CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Research Papers in Economics

Working Paper Series

Working Paper WP03/01 April , 2003

THE FAST DECAY PROCESS IN RECREATIONAL DEMAND ACTIVITIES AND THE USE OF ALTERNATIVE COUNT DATA MODELS

by

Rakhal Sarker
Yves Surry

UNIVERSITY
of GUELPH

Department of Agricultural Economics and Business

University of Guelph
Guelph, Ontario N1G 2W1
Canada

Copyright 2003 by Rakhal Sarker and Yves Surry. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

THE FAST DECAY PROCESS IN RECREATIONAL DEMAND ACTIVITIES AND THE USE OF ALTERNATIVE COUNT DATA MODELS

Rakhal Sarker
Department of Agricultural Economics and Business
University of Guelph, Ontario, Canada
(e.mail: rsarker@uoguelph.ca)

Yves Surry
Institut National de la Recherche Agronomique
Station d'Economie et Sociologie Rurales
Rennes, France
(e.mail : surry@roazhon.inra.fr)

The senior authorship is equally shared. This paper was presented at the IX th Congress of the European Association of Agricultural Economists, Zaragoza, Spain, August 28-30, 2002

THE FAST DECAY PROCESS IN RECREATIONAL DEMAND ACTIVITIES AND THE USE OF ALTERNATIVE COUNT DATA MODELS

Abstract

Since the early 1990s, researchers have routinely used count data models (such as the Poisson and negative binomial) to estimate the demand for recreational activities. Along with the success and popularity of count data models in recreational demand analysis during the last decade, a number of shortcomings of standard count data models became obvious to researchers. This had led to the development of new and more sophisticated model specifications. Furthermore, semi-parametric and non-parametric approaches have also made their way into count data models.

Despite these advances, however, one interesting issue has received little research attention in this area. This is related to the fast decay process of the dependent variable and the associated long tail. This phenomenon is observed quite frequently in recreational demand studies; most recreationists make one or two trips while a few of them make exceedingly large number of trips. This introduces an extreme form of overdispersion difficult to address in popular count data models. The major objective of this paper is to investigate the issues related to proper modelling of the fast decay process and the associated long tails in recreation demand analysis. For this purpose, we introduce two categories of alternative count data models. The first group includes four alternative count data models, each characterised by a single parameter while the second group includes one count data model characterised by two parameters. This paper demonstrates how these alternative models can be used to properly model the fast decay process and the associated long tail commonly observed in recreation demand analysis. The first four alternative count data models are based on an adaptation of the geometric, Borel, logarithmic and Yule probability distributions to count data models while the second group of models relied on the use of the generalised Poisson probability distribution.

All these alternative count data models are empirically implemented using the maximum likelihood estimation procedure and applied to study the demand for moose hunting in Northern Ontario. Econometric results indicate that most of the alternative count data models proposed in this paper are able to capture the fast decay process characterising the number of moose hunting trips. Overall they seem to perform as well as the conventional negative binomial model.and better than the Poisson specification. However further investigation of the econometric results reveal that the geometric and generalised Poisson model specifications fare better than the modified Borel and Yule regression models.

Keywords : fast decay process ; recreational demand; count data models ; Borel, Yule, logarithmic and generalised Poisson regression models.

THE FAST DECAY PROCESS IN OUTDOOR RECREATIONAL ACTIVITIES AND THE USE OF ALTERNATIVE COUNT DATA MODELS

INTRODUCTION

To measure the nonmarket values of various recreational activities using the travel cost method, economists generally use annual demand for trips. The data for individual recreationists are often collected from recreation sites where only the number of participants consuming positive quantities are represented. Data for non-participants or for participants consuming zero quantities are not readily available. Since observed trips are nonnegative and occurs in integer quantities, the dependent variable is truncated and censored. Failure to address censoring and truncation issues in econometric analysis can lead to biased estimates.

Early attempts to address these issues include estimation of continuous demand models with truncated error distributions (Shaw). Smith and Kaoru, and Hanneman estimated random utility models in which recreational choice is represented as purely discrete. Finally, Heckman and Bocksteal *et al.* made attempts to combine continuous and discrete models to address truncation and censoring. In particular, discrete models were used to predict the probability of participation while the continuous models were used to estimate the quantity demanded of the selected goods or services, given participation. Problems encountered in above attempts to address truncation and censoring along with the realization that demand for recreation trips can be modelled more parsimoniously as a non-negative integer valued variable, led researchers to employ count data models (Smith, Hellerstein and Mendelsohn). Since the early 1990s, researchers have routinely used count data models such as the Poisson, negative binomial (denoted negbin hereafter) models to estimate the demand for recreational fishing (Grogger and Carson, Woodward *et al.*), big game hunting (Creel and Loomis, 1990; Yen and Adamowicz, Offenbach and Goodwin, and Sarker and Surry, among others), water fowl hunting (Cooper, Cooper and Loomis), recreational boating (Ozuna and Gomez), canoeing (Hellerstein), hiking (Englin and Shonkwiler), whitewater rafting (Bowker *et al.*) and rock climbing (Shaw and Jakus).

Along with the success and growing popularity of count data models in recreational demand analysis during the last decade, a number of their shortcomings became obvious to researchers. The major inadequacies relate to the problem of treating zeros, institutional constraints, visitation of multiple sites and over-dispersion adequately in count data models (Habb and McConnell; Creel and Loomis, 1992; and Hausman *et al.*, 1995).

In a typical recreation demand application of the benchmark Poisson model, the estimated model underpredicts the true frequency of zeroes, overpredicts the true frequency of other small values and underpredicts the true frequency of large counts. A manifestation of this phenomenon is the existence of a variance larger than its mean. This is well recognized in the count data literature as overdispersion and is caused by some form of unobserved heterogeneity in population parameter. Three alternative approaches have been used to capture different forms of heterogeneity in the Poisson model by allowing the variance of the distribution to vary across counts. Following the parametric tradition, King, and Winkelmann and Zimmermann proposed generalized count data models developed by exploiting the properties of the Katz family of probability distributions to tackle the problem of overdispersion. Note that these generalizations modify only the variance function but not the conditional mean. Recently, Cameron and Johansson proposed another parametric approach that simultaneously affects the specification of all conditional moments. In particular, they consider generalisations of the Poisson count data model based on a squared polynomial series expansion which permits flexible modelling of conditional moments and allows us to escape the restrictive framework of commonly used parametric count data models. Developments in nonparametric and semi-parametric econometrics during the last decade have also made their way into count data models. For example, Gurmu *et al.* proposed a specification in which the distribution of the variance is estimated non-parametrically using Laguerre series expansion estimators. The Laguerre polynomials are useful for count data models because they are based on gamma random variables commonly used in parametric models. Gurmu extends the methods proposed in Gurmu *et al.* to the case of hurdle count models. Hurdle models are considered as refinements of models with truncation and censoring. While parametric variants of hurdle count models have been very useful in empirical work for handling 'excess zero' problem, the treatment of unobserved heterogeneity has been problematic in this class of model. Gurmu's analysis shows how one can incorporate additional functional form flexibility by using series expansions to model unobserved heterogeneity. The proposed method nests the Poisson and negative binomial hurdle models and permits non-Gamma distributions for unobservables. Finally, Cooper proposed two nonparametric approaches, the pool adjacent violators and the kernel smoothing, to travel cost analysis of recreation demand.

These developments in count data models are, indeed, exciting and have contributed to their growing popularity in recreation demand analysis. Despite these advances, however, one interesting issue has received little research attention. This is related to the fast decay process of the dependent variable and the associated long tail. This phenomenon is observed quite frequently in recreational demand studies; most recrationists make one or two trips while a few recreationists make exceedingly large number of trips[1] This introduces an extreme form of overdispersion very difficult to address in popular count data models. The major objective of this paper is to investigate the issues related to proper modelling of the fast decay process and the associated long tails in recreation demand analysis. Although nonparametric approach makes no precise assumptions about functional form and allows the data to 'speak for themselves', good estimates of a nonparametric model can be obtained only with a very large amount of data (Delgado and Robinson). While semi-parametric approach provides a compromise between parametric and nonparametric approaches and can reduce the potential for misspecification, it requires complex and delicate modelling efforts and careful fitting to the data. Even with careful modelling, the interpretation of the results remains open (Creel). Moreover, Cooper's empirical analysis of waterfowl hunting shows that with proper econometric specification the parametric approach generates more reliable results than the semi-parametric or nonparametric approaches to recreation demand analysis.

