Bond University

Research Repository

Bayesian truncated poisson regression with application to dutch illegal immigrant data

Puza, Borek D.; Johnson, Helen L.; O'Neill, Terence J.; Barry, Simon C.

[Link to output in Bond University research repository.](#)

Bayesian Truncated Poisson Regression With Application to Dutch Illegal Immigrant Data

BOREK D. PUZA, HELEN L. JOHNSON, AND TERENCE J. O'NEILL (22 OCTOBER 2007)

This paper presents a Bayesian approach to the regression analysis of truncated data, with a focus on zero-truncated counts from the Poisson distribution. The approach provides inference not only on the regression coefficients but also on the total sample size and the parameters of the covariate distribution. The theory is applied to some illegal immigrant data from The Netherlands. Several models are fitted with the aid of Markov chain Monte Carlo methods and assessed via posterior predictive p-values. Inferences are compared with those obtained elsewhere using other approaches.

KEY WORDS: Markov chain Monte Carlo methods; Metropolis Hastings algorithm; Posterior predictive p-value.

Borek D. Puza is a Senior Lecturer in Statistics in the School of Finance and Applied Statistics, Australian National University, Canberra, ACT 0200, Australia (email: borek.puza@anu.edu.au). Helen L. Johnson is a Lecturer in Statistics in the School of Mathematical Sciences, Queensland University of Technology, Brisbane, GPO Box 2434, Brisbane 4001, QLD, Australia (email: h.johnson@qut.edu.au). Terence J. O'Neill is a Professor in Statistics and Head of School in the School of Finance and Applied Statistics, Australian National University, Canberra, ACT 0200, Australia (email: terry.oneill@anu.edu.au).

1. INTRODUCTION

In statistics, *truncation* occurs when only those values which lie in a certain region are observed. This phenomenon is related to but differs from *censoring*, whereby particular sample values are known only to lie in a certain region. Thus, under censoring the number of unobserved values is known, whereas under truncation that number is unknown. Truncation and censoring may both be thought of as examples of non-ignorable non-response or more generally as examples of biased sampling.

This paper focuses on the regression analysis of data from a truncated Poisson distribution where the response is observed if and only if it is non-zero. This is an example of truncation where each unobserved response is exactly known (i.e. zero). This situation may be contrasted with truncated Gaussian regression where each unobserved response is known only to lie in a certain region.

Previously, the analysis of truncated count data has largely been based on maximum likelihood (ML) methods. Dahiya and Gross (1973) examined ML estimation of the mean of a Poisson distribution truncated at zero, as well as the total sample size (in this case the number of observed values plus the number of zeros). Blumenthal, Dahiya, and Gross (1978) extended these results with the aim of obtaining an improved estimator of the total sample size and investigated the asymptotic properties of that estimator. Later, Scollnik (1997) addressed the same inferential issues from a Bayesian perspective with the aid of Markov chain Monte Carlo (MCMC) methods. However, none of these three papers considered covariate information. Recently, Ibrahim, Chen and Lipsitz (2002) proposed Bayesian generalized linear models for analysing data with missing covariates but did not consider the issue of truncation.

The regression analysis of truncated data has been the subject of research in a number of contexts. Amemiya (1984) studied a form of the Tobit model for the regression analysis of truncated Gaussian responses. Also, Shaw (1988) and Grogger and Carson (1991) considered several regression models for truncated data from the Poisson and negative binomial distributions. More recently, O'Neill and Barry (1995a) proposed a truncated model for grouped binary data and extended it to grouped ordinal data (O'Neill and Barry, 1995b).

Two recent papers which deal specifically with the regression analysis of zero-truncated Poisson data are Van der Heijden, Cruyff and Van Houwelingen (2003) and Van der Heijden, Bustami, Cruyff, Engbersen, and Van Houwelingen (2003). The former paper proposes the use of a Horvitz-Thompson estimator for the total sample size and thereby calculates estimates of two offender populations in The Netherlands: drunk drivers and persons who illegally possess a firearm. The latter paper presents the theory used in the former paper in more detail and applies it to a third data set so as to estimate the number of illegal immigrants in The Netherlands during 1995.

In this paper we propose a Bayesian alternative to the Horvitz-Thompson approach in the two papers by Van der Heijden et al. Arguably, the Bayesian approach is highly suitable for analysing the type of truncated data in those two papers. It provides a convenient platform for inference not only on the regression coefficients and total sample size but also on several other quantities, such as the parameters of the covariate distribution itself. The covariate distribution is not a focus of any inference in Shaw (1988) or Grogger and Carson (1991) nor in the two papers by Van der Heijden et al. It should also be noted that the theory in the latter two papers does not lead to the maximum likelihood estimates (MLEs) of any model parameters, only to

approximations thereof. Moreover, that theory leads to confidence intervals (CIs) for the total sample size which are based on an asymptotic normal approximation to the essentially skewed distribution of the authors' proposed Horvitz-Thompson estimator of that quantity.

Most importantly, in each of the models considered in Van der Heijden, Bustami et al. (2003) there remains significant unobserved heterogeneity. One primary goal of the present paper is to reanalyse the data in that paper, to find a model which fits them satisfactorily - so that there remains no heterogeneity - and hence to derive convincing point and interval estimates of the total number of Dutch illegal immigrants in 1995. We believe that this goal has been achieved.

In Section 2 we review the ML approach to the analysis of truncated data and point out some problems with that approach. The Bayesian approach is discussed in Section 3, and Sections 4 and 5 narrow down the class of models to be considered to the case of a categorical covariate distribution, a Poisson response with truncation at zero, and a priori ignorance regarding all model parameters. In Section 6 the Bayesian theory is applied to the Dutch illegal immigrant data in Van der Heijden, Bustami et al. (2003), and inferences are compared with the results obtained in that paper.

The Bayesian model used in Section 6 is assessed in Section 7 using posterior predictive p-values and rejected as not fitting the Dutch illegal immigrant data very well. Sections 8 and 9 feature two alternative Bayesian models, each involving a hidden covariate. A fourth model is then fitted in Section 10 and shown to pass a suitable posterior predictive p-value check in Section 11. Some final inferences are drawn from the fourth model in Section 12, and Section 13 contains a summary and discussion. Several technical details are relegated to the Appendix.

