Economics Bulletin

Volume 31, Issue 4

Reducing the bias of the maximum likelihood estimator for the Poisson regression model

David E Giles
University of Victoria, Canada

Hui Feng
Department of Economics, Business and Mathematics, King''''s University College, UWO

Abstract

We derive expressions for the first-order bias of the MLE for a Poisson regression model and show how these can be used to adjust the estimator and reduce bias without increasing MSE. The analytic results are supported by Monte Carlo simulations and three illustrative empirical applications.

We are grateful to the Editor and an anonymous referee for helpful comments on an earlier version of this work.
Citation: David E Giles and Hui Feng, (2011) "Reducing the bias of the maximum likelihood estimator for the Poisson regression model", *Economics Bulletin*, Vol. 31 No. 4 pp. 2933-2943.
Contact: David E Giles - dgiles@uvic.ca, Hui Feng - hfeng8@uwo.ca.
Submitted: December 31, 2010. **Published:** October 18, 2011.

1. Introduction

The problem of modeling "count" data arises frequently in economics. These data are non-negative integers, so the linear regression model is discarded in favor of an appropriate discrete probability distribution and covariates are introduced through its mean. The simplest count model is based on the Poisson distribution, despite the limitations implied by the equivalence of its mean and variance. This model provides the starting point for other models (such as the negative binomial) that allow for over-dispersion, as well as models that allow for count "inflation", especially with respect to zero counts. So, the central role of the Poisson regression model means that its performance is of considerable interest.

The Poisson regression model is generally estimated by using the maximum likelihood estimator (MLE). It is well known that the likelihood function for this model is concave, so the MLE is unique. This estimator for the Poisson regression model possesses its usual desirable asymptotic properties, but surprisingly little is known about its finite-sample properties, *once covariates are introduced into the model*. Using Monte Carlo simulation experiments, for models with one covariate, Breslow (1990, p.568) reported biases in the range 1.2% to 1.9% when the sample size (n) = 36, 72; and Brännäs (1991, 234-235) reported biases in the range -2% to 1% when n = 50.

Recently, Chen and Giles (2011) derived analytic approximations for the bias and mean squared error (MSE) for the MLE of the Poisson regression model when the regressors are stochastic. However, their approach yields expressions that are quite unwieldy, and are not readily simplified to the case of non-random covariates. Our methodology is based on work by Cox and Snell (1968) and others, and is fundamentally different from that used by Chen and Giles (2011). We develop a simple analytic expression for the bias, to $O(n^{-1})$, of the MLE in the Poisson regression model with non-random covariates. The approach that we use, and this order of approximation, is standard in the statistics literature (*e.g.*, Giles, 2011; Schwartz *et al.*, 2011). It yields a simple and tractable expression for the bias, even though the MLE cannot be expressed in closed form. However, this approach is not suited to the situation of random regressors. We use the estimated bias to "bias-correct" the MLE, and find that dramatic reductions in bias can be achieved in small samples, without any increase in MSE. Our approach is "corrective", rather than "preventive". Firth (1993) provides an analytic approach to bias reduction of the latter type, but in practice little difference has been found in the performances of the two approaches.

Section 2 summarizes the methodology used to determine the finite-sample bias of the MLE, and this is applied to the Poisson regression model in section 3. Section 4 provides simulation evidence relating to the quality of the "bias-corrected" MLE. Although we are concerned here mainly with presenting a new theoretical result, its practical usefulness is also important. So, we present three illustrative empirical applications in section 5; and our conclusions are in section 6.

2. Bias Reduction

Let $l(\theta)$ be a log-likelihood function that is regular with respect to all derivatives up to and including the third order, and is based on a sample of n observations and a $(p \times 1)$ parameter vector, θ. The joint cumulants of the derivatives of $l(\theta)$, which are assumed to be $O(n)$, are:

$$k_{ij} = E(\partial^2 l / \partial\theta_i \partial\theta_j) \quad ; \quad i, j = 1, 2, \ldots, p \tag{1}$$

$$k_{ijl} = E(\partial^3 l / \partial\theta_i \partial\theta_j \partial\theta_l) \quad ; \quad i, j, l = 1, 2, \ldots, p \tag{2}$$

$$k_{ij,l} = E[(\partial^2 l / \partial\theta_i \partial\theta_j)(\partial l / \partial\theta_l)] \quad ; \quad i, j, l = 1, 2, \ldots, p \tag{3}$$

and denote:

$$k_{ij}^{(l)} = \partial k_{ij} / \partial\theta_l \quad ; \quad i, j, l = 1, 2, \ldots, p. \tag{4}$$

