University of Victoria
Department of Economics

Econometrics Working Paper EWP0909

ISSN 1485-6441

Almost Unbiased Estimation of the Poisson Regression Model

David E. Giles
Department of Economics, University of Victoria
Victoria, BC, Canada V8W 2Y2

&

Hui Feng
Department of Economics, Business & Mathematics
King's University College, University of Western Ontario
London, ON, Canada N6A 2N3

December, 2009

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 an empirical application.

Keywords Poisson regression; maximum likelihood estimation; bias reduction

JEL Classifications C13; C25

Author Contact:
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; Phone: (250) 721-8540; FAX: (250) 721-6214

1. Introduction

The problem of modeling "count" data arises frequently in economics. The 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 and most widely used count model is based on the Poisson distribution, despite the limitations implied by the equivalence of its mean and variance. Cameron and Trivedi (1998) and Winkelman (2000) provide excellent discussions of modeling count data.

Although the maximum likelihood estimator (MLE) possesses its usual desirable asymptotic properties for such models, 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) 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. Recently, Chen and Giles (2009) derived analytic approximations for the bias and mean squared error (MSE) for the MLE in this context when the regressors are stochastic. However, their approach yields expressions that are quite unwieldy, and are not readily simplifed to the case of non-random covariates. Here, 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. We then use the estimated bias to "bias-correct" the MLE. The methodology is based on work by Cox and Snell (1968) and others, and is fundamentally different from that used by Chen and Giles (2009). We find that dramatic reductions in bias can be achieved in small samples, without any increase in MSE.

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, and an empirical example is given in section 5.

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:

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

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:

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

where k^{ij} is the $(i,j)^{\text{th}}$ element of the inverse of the information matrix, $K = \{-k_{ij}\}$. This result extends earlier work by Bartlett (1953), Haldane and Smith (1956), Shenton and Bowman (1963) and others. Equation (5) also holds for non-independent data, and can be written as:

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

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

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

A "bias-corrected" MLE for θ is:

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

where $\hat{K} = (K)|_{\hat{\theta}}$ and $\hat{A} = (A)|_{\hat{\theta}}$. The estimator $\tilde{\theta}$ is "almost unbiased" - its bias is $O(n^{-2})$.

3. The Poisson regression model

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

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

where

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

and x_i is a $(p \times 1)$ vector of covariates, x_i.

Assuming independent sampling, the log-likelihood is

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

the likelihood equations are

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

and

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

with typical element

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

As (12) does not involve the y data,

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

and the information matrix is

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

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, if it exists. From (12) and (13):

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

and

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

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

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

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

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. 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). If 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:

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

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

The biases of the MLEs, and the bias-adjusted estimators, follow from (6) and (10).

4. Numerical evaluations

We report the results of a Monte Carlo experiment that investigates the usefulness of our bias corrections, which are valid only to $O(n^{-1})$. The actual 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, and we limit attention to the univariate and bivariate (with intercept) models. In the two-covariate case we consider various degrees of correlation (ρ) between the regressors. The results in Tables 1 to 3 are *percentage* biases, defined as $100\times$(Bias / | β_s |), and percentage MSEs, defined as $100\times$(MSE / β_s^2).

The magnitudes of the reported biases for the (uncorrected) MLEs are consistent with those reported by other authors in 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 – often by one or two orders of magnitude.

Although this gain comes at the expense of increased variability, the percentage MSEs are either slightly reduced or essentially unaltered by bias-adjusting the estimators.

5. Empirical application

We present a simple illustration of the application of our proposed bias correction by modeling the number of banking crises in a sub-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 sample characteristics and the Poisson regression results are in Table 4. Banking crises are significantly more prevalent under pegged exchange rates than under floating rates. (This also holds for the full sample of 167 countries, in contrast to the descriptive results of Ghosh *et al.*, 2002, p.169.) Here, the bias adjustments modify the point estimates of the coefficients by 10.8%, 12.3% and -12.7% respectively.

6. Conclusions

We have derived an analytic expression for the first-order bias of the MLE in a Poisson regression model. Almost-unbiased MLEs for the coefficients are then constructed by subtracting the estimated biases from the original MLEs. Monte Carlo evidence shows that this results in dramatic reductions in bias in small samples, and although it increases the variability of the estimators, the MSE is not adversely affected. This analytic bias correction is recommended for the Poisson regression model with a sample of size 200 or less.