2. MAXIMUM LIKELIHOOD ANALYSIS OF TRUNCATED DATA

Consider a situation where events of a certain type occur randomly according to some process, and let N denote the total number of such events which occur in a specified time period and location under study. We call N the *total sample size*. When an event occurs it is associated with certain characteristics which may be described by the value of a covariate x, whose probability density function (pdf) $f(x)$ depends on a parameter θ. Also, the event is associated with the value of a response variable y whose conditional pdf $f(y \mid x)$ depends on a parameter β. Each of x, y, θ and β may be a scalar or a column vector. For the time being, θ, β and N are to be thought of as unknown constants, but later we will treat them as random variables. Suppose that:

- We observe an event with value (x, y) if and only if y is in some specified region, R.
- We observe n such events, with their (x, y) values denoted $(x_1, y_1), ..., (x_n, y_n)$.
- On the basis of these n data pairs, we wish to make inferences regarding β, θ and N.

First observe that the pdf of an observed (or *truncated*) data pair (x, y) is

$$f(x, y \mid y \in R) = \frac{f(x, y)}{P}, \quad y \in R,$$

where $f(x,y)=f(x)f(y\,|\,x)$ and $P=P(y\in R)=P(\beta,\theta)$ (a function of β and θ). Hence the joint pdf of the truncated data pairs - conditional on there being n of them - may be written

$$f(X,Y\,|\,Y\in R^n,n)=\frac{f(X,Y\,|\,n)}{P^n},\quad Y\in R^n,$$

where: $X=(x_1,...,x_n)'$ and $Y=(y_1,...,y_n)'$; $f(X,Y\,|\,n)=\prod_{i=1}^{n}f(x_i,y_i)$, where $f(x_i,y_i)$ denotes $f(x,y)$ evaluated at $x=x_i$ and $y=y_i$; and $Y\in R^n$ means "$y_i\in R$ for all $i=1,...,n$".

Now, the observed (or *truncated*) sample size n has a binomial distribution with parameters N and P. Hence the pdf of the *truncated data*, defined as $D=(X,Y,n\,|\,Y\in R^n)$, may be written

$$f(D)=\frac{f(X,Y\,|\,n)}{P^n}\times\binom{N}{n}P^n(1-P)^{N-n},\quad Y\in R^n. \tag{1}$$

Consequently, the likelihood function may be taken as

$$L(\beta,\theta,N)=f(X,Y\,|\,n)\frac{N!}{(N-n)!}(1-P)^{N-n}, \tag{2}$$

keeping in mind that P and $f(X,Y\,|\,n)$ are implicitly functions of β and θ. The MLEs of β, θ and N can be obtained as follows. First, find the conditional MLEs $\hat{\beta}(N)$ of β and $\hat{\theta}(N)$ of θ which maximize (2) for each fixed $N=n,n+1,n+2,...$ Then $\hat{N}$, the MLE of N, is the value which maximizes $L(\hat{\beta}(N),\hat{\theta}(N),N)$. The MLEs

of β and θ are then $\hat{\beta} = \hat{\beta}(\hat{N})$ and $\hat{\theta} = \hat{\theta}(\hat{N})$. Each pair $(\hat{\beta}(N), \hat{\theta}(N))$ can be obtained using a Newton-Raphson or EM algorithm.

Unfortunately, the above approach becomes problematic when it comes to estimating variances. This is because N is a discrete quantity. Consequently, special techniques are required, either along the lines of Dahiya and Gross (1973) and Blumenthal et al. (1978) which involve asymptotic expansions (in a simpler context with no covariates), or along the lines of Van der Heijden, Bustami et al. (2003) which involve the Horvitz-Thompson estimator. Our situation differs from those considered previously - and is more complicated - because we are additionally concerned with estimating the parameter θ which characterizes the covariate distribution.

Arguably, the Bayesian approach is highly suitable in the present context. As we will see, this approach permits easy calculation of both point and interval estimates for β, θ and N. It also provides a convenient platform for inference on other quantities: such as functions of β, θ and N; and such as the number of the $N - n$ unobserved values with particular characteristics. After a description of the Bayesian approach we will apply it to some Dutch illegal immigrant data and compare our results with the analysis of the same data in Van der Heijden, Bustami et al. (2003).

3. THE BAYESIAN APPROACH

The Bayesian approach to inference on β, θ and N is to treat these quantities as random variables, to specify a joint prior pdf for them, and to multiply that pdf by the pdf of the truncated data at (1), rewritten as $f(D|\beta,\theta,N)$. The resulting joint posterior pdf is given by

$$f(\beta,\theta,N\,|\,D) \propto f(\beta,\theta,N) \times f(X,Y\,|\,n,\beta,\theta)\frac{N!}{(N-n)!}(1-P)^{N-n}, \tag{3}$$

where $f(\beta,\theta,N)$ is the prior pdf, where $f(X,Y\,|\,n)$ at (1) has been rewritten $f(X,Y\,|\,n,\beta,\theta)$, and where $P = P(y \in R\,|\,\beta,\theta) = P(\beta,\theta)$ is the same function of β and θ as P in Section 2.

The equations necessary for inference based on (3) are typically intractable but can be solved to any degree of precision with the aid of MCMC methods (Gilks, Richardson and Spiegelhalter, 1996). In particular, a Metropolis-Hastings (MH) algorithm can be used to obtain a large random sample from (3), namely $(\beta^{(1)},\theta^{(1)},N^{(1)}),...,(\beta^{(J)},\theta^{(J)},N^{(J)}) \sim iid\ f(\beta,\theta,N\,|\,D)$.

This sample can then be used to perform Monte Carlo inference on any functional $\gamma = \gamma(\beta,\theta,N)$. For example, the posterior mean $\hat{\gamma} = E(\gamma\,|\,D)$ may be estimated by $\bar{\gamma} = J^{-1}\sum_{j=1}^{J}\gamma^{(j)}$, where $\gamma^{(j)} = \gamma(\beta^{(j)},\theta^{(j)},N^{(j)})$; and a $1-\alpha$ CI for $\hat{\gamma}$ is $(\bar{\gamma} \pm z_{\alpha/2}s_\gamma/\sqrt{J})$, where z_δ denotes the upper δ-quantile of the standard normal distribution and $s_\gamma^2 = (J-1)^{-1}\sum_{j=1}^{J}(\gamma_j - \bar{\gamma})^2$. The $1-\alpha$ central posterior density region (CPDR) for γ itself may be estimated by the $\alpha/2$ and $1-\alpha/2$ empirical quantiles of $\gamma^{(1)},...,\gamma^{(J)}$.