Cox and Snell (1968) showed that when the sample data are independent (but not necessarily identically distributed) the bias of the s^{th} element of the MLE of θ ($\hat{\theta}$) is:

$$Bias(\hat{\theta}_s) = \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{l=1}^{p} k^{si} k^{jl} [0.5 k_{ijl} + k_{ij,l}] + O(n^{-2}) \; ; \; s = 1, 2, \ldots, p \tag{5}$$

where k^{ij} is the $(i,j)^{th}$ element of the inverse of the information matrix, $K = \{-k_{ij}\}$. Using the usual "vec" notation to denote the stacking of the columns of a matrix, equation (5) can be written as:

$$Bias(\hat{\theta}) = K^{-1} A \, vec(K^{-1}) + O(n^{-2}), \tag{6}$$

where

$$A = [A^{(1)} \,|\, A^{(2)} \,|\, \ldots\ldots \,|\, A^{(p)}] \tag{7}$$

$$A^{(l)} = \{a_{ij}^{(l)}\} \; ; \quad i, j, l = 1, 2, \ldots, p \tag{8}$$

and

$$a_{ij}^{(l)} = k_{ij}^{(l)} - (k_{ijl} / 2), \text{ for } i, j, l = 1, 2, \ldots, p. \tag{9}$$

Equation (6) provides the bias, to $O(n^{-1})$, for the MLE. This is the standard order of magnitude for such evaluations. A "bias-corrected" MLE for θ can then be obtained as:

$$\tilde{\theta} = \hat{\theta} - \hat{K}^{-1} \hat{A} \, vec(\hat{K}^{-1}), \tag{10}$$

where $\hat{K} = (K)|_{\hat{\theta}}$ and $\hat{A} = (A)|_{\hat{\theta}}$. The advantage of the estimator $\tilde{\theta}$ is that it has a bias that is $O(n^{-2})$. This bias correction is valid for any n, but becomes redundant, of course as $n \to \infty$.

3. The Poisson Regression Model

The Poisson regression model assumes that the count data (y_i) follow the Poisson distribution:

$$\Pr.[Y = y_i \,|\, x_i] = e^{-\lambda_i} \lambda_i^{y_i} / y_i! \quad ; \quad y_i = 0, 1, 2, 3, \ldots\ldots$$

where

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

and x_i is a $(p \times 1)$ vector of covariates, x_i. The i^{th} observation on the vector of marginal effects is $\lambda_i \beta$, so these effects have the same sign(s) as the parameter(s).

With independent sampling, the log-likelihood function is

$$l = \sum_{i=1}^{n} [-\lambda_i + y_i x_i' \beta - \log(y_i)],$$

the likelihood equations are

$$(\partial l / \partial \beta) = \sum_{i=1}^{n} (y_i - \lambda_i) x_i = 0, \qquad (11)$$

and

$$(\partial^2 l / \partial \beta \partial \beta') = -\sum_{i=1}^{n} \lambda_i x_i x_i',$$

with typical element

$$(\partial^2 l / \partial \beta_j \partial \beta_l) = -\sum_{i=1}^{n} \lambda_i x_{ij} x_{il} \quad ; \quad j, l = 1, 2, \ldots\ldots, p. \qquad (12)$$

As (12) does not involve the y data,

$$k_{jl} = E(\partial^2 l / \partial \beta_j \partial \beta_l) = -\sum_{i=1}^{n} \lambda_i x_{ij} x_{il} \quad ; \quad j, l = 1, 2, \ldots\ldots, p. \qquad (13)$$

and the information matrix is

$$K = \sum_{i=1}^{n} \lambda_i x_i x_i'. \qquad (14)$$

There is no closed-form solution to (11), so the MLE for β must be obtained numerically. However, as the Hessian is negative definite for all x and β, the MLE ($\hat{\beta}$) is unique. From (12) and (13):

$$k_{jlr} = E(\partial^3 l / \partial \beta_j \partial \beta_l \partial \beta_r) = -\sum_{i=1}^{n} \lambda_i x_{ij} x_{il} x_{ir} \qquad (15)$$

and

$$k_{jl}^{(r)} = (\partial k_{jl} / \partial \beta_r) = -\sum_{i=1}^{n} \lambda_i x_{ij} x_{il} x_{ir} \; ; \quad j, l, r = 1, 2, \ldots\ldots, p. \qquad (16)$$

To make matters more transparent, consider the case of a single covariate and an intercept. Then x_i is a scalar observation and

$$l = \sum_{i=1}^{n} [-\lambda_i + y_i(\beta_1 + \beta_2 x_i) - \log(y_i)],$$

where $\lambda_i = \exp(\beta_1 + \beta_2 x_i)$, for $i = 1, 2, \ldots, n$.