[1] The fast decay phenomenon can be noted in the recreational data set used in the Ozuna and Gomez's study. The total number of trips for recreational boating range from zero to 88 while the frequency of trips (which totals 659) goes from a high 417 observations for zero trips to 68 and 38 respectively for one and two trips. A few receationists, however, took 8 or more boating trips. As a result, the mean is equal to 2.34 and its standard deviation is 6.29

In light of these observations and of the fact that the data sets available for recreation demand analysis are often small, we employ the parametric approach to address the fast decay process and the associated long tail of the distribution. To this end, we introduce three categories of alternative count data models. Included in the first category is a generalisation of the negative binomial regression model in which the variance is posited to be an increasing and non-linear function of its conditional mean. The second group includes four alternative count data models, each characterised by a single parameter. The final group includes three types of hybrid generalized Poisson models.

The inadequacies of the conventional count data models to deal with the issues related to the fast decay process are highlighted in Section 2. This section also focuses on appropriate generalisation of the negbin specifications to accommodate the fast decay process. The other alternative count data models capable of capturing the features of the fast decay process along with their basic properties are presented in Sections 3 and 4. Section 5 addresses various estimation issues and modelling questions (i.e., model performance and selection) encountered in the empirical implementation of these alternative count data models. Section 6 reports the results of an empirical application of proposed alternative count data models and discusses their policy implications. The final section of the paper summarizes the major findings of this study and offers some concluding remarks.

FAST DECAY PROCESS AND CONVENTIONAL COUNT DATA MODELS

A common observation in recreation demand studies is that a vast majority of the participants make at least one or two trips and the number of recreational trips higher than two falls rapidly. However, only a few overly enthusiastic recreationists make exceedingly large number of trips. Such idiosyncratic behaviour of recreationists generates trip data with some special features; the frequency of trips fall sharply after one or two trips but the distribution contains a long tail. This is called the fast decay process. As a result, the variance will be greater than the mean (over-dispersion) and is likely be an increasing (and possibly non-linear) function of its mean. We revisit the popular count data models - i.e. Poisson and the negbin models - and comment on their ability to capture the fast decay process and the associated over-dispersion. Then we examine a ″generalized″ form of the negbin model that is capable of capturing the fast decay process.

The most widely used single parameter count data model in recreational demand analysis is the Poisson distribution. The basic Poisson model assumes that Y_i, the *ith* observation of the number of recreational trips follows a Poisson distribution given by

$$\Pr ob(Y_i = k) = \frac{e^{-l} l^k}{k!} \qquad (1)$$

where l is the Poisson parameter to be estimated and $k = 0, 1, 2\ldots n$.

A count data regression based on the Poisson distribution is specified by letting l to vary over observations according to a specific function of a set of explanatory variables. The most commonly used specification for l is $\lambda_i = \exp(\boldsymbol{X'_i} b)$ where $\boldsymbol{X_i}$ is a matrix of explanatory variables including a constant and b is a conformable vector of unknown parameters to be estimated. The basic Poisson model captures the discrete and nonnegative nature of the dependent variable and allows inference on the probability of trip occurrence. However, this specification also implies that the variance of the distribution is equal to its mean. This is a restrictive property not often met in reality. In particular, when the dependent variable is characterised by a fast decay process, the so-called equidispersion property of the Poisson distribution is flagrantly violated. If this is not recognised and accounted for in modelling demand for recreation, the estimated parameters will be biased and inconsistent (Grogger and Carson). The Poisson distribution admits the fast decay process only when the estimated value of the parameter l is less than one.

An alternative to the Poisson model has been proposed about two decades ago by Hausman *et al.* (1984) to deal with over-dispersion in count data models. This alternative can be justified on the grounds that measurement errors and/or omission of explanatory variables could introduce additional heterogeneity and hence, over-dispersion in the data. Under these conditions, it can be assumed that the dependent variable is measured with a multiplicative error capturing unobserved heterogeneity and this error term is uncorrelated with the explanatory variables. If the error term, e_i, follows a Gamma distribution, a two-parameter negative binomial model can be defined as:

$$\Pr ob(Y_i = k; k = 0,1,2,...n) = \frac{\Gamma(k+n)}{\Gamma(k+1) \bullet \Gamma(n)} \bullet \left[\frac{n}{n+l}\right]^n \bullet \left[\frac{l}{n+l}\right]^k \qquad (2)$$

The expected value and the variance of this distribution are λ and [λ+λ^2/ν], respectively. The parameter *n* is non-negative and called the precision parameter. Note also that the variance is a quadratic function of its mean. To make sure that the mean λ is non-negative, the model is parameterised by assuming λ_i =exp($\boldsymbol{X_i}'$β) where $\boldsymbol{X_i}$ is a vector of explanatory variables. A wide range of model specifications can be generated by setting the parameter ν as a function of the explanatory variables, $\boldsymbol{X_i}$, such that:

$$n_i = \frac{(l_i)^m}{a} = \left[\frac{1}{a}\right]\left[\exp(X_i'b)\right]^m \quad \forall \;\; \alpha > 0 \tag{3}$$

where *m* is an arbitrary constant. By replacing n_i in the variance by equation (3) results in a generalised form of the variance such as (Cameron and Trivedi, 1987):

$$Var(Y_i|X_i) = E[Y_i|X_i] + aE[Y_i|X_i]^{2-m} = l_i + a \bullet (l_i)^{2-m} \tag{4}$$

The associated probability distribution is now a "generalised" density function given by

$$\Pr ob(Y_i = k; k = 0,1,2,...) = \frac{\Gamma\left[k + \left[\frac{(l_i)^m}{a}\right]\right]}{\Gamma(k+1) \bullet \Gamma\left[\frac{(l_i)^m}{a}\right]} \bullet \left\{ \frac{(a\,l_i)^k (l_i)^{\frac{m(l_i)^m}{a}}}{\left[(l_i)^m + a\,l_i\right]^{(\frac{(l_i)^m}{a}+k)}} \right\} \tag{5}$$

A closer look at expressions (4) and (5) reveals that different forms of overdispersion can be captured in this model depending on the values taken by the parameters *m* and *a*. Moreover, it provides a convenient formulation for nesting popular count data models through linking the conditional mean and variance of the dependent variable as discussed below.

- A value of a = 0 yields the Poisson model where variance and mean are equal.

- If m=1, $Var(Y_i | X_i)=E[Y_i | X_i] \bullet (1+\alpha)=\lambda_i \bullet (1+\alpha)$. This specification is called negbin type I. It assumes a constant relationship between conditional mean and variance.

- When m = 0, the precision parameter ν_i is a constant and equals to $1/a$. The variance of the distribution is equal to $l \cdot (1+a\, l)$. This specification is known as negbin type II.

- A fast decay process is obtained when the parameter a assumes values greater than or equal to one.

- When $m \neq 1$ and/or $\neq$ 0, we have several types of specification to represent over-dispersion. For example, when m < 1, the conditional variance increases with the mean at an increasing rate. On the other hand, when 1 < m < 2, the variance increases with the mean but at a decreasing rate. When m > 2, the derivative of $Var(Y_i | X_i)$ with respect to the conditional mean becomes negative[2] and the conditional variance becomes a decreasing function of the mean $E(Y_i | X_i)$.

From the above, it can be seen that a fast decay process can be captured in a negbin binomial model with $a \geq 1$ and m < 1. This explains why the negbin II specification provides a better representation of over-dispersion than negbin I and has been extensively used for modelling the demand for recreational activities. However, the "generalized" version of this model presented in equation (5) is yet to receive wide application because its probability distribution and the associated likelihood function are both highly non-linear with respect to the parameters. Such nonlinearities make it difficult to obtain convergent estimates of parameters from this model (Saha and Dong).

ALTERNATIVE COUNT DATA MODELS AND ONE-PARAMETER PROBABILITY DISTRIBUTIONS

A set of four alternative count data models and their basic features are presented in this section. Each model is characterized by one-parameter probability distribution and is suitable for representing the fast decay process and the long tail. These models are

---

[2] $\frac{\partial Var(Y_i)}{\partial E(Y_i)} = 1 + a(2-m)\, l_i^{1-m}$. This first order derivative is negative for $\lambda_i < [\{-1\} / \{\alpha(2-m)\}]^{1/(1-m)}$

presented in the same spirit as the original introduction of count data models in recreational demand analysis.