4. A PRIORI IGNORANCE REGARDING THE TOTAL SAMPLE SIZE

In the special case of a priori ignorance regarding *N*, one suitable choice of joint prior is given by:

$$f(\beta,\theta,N) = f(\beta,\theta)f(N)$$

$$f(N) \propto 1/N, \ N = 1,2,3,...,$$

where $f(\beta,\theta)$ is the joint prior of β and θ. This choice yields the convenient factorisation

$$f(\beta,\theta,N \mid D) = f(\beta,\theta \mid D)f(N \mid D,\beta,\theta),$$

where:

$$f(\beta,\theta \mid D) \propto f(\beta,\theta)\frac{f(X,Y \mid n,\beta,\theta)}{P^n} \tag{4}$$

$$f(N \mid D,\beta,\theta) = \binom{N-1}{n-1}P^n(1-P)^{N-n}, \ N = n,n+1,n+2,.... \tag{5}$$

By the method of composition (Tanner, 1993), equations (4) and (5) permit the Monte Carlo sample from (3) to be obtained more simply. First apply an MH algorithm based on (4) (rather than on (3)) so as to yield $(\beta^{(1)},\theta^{(1)}),...,(\beta^{(J)},\theta^{(J)}) \sim iid \ f(\beta,\theta \mid D)$. Then for each $j = 1,...,J$, sample $N^{(j)}$ from the negative binomial distribution with parameters n and $P(\beta^{(j)},\theta^{(j)})$. The J resulting triplets $(\beta^{(j)},\theta^{(j)},N^{(j)})$, $j = 1,...,J$, then constitute a random sample from $f(\beta,\theta,N \mid D)$.

Note that (4) defines exactly the same posterior for β and θ as would be implied by sampling repeatedly until n positive values of y have been obtained. In that case, N would be the total number of pairs (x,y) sampled and not of direct interest. In contrast, the actual situation under consideration is that of N already-existing pairs

(x, y), where only the n pairs with $y > 0$ are observed. This appealing equivalence results from using the prior $f(N) \propto 1/N$, $N = 1, 2, 3, ...$

5. CATEGORICAL COVARIATES AND A POISSON RESPONSE

To provide a concrete example, first suppose that each covariate x is a column vector of length $K + 1$ which takes on one of the known values $v_1, ..., v_T$ with corresponding unknown probabilities $\pi_1, ..., \pi_T$, where these are parameterized by $\theta = (\theta_2, ..., \theta_T)'$ according to the logistic formulation $\theta_t = \log(\pi_t / \pi_1)$. Secondly, suppose that the response y for an event with covariate x is Poisson with mean $\lambda(x) = \exp(x'\beta)$, where $\beta = (\beta_0, ..., \beta_K)'$ is the vector of regression coefficients. Thirdly, suppose that an event is observed if and only if its response y is positive.

Then, in the classical framework, x has a categorical distribution with pdf $f(x) = \sum_{t=1}^{T} \pi_t I(x = v_t)$, $x = v_1, ..., v_T$, where $I(\cdot)$ is the standard indicator function, $\pi_1 = \pi_1(\theta) = P(x = v_1) = 1/\{1 + \exp(\theta_2) + ... + \exp(\theta_T)\}$, and $\pi_t = P(x = v_t) = \pi_1 \exp(\theta_t)$, $t = 2, ..., T$. Also, $f(y \mid x) = \exp(-\lambda(x))\lambda(x)^y / y!$, $y = 0, 1, 2, ...$, and $R = \mathbb{N}_+ \equiv \{1, 2, 3, ...\}$, so that $P = P(y > 0) = \sum_{t=1}^{T} \pi_t p_t$, where $p_t = P(y > 0 \mid x = v_t) = 1 - \exp\{-\exp(v_t'\beta)\}$.

In a Bayesian setting, some expressions above require rewriting with an appropriate conditioning on β and θ. Thus, we should replace $f(x)$ by $f(x \mid \theta)$, $P(x = v_t)$ by $P(x = v_t \mid \theta)$, $f(y)$ by $f(y \mid x, \beta)$, $P(y > 0)$ by $P(y > 0 \mid \beta, \theta)$, and $P(y > 0 \mid x = v_t)$ by $P(y > 0 \mid x = v_t, \beta)$.

Suppose that there is a priori ignorance regarding all unknown quantities and it is reasonable to set $f(\beta, \theta, N) = f(\beta, \theta)f(N)$, where $f(N) \propto 1/N$, $N = 1, 2, 3, ...$ (see

Section 4) and $f(\beta,\theta) \propto 1, \beta \in \Re^{K+1}, \theta \in \Re^{T-1}$ (so that all the components of β and θ are a priori uniform over $\Re \equiv (-\infty,\infty)$, independently of one another and of N).

Given the above specifications, the pdfs at (3), (4) and (5) are completely defined. After some algebra, a convenient way to write the pdf at (4) is $f(\beta,\theta \mid D) \propto \exp\{l(\beta,\theta)\}$, where

$$l(\beta,\theta) = \sum_{t=1}^{T} n_t \left\{ \overline{y}_t \eta_t(\beta) + \theta_t - \mu_t(\beta) \right\} - n \log\left(\frac{1}{\pi_1(\theta)} - \sum_{t=1}^{T} \exp\left\{ \theta_t - \mu_t(\beta) \right\} \right),$$

and where: $\theta_1 \equiv 0$; $n_t = \sum_{i=1}^{n} I(x_i = v_t)$ is the number of observed events with covariate v_t; $\overline{y}_t = n_t^{-1} \sum_{i=1}^{n} y_i I(x_i = v_t)$ is the average response for the observed events with covariate v_t; $\eta_t(\beta) = v_t'\beta$ is the linear predictor for a value with covariate v_t; $\mu_t(\beta) = \exp\{\eta_t(\beta)\}$ is the mean of the postulated Poisson distribution for the response of an event with covariate v_t; and $\pi_1(\theta) = 1/\sum_{t=1}^{T} \exp(\theta_t)$ is the probability that an event has covariate v_1.

6. APPLICATION TO DUTCH ILLEGAL IMMIGRANT DATA

To illustrate the model in Section 5, we will apply it to some data taken from Van der Heijden, Bustami et al. (2003), as given in Table 1. In their paper these data are discussed in detail and only a brief description will be given here. According to Dutch police records, in 1995 there were 4392 persons who were apprehended and identified as illegal immigrants in The Netherlands. Of these, only 2512 were effectively expelled from the country. The remaining 1880 consisted of 398 females and 1482 males. In turn there were 366 females and 1279 males who were each

apprehended once, 24 females and 159 males who were each apprehended twice, and so on.

Of primary interest is the total number of apprehensible but not effectively expellable illegal immigrants in the Netherlands during 1995 - hereafter referred to simply as the total number of illegal immigrants. This question was addressed by Van der Heijden, Bustami et al. (2003) using several different models with varying amounts of covariate information. However, none of these models was deemed to fit the data very well, and none of the models made an attempt to estimate the covariate distribution. Our aim here is to find a satisfactory model whose solution yields plausible inferences regarding the quantity of primary interest as well as several others.