Table 1: 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) Standard normal regressor: $\boldsymbol{\beta_1 = 1, \beta_2 = 0.5}$				
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) Standard normal regressor: $\boldsymbol{\beta_1 = 1, \beta_2 = -0.5}$				
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]

Table 2: Percentage biases and MSEs of the MLEs and bias-corrected MLEs – intercept and one regressor

n	$\%Bias(\hat{\beta}_1)$ [$\%MSE(\hat{\beta}_1)$]	$\%Bias(\tilde{\beta}_1)$ [$\%MSE(\tilde{\beta}_1)$]	$\%Bias(\hat{\beta}_2)$ [$\%MSE(\tilde{\beta}_2)$]	$\%Bias(\tilde{\beta}_2)$ [$\%MSE(\breve{\beta}_2)$]
	(a) Uniform (0,1) regressor: $\boldsymbol{\beta_1}$ = 1, $\boldsymbol{\beta_2}$ = 0.5			
10	-2.0161	0.1518	-4.2240	-0.4868
	[13.3546]	[12.6881]	[171.7988]	[162.4338]
25	-1.2173	-0.0122	0.1425	0.0183
	[4.5360]	[4.4315]	[49.1741]	[48.2918]
50	-0.6460	0.0809	0.3245	-0.2397
	[2.5416]	[2.5093]	[22.4297]	[22.2247]
100	-0.2867	0.0601	0.0217	-0.2093
	[1.2957]	[1.2884]	[11.9973]	[11.9481]
200	-0.1835	0.0009	0.1795	0.0252
	[0.6854]	[0.6834]	[6.1290]	[6.1167]
	(b) Uniform (0,1) regressor: $\boldsymbol{\beta_1}$ = 1, $\boldsymbol{\beta_2}$ = -0.5			
10	-1.4781	0.2296	-16.8178	-0.6932
	[19.2841]	[17.9263]	[345.0662]	[304.6385]
25	-1.2899	0.0220	-2.9498	-0.1476
	[5.7814]	[5.6152]	[85.8775]	[83.0689]
50	-0.7636	0.0752	-0.8159	-0.2868
	[3.2593]	[3.2054]	[37.9555]	[37.3819]
100	-0.3379	0.0706	-0.5556	-0.2675
	[1.6804]	[1.6679]	[20.2826]	[20.1425]
200	-0.2300	-0.0038	-0.0002	0.0557
	[0.8900]	[0.8864]	[10.2080]	[10.1758]

Table 3: 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) Standard normal regressors: β_1 = 1, β_2 = 0.5, β_3 = 0.5, ρ = 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) Standard normal regressors: β_1 = 1, β_2 = 0.5, β_3 = 0.5, ρ = 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]

Table 4: Banking crisis application

(a) Data characteristics

BCRISES:	0	1	2	3	Mean: 0.875; variance: 0.760
Frequency:	13	11	7	1	

(b) Poisson regression results (z-statistics in parentheses; bias-adjusted estimates in **bold**)

E[BCRISES] = -0.7659 + 0.2561[CCRISES×DINCHI] + 0.8727 DPEG ; R^2 = 0.4454

(-2.45) (2.00) (2.15)

-0.6831 **0.2875** **0.7616**

References

Bartlett, M. S., 1953, Approximate confidence intervals II. More than one unknown parameter, Biometrika 40, 306-317.

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. and P. K. Trivedi, 1998, Regression Analysis of Count data, Cambridge University Press, Cambridge.

Chen, Q. and D. E. Giles, 2009, Finite-sample properties of the maximum likelihood estimator for the Poisson regression model with random covariates, Communications in Statistics – Theory and Methods, in press.

Cox, D. R. and E. J. Snell, 1968, A general definition of residuals, Journal of the Royal Statistical Society, B 30, 248-275.

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

Haldane, J. B. S. and S. M. Smith, 1956, The sampling distribution of a maximum likelihood estimate, Biometrika 43, 96-103.

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

Shenton, L. R. and K. Bowman, 1963, Higher moments of a maximum-likelihood estimate, Journal of the Royal Statistical Society, B 25, 305-317.

Toomet, O. and A. Henningsen, 2008, maxLik: Maximum likelihood estimation, http://CRAN.R-project.org ; http://maxLik.org

Winkelman, R., 2000, Econometric Analysis of Count Data, 3rd ed., Springer, Berlin