The geometric distribution

The geometric probability distribution is characterised by the parameter λ assumed to be positive and can be adopted to analyse recreational demand (Mullahy). Its density function can be obtained as a special case of expression (2) because it is equivalent to the negbin probability distribution when the its precision parameter n is equal to one. The mean and variance of this distribution are l and $l \cdot (1+l)$, respectively (see Appendix 1). Since the variance is a quadratic function of the mean, the geometric distribution allows for over dispersion in data and can represent the fast decay process. The left-truncated geometric model can be specified as:

$$\Pr ob(Y = k; k = 0,1 \ldots n \,|\, k > 0) = \frac{\Pr ob(Y = k)}{1 - \Pr ob(Y = 0)} = \frac{l^{k}(1+l)^{-(k+1)}}{(1-(1+l)^{-1})} = l^{k-1}(1+l)^{-k} \qquad (6)$$

The geometric distribution is unimodal for k = 0. The model is parameterised as l_i=exp($\boldsymbol{X_i}'b$), where $\boldsymbol{X_i}$ is the matrix of explanatory variables.

Borel distribution

Another count data model capable of capturing the fast decay process is the Borel distribution (see Appendix 1). It was originally developed in the context of queuing theory[3]. The Borel distribution cannot be used in its original form to model the demand

[3] Conceptualised first by Borel and then extended by Tanner, this probability distribution described the distribution of the total number of customers served before a queue vanishes given a simple queue with random arrival times of customers (at constant rate) and a constant time occupied in serving each customer (Johnson *et al.*). To the best of our knowledge, none has previously applied the Borel distribution in recreational demand analysis.

for recreational trips because it admits only positive values for the random variable (i.e., the number of trips) and thus excludes zeroes. This could be overcome by shifting the Borel distribution to the left so that it supports *0, 1, 2…*(i.e. by obtaining $Z = Y-1$). The resulting probability distribution can be called a modified Borel distribution and defined as follows:

$$\Pr ob(Z = k; k = 0,1,....n) = \frac{(k+1)^{k-1} l^k \exp[-l(k+1)]}{k!} \tag{7}$$

where the parameter l is positive and smaller than one. The modified Borel probability distribution is unimodal for $k = 0$. The left-truncated modified Borel distribution is as follows:

$$\Pr ob(Z = k; k = 0,1,....n | k > 0) = \frac{(k+1)^{k-1} l^k \exp[-l(k+1)]}{k![1-\exp(-l)]} \tag{8}$$

The mean m of this distribution is $l/(1-l)$, while the variance is equal to $\lambda/(1-\lambda)^{3.}$ or to $m(1+m)^2$ when expressed in terms of m. The variance of the modified Borel distribution is then a polynomial of degree three of its mean and thus it can allow a richer kind of over-dispersion and a better representation of the fast decay process than the geometric model. To model the recreational demand for trips, a modified Borel regression can be parameterised such that $l=1/(1+\exp(-\boldsymbol{X}'b))$ where $\boldsymbol{X}$ is a matrix of explanatory variables.

The logarithmic distribution

The third alternative is based on the logarithmic distribution developed by Fisher *et al.*. The density function and the basic characteristics of this distribution are presented in Appendix 1. The logarithmic distribution is unimodal for $Y = 1$ and it is the limiting distribution of a left-truncated negative binomial distribution when the precision parameter v approaches 0 (or α tends to +∝). In its original form, this distribution also excludes zero values. Consequently, it cannot be used for modelling the demand for recreational activities. For the purpose of this paper, we use a random variable, $Z = Y - 1$, and develop a modified logarithmic distribution, which can be written as:

$$\Pr ob(Z = k; k = 0,1,..n) = \frac{-\,l^{k+1}}{(k+1)\ln(1-l)} \tag{9}$$

where, l is the parameter of this distribution. It is positive but smaller than one.

The mean of the modified logarithmic distribution, *E(Z)* is given by

$$E(Z) = E(Y) - 1 = \frac{l\left[\ln(1-l) - 1\right] - \ln(1-l)}{(1-l)\ln(1-l)} \tag{10}$$

while the variance is linked to its mean through the following non-linear relationship:

$$E(Z)= [E(Z)+1]^2*[-\ln(1-l)-1] \tag{11}$$

Unlike its original form, the modified logarithmic distribution is characterised by a variance that is always greater than the mean regardless of the values taken by λ over the range [0, 1]. The left-truncated probability density for the modified logarithmic model can be defined as:

$$\Pr ob(Z = k; k = 0,1,....n \mid k > 0) = \frac{\Pr ob(Z = k)}{1 - \Pr ob(Z = 0)} = \frac{-\,l^{k+1}}{(k+1)[\ln(1-l) + l]} \tag{12}$$

While the logarithmic distribution can handle both under- and over-dispersed data generating processes, the over-dispersion is satisfied only when the variance to mean ratio is greater than unity. The later condition is readily satisfied by the modified logarithmic distribution.

Assuming that the parameter of the distribution λ is a logistic function of the explanatory variables, the mean of the modified logarithmic distribution can be derived as

$$E(Z) = \frac{\exp(X'b)}{\log[1+\exp(X'b)]} - 1 \tag{13}$$

The Yule distribution

The last one-parameter count data model suggested in this paper is based on the Yule distribution (see Appendix 1). Like the Borel and logarithmic probability distributions, the Yule distribution in its original form does not accommodate zero values and hence cannot be employed to analyse the demand for recreational trips. To overcome this problem, we shift the Yule distribution to the left so that it has support *0, 1, 2,..n* (i.e. by obtaining *Z* = *Y*-1). The resulting probability distribution is a modified Yule distribution defined as

$$\Pr ob(Z = k; k = 0,1,....n) = h\mathrm{B}(k+1, h+1) = \frac{h\Gamma(k+1)\Gamma(h+1)}{\Gamma(k+h+2)} \tag{14}$$

Where, *B(.)* and *G(.)* are the Beta and Gamma functions, respectively and the parameter h is greater than one. The modified Yule distribution is unimodal for k = 0 (Johnson *et al.*). The mean is now equal to E(Z)=1/(h-1) while its variance is linked to its mean as follows:

$$V(Z) = s^2 = \frac{[E(Z)+1]^2}{1-E(Z)} \tag{15}$$

It can be seen from equation (15) that the variance exists only if the mean is smaller than one[4]. This property of the modified Yule distribution can be viewed as a weakness in case of recreational demand analysis because the average number of trips is usually greater than 1. To model the recreational demand for trips, a modified Yule regression can be parameterised such that the distribution parameter[5] *h* is equal to exp(-***X'b***)+1 where ***X*** is the matrix of explanatory variables.

Although the Borel, logarithmic and Yule distributions have been modified to capture zero counts, these distributions can also be used in their original form to estimate the demand for recreational activities. If the recreation decisions can be viewed as an outcome of a two-stage decision making process where the consumer decides first whether or not to participate and then decides how many trips to take. Following Haab and McConnell, such a decision making process can be modelled as:

$$\Pr ob(Y_i) = \begin{cases} w_i \;\; for\, k = 0 \\ (1-w_i)g(k) \;\; for\, k = 1, 2, \ldots n \end{cases} \tag{16}$$

where w_i is an indicator that represents the participation decision; w_i is a function of the variables affecting the participation decision and its domain lies between zero and one. The density function *g(k)* can be approximated by a Borel, logarithmic or a Yule distribution. The expected value of *Y* is then $(1-w_i)E(Y)$. The above model can also be viewed as a Borel, logarithmic or Yule probability distribution with added zeroes and w_i is a varying parameter of some of the explanatory variables.

[4] See also the Cauchy probability distribution for which the variance is not also defined.

[5] We implicitly assume that the parameter *h* is equal to the inverse of the logistic function of the explanatory variables.

Each of the four count data models presented above captures over-dispersion through a variance that is an increasing function of the mean. Note, however, the relationship between the mean values and the variance differs across the models. A simple numerical simulation was performed to obtain the probability distributions associated with each of the four count data models presented above. Figure 1 provides a pictorial view of the results on a linear scale while Table 2 reports some useful indicators defining the shape of each probability distribution. A number of interesting features emerge from Figure 1 and Table 1. First, irrespective of the mean values, the modified Yule has the highest mode, followed by the modified Borel, modified logarithmic and geometric distributions. Second, the nature of the decay process is more pronounced in case of modified Borel, logarithmic and Yule distributions than in case of a geometric distribution.