Table 1. Dutch illegal immigrant data

	Frequency of apprehension						
	1	2	3	4	5	6	Total
Females	366	24	6	1	1	0	398
Males	1279	159	31	12	0	1	1482
Persons	1645	183	37	13	1	1	1880

The data in Table 1 may be taken as representing $n = 1880$ data pairs (x_i, y_i), $i = 1,...,n$, where each y_i is a positive integer, and where each x_i is either $v_1 = (1,0)'$ (in the case of a female) or $v_2 = (1,1)'$ (in the case of a male). Thus $T = 2$, $n_1 = 398$ and $n_2 = 1482$. It is also useful at this stage to define $n_{ty} = \sum_{i=1}^{n} I(x_i = v_t, y_i = y)$ (conditional on $Y > 0$), and $n_{\bullet y} = \sum_{t=1}^{T} n_{ty}$ (i.e. the total number of persons who were

apprehended exactly y times). Thus, $n_{24} = 12$, $n_{\bullet 2} = 183$, etc. The regression parameter here is $\beta = (\beta_0, \beta_1)'$, and so $K = 1$.

Assuming suitability of the model in Section 5, the number of times a female is apprehended follows a Poisson distribution with mean $\mu_1 = \exp(v_1'\beta) = \exp(\beta_0)$, and the number of times a male is apprehended follows a Poisson distribution with mean $\mu_2 = \exp(v_2'\beta) = \exp(\beta_0 + \beta_1)$. The probabilities of an illegal immigrant being female and male are, respectively, $\pi_1 = 1/(1 + \exp(\theta_2))$ and $\pi_2 = 1 - \pi_1$. We wish to make inferences regarding β_0 (the intercept), β_1 (the male indicator coefficient), θ_2 (the logodds of being male), N (the total number of illegal immigrants), and some other quantities, such as m_2, the number of unapprehended males.

An MH algorithm was applied according to Sections 4 and 5 so as to obtain a sample of size $J = 10000$ from $f(\beta_0, \beta_1, \theta_2 | D)$. Let this sample be denoted $(\beta_0^{(1)}, \beta_1^{(1)}, \theta_2^{(1)}), ..., (\beta_0^{(J)}, \beta_1^{(J)}, \theta_2^{(J)})$. Then for each $j = 1, ..., J$, a value $N^{(j)}$ was sampled from the negative binomial distribution with parameters n and $P^{(j)} = \pi_1^{(j)} p_1^{(j)} + \pi_2^{(j)} p_2^{(j)}$, where $\pi_1^{(j)} = 1/(1 + \exp(\theta_2^{(j)}))$, $\pi_2^{(j)} = 1 - \pi_1^{(j)}$, $p_1^{(j)} = 1 - \exp(-\exp(\beta_0^{(j)}))$ and $p_2^{(j)} = 1 - \exp(-\exp(\beta_0^{(j)} + \beta_1^{(j)}))$.

Taking $N^{(1)}, ..., N^{(J)}$ as a sample from $f(N | D)$, the posterior mean of N, $\hat{N} = E(N | D)$, was estimated by $\bar{N} = (N^{(1)} + ... + N^{(J)})/J = 7368.7$. Also, the 95% CPDR for N was estimated as [6592, 8308], whose bounds are the 0.025 and 0.975 empirical quantiles of $N^{(1)}, ..., N^{(J)}$.

These inferences are similar to those in Table 5 of Van der Heijden, Bustami et al. (2003), where a non-Bayesian model with only gender as a covariate yielded the point estimate 7319 (≈ 7369) and the 95% CI [6504, 8134] ($\approx$ [6592, 8308]).

We see that our inferences are slightly higher. This may be due to the posterior density of N being skewed to the right, as is evident from histograms of $N^{(1)},...,N^{(J)}$ (not shown here). By contrast, the approach of Van der Heijden, Bustami et al. (2003) involves approximating the distribution of a Horvitz-Thompson-type estimator with a normal distribution. We hypothesize that that distribution is in fact right-skewed and that if this were taken into account the estimates in Van der Heijden, Bustami et al. (2003) would be slightly lower and more in line with our inference.

In the same way as for N we also estimated β_0, β_1 and θ_2 by -1.57, 0.48 and 0.90, with 95% CPDRs (-1.90, -1.28), (0.17, 0.83) and (0.56, 1.19), respectively. From our results regarding β_1 we deduce that there is a difference between males and females: the ratio of the Poisson mean for males to the Poisson mean for females is estimated to be about $\exp(0.48) = 1.6$, and there appears to be a 95% probability (a posteriori) that that ratio is between about $\exp(0.17) = 1.2$ and $\exp(0.83) = 2.3$. In comparison, using a non-Bayesian model with more covariates, Van der Heijden, Bustami et al. (2003) estimated β_1 by 0.397 with standard error 0.163 (see Table 4 in that paper).

The usefulness of the Bayesian approach is evidenced by the ease with which the Monte Carlo sample can be used to make inferences regarding a host of other quantities of potential interest. In particular, suppose that we are interested in m_2, defined as the number of unapprehended males amongst the $N - n$ unapprehended persons. According to our model, $(m_2 \mid \beta, \theta, N, n) \sim Bin(N - n, \psi_2)$, where $\psi_2 = P(x = v_2 \mid y = 0, \beta, \theta) = \pi_2(1 - p_2)/(1 - P)$. Hence we simulate $m_2^{(j)} \sim Bin(N^{(j)} - n, \psi_2^{(j)})$ for each $j = 1,...,J$, where $\psi_2^{(j)} = \pi_2^{(j)}(1 - p_2^{(j)})/(1 - P^{(j)})$.

The resulting values, $m_2^{(1)},...,m_2^{(J)}$, constitute a random sample from the posterior distribution of m_2, and their average, $\bar{m}_2 = 3738.2$, provides an estimate of the posterior mean of m_2. Also, the empirical 0.025 and 0.975 quantiles of $m_2^{(1)},...,m_2^{(J)}$ are 3211 and 4317, which provide estimates of the bounds of the 95% CPDR for m_2. A possibly better approach to point estimation of m_2 is to note that $E(m_2 \mid \beta, \theta, N, n) = (N-n)\psi_2$. This fact leads to the Rao-Blackwell estimate, $J^{-1}\sum_{j=1}^{J}(N^{(j)}-n)\psi_2^{(j)} = 3738.2$ (which in this case is identical to $\bar{m}_2$ to one decimal place).