From (9) and (14) – (16),

$$K = \begin{bmatrix} \sum_{i=1}^{n} \lambda_i & \sum_{i=1}^{n} x_i \lambda_i \\ \sum_{i=1}^{n} x_i \lambda_i & \sum_{i=1}^{n} x_i^2 \lambda_i \end{bmatrix},$$

$$k_{11}^{(1)} = k_{111} = -\sum_{i=1}^{n} \lambda_i$$

$$k_{12}^{(1)} = k_{21}^{(1)} = k_{11}^{(2)} = k_{121} = k_{211} = k_{112} = -\sum_{i=1}^{n} x_i \lambda_i$$

$$k_{22}^{(1)} = k_{12}^{(2)} = k_{21}^{(2)} = k_{221} = k_{212} = k_{122} = -\sum_{i=1}^{n} x_i^2 \lambda_i$$

$$k_{22}^{(2)} = k_{222} = -\sum_{i=1}^{n} x_i^3 \lambda_i$$

$$a_{11}^{(1)} = -0.5\sum_{i=1}^{n} \lambda_i$$

$$a_{12}^{(1)} = a_{11}^{(2)} = -0.5\sum_{i=1}^{n} x_i \lambda_i$$

$$a_{22}^{(1)} = a_{12}^{(2)} = -0.5\sum_{i=1}^{n} x_i^2 \lambda_i$$

$$a_{22}^{(2)} = -0.5\sum_{i=1}^{n} x_i^3 \lambda_i$$

and

$$Bias(\hat{\beta}_1) = \left[\frac{(\sum_{i=1}^{n} x_i\lambda_i)^2(\sum_{i=1}^{n} x_i^2\lambda_i) + n\bar{\lambda}(\sum_{i=1}^{n} x_i\lambda_i)(\sum_{i=1}^{n} x_i^3\lambda_i) - 2n\bar{\lambda}(\sum_{i=1}^{n} x_i^2\lambda_i)^2}{2[n\bar{\lambda}(\sum_{i=1}^{n} x_i^2\lambda_i) - (\sum_{i=1}^{n} x_i\lambda_i)^2]^2} \right] + O(n^{-2}) \qquad (17)$$

$$Bias(\hat{\beta}_2) = \left[\frac{3n\bar{\lambda}(\sum_{i=1}^{n} x_i\lambda_i)(\sum_{i=1}^{n} x_i^2\lambda_i) - 2(\sum_{i=1}^{n} x_i\lambda_i)^3 - (n\bar{\lambda})^2(\sum_{i=1}^{n} x_i^3\lambda_i)}{2[n\bar{\lambda}(\sum_{i=1}^{n} x_i^2\lambda_i) - (\sum_{i=1}^{n} x_i\lambda_i)^2]^2} \right] + O(n^{-2}) \quad . \qquad (18)$$

where $\bar{\lambda} = \frac{1}{n}\sum_{i=1}^{n} \lambda_i$.

These biases may be positive or negative, depending on the sample data and the true parameter values. They apply for n of any magnitude, and vanish as $n \to \infty$. Bias-corrected MLEs are

$$\tilde{\beta}_s = \hat{\beta}_s - Bi\hat{a}s(\hat{\beta}_s) \qquad ; \qquad s = 1, 2$$

where $Bi\hat{a}s(\hat{\beta}_s)$ is obtained by replacing λ_i by $\hat{\lambda}_i = \exp(\hat{\beta}_1 + \hat{\beta}_2 x_i)$ in (17) and (18).