It is not difficult to implement the four count data models for studying recreational demand and it is also relatively easy to generate the estimates of consumers' surplus per trip for the geometric, modified Borel and modified Yule models. Due to specific parameterisation of the regression models, each yields a semi-logarithmic demand function. The associated consumer surplus per trip is equal to -1/price coefficient. However, the derivation of the consumer surplus for the modified logarithmic count data model involves a more complicated procedure[6].

[6] For this model, expression (14) defining the demand for recreational activities is quite complex at first sight because it involves a ratio of an exponential function over a logarithmic function. The computation of the consumer surplus, therefore, requires some special treatment using the *LogIntegral* function. For this purpose, we derive the following expression of consumer surplus for this distribution

$$CS_i = \int_P^{Pch}\left[\frac{\exp(a+bp)}{Log[1+\exp(a+bp)]}-1\right]dP_i = \left[-P+\frac{LogIntegral(1+\exp(a+bp))}{b}\right]_P^{Pch}$$

THE GENERALISED POISSON DISTRIBUTION AS AN ALTERNATIVE COUNT:

An assumption implicit in most count data analysis is that the occurrence of one count is independent of the occurrence of another count. While this may be a reasonable assumption for modelling many physical processes, it is not so in social sciences. For example, the independent occurrence assumption may not be plausible when one is dealing with the number of visits to a doctor (Pohlmeier and Ulrich) or the number of trips to a recreation site (Creel and Loomis, 1990; Grogger and Carson). Since the generalised Poisson distribution allows for the probability of an event to depend on the number of events already occurred (Consul and Shoukri), this distribution may be particularly useful in recreational demand analysis. Introduced by Consul and Jain (1973a, 1973b), the generalised Poisson distribution has recently been used in a regression context by Consul and Famoye, 1992; Famoye and Santos Silva. We concentrate on only those aspects of the generalised Poisson (GP) distribution relevant for recreational demand analysis.

Following Consul, the generalised Poisson distribution can be defined as:

$$\text{Prob}(Y = k;\ k = 0, 1, 2\ldots..n) \;=\; \begin{cases} \dfrac{l\,(l + dk)^{k-1} \exp\left[-(l + dk)\right]}{k!} \\ 0 \;\; for \;\; y > m \;\; when \;\; d < 0 \end{cases} \tag{17}$$

6(continued) where P_{ch} is the choke price defined as the price at which the quantity demanded approaches zero. *LogIntegral* is defined as follows:

$$LogIntegral(z) = \int_0^z \frac{dt}{Log(t)}$$

The consumer surplus per trip is then obtained by dividing $CS(P_i)$ by the expected number of trips.

where, $l > 0$, max (-1, $-l/m$) < d < 1 and $m \geq 4$ is the largest positive integer for which $l + d\ m > 0$ when d is negative. The mean and variance of the generalised Poisson distribution are m = E(Y) = $\lambda/(1-\delta)$ and s^2 = V(Y) = $\lambda/(1-\delta)^3 = \mu/(1-\delta)^2$, respectively. Note that the variance is greater than, equal to, or less than the mean if d is positive, zero or negative. Moreover, when δ is positive, both the mean and variance increase as the value of d increases. However, the variance increases faster than the mean. This property is very useful in recreational demand studies where the dependent variable is characterised by over-dispersion. The GP distribution also admits under-dispersion and equidispersion.

Using the fact that $m = l/(1-d) = l\,r$, it is also possible to express the GP distribution as a function if its mean. The resulting distribution is as follows:

$$Prob(Y = k;\ k = 0,1,2...,\ n) = \begin{cases} \dfrac{m\left[m + (r-1)k\right]^{k-1} r^{-k}\ e^{-[((m+(r-1)k)/r]}}{k!} \\ 0 \ for\ \ y > m\ \ when\ \ r < 1 \end{cases} \qquad (18)$$

where, $r \geq$ max (1/2, $1-m/4$) and m is the largest positive integer for which $m+m\ (r-1) > 0$ when r is less than one. The variance is given by σ^2 = V(Y) = $\rho^2 \bullet \mu$. When r = 1, the GP is equivalent to the Poisson model while the modified Borel probability distribution is obtained if $r = 1/(1-l)$ or if $r = (1+m)$. Any values of $r > 1$ represents count data process with over-dispersion and $0.5 \leq r < 1$ characterises count data with under-dispersion when $m > 2$. Note that, the variance is proportional to its mean, thus implying a constant variance to mean ratio (like in the case of negbin I). This is not a desirable property for a model to capture the fast decay process.

To gain additional insights about the ability of the GP model to capture over-dispersion associated with the fast decay process, a simulation exercise was performed for the one-parameter GP probability distribution using different values of the mean, m and of the parameter $r \geq 1$. Figure 2 presents a graphical representation of these results. As expected, the unimodality of the GPD is preserved for a value of r equal to one and a mean, $m \leq 0.5$. This result is not surprising because this case corresponds to the standard Poisson distribution. Secondly, the graphical results show that the L-shaped distribution is well represented for $m = 0.5$ and $r \geq 1$. It appears from this simulation exercise that the GP Poisson distribution admits a fast decay process only under some restrictive conditions[7].

To overcome this problem, a restricted version of the generalised Poisson (denoted RGP hereafter) distribution can be formulated by making the parameter d proportional to λ, such that $d = al$. Substituting this expression for d in expression (17) yields the following RGP distribution:

$$\Pr ob(Y = 0; k = 0,1,....n) = \left\{ \begin{array}{c} \dfrac{l^k (1+ak)^{k-1} \exp\left[-l(1+ak)\right]}{k!} \\ 0 \quad for \;\; x > m \;\; when \;\; a < 0 \end{array} \right\} \quad \text{for } k = 0, 1, 2\ldots n \qquad (19)$$

[7] As demonstrated by Consul and Famoye (1986), the GP distribution is unimodal for all values of r at k=0 if $m < r \exp[(r-1)/r]$. This condition is satisfied in terms of the original parameters of the GP distributon if $l < \exp(-d)$.

The domain of the parameter a is given by max $(-l^{-1}, -1/4) \leq a \leq l^{-1}$ (Famoye). If $a = 0$, the RGP distribution reduces to the Poisson distribution while for $a = 1$ and $l < 1$, we get a modified Borel distribution. The mean and variance associated with the RGP distribution are: $\mu = E(Y) = \lambda/(1-\alpha\lambda)$ and $\sigma^2 = V(Y) = \lambda/(1-\alpha\lambda)^3 = \mu \bullet [1 + \alpha\ \mu]^2$ respectively.

An alternative specification of the RGP distribution can also be obtained if the parameter l is expressed as a function of the mean (m). This yields a one-parameter probability distribution such as:

$$\Pr ob(Y = 0; k = 0,1,\ldots n) = A^k \frac{(1+ak)^{k-1}\exp\left[-A(1+ak)\right]}{k!} \tag{20}$$

where $A = \dfrac{m}{1+am}$.

Over-dispersion is obtained when $a > 0$. It is interesting to note that the variance of this model is a third degree polynomial function of its mean. This permits a richer type of over-dispersion and is likely to model the fast decay process efficiently. Figure 3 provides a graphical representation of the one-parameter RGP distribution for a wide range of values of the mean m and the parameter a. Clearly, we obtain well-defined L-shaped distribution regardless of the values of the mean, m and the parameter a. Therefore, the restricted generalised Poisson distribution can represent the fast decay process.

If we assume that the mean, m is an exponential function of the explanatory variables so that $\mu_i = \exp(\boldsymbol{X_i}'b)$, then it is possible to define generalised Poisson regression models which can be estimated either in restricted or unrestricted forms (Consul and Famoye, 1992; Famoye). Santos Silva has shown that the two (restricted and unrestricted) forms of the GP regression model can be nested through a hybrid generalised Poisson model. To do so, the parameter a is linked to the covariates $\boldsymbol{X_i}$, so that a_i = $\alpha_i(\boldsymbol{X_i}, q, b)$= q_0 exp[q_1 ($\boldsymbol{X}'b$)] and that m_i =exp($\boldsymbol{X_i}'b$). Incorporating these expressions in equation (20), a hybrid generalised Poisson (denoted HGPI) regression model can be defined as:

$$\Pr ob(Y = k; k = 0,1,...n) = [A(X,b,q)]^k \bullet \frac{[1+a(X,q,b)k]^{k-1} \exp[-A(X,b,q)\bullet(1+a(X,b,q)k)]}{k!} \tag{21}$$

where A(X,$\boldsymbol{b}$, q) is now equal to $\dfrac{\exp(X'\boldsymbol{b})}{1+q_0\exp[(1+q_1)(X'\boldsymbol{b})]}$

A closer look at expression (21) reveals that, depending upon the form taken by the function $\alpha(\boldsymbol{X},q,b)$, the following model specifications can be obtained as special cases.