7. MODEL CHECK VIA POSTERIOR PREDICTIVE P-VALUES

None of the models in Van der Heijden, Bustami et al. (2003) fit the illegal immigrant data adequately (see p319 in that paper). It is therefore of interest to see whether the same conclusion applies for the Bayesian model fitted in Section 6. One way to address this issue is via posterior predictive p-values (Gelman, Carlin, Stern and Rubin, 1995, pp169-174). To this end, let the discrepancy quantity be

$$\Delta = \Delta(W,\beta,\theta) = \sum_{t=1}^{T}\sum_{y=1}^{M} r_{ty}^2 ,$$

where $T = 2$, $M = 6$ and:

$$r_{ty} = \frac{n_{ty} - e_{ty}}{s_{ty}} \qquad \text{(the ty-th standardized residual)}$$

$$W = \{n_{ty} : t = 1,...,T; y = 1,...,M\} \qquad \text{(the set of 12 counts in Table 1)}$$

$$e_{ty} = E(n_{ty} \mid \beta, \theta, n_t), \quad s_{ty}^2 = V(n_{ty} \mid \beta, \theta, n_t).$$

Here, e_{ty} and s_{ty}^2 are the mean and variance of the ty-th frequency n_{ty}, given the model parameters, and given that $n_1 = 398$ females and $n_2 = 1482$ males were observed. The ppp-value may be defined as $\Pi = P(\tilde{\Delta} \geq \Delta \mid D)$, where $\tilde{\Delta} = \Delta(\tilde{W}, \beta, \theta)$ and $\tilde{W}$ is an independent future replicate of W given $n_1, ..., n_T$. Although Π is difficult to calculate analytically, it can readily be estimated via Monte Carlo. Details can be found in Appendix A.1. Thereby we estimate Π by $\hat{\Pi} = 0.0003$, with 95% CI (0, 0.00064).

Because Π is very small, we conclude that the model used in Section 6 provides a poor fit to the data. Further details of this analysis are presented in Table 2 which shows the observed frequencies n_{ty} in Table 1 alongside the estimated expected frequencies e_{ty} and estimated standardized residuals r_{ty} for both females ($t = 1$) and males ($t = 2$).

In Table 2 the residuals greater than 3 in absolute value are marked by an asterisk, suggesting points of poor fit. For example, $n_{24} = 12$ males were apprehended 4 times. According to the fitted model, we would expect only about $e_{24} = 2.0$ of the $n_2 = 1482$ males to be apprehended 4 times. The discrepancy between 12 and 2 appears to be due to more than just chance, and this intuitive assessment is corroborated by the estimated standardized residual $r_{24} = 7.24$ being large.

Table 2:   Observed and estimated counts

y	Observed			Expected			Standardized residual	
	Female	Male	Total	Female	Male	Total	Female	Male
0	0	0	0	1750.5	3738.2	5488.7		
1	366	1279	1645	357.7	1247.7	1605.4	1.32	2.21
2	24	159	183	37.5	208.8	246.3	-2.27	-3.71*
3	6	31	37	2.7	23.4	26.1	2.17	1.62
4	1	12	13	0.15	2.0	2.1	2.44	7.24*
5	1	0	1	0.0066	0.13	0.14	14.03	-0.36
6	0	1	1	0.00025	0.0075	0.0080	-0.01	11.79*
Total	398	1482	1880	2148.5	5220.2	7368.7		

8. A HIDDEN COVARIATE MODEL

Having established that the model used in Section 6 does not fit the data very well, we now consider alternatives. One possibility is to postulate a hidden covariate with two possible values. To make things concrete, let us hypothesize that each individual has an unknown *susceptibility* to being apprehended and is either *cautious* or *careless*, although we don't know which. Then we may take the logarithm of the mean of the ith person's response y_i as

$$x_i'\beta = \beta_0 x_{0i} + \beta_1 x_{1i} + \beta_2 x_{2i},$$

where $x_i = (x_{0i}, x_{1i}, x_{2i})'$ is the value of the ith person's covariate vector and equals one of the following:

$$v_1' = (1,0,0) \qquad \text{(if they are a cautious female)}$$

$$v_2' = (1,0,1) \qquad \text{(if they are a careless female)}$$

$$v_3' = (1,1,0) \qquad \text{(if they are a cautious male)}$$

$$v_4' = (1,1,1) \qquad \text{(if they are a careless male).}$$

This specification corresponds to a model in the class of models already considered in Section 5, with $K = 2$, $\beta = (\beta_0, \beta_1, \beta_2)'$, $T = 4$, and $\theta = (\theta_2, \theta_3, \theta_4)'$. However, in this case we do not know n_1, n_2, n_3 and n_4 (the numbers of cautious females, careless females, cautious males and careless males, respectively), only that the total number of females is $n_1 + n_2 = 398$, and the total number of males is $n_3 + n_4 = 1482$. Also, we do not know n_{23}, the number of careless females who were apprehended three times, only that the total number of females apprehended three times is $n_{13} + n_{23} = 6$ (see Table 1).

As before we will take the priors on $\beta_0, \beta_1, \theta_2, \theta_3$ and θ_4 as independent and uniform over $\Re$. Also, for reasons of identifiability, we will restrict β_2 (the *susceptability coefficient*) to be positive by using the prior $f(\delta_2) \propto 1, \delta_2 \in \Re$, where $\delta_2 = \log \beta_2$ (instead of $f(\beta_2) \propto 1, \beta_2 \in \Re$).

Inference can once again proceed via Monte Carlo after first implementing a suitable MH algorithm. In this case our target is a random sample from the joint posterior distribution of

$$\beta_0, \beta_1, \beta_2; \quad \theta_2, \theta_3, \theta_4; \quad n_{21}, n_{22}, n_{23}, n_{24}, n_{25}; \quad n_{41}, n_{42}, n_{43}, n_{44}, n_{46}. \tag{6}$$

Note that n_{16} and n_{26} must be zero because no female was apprehended 6 times (see Table 1), and this is why n_{26} does not appear in (6). Likewise, $n_{35} + n_{45} = 0$ implies that $n_{45} = 0$.

We see that a suitable MH algorithm is very similar to the one used in Section 6 but with 10 additional Gibbs steps, one for each n_{ty} listed in (6). See Appendix A.2 for details regarding these steps. From the output of the MH algorithm we estimated the posterior mean and 95% CPDR for β_2 as 3.18 and (2.00, 5.66). This implies a clear distinction between the hypothesized careless and cautions persons, with the ratio of the two expected numbers of apprehensions being about exp(3.18) = 24, and almost certainly more than exp(2.00) = 7.4.