Extending the above discussion to the case where the conditional mean is a function of two covariates and an intercept,

$$l = \sum_{i=1}^{n} [-\lambda_i + y_i(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i}) - \log(y_i)],$$

where $\lambda_i = \exp(\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i})$, for i = 1, 2, …, n. Then:

$$K = \begin{bmatrix} \sum_{i=1}^{n}\lambda_i & \sum_{i=1}^{n}\lambda_i x_{2i} & \sum_{i=1}^{n}\lambda_i x_{3i} \\ \sum_{i=1}^{n}\lambda_i x_{2i} & \sum_{i=1}^{n}\lambda_i x_{2i}^2 & \sum_{i=1}^{n}\lambda_i x_{2i}x_{3i} \\ \sum_{i=1}^{n}\lambda_i x_{3i} & \sum_{i=1}^{n}\lambda_i x_{2i}x_{3i} & \sum_{i=1}^{n}\lambda_i x_{3i}^2 \end{bmatrix}; \qquad A^{(1)} = -0.5K$$

$$A^{(2)} = -0.5\begin{bmatrix} \sum_{i=1}^{n}\lambda_i x_{2i} & \sum_{i=1}^{n}\lambda_i x_{2i}^2 & \sum_{i=1}^{n}\lambda_i x_{2i}x_{3i} \\ \sum_{i=1}^{n}\lambda_i x_{2i}^2 & \sum_{i=1}^{n}\lambda_i x_{2i}^3 & \sum_{i=1}^{n}\lambda_i x_{2i}^2 x_{3i} \\ \sum_{i=1}^{n}\lambda_i x_{2i}x_{3i} & \sum_{i=1}^{n}\lambda_i x_{2i}^2 x_{3i} & \sum_{i=1}^{n}\lambda_i x_{2i}x_{3i}^2 \end{bmatrix}; \qquad A^{(3)} = -0.5\begin{bmatrix} \sum_{i=1}^{n}\lambda_i x_{3i} & \sum_{i=1}^{n}\lambda_i x_{2i}x_{3i} & \sum_{i=1}^{n}\lambda_i x_{3i}^2 \\ \sum_{i=1}^{n}\lambda_i x_{2i}x_{3i} & \sum_{i=1}^{n}\lambda_i x_{2i}^2 x_{3i} & \sum_{i=1}^{n}\lambda_i x_{2i}x_{3i}^2 \\ \sum_{i=1}^{n}\lambda_i x_{3i}^2 & \sum_{i=1}^{n}\lambda_i x_{2i}x_{3i}^2 & \sum_{i=1}^{n}\lambda_i x_{3i}^3 \end{bmatrix}.$$

The biases of the MLEs, and the bias-adjusted estimators, follow by inserting these expressions into equations (6), (7) and (10). Again, the directions of the biases are data-dependent.

4. A Simulation Experiment

We have undertaken a Monte Carlo experiment to evaluate the effectiveness of our bias corrections. Recalling that our bias expressions are valid to $O(n^{-1})$, the exact biases and MSEs of the MLEs and bias-corrected MLEs have been simulated using code written for the *R* statistical software environment (R 2008). The log-likelihood function was maximized using the Newton-Raphson method in the *maxLik* package (Toomet and Henningsen 2008). Each part of our experiment uses 100,000 Monte Carlo replications. In the two-covariate case we consider various degrees of correlation (ρ) between the regressors. The results in Tables I and II are for regressors that are standard normal, but fixed in repeated samples. Similar results were obtained for (fixed) regressors distributed uniformly on [0 , 1], and these are available on request. Tables I and II report *percentage* biases and MSEs, defined as $100\times(\text{Bias} / |\beta_s|)$ and $100\times(\text{MSE} / \beta_s^2)$.

The magnitudes of the reported biases for the (uncorrected) MLEs are consistent with those reported by other authors in their simulation experiments, as discussed in section 1. They are quite small, except for very small sample sizes. The effectiveness of our bias correction is clear in all of the cases tabulated. The percentage biases themselves are substantially reduced by applying this correction – often by one or two orders of magnitude. For example, in Table I(b) when n = 25, the percentage bias of the original MLE for the intercept coefficient is -1.5%, and this is reduced to -0.005% by bias-correcting. Not surprisingly, the reduction in bias comes at the expense of some increased variability. However, the percentage MSEs are also either slightly reduced or essentially unaltered by bias-adjusting the estimators. Taking the previous example from Table I(b), the %MSE falls marginally, from 1.7% for the original MLE to 1.6% for the bias-corrected MLE. It is also clear from these results that the known *asymptotic* unbiasedness of the MLE actually reveals itself quite quickly – namely, by the time that the sample size reaches 200. So there is no point in considering larger sample sizes here. This information is useful to a practitioner who is estimating a Poisson regression model with a modest sample size of, say, 500 observations. Reliable maximum likelihood estimation of the Poisson regression model does not require thousands of cross-section observations.