- When $\alpha_i(\boldsymbol{X_i}, q, b)$ is a constant (when q_1 = 0), one obtains the restricted GP model(denoted HGPII)[8]. In addition, if q_0 =1, the model is reduced to the modified Borel regression model.

- If $\alpha_i(\boldsymbol{X_i}, q, b)$ is proportional to exp(-$\boldsymbol{X_i}'b$) (which is obtained when q_1 = -1), we obtain the GP regression model(denoted HGPIII)[9].

- Finally, if $\alpha_i(\boldsymbol{X_i}, q, b)$ is equal to zero (which is obtained when q_0 = 0), we obtain the standard Poisson regression model.

A left- truncated HGP model could also be defined by adjusting expression (21) with the prob(Y_i > 0) setting equal to 1-exp[-A($\boldsymbol{X}$,b, q)]. The resulting left- truncated HGP model is:

$$\Pr ob(Y = k; k > 0) = [A(X,b,q)]^k \bullet \frac{[1+a(X,b,q)k]^{k-1} \exp[-A(X,b,q)\bullet(1+a(X,q,b)k)]}{(k!)\bullet[1-\exp[-A(X,b,q)]]} \tag{22}$$

[8] q_0 is then equal to the parameter a in expression (20).

HGP demand models for recreational activities are obtained through $\mu = \exp(\boldsymbol{X'b})$ with the covariates $\boldsymbol{X}$ including price (i.e., travel costs). It is also interesting to note that the consumer surplus per trip obtained from this HGP model is equal to -1/(price coefficient). This is a major advantage of the HGP models.

ESTIMATION ISSUES AND MODEL EVALUATION

This section deals with two important aspects of practical application of alternative count data models. First, how to obtain coefficient estimates from each model that are consistent and unique. Second, how to evaluate the performance of one model relative to others.

The first issue is related to whether or not a closed form and well-behaved (log) likelihood function can be obtained for each model. These in turn, depend on if the log likelihood function is globally optimal for each model. A cursory look at the mathematical expressions of the log likelihood functions of alternative count data models presented in appendix 2 seem to suggest that obtaining globally optimum parameter estimates for some models may not be possible because their log likelihood functions are highly non-linear. However, this impression must be tempered because most of these models have been studied thoroughly by statisticians. For example, Consul and Famoye (1992), Famoye and Santo da Silva addressed these estimation issues for the unrestricted and restricted versions of the generalised Poisson model and found that the maximum likelihood estimation (ML) estimation procedure yields efficient parameter estimates. The results also hold for the modified Borel regression specification because it is nested

---

[9] In this case, q_0 is equal to ρ-1 and replacing it in (21) yields expression (18).

in the generalized Poisson model. The ML estimation of the geometric model is also straightforward since it is nested in negbin model for which the ML estimator has been extensively studied in the econometric literature[10] To obtain global convergence of the log likelihood function of the modified logarithmic regression models we relied on Johnson *et al.* (p. 294). For the modified Yule regression model, the first and second derivatives of the log likelihood function with respect to b produced highly non-linear expressions that required to be solved through the Newton-Raphson gradient algorithm.

Once the ML estimates of the parameters are obtained, negative inverse of the matrix of second derivatives of the log likelihood function with respect to the parameters can be used to estimate the asymptotic variance-covariance matrix of the parameters. These estimates can be used to form Wald (W) or likelihood ratio (LR) tests for testing relevant

---

[10] As pointed out earlier; due to high nonlinearities, estimating the ″generalised″ negbin model is frown with difficulties. However, adopting the iterative procedure suggested by Dong and Saha eased somewhat this task, allowing to obtain global convergence of the underlying maximum likelihood function and to generate ML estimates of parameters associated with the ″generalised″ negbin model. This procedure consists of the following steps (Dong and Saha, p. 426):

i) obtain estimated values of the parameters a and m in expressions (4) and (5) by running the following equation using non linear least squares :

$$\frac{\left(y_i - \hat{l}_i\right)^2 - y_i}{\hat{l}_i} = a\left(\hat{l}_i\right)^{1-m} + u_i$$

where $\hat{l} = \exp(X'\hat{b})$ with $\hat{b}$ being the estimate of b from the Poisson model, y_i is the observed number of trips while u_i is the error term. The estimates of a and m are used as starting values in maximising the log likelihood function of the″generalised″ negbin model .

ii)holding the parameters a and m at their estimated values, optimise the log likelihood function of the″generalised″ negbin model with respect to the parameters b and obtain estimated values of b, denoted by $\hat{b}$.

iii) using $\hat{b}$, re-estimate expression (4) generating new values for a and m. Use these new estimates of a and m and $\hat{b}$ as starting values to maximise the log likelihood function of the″generalised″ negbin model .

hypotheses. Econometric implementation of all the count data models has been conducted using the TSP program (Hall and Cummins).

The performance of all count data models proposed in this paper is evaluated using several indicators ranging from pseudo-R^2 to information-based statistics. Cameron and Trivedi (1997) provide a good overview of such indicators. The choice of indicators in this research has been influenced by two considerations; how each count data model fit individual observations (frequencies) and more importantly, how well each model the fast decay process. The overall goodness of fit of each model has been evaluated using the Chi-square test as well pseudo-R^2 . The Chi-square goodness-of-fit test is given by

$$c^2 = \sum_{j=1}^{J} \frac{\left((n\bar{p}_j - n\hat{p}_j\right)^2}{n\hat{p}_j} \tag{23}$$

where J is the number of cells, n is the number of observations, $\bar{p}_j$ is the observed relative frequency, and $\hat{p}_j$ is the estimated relative frequency (probability) of cell j. The pseudo-R^2 we use is the R^2_{LRT} measure proposed by Maddala (1983) and Magee (1990) and defined as follows:

$$R^2_{LRT} = 1 - \exp(-LRT / n) \tag{24}$$

where, n is the total number of observations and LRT is the likelihood ratio test statistic for the joint significance of slope parameters. This measure takes values between 0 and 1 and is invariant to units of measurement. It becomes larger as the goodness of fit of the model improves. It is also a more general goodness-of-fit measure in the sense that R^2_{LRT} is equal to R^2_{OLS} in a linear model (Cameron and Windmeijer, 1997).

AN EMPIRICAL APPLICATION OF ALTERNATIVE COUNT DATA MODELS: THE DEMAND FOR MOOSE HUNTING IN NORTHERN ONTARIO

The data used in the illustrative applications of various count data models discussed earlier in this paper relate to the 1992 moose hunting season at the wildlife management Unit (WMU) #21A located in Northern Ontario. It is a popular WMU for moose hunting because of its remoteness and moose population density. During the 1992 season some 1286 hunters received moose validation tags to hunt an adult moose at this site and about 99% of these hunters were from Ontario[11]. Most of the data came from the Ontario Ministry of Natural Resources. Data include the number of moose hunting trips made by each hunter to the WMU #21A and the travel cost per hunter (which includes vehicle related costs, a licence fee of $26.50 per hunter per season, equipment costs, costs of food and lodging and time cost). The income variable consists of 1991 average employment income and other income at the Enumeration Area (EA) level. This information is based on 1991 census data and was adjusted to 1992 level using consumer price index (CPI). Further details on this data set can be found in Appendix 3.

The general specification of the travel cost model adopted for the *ith* moose hunter is:

$$Y_i = \exp(b_0 + b_1\, COST_i + b_2\, INCOME_i) \tag{25}$$

where the *bi's* are parameters to estimate, *COST* and *INCOME* represent respectively the travel cost (price) and income of moose hunter *i*. It is expected that $b_1 < 0$ and $b_2 > 0$. The dependent variable is the number of moose hunting trips taken to WMU21A and is

[11] About 200 hunters gave information on the number of hunting trips made to the WMU21A, duration of each trip, number of hunters in each group and the number of moose hunted. Several inconsistent responses were dropped annd the final sample includes 194 hunters

truncated at zero. During the 1992 hunting season, hunters in Ontario made 2.35 moose hunting trips on an average to the WMU21A. Note that the data exhibit a quick decay process; about 78% of the sample hunters made only one trip during this season and the number of trips higher than one falls rapidly. Hence, the ratios of the frequency of one trip over two and three trips are equal to 15.2 and 50.68, respectively. However, a few hunters made more than 10 trips to this hunting site. Moreover, the variance of the dependent variable is 12.89. Clearly, the equidispersion property of the Poisson distribution is not satisfied.