Also, we estimate the posterior mean and 95% CPDR for β_1 as -0.21 and (-1.18, 0.67), and the posterior probability that β_1 is positive as 32%. These results are very notable because they suggest that there may be no difference between males and females, and that the difference found in Section 6 was only apparent and possibly just an artefact of using a poorly fitting model.

9. A HIDDEN COVARIATE MODEL WITH UNEQUAL RATIOS

The hidden covariate model in Section 8 assumes that the ratio of the Poisson mean for careless females to the Poisson mean for cautious females is exactly the same as the ratio of the Poisson mean for careless males to the Poisson mean for cautious males (and that both ratios are equal to $\exp(\beta_2)$). This suggests that we also consider a model where the two ratios may be unequal. To this end suppose that the logarithm of the mean of the ith person's response y_i is

$$x_i'\beta = \beta_0 x_{0i} + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} \,,$$

where $x_i = (x_{0i,} x_{1i,} x_{2i}, x_{3i})'$ is the value of the person's covariate vector and equals one of the following:

$v_1' = (1,0,0,0)$ (if they are a cautious female)

$v_2' = (1,0,1,0)$ (if they are a careless female)

$v_3' = (1,1,0,0)$ (if they are a cautious male)

$v_4' = (1,1,0,1)$ (if they are a careless male).

The only difference between the model here and the one in Section 8 is that $K = 3$, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)'$, $\exp(\beta_2)$ is the ratio of the Poisson mean for careless females to the Poisson mean for cautious females, and $\exp(\beta_3)$ is the ratio of the Poisson mean for careless males to the Poisson mean for cautious males. These two ratios are now not necessarily the same, as they were in Section 8. For reasons of identifiability we will use the prior $f(\beta_0, \beta_1, \delta_2, \delta_3) \propto 1$, $\beta_0, \beta_1, \delta_2, \delta_3 \in \Re$, where $\delta_k = \log \beta_k$ (instead of $f(\beta_0, \beta_1, \beta_2, \beta_3) \propto 1$, $\beta_0, \beta_1, \beta_2, \beta_3 \in \Re$).

Applying an MH algorithm similar to the one in Section 8, we obtained a random sample from the joint posterior distribution of the now 17 unknown parameters (the 16 parameters in (6) plus β_3). Thereby we estimated the posterior mean and 95% CPDR for β_1 as 1.11 and (-2.03, 4.51), and the posterior probability of β_1 being positive as 74%. These results again suggest that there is no difference between males and females.

10. A HIDDEN COVARIATE NO-GENDER MODEL

Section 7 showed that a model with only gender as a covariate does not fit the data very well. Sections 8 and 9 then showed that introducing a hidden covariate makes gender statistically insignificant - both when that hidden covariate is assumed to be the same for males and females and when it is assumed to be different. This suggests that we now fit a model with only the hidden covariate and no gender. To this end, suppose that for both males and females the log-mean of the ith person's response y_i is

$$x_i'\beta = \beta_0 x_{0i} + \beta_1 x_{1i},$$

where $_i x_i = (x_{0i}, x_{1i})'$ is the value of the person's covariate vector and equals one of the following:

$$v_1' = (1,0) \qquad \text{(if they are cautious)}$$
$$v_2' = (1,1) \qquad \text{(if they are careless)}.$$

In this case the unknown quantities are $\beta_0, \beta_1, \theta_2, n_{21}, n_{22}, n_{23}, n_{24}, n_{25}, n_{26}$, where n_{2y} is the number of careless persons (both male and female) who were apprehended y times. A suitable MH algorithm involves six Gibbs steps, as detailed in Appendix A.3. Applying this MH algorithm, we obtained a random sample from the posterior distribution of the now 9 unknown quantities. Thereby we estimated the posterior mean and 95% CPDR for β_1 as 2.31 and (1.89, 3.02). Thus the hidden covariate is statistically significant when gender has been removed from the model. The corresponding inferences for β_0 are -3.13 and (-3.17, -1.59), and those for θ_2 are -3.58 and (-4.47, -2.86).

11. CHECK OF THE HIDDEN COVARIATE NO-GENDER MODEL

It remains to check the fit of the covariate model with no gender considered in Section 10. To this end we may define the discrepancy quantity as

$$\Delta = \Delta(W, \beta, \theta) = \sum_{y=1}^{M} r_y^2 ,$$

where $M = 6$ and:

$r_y = \dfrac{n_{\bullet y} - e_y}{s_y}$ (the y-th standardized residual)

$W = \{n_{\bullet 1}, ..., n_{\bullet M}\}$ (the set of person frequencies in Table 1)

$e_y = E(n_{\bullet y} \mid \beta, \theta, n)$, $s_y^2 = V(n_{\bullet y} \mid \beta, \theta, n)$.

The ppp-value, $\Pi = P(\tilde{\Delta} \geq \Delta \mid D)$, may be estimated via Monte Carlo (see Appendix A.4) and we find that $\hat{\Pi} = 0.46$. This provides no evidence against the model fitting well.

Table 3 provides estimates of the expected counts and standardized residuals used in the calculation of $\hat{\Pi}$. We see that all six standardized residuals $r_1, ..., r_6$ are less than unity in absolute value. This is in contrast to the residuals in Table 6 of Van der Heijden, Bustami et al. (2003) which include three that are greater than 3 in absolute value, indicating a lack of fit. Those residuals are reproduced in the last column of our Table 6 and the three extremes are asterisked.

Table 3: Observed and estimated counts for the hidden covariate no-gender model

		Results for the hidden covariate model in Section 7		Results in Table 6 of Van der Leijden, Bustami et al. (2006)	
y	Observed	Estimated	Residuals	Estimated	Residuals
0	0	13433.8		10810.4	
1	1645	1643.8	0.06	1612.6	0.81
2	183	184.7	-0.10	233.7	-3.32*
3	37	38.2	-0.14	30.1	1.25
4	13	10.2	0.96	3.2	5.42*
5	1	2.5	-0.90	0.3	1.31
6	1	0.6	0.81	0.0	6.57*
Total	1880	15313.8		12690.3	

12. FURTHER INFERENCES USING THE BEST MODEL

Having found a model in Section 10 which passed the posterior predictive model check in Section 11, we now provide further inference based on that model. Using the output from the MH algorithm in Section 10, our Monte Carlo estimate of $\mu_1 = \exp(v_1'\beta) = \exp(\beta_0)$, the mean of the Poisson distribution for cautious illegal immigrants, is 0.13 with 95% CPDR (0.04, 0.20); and our estimate of $\mu_2 = \exp(v_2'\beta) = \exp(\beta_0 + \beta_1)$, the mean of the Poisson distribution for careless illegal immigrants, is 1.23 with 95% CPDR (0.78, 1.94).