Table I: Percentage biases and MSEs of the MLEs and bias-corrected MLEs – intercept and one regressor

n	$\%Bias(\hat{\beta}_1)$ [$\%MSE(\hat{\beta}_1)$]	$\%Bias(\tilde{\beta}_1)$ [$\%MSE(\tilde{\beta}_1)$]	$\%Bias(\hat{\beta}_2)$ [$\%MSE(\hat{\beta}_2)$]	$\%Bias(\tilde{\beta}_2)$ [$\%MSE(\tilde{\beta}_2)$]
		(a) $\beta_1 = 1, \beta_2 = 0.5$		
10	-3.8408	-0.0703	2.3409	0.1265
	[6.0244]	[5.5523]	[10.8055]	[10.5440]
25	-1.4891	-0.0514	0.8831	0.1469
	[1.9747]	[1.9061]	[4.7102]	[4.6645]
50	-0.7269	-0.0053	0.3854	0.0107
	[0.9268]	[0.9104]	[2.8139]	[2.8042]
100	-0.2792	0.0035	-0.0223	0.0016
	[0.4382]	[0.4360]	[1.0186]	[1.0199]
200	-0.1312	0.0113	-0.0385	-0.0171
	[0.2137]	[0.2132]	[0.5105]	[0.5106]
		(b) $\beta_1 = 1, \beta_2 = -0.5$		
10	-4.1469	0.0538	-0.9386	0.0582
	[4.8242]	[4.2403]	[10.9303]	[10.3340]
25	-1.5336	-0.0046	-0.2357	0.1412
	[1.6930]	[1.6184]	[4.7186]	[4.6419]
50	-0.7079	0.0161	-0.0695	0.0106
	[0.8140]	[0.7977]	[2.4986]	[2.4819]
100	-0.3646	-0.0007	-0.0301	0.0325
	[0.4023]	[0.3982]	[1.2859]	[1.2832]
200	-0.1764	0.0029	-0.0546	-0.0185
	[0.2003]	[0.1993]	[0.6744]	[0.6739]

5. Empirical Applications

Although our primary objective in this paper is the derivation and evaluation of a simple bias correction formula for the MLE in the context of the Poisson regression model, we present results here that illustrate its usefulness in practice. These applications are not intended to be "full blown" empirical studies of the phenomena in question. Rather, they are intended to demonstrate the extent to which estimates based on actual data may change as a result of implementing our bias correction for this particular MLE.

Table II: Percentage biases and MSEs of the MLEs and bias-corrected MLEs – intercept and two regressors

n	$\%Bias(\hat{\beta}_1)$ [$\%MSE(\hat{\beta}_1)$]	$\%Bias(\tilde{\beta}_1)$ [$\%MSE(\tilde{\beta}_1)$]	$\%Bias(\hat{\beta}_2)$ [$\%MSE(\hat{\beta}_2)$]	$\%Bias(\tilde{\beta}_2)$ [$\%MSE(\tilde{\beta}_2)$]	$\%Bias(\hat{\beta}_3)$ [$\%MSE(\hat{\beta}_3)$]	$\%Bias(\tilde{\beta}_3)$ [$\%MSE(\tilde{\beta}_3)$]
	(a) $\beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5, \rho = 0.1$					
10	-5.4984	-0.0506	-1.8594	0.0248	3.7403	0.1129
	[7.6977]	[6.8300]	[5.7465]	[5.5245]	[17.5002]	[16.6353]
25	-1.9800	-0.0695	1.4561	0.0965	0.1200	0.0327
	[2.3178]	[2.1990]	[5.6222]	[5.4589]	[2.8013]	[2.7645]
50	-0.9785	-0.0244	0.3678	0.0475	0.3463	-0.0076
	[1.0203]	[0.9980]	[1.9837]	[1.9766]	[2.6724]	[2.6769]
100	-0.3575	-0.0041	0.0408	-0.0008	-0.0420	0.0040
	[0.4692]	[0.4660]	[1.0499]	[1.0491]	[0.6926]	[0.6921]
200	-0.1705	0.0053	-0.0589	-0.0024	0.0075	-0.0214
	[0.2221	[0.2213]	[0.5039]	[0.5052]	[0.4687]	[0.4690]
	(b) $\beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0.5, \rho = 0.9$					
10	-4.6490	0.0068	-0.5608	-0.2005	3.4684	0.2322
	[7.6072]	[6.9995]	[29.5010]	[25.8512]	[56.3052]	[54.6193]
25	-1.8510	-0.0604	2.5693	0.1463	-1.1418	-0.0510
	[2.4696]	[2.3565]	[18.5975]	[18.1754]	[11.1513]	[11.0251]
50	-0.8136	-0.0023	0.2679	0.0137	0.1354	-0.0374
	[1.0489]	[1.0335]	[8.1901]	[8.1873]	[9.5613]	[9.5766]
100	-0.2673	0.0191	0.0848	-0.0229	-0.0794	0.0002
	[0.4627]	[0.4605]	[4.2837]	[4.2821]	[3.1600]	[3.1581]
200	-0.1407	-0.0038	-0.0563	0.0480	0.0545	-0.0401
	[0.2170]	[0.2165]	[2.2196]	[2.2208]	[2.1409]	[2.1421]