The results for recreational moose hunting trips in Ontario for all alternative left-truncated count data models proposed in this paper are presented in Table 2. In addition, Poisson and negbin II models have been estimated for comparative purposes. The standard errors of the coefficients were estimated using the Eicker-White procedure. This procedure generates heteroskedasticity-consistent variance-covariance matrix when the heteroskedasticity is of unknown form (White).

The econometric results indicate that in terms of explanatory power and/or goodness of fit all count data models proposed in this paper perform at least as well as the negbin II model. Thus, based on estimated values of R^2_{LRT}, the modified Borel model provides the best fit followed by the geometric and the two negative binomial models. The other alternative count data models have R^2_{LRT} ranging from 0.64 for the modified Yule model to 0.77 for HGPI model.

Based on the Chi-squared goodness-of-fit test, the null hypothesis that the demand for moose hunting in northern Ontario is represented by Borel and Yule regression models is rejected at 5% and 10% levels of error probability (Table 3). The null hypothesis is also rejected for the Poisson specification. On the other hand, test results confirm that two negative binomial, geometric and modified logarithmic specifications fit the data well. Only the HGPI and HGPII specifications could not be rejected at 10% level of significance.

The estimated price coefficients (b_1) have expected signs and are statistically significant in all cases. However, there is no uniformity in terms of its magnitude across alternative count data models. While the ″generalised″ negative binomial, geometric, modified logarithmic, HGPI and HGPII models have an estimated price coefficient similar in value to the one estimated for the negbin II specification, the estimated price coefficients are very different for the three remaining alternative count data models (Modified Borel, modified Yule and HGPIII). The estimated income coefficients (b_2) are not statistically significant at a 5% level of significance regardless of model specification and two of them even have the wrong signs.

An attractive feature of alternative count data models is that some of them can be nested with the others enabling us to test them using a LR ratio test or Wald test. A Wald test applied to the negbin II model shows that the estimated precision parameter (a) is not significantly different form one, indicating that the geometric and negbin II specifications yield similar results for this sample. Similarly, a Wald and/or LR ratio tests applied to the ″generalised″ negbin model reveals that the estimated parameters m and a are not significantly different from zero and one, respectively. This confirms the former result that we can accept a geometric model specification at the expense of the negbin II or its

″generalised″ version. LR tests applied to the HGP and Poisson models indicate that we can safely reject the HGPIII and Poisson specifications. On the other hand, we cannot reject the null hypothesis of an admissible restricted HGPII model (corresponding to the case where θ_1 = -1). Similarly, we cannot accept the null hypothesis that the restrictive HGPII specification is a modified Borel model. The results of these tests suggest that the geometric and the restricted generalised Poisson regression (HGPII) models are viable alternatives for capturing the fast decay process in the demand for moose hunting in Northern Ontario.

Finally, the reliability of alternative count data models can also be judged by looking at the estimated benefits they generate. For this purpose, estimated consumer surplus (CS) per moose hunting trip along with their standard errors and 95% confidence intervals computed for left-truncated count data models and reported in Table 4. The results indicate that for five out of eight alternative count data models the estimated consumer surplus vary from $CDN 168 to $CDN 203 per moose hunting trip. These values fall within the 95% confidence interval [$CDN 156 $CDN 216] obtained for the negbin II model. Finally, the estimated consumer surplus values are smaller for the modified logarithmic, modified Borel and modified Yule specifications and are less reliable.

CONCLUDING REMARKS:

Our analysis in this paper was motivated by the fact that a vast majority of the participants in many recreational activities make at least one or two trips. While the number of trips higher than two fall rapidly, a few overly enthusiastic recreationists make exceedingly large number of trips. Such behaviour of recreationists generates trip data with some special features; the frequency of trips fall sharply after one or two trips but the distribution contain a long tail. Despite considerable progress in count data modelling of recreational demand activities during past two decades, this issue has eluded researchers. The objectives of this paper were to address the issue of fast decay process in recreational demand activities more formally and demonstrate how it can be represented through appropriate count data models.

Although recent advances in semi- and nonparametric approaches could have been used to capture this phenomenon, we decided to investigate the issue through the parametric approach. Accordingly, we proposed a set of eight alternative count data models that can be used with some modifications to capture the fast decay process. Included in this set a generalisation of the negbin I and II regression models (in which the variance is an increasing and non-linear function of its conditional mean), geometric, Borel, logarithmic and Yule probability distributions and three different versions of the generalised Poisson distribution. The characteristics of the probability distributions and the log likelihood functions for each of these count data models have been studied and their ability to capture the fast decay process investigated through some simulation exercises. The results suggest that the alternative count data models mimic the fast decay process much better than conventional count data models.

Finally, an illustrative application of alternative count data models proposed in this paper is presented. The empirical application concentrates on the demand for moose hunting trips in Northern Ontario. The results suggest a satisfactory performance of five out of eight alternative count data models (including the ″generalised″ negbin, geometric and the three generalised Poisson specifications). The estimated benefits (measured by consumer surplus per hunting trip) obtained from these specifications are more reliable and can be compared to those obtained from the standard negative binomial model.

A number of directions for future research in this area can be suggested. First, it would be beneficial to apply these alternative count data models to other situations dealing with other recreational activities. Second, the performance of these models can be compared and contrasted to those of semi-parametric and non-parametric models. Finally, research is needed to develop a generalised regression framework that could allow nesting of all alternative count data models along with the traditional count data models (Famoye, and Kaufman Jr.).

REFERENCES

Bockstael, N.E., I.E. Strand, K.E. McConnell and F. Arsanjani. "Sample selection biais in the estimation of recreation demand functions: an application to sportfishing". *Land Economics,* 66 (1990): 40-49.

Borel E. "Sur l'emploi du théorème de Bernouilli pour faciliter le calcul d'une infinité de coefficients. Applications au problème de l'attente à un guichet". *Comptes Rendus de l'Académie des Sciences* . 214 (1942): 452-456

Bowker J. M., English D. B. K. and Donavan J. A. "Toward a value for guided rafting on southern rivers". *Journal of Agricultural and Applied Economics*. 28 (1996): 423-432.

Cameron, A. C. and P. K. Trivedi. "Econometric models based on count data : comparisons and applications of some estimators". *Journal of Applied Econometrics*. 1 (1987): 29-53.

Cameron A. C. and P. K. Trivedi. *Count data regression*. Cambridge University Press, 1997.

Cameron, A. C. and Windmeijer, F. a. G. "An R-squared measure of goodness of fit for some common nonlinear regression models". *Journal of Econometrics*. 77 (1997): 329-342

Cameron A. C. and P. Johansson. "Count data regression using series expansion with applications". *Journal of Applied Econometrics*. 12 (1997): 203-224

Consul, P. C. *Generalized Poisson Distributions: Pproperties and Applications*. Marcel Dekker Inc. New York, 1989.

Consul, P. C. and G. C. Jain. "A generalization of the Poisson distribution". *Technometrics*. 15 (1973a): 791-799

Consul, P. C. and G. C. Jain. "On some interesting properties of the generalized Poisson distribution". *Biometrische Zeitschrift*. 15 (1973b): 495-500

Consul, P. C. and Shoukri, M. M. "The generalised Poisson distribution when the sample mean is larger than the sample variance". *Communications in Statistics, B - Simulation and Computation* 14(1985): 667-681.

Consul, P. C., and Famoye F. "On the unimodality of Generalized Poisson distribution". *Statistica Neerlandica,* 40 (1986): 117-121.

Consul, P. C., and Famoye F. "Generalized Poisson regression model". *Communications in Statistics - Theory and Methods*. 21(1992): 89-109.

Cooper J. C. "Nonparametric and semi-nonparametric recreational demand analysis". *American Journal of Agricultural Economics*. 82 (2000): 451-462.

Cooper J. C. and Loomis J. "Testing whether Waterfowl Hunting Benefits Increase with Greater Water Deliveries to Wetlands". *Environmental and Resource Economics*. 3(1993): 45-61.