Using further simulations, we also estimate the posterior mean and 95% CPDR for N as 15314 and [9574, 32656]. These results may be compared with 12690 and [7186,

18194] which are the corresponding estimates in Van der Heijden, Bustami et al. (2003, Table 5) under their best model which takes into account gender, age and nationality but still exhibits some unobserved heterogeneity. Our inference is somewhat higher, and this is in agreement with the suggestion in Van der Heijden, Bustami et al. (2003, page 319) that "12690 should be interpreted as an underestimate". For an explanation of why remaining unobserved heterogeneity could cause N to be underestimated, see the last paragraph on page 297 of Van der Heijden, Cruyff et al. (2003).

Also, our estimate of n_1, the number of *cautious* illegal immigrants in the sample, is 1579; and so our estimate of n_2, the number of *careless* illegal immigrants in the sample, is 1880 - 1579 = 301. The 95% CPDR for n_2 is estimated as [102,602].

After further simulations, we also estimate m_1, the number of unapprehended *cautious* illegal immigrants, as 13274, and so we estimate m_2, the number of unapprehended *careless* illegal immigrants, as 15314 - 13274 = 160. Our 95% CPDR estimate for m_2 is [16,500].

Thus it appears that the vast majority of illegal immigrants are cautious and have only a small probability of being apprehended, namely about $1-\exp(-0.13)=12\%$. At the same time there exists a small but significant proportion of careless persons whose probability of apprehension is higher at about $1-\exp(-1.23)=71\%$. That proportion is about 301/1880 = 16% in the sample, about 160/(15314 - 1880) = 1% in the non-sample, and about (301 + 160)/15314 = 3% overall.

13. SUMMARY AND DISCUSSION

In this paper we have developed a Bayesian modeling framework for the analysis of truncated data and highlighted the advantages of that framework relative to the classical frequentist approach. We focused on a specific class of Bayesian models involving a zero-truncated Poisson response, categorical covariates, and a priori ignorance regarding all model parameters. An outline was given of how any model in that class can be solved with the aid of Markov chain Monte Carlo methods.

Several such Bayesian models were applied to some Dutch illegal immigrant data which had previously been analyzed by Van der Heijden, Bustami et al. (2003) using approximate maximum likelihood methods and a Horvitz-Thomspon estimator. The best model in that paper exhibited residual unexplained heterogeneity, but we were able to find a Bayesian model which passed a suitable goodness-of-fit test (with a posterior predictive p-value of 0.46). That model includes a hidden covariate which can be used to categorize illegal immigrants as either *careless* or *cautious* in relation to their susceptibility to apprehension. A notable finding was that with the inclusion of this hidden covariate, the previously apparent difference between males and females disappeared.

The use of hidden covariates is akin to the non-Bayesian approach in Böhning, Dietz, Kuhnert and Schön (2005) and Böhning and Schön (2005) where it is shown how an unacceptable goodness-of-fit can be addressed by modeling with mixtures. That is, our best model is in fact a mixture model with two components - one for careless illegal immigrants and one for cautious illegal immigrants. The advantage of the Bayesian approach is that it also permits convenient inference on the parameters of

the covariate distribution itself (θ in our notation), something which is not considered in Böhning et al. (2005) or Böhning and Schön (2005).

It should be pointed out that "Algorithm 1 for the Poisson" in Böhning et al. (2005) (a version of the EM algorithm) is not actually guaranteed to find the true MLEs of N and β (in our notation). By applying that algorithm in simple situations we found that it sometimes terminates close to, but not exactly at, the true MLEs of N and β (obtained by way of trial and error). This is likely due to N being a discrete quantity.

APPENDIX

A.1. The Monte Carlo procedure in Section 7

Observe that $(n_{ty} \mid \beta, \theta, n_t) \sim Bin(n_t, \rho_{ty})$, where

$$\rho_{ty} = P(y_i = y \mid x = v_t, \beta, \theta, y_i > 0) = \frac{\exp(-\exp(v_t'\beta) + v_t'\beta y)}{y!(1 - \exp(-\exp(v_t'\beta)))}, \quad y = 1, 2, \ldots,$$

so that $e_{ty} = n_t \rho_{ty}$ and $s_{ty}^2 = n_t \rho_{ty}(1 - \rho_{ty})$. The ppp-value $\Pi = P(\tilde{\Delta} \geq \Delta \mid D)$ may be estimated by

$$\hat{\Pi} = \frac{1}{J} \sum_{j=1}^{J} I(\tilde{\Delta}^{(j)} \geq \Delta^{(j)}), \quad \text{with 95\% CI} \quad \left(\hat{\Pi} \pm 1.96 \sqrt{\frac{\hat{\Pi}(1 - \hat{\Pi})}{J}} \right),$$

where $\Delta^{(j)} = \Delta(W, \beta^{(j)}, \theta^{(j)})$, $\tilde{\Delta}^{(j)} = \Delta(\tilde{W}^{(j)}, \beta^{(j)}, \theta^{(j)})$, and $\tilde{W}^{(j)}$ is a value sampled from the distribution of W given $\beta = \beta^{(j)}$, $\theta = \theta^{(j)}$ and $n_1, \ldots, n_T$. Here, $\beta^{(j)}$ and $\theta^{(j)}$ are taken from the output of the MH algorithm in Section 6.

Each $\tilde{W}^{(j)}$ can be obtained as follows. Repeatedly sample from the Poisson distribution with mean $\exp(\beta_0^{(j)})$ until $n_1 = 398$ positive values have been obtained. Then for each $y = 1,...,M$, let $\tilde{n}_{1y}^{(j)}$ be the number of these n_1 values which are equal to y. Likewise, repeatedly sample from the Poisson distribution with mean $\exp(\beta_0^{(j)} + \beta_1^{(j)})$ until $n_2 = 1482$ positive values have been obtained, and then for each $y = 1,...,M$, let $\tilde{n}_{2y}^{(j)}$ be the number of these n_2 values which are equal to y. Finally form $\tilde{W}^{(j)} = \{\tilde{n}_{ty}^{(j)} : t = 1,...,T; y = 1,...,M\}$.

A.2. The 10 Gibbs steps in Section 8

For $t = 2$ and each $y = 1,2,3,4,5$ the Gibbs step is to sample $n_{2y} \sim Binomial(n_{1y} + n_{2y}, \rho_{2y})$, where

$$\rho_{2y} = P(x = v_2 \mid y, \beta, \theta, x = v_1 \text{ or } v_2) = \frac{P(x = v_2 \mid y, \beta, \theta)}{P(x = v_1 \mid y, \beta, \theta) + P(x = v_2 \mid y, \beta, \theta)}$$
$$= \frac{\pi_2 f(y \mid v_2, \beta)}{\pi_1 f(y \mid v_1, \beta) + \pi_2 f(y \mid v_2, \beta)},$$

and where $\pi_t = \exp(\theta_t)/\{1 + \sum_{t=2}^{4} \exp(\theta_t)\}$, $f(y \mid v_t, \beta) = \exp(\mu_t)\mu_t^y / y!$ and $\mu_t = \exp(v_t'\beta)$.