5.1 Supreme Court Judge Appointments

We begin with a simple example that uses data on the number of U.S. Supreme Court judges appointed by U.S. Presidents, as reported by Voinov *et al.* (2010, Table 2). We use data only for those Presidents representing the Democrat or Republican parties, resulting in 38 observations. The independence of the sample observations is confirmed by tests conducted by Voinov *et al.* (2010), who also show that a Poisson process cannot be rejected. The characteristics of our sample are as follows, and it can be seen that the sample data are equi-dispersed:

JUDGES:	0	1	2	3	4	5	6	Mean: 2.0
Frequency:	6	7	14	5	4	1	1	Variance: 2.1

We have estimated a very simple Poisson regression that explains the number of appointments in terms of a constant a dummy variable, REP, which is unity if the President was a Republican, and zero if he was a Democrat. The results appear below (z-statistics are in parentheses; bias-adjusted estimates are in **bold**):

$$E[\text{JUDGES}] = \underset{\substack{(3.03)\\ \mathbf{0.5597}}}{0.5436} + \underset{\substack{(1.24)\\ \mathbf{0.2840}}}{0.2893}\,\text{REP} \; ; \; R^2 = 0.0411$$

We see that there is no significant Republican-Democrat effect in the appointment of Supreme Court judges; and in this example the effect of bias-correcting the estimated coefficients is negligible, notwithstanding the modest sample size.

5.2 Atlantic Hurricanes

Modeling and predicting the occurrence of hurricanes is of considerable economic importance. For example, in the case of Hurricane Katrina in 2005, the cost of property damage alone was estimated at $81 billion. The possible role of an El Niño effect on hurricane frequency has received considerable attention. It is believed that warm El Niño events are associated with an increase in the number of tropical storms and hurricanes in the eastern Pacific Ocean, and a decrease in the Atlantic Ocean, Gulf of Mexico and the Caribbean Sea. We have estimated a Poisson regression model for the number of North Atlantic hurricanes in the years 1990 to 2010. The data (Unisys 2011) have the following characteristics:

HURRICANES:	3	4	5	6	7	8	9	10	11	12	15	Mean: 7.1
Frequency:	3	4	1	1	1	4	3	1	1	1	1	Variance: 10.9

The estimated Poisson regression model (with z-statistics in parentheses, and bias-adjusted estimates in **bold**) is:

$$E[\text{HURRICANES}] = \underset{\substack{(24.42)\\ \mathbf{2.1215}}}{2.1178} - \underset{\substack{(-2.23)\\ \mathbf{-0.7973}}}{0.8650}\,\text{DMOD} - \underset{\substack{(-2.79)\\ \mathbf{-0.8675}}}{0.9138}\,\text{DSTRONG} \; ; \; R^2 = 0.4208$$

DMOD and DSTRONG are dummy variables with the value unity if the year was characterized by a moderate or strong El Niño effect (Null 2011) respectively, and zero otherwise. The results exhibit the anticipated negative El Niño effect, with the relative magnitudes of the dummy variable coefficients also being as expected. Bias-correcting these two estimated coefficients increases their values – by 7.8% and 5.1%, respectively. That is, the estimated impact of the El Niño effect is reduced somewhat.