Creel M. "Welfare estimation using the Fourier form : simulation evidence for the recreational demand case". *Review of Economics and Statistics*. 78 (1997): 88-94

Creel, M. E. and J. B. Loomis. "Modeling hunting demand in the presence of bag limit, with tests of alternative specifications". *Journal of Environmental and Management*. 22 (1992): 93-113.

Creel, M. E. and J. B. Loomis. "Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California". *American Journal of Agricultural Economics*. 72 (1990): 434-441.

Delgado, M. and P. Robinson. "Nonparametric and sem-parametric methods for economics research". *Journal of Economic Surveys*, 6 (1992): 201-249.

Englin, J. and Shonkwiler, J. S. "Estimating Social Welfare Using Count Data Models: An Application to Long-Run Recreation Demand under Conditions of Endogenous stratification and Truncation". *Review of Economics and Statistics*. 77(1995): 104-12..

Famoye F. and E. H. Kaufman Jr. "Generalized negative binomial regression model". *Journal of Applied Statistical Science*. 1 (forthcoming).

Famoye F. "Restricted generalized Poisson regression model". *Communications in Statistics - Theory and Methods*. 22(1993): 1335-1354.

Fisher, R. A., Corbet, S. A. and William, C. B. "The relation between the number of species and the number of individuals in a random sample of animal population". *Journal of Animal Ecology*. 12 (1943): 42-58.

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

Gurmu, S. "Semi-parametric estimation of hurdle of hurdle regression models with an application to mediaid utilization". *Journal of Applied Econometrics*. 12 (1997): 225-243

Gurmu S., P. Rilstone and S. Stern. "Semi-parametric estimation of count regression models". *Journal of Econometrics*, 88(1999): 123-150.

Habb, T. C. and K. E. McConnell. "Count data models and the problem of zero in recreation demand analysis". *American Journal of Agricultural Economics*, 78 (1996): 89-103.

Hall, B. and Cummins C. Times Series Processor Version 4.5. Reference Manual. TSP International, Palo Alto, USA, 1999.

Hanneman,W. M. Discrete-continuous models of consumer behaviour. *Econometrica*, 52: 541-561.

Hausman, J. A., B. H. Hall and Z. Griliches. "Econometric models for count data with an economic application to the patent-R&D relationship". *Econometrica*, 52 (1984): 1251-1271.

Hausman J. A., Leonard G. K. and D. McFadden. "A utility-consistent, combined discrete choice and count data model assessing recreational use losses due to natural resource damage". *Journal of Public Economics,* 56 (1995): 1-30.

Heckman, J. Sample selection bias as a specification error. *Econometrica*, 47(1979): 153-161

Hellerstein, D. M. "Using count data models in travel cost analysis with aggregate data". *American Journal of Agricultural Economics,* 73 (1991): 860-866.

Hellerstein, D. M. and R. Mendelsohn. "A theoretical foundation for count data models". *American Journal of Agricultural Economics*, 75 (1993): 604-611.

Johnson N. L;, S. Kotz and A. W. Kemp. *Univariate Discrete Distributions*. Second edition. John Wiley and Sons : New York., 1992.

King, G. "Variation in specification of event count models: from restrictive assumptions to a generalised estimator". *American Journal of Political Science*, 33 (1989): 762-784.

Maddala, G.S. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press, 1983.

Magee, L. "R^2 Measures based on Wald and Likelihood ratio joint significance tests." *The American Statistician*, 44(1990): 250-253.

Mullahy, J. "Specification and testing of some modified count data models". *Journal of Econometrics*, 33 (1986): 341-365.

Offenbach, L. A. and B. K. Goodwin. "A travel-cost analysis of the demand for hunting trips in Kansas" *Review of Agricultural Economics*, 16 (1994): 55-61.

Ozuna T. and I A. Gomez. "Specification and testing of count data recreation demand functions". *Empirical Economics,* 20 (1995): 543-550.

Pohlmeier, W. and V. Ulrich. "An econometric model of the two-part decision making process in the demand for health care". *Journal of Human ressources,* 30 (1995): 339-361

Saha A. and D. Dong. "Estimating nested count data models". *Oxford Bulletin of Economics and Statistics*, 59(1997): 423-430.

Santos-Silva J. M. C. "Generalized Poisson regression for positive count data". *Communications in Statistics - Simulation Methods*, 26(1997): 1089-1102.

Sarker R. and Y. Surry. "Economic value of big game hunting : the case of moose hunting in Ontario". *Journal of Forest Economics,* 4(1998): 29-60.

Shaw, W. D. "On-site sample's regression: problems of non-negative integers, truncation and endogenous stratificiation". *Jounral of Econometrics*, 37(1988): 211-223.

Shaw, W. D. and P. Jakus. "Travel cost models of the demand for rock climbing". *Agricultural and Resource Economics Review*, 25(1996):133-143.

Smith, V. K. "Selection and recreational demand". *American Journal of Agricultural Economics*, 70(1988): 29-36.

Smith, V. K. and Kaoru, Y. "Signals or noise: explaining the variation in recreation benefit estimates". *American Journal of Agricultural Economics*, 72(1990): 419-433

Tanner J. C. "A problem of interference between two queues". *Biometrica* , 40 (1953): 58-69.

White, H. "A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity". *Econometrica*, 48(1980): 817-838.

Winkelman, R. and K. F. Zimmerman. "A new approach for modelling economic count data". *Economic Letters*, 37(1991): 467-474.

Woodward R. T., Gillig D., Griffin W. L. and Owuna T. "The welfar eimpacts of unanticipated trip limitations in travel cost methods". *Land Economics,* 77(2001): 327-339

Yen S. T. and W. Adamowicz. "Participation, trip frequency and site choice: a multinomial-Poisson hurdle model of recreational demand". *Canadian Journal of Agricultural Economics*, 42 (1994): 65-76.

Appendix 1: Characteristics of the alternative probability distributions used in this study

Designation	Range of support values	Probability distribution	Expectation	Variance	Specific remarks
Geometric	0, 1, 2, 3, …n	$\Pr ob(Y=k)=\lambda^{k}(1+\lambda)^{-(k+1)}$	λ	λ(1+λ)	bV(Y) =E(Y)*[1+E(Y)]
Borel	1,2,3,…….n	$\Pr ob(Y=k)=\dfrac{k^{k-2}\lambda^{k-1}e^{-\lambda k}}{(k-1)!}$	1/(1-λ)	$\lambda/(1-\lambda)^3$	i) 0 < λ < 1 ii)$V(Y)=[E(Y)]^2*[E(Y)-1]$
Logarithmic	1,2,3,…….n	$\Pr ob(Y=k)=\dfrac{-\lambda^{k}}{k\ln(1-\lambda)}$	$\dfrac{-\lambda}{(1-\lambda)\ln(1-\lambda)}$	$\dfrac{-\lambda(\ln(1-\lambda)+\lambda)}{(1-\lambda)^2(\ln(1-\lambda))^2}$	i)0 < λ < 1 ii)$V(Y)=E(Y)*\left[\dfrac{1}{1-\lambda}-E(Y)\right]$
Yule	1,2,3,…….n	$\Pr ob(Y=k)=\eta\mathrm{B}(k,\eta+1)$ $=\dfrac{\eta\Gamma(k)\Gamma(\eta+1)}{\Gamma(k+\eta+1)}$	$\dfrac{\eta}{\eta-1}$	$\dfrac{\eta^2}{(\eta-1)^2(\eta-2)}$	i)η > 1 for the existence of the mean ii)η > 2 for the existence of the variance iii)$V(Y)=\dfrac{[E(Y)]^2[E(Y)-1]}{2-E(Y)}$
Generalised Poisson	0, 1, 2, 3, …n	$Prob(Y=k)=\dfrac{\lambda(\lambda+\delta k)^{k-1}e^{-(\lambda+\delta k)}}{k!}$	$\dfrac{\lambda}{1-\delta}$	$\dfrac{\lambda}{(1-\delta)^3}$	i) ? > 0 and *d* < 1 ii)δ = 0, Poisson iii)? = d < 1, Modifed Borel

Notes: B(η, k) and Γ(k) designates the Beta and Gamma functions, respectively. They are linked to each other by the following relationship: B(η, k) = Γ(η)Γ(k)/Γ(η+k).

