	-1.0942	-1.0941	0.1030	0.1026		50	-0.3148	-0.2989	-0.2996	0.1032	0.0995
	100	-1.0949	-1.0975	-1.0974	0.0624	0.0622		100	-0.3082	-0.3004	-0.3006	0.0661	0.0648
1.1	25	1.0793	1.0900	1.0896	0.1799	0.1796	0.1	25	0.0956	0.1011	0.1010	0.0478	0.0475
	50	1.0966	1.1018	1.1018	0.1027	0.1026		50	0.0979	0.1006	0.1006	0.0296	0.0295
	100	1.0974	1.1000	1.1000	0.0623	0.0622		100	0.0984	0.0998	0.0998	0.0193	0.0192
2.0	25	1.9929	1.9944	1.9944	0.2414	0.2426	0.2	25	0.1968	0.1995	0.1994	0.1105	0.1232
	50	1.9990	1.9998	1.9998	0.1223	0.1224		50	0.1988	0.2001	0.2001	0.0599	0.0625
	100	1.9998	2.0002	2.0002	0.0623	0.0623		100	0.1991	0.1998	0.1998	0.0312	0.0321
Uniform (0, 1)													
β_0	n	$\hat{\beta}$	$\tilde{\beta}_{BC}$	$\hat{\beta}_{BC}$	s.e.($\hat{\beta}$)	s.d.($\hat{\beta}$)	β_0	n	$\hat{\beta}$	$\tilde{\beta}_{BC}$	$\hat{\beta}_{BC}$	s.e.($\hat{\beta}$)	s.d.($\hat{\beta}$)
-2.7	25	-3.0189	-2.7297	-2.7750	1.4414	1.2593	0.7	25	0.6812	0.7088	0.7084	0.2912	0.2889
	50	-2.8442	-2.7124	-2.7223	0.8129	0.7686		50	0.6858	0.6996	0.6995	0.1972	0.1961
	100	-2.7868	-2.7229	-2.7259	0.5105	0.4949		100	0.6875	0.6944	0.6943	0.1364	0.1357
-1.3	25	-1.3905	-1.2772	-1.2835	0.7127	0.6850	2.0	25	1.9927	2.0029	2.0029	0.1678	0.1673
	50	-1.3522	-1.2969	-1.2987	0.4500	0.4407		50	1.9940	1.9991	1.9991	0.1154	0.1151
	100	-1.3424	-1.3149	-1.3156	0.2997	0.2949		100	1.9948	1.9974	1.9974	0.0804	0.0803

References

- Brännäs, K. (1992). Finite sample properties of estimators and tests in Poisson regression models. *Journal of Statistical Computation and Simulation* 41: 229-241.
- Breslow, N. (1990). Tests of hypotheses in over-dispersed Poisson regression and other quasi-likelihood models. *Journal of the American Statistical Association* 85: 565-571.
- Cameron, A. C., Trivedi, P. K. (1986). Econometric models based on count data: comparisons and applications of some estimators and tests. *Journal of Applied Econometrics* 1: 29-53.
- Chen, Q., Giles, D. E. (2009). The finite sample properties of the maximum likelihood estimator for the binary logit model with random covariates. University of Victoria, Econometrics Working Paper EWP0906.
- Gourieroux, C., Monfort, A., Trognon, A. (1984). Pseudo maximum likelihood methods : Applications to Poisson models. *Econometrica* 52: 701-720.
- King, G. (1988). Statistical models for political science event counts: bias in conventional procedures and evidence for the exponential Poisson model. *American Journal of Political Science* 32: 838-863.
- R (2008) The R Project for Statistical Computing, <http://www.r-project.org>
- Rilstone, P., Srivatsava, V. K., 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, Y. P. (ed.) *Recent Advances in Statistical Methods*. Hackensack NJ: World Scientific.
- 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.
- Ullah, A., (2004). *Finite Sample Econometrics*. Oxford: Oxford University Press.