5.3 Banking Crises

Finally, we apply our bias correction to a model for the number of banking crises in a sample of 32 IMF-member countries over the period 1970 to 1999. The data for the banking crises are from Ghosh *et al*. (2002). From the raw data we have constructed a data-set for the number of such crises, and other indicators, for each country. This is available on request. The variables used are the number of banking crises (BCRISES); the number of currency crises

(CCRISES); a dummy variable (DPEG) which is unity if there were one or more banking crises under a (*de jure*) pegged exchange rate regime; and a dummy variable (DINCHI) which is unity if the observation is for an upper or upper-middle income country. The characteristics of the sample are:

BCRISES:	0	1	2	3	Mean: 0.9
Frequency:	13	11	7	1	Variance: 0.8

The Poisson regression results appear below (z-statistics are in parentheses; bias-adjusted estimates are in **bold**):

$$E[\text{BCRISES}] = -0.7659 + 0.2561[\text{CCRISES}\times\text{DINCHI}] + 0.8727\ \text{DPEG}\ ;\ \ R^2 = 0.4454$$

	(-2.45)	(2.00)	(2.15)
	-0.6831	**0.2875**	**0.7616**

We see that banking crises are significantly more prevalent under pegged exchange rates than under floating rates. (This also holds for all 167 IMF-member countries, in contrast to the descriptive results of Ghosh *et al.*, 2002, p.169.) The marginal effect for DPEG dummy (averaged over the sample) is 0.788, from the MLE results, and 0.725 using the bias-corrected estimate. In this example, the bias adjustments modify the point estimates of the coefficients by 10.8%, 12.3% and -12.7% respectively; and the marginal effect for the exchange rate dummy by -8.0%. We do not report results for the full sample of 167 observations, as for that sample size the bias correction effects are negligible (as would be anticipated from the results in section 4).

6. Conclusions

We have derived an analytic expression for the first-order bias of the MLE in the Poisson regression model. Almost-unbiased MLEs for the coefficients can then be constructed by subtracting the estimated biases from the original MLEs. We have presented Monte Carlo evidence that shows that this bias correction can result in substantial reductions in bias in small samples, and although it increases the variability of the estimators, the MSE is not adversely affected. We have also shown that there is no need to bias-correct the MLE for the Poisson regression model with samples of size 200 or more.

This demonstrated robustness of the asymptotic properties of this MLE, to reductions in the sample size to quite modest levels, should be of considerable comfort to practitioners. The illustrative empirical examples that we have provided also show the extent to which point estimates of the parameters of a Poisson regression model may change as a result of using our bias correction in small samples.

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.

Chen, Q. and D. E. Giles, (2011) “Finite-Sample Properties of the Maximum Likelihood Estimator for the Poisson Regression Model with Random Covariates” *Communications in Statistics – Theory and Methods* **40**, 1000-1014.

Cox, D. R. and E. J. Snell, (1968) “A General Definition of Residuals” *Journal of the Royal Statistical Society*, *B* **30**, 248-275.

Firth, D., (1993) “Bias Reduction of Maximum Likelihood Estimates” *Biometrika* **80**, 27-38.

Ghosh, A. R., A-M. Gulde and H. C. Wolf, (2002) *Exchange Rate Regimes: Choices and Consequences*, MIT Press, Cambridge MA.

Giles, D. E., (2011) **“**Bias Reduction for the Maximum Likelihood Estimator of the Parameters in the Half-Logistic Distribution” *Communications in Statistics – Theory & Methods*, in press.

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.

Null, J. (2011) “El Niño and La Niña Years and Intensities Based on Oceanic Niño Index (ONI)” http://ggweather.com/enso/oni.htm

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

Schwartz, J., R. T. Godwin and D. E. Giles, (2011) “Improved Maximum Likelihood Estimation of the Shape Parameter in the Nakagami Distribution” *Journal of Statistical Computation and Simulation*, in press.

Toomet, O. and A. Henningsen, (2008) “maxLik: Maximum Likelihood Estimation” http://CRAN.R-project.org and http://maxLik.org

Unisys (2011) “Atlantic Tropical Storm Tracking by Year” http://weather.unisys.com/hurricane/atlantic/

Voinov, V., E. Voinov and R. Rakhimova, (2010) “Poisson Versus Binomial: Appointment of Judges to the U.S. Supreme Court” *Australian and New Zealand Journal of Statistics*, **52**, 261-274.