For $t = 4$ and each $y = 1,2,3,4,6$ the Gibbs step is to sample $n_{4y} \sim Binomial(n_{3y} + n_{4y}, \rho_{4y})$, where

$$\rho_{4y} = P(x = v_4 \mid y, \beta, \theta, x = v_3 \text{ or } v_4) = \frac{\pi_4 f(y \mid v_4, \beta)}{\pi_3 f(y \mid v_3, \beta) + \pi_4 f(y \mid v_4, \beta)}.$$

A.3. The 6 Gibbs steps in Section 10

For each $y = 1,2,3,4,5,6$ the Gibbs step is to sample $n_{2y} \sim Binomial(n_{\bullet y}, \rho_{2y})$, where

$$\rho_{2y} = P(x = v_2 \mid y, \beta, \theta) = \frac{\pi_2 f(y \mid v_2, \beta)}{\pi_1 f(y \mid v_1, \beta) + \pi_2 f(y \mid v_2, \beta)},$$

and where: $n_{\bullet y} = n_{1y} + n_{2y}$, $\pi_t = \exp(\theta_t)/(1 + \exp(\theta_t))$, $f(y \mid v_t, \beta) = \exp(\mu_t)\mu_t^y / y!$ and $\mu_t = \exp(v_t'\beta)$. For reasons of identifiability, we use the prior $f(\delta_1) \propto 1, \delta_1 \in \Re$, $\delta_1 = \log \beta_1$.

A.4. The Monte Carlo procedure in Section 11

Observe that $(n_{\bullet y} \mid \beta, \theta, n) \sim Bin(n, \rho_y)$, where

$$\rho_y = f(y \mid y > 0, \beta, \theta) = \frac{1}{P}\{\pi_1 f(y \mid v_1, \beta) + \pi_2 f(y \mid v_2, \beta)\},$$

so that $e_y = n\rho_y$ and $s_y^2 = n\rho_y(1 - \rho_y)$. The ppp-value may be defined as $\Pi = P(\tilde{\Delta} \geq \Delta \mid D)$, where $D = (Y \mid Y > 0)$ (the data), $\tilde{\Delta} = \Delta(\tilde{W}, \beta, \theta)$ and $\tilde{W}$ is a future replicate of W given n. A Monte Carlo estimate of Π is $\hat{\Pi} = J^{-1} \sum_{j=1}^{J} I_j$, where $I_j = I(\tilde{\Delta}_j \geq \Delta_j)$, $\Delta^{(j)} = \Delta(W, \beta^{(j)}, \theta^{(j)})$ and $\tilde{\Delta}^{(j)} = \Delta(\tilde{W}^{(j)}, \beta^{(j)}, \theta^{(j)})$. To generate $\tilde{W}^{(j)}$, first sample $n_1^{(j)} \sim Bin(n, \pi_1^{(j)})$, where $\pi_1^{(j)} = \exp(\theta_1^{(j)})/(1 + \exp(\theta_1^{(j)}))$. Then repeatedly sample from the $Poisson(\exp(\beta_0^{(j)}))$ distribution until $n_1^{(j)}$ positive values have been obtained. Then repeatedly sample from the $Poisson(\exp(\beta_0^{(j)} + \beta_1^{(j)}))$ distribution until $n_2^{(j)} = n - n_1^{(j)}$ positive values have been obtained. Then for each $y = 1,...,6$, let $n_{\bullet y}^{(j)}$ be the number of the n values obtained which are equal to y. Finally, form $\tilde{W}^{(j)} = \{n_{\bullet 1}^{(j)},...,n_{\bullet 6}^{(j)}\}$.

REFERENCES

Amemiya, T. (1984), "Tobit Models: A Survey," *Journal of Econometrics*, 24, 3-61.

Böhning, D., Dietz, E., Kuhnert, R., and Schön, D. (2005), "Mixture models for capture-recapture count data," *Statistical Methods and Applications*, 14, 29-43.

Böhning, D., and Schön, D. (2005), "Nonparametric Maximum Likelihood Estimation of Population Size Based on the Counting Distribution," *Applied Statistics*, 54, 721-737.

Blumenthal, S., Dahiya, R. C., and Gross, A. J. (1978), "Estimating the Complete Sample Size From an Incomplete Poisson Sample," *Journal of the American Statistical Association*, 73, 182-187.

Dahiya, R. C., and Gross, A. J. (1973), "Estimating the Zero Class From a Truncated Poisson Sample," *Journal of the American Statistical Association*, 68, 731-733.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, New York: Chapman and Hall.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds) (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall.

Grogger, J. T., and Carson, R. (1991), "Models for Truncated Counts," *Journal of Applied Econometrics*, 6, 225-238.

Ibrahim, J. G., Chen, M. H., and Lipsitz, S. R. (2002), "Bayesian Methods for Generalized Linear Models with Covariates Missing at Random," *Canadian Journal of Statistics*, 30, 55-78.

O'Neill, T. J., and Barry, S. C. (1995a), "Truncated Logistic Regression," *Biometrics*, 51, 533-541.

O'Neill, T. J., and Barry, S. C. (1995b), "Group Truncated Ordinal Regression," *Statistics and Probability Letters*, 22, 195-203.

Scollnik, D. P. M. (1997), "Inference Concerning the Size of the Zero Class From an Incomplete Poisson Sample," *Communications in Statistics: Theory and Methods*, 26, 221-236.

Shaw, D. (1988), "On-Site Samples' Regression: Problems of Non-Negative Integers, Truncation, and Endogenous Stratification," *Journal of Econometrics*, 37, 211-223.

Tanner, M. A. (1993), *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions*, New York: Springer-Verlag.

Van der Heijden, P. G. M., Cruyff, M., and Van Houwelingen, H. C. (2003), "Estimating the Size of a Criminal Population From Police Records Using the Truncated Poisson Regression Model," *Statistica Neerlandica*, 57, 289-304.

Van der Heijden, P. G. M., Bustami, R., Cruyff, M. J. L. F., Engbersen, G., and Van Houwelingen, H. C. (2003), "Point and Interval Estimation of the Population Size Using the Truncated Poisson Regression Model," *Statistical Modelling*, 3, 305-322.