Appendix 2: Log Likelihood functions associated with the alternative count data models proposed in this paper

Designation	Untruncated cases	Left-truncated cases	Parameters of the distributions
Geometric	$\sum_{i=1}^{n}\{k\ln(l_i)-(k+1)\ln(1+l_i)\}$	$\sum_{i=1}^{n}\{(k-1)\ln(l_i)-k\ln(1+l_{i)}\}$	$\lambda_i = \exp(X_i'\beta)$
Modified Borel	$\sum_{i=1}^{n}\begin{Bmatrix}(k-1)\ln(k+1)+k\ln(l_i)\\ -(k+1)\,l_i-\ln(k!)\end{Bmatrix}$	$\sum_{i=1}^{n}\begin{Bmatrix}(k-1)\ln(k+1)+k\ln(l_i)\\ -(k+1)l_i-\ln(k!)-\ln[1-\exp(-l_i)]\end{Bmatrix}$	$l_i=\frac{\exp(X_i'b)}{1+\exp(X_i'b)}$
Modified Logarithmic	$\sum_{i=1}^{n}\begin{Bmatrix}(k+1)\ln(l_i)-\ln(k+1)\\ -\ln(-\ln(1-l_i))\end{Bmatrix}$	$\sum_{i=1}^{n}\{(k+1)\ln(k)-\ln(k+1)-\ln[-\ln[(1-l_i)+l_i]]\}$	$l_i=\frac{\exp(X_i'b)}{1+\exp(X_i'b)}$
Modified Yule	$\sum_{i=}^{n}\begin{Bmatrix}\ln(h_i)+\ln[\Gamma(k+1)]\\ +\ln[h_i+1]-\ln[\Gamma(k+h_i+2)]\end{Bmatrix}$	$\begin{cases}\sum_{i=1}^{n}\{\ln(h_i)+\ln[\Gamma(k+1)]+\ln[\Gamma(h_i+1)]\\ -\ln[\Gamma(k+h_i+2)]+\ln(h_i+1)\}\end{cases}$	$h_i=\frac{1+\exp(X_i'b)}{\exp(X_i'b)}$
Hybrid generalised Poisson	$\sum_{i=1}^{n}.\begin{Bmatrix}k\ln(A_i)+(k-1)\ln(1+a_ik)\\ -A_i(1+a_ik)-\ln(k!)\end{Bmatrix}$	$\begin{cases}\sum_{i=1}^{n}\{k\ln(A_i)+(k-1)\ln(1+a_ik)\\ -A_i(1+a_ik)-\ln(k!)-\ln(1-\exp(-m_i))\}\end{cases}$	$\mu_i = \exp(X_i'\beta)$ $\alpha_i = q_0{*}\exp[q_1(X_i'\beta)]]$ $A_i=\frac{\exp(X_i'b)}{1+q_0\exp[(1+q_1)(X_i'b)]}$

Note: $\Gamma(k)$ designates the Gamma function.

Appendix 3: The frequency distribution of the number of moose hunting trips

Number of trips	Frequency
1	152
2	10
3	3
4	6
5	4
6	2
7	1
8	0
9	4
10	1
12	5
13	4
15	3
20	3
Total	194
Mean	2.35
Standard deviation	12.89
Mode	152
Median	1
Ratio I	15.2
Ratio II	50.68

Notes: Ratio I is defined as the ratio of the frequeny of one trip over the frequency of two trips while Ration II is expressed as the ration of the frequency of one trip over the frequency of three trips.

Table 1: On some characteristics of the one parameter probability distributions

		Geometric	Modified Borel	Modified logarithmic	Modified Yule
		Mode			
Means	0.5	0.6667	0.71653	0.70191	0.75000
	1	0.5000	0.60653	0.56934	0.66667
	2	0.3333	0.53134	0.44700	0.60000
	4	0.2000	0.44933	0.34852	0.55556
	8	0.1111	0.41111	0.27892	0.52941
		Ratio I = probability of zero count/ probability of one count			
Means	0.5	3.00000	4.18684	3.77301	5.00000
	1	2.00000	3.29744	2.79591	4.00000
	2	1.50000	2.92160	2.35018	3.50000
	4	1.25000	2.78193	2.15035	3.25000
	8	1.12506	2.73648	2.06396	3.12500
		Ratio II = probability of zero count/ probability of two counts			
Means	0.5	9.00000	11.68640	10.67670	15.0000
	1	4.00000	7.24875	5.86284	10.0000
	2	2.25000	5.69050	4.14250	7.87500
	4	1.56250	5.15941	3.46801	6.90625
	8	1.26562	4.99221	3.19495	6.44531

Table 2 : Econometric results with truncated sample

	Conventional count data models		Alternative count data models							
								Hybrid generalised Poisson		
Parameters	Poisson	Negative binomial II	Geometric	Modified Logarith.	Modified Borel	Modified Yule	Generalised negative binomial	HGPI	HGPII	HGPIII
b_0	3.507 (10.300)	2.710 (3.409)	2.845 (4.058)	3.618 (3.783)	2.984 (1.951)	5.103 (1.690)	2.858 (2.892)	3.011 (3.359)	2.720 (3.515)	3.794 (7.456)
b_1	-0.416 (15.488)	-0.554 (11.609)	-0.533 (9.454)	-0.674 (9.732)	-0.861 (3.514)	-1.341 (3.661)	-0.585 (12.108)	-0.596 (6.068)	-0.551 (6.801)	-0.493 (9.435)
b_2	-0.127 (1.691)	0.137 (0.660)	0.101 (0.624)	0.316 (1.412)	0.646 (1.772)	1.149 (1.548)	0.116 (0.516)	0.129 (0.591)	0.204 (0.959)	-0.175 (1.601)
r										1.184 (1.792)
q_0								0.499 (2.070)	0.344 (2.460)	
q_1								-0.223 (1.226)		
a		1.366 (1.714)					1.571 (1.590)			
m							0.106 (0.999)			
LogL	-191.52	-152.51	-152.68	-152.76	-155.56	-154.61	-152.39	-153.13	-153.75	-185.40
R^2_{LRT}	0.956	0.782	0.803	0.740	0.889	0.640	0.773	0.768	0.766	0.676

The figures in parentheses are "t" values. LogL designates the value of the likelihood function. R^2_{LRT} is Maddala's pseudo-R^2.

Table 3: Recreational moose hunting trips in Ontario: actual and predicted numbers using model specifications with left truncated sample

Counts	Actual values	Predicted values									
		Conventional count data models		Alternative count data models							
									Hybrid Generalised Poisson (HGP)		
		Poisson	Negbin II	Geom.	Mod. log..	Mod. Borel	Mod. Yule.	Gen. negbin.	HGPI	HGPII	HGPIII
1	152	147	152	152	152	149	147	152	152	150	150
2	10	11	6	6	6	4	4	7	7	8	11
3	3	5	3	4	4	3	1	3	2	3	6
4	6	4	3	2	2	3	0	2	3	1	0
5	4	0	3	3	3	1	0	3	3	2	3
6	2	4	2	2	2	0	0	2	2	2	2
7	1	3	1	1	1	1	1	1	0	3	2
8	0	0	4	4	4	0	1	4	1	0	0
9	4	2	0	0	0	0	1	0	4	1	17
10	1	17	0	0	0	0	0	0	0	4	2
11	0	0	7	8	8	0	1	8	0	0	0
12	5	0	1	10	10	1	0	0	2	0	0
>12	6	1	12	2	2	32	39	12	18	20	1
	-------	------	------	------	------	------	------	------	------	------	------
Total	194	194	194	194	194	194	194	194	194	194	194

Pearson χ^2 Statistic	
Poisson	33.38
Negbin II	8.63
Geometric	8.63
Modified logarithmic	8.63
Modified Borel	26.32
Modified Yule	61.28
Generalised negbin	8.95
HGPI	9.83
HGPII	10.83
HGPIII	11.93

For all the model specifications but the modified Borel and Yule ones, the computation of the Pearson statistic has required to aggregate the various into six cells defined as follows: {1},{2},{3,4},{5,6,7}, {8,9,10} and {11,>11}. As a result, the critical χ^2 value with five degrees of freedom is 11.07 for a significance level of 5% and 9.24 for a 10% significance level. In the case of the modified Borel and the modified Yule model specifications, the various counts have been aggregated into five cells defined as follows: {1}, {2}, {3,4}, {5,6,7} and {8, >8}. Consequently, the critical χ^2 value with four degrees of freedom is 9.49 for a significance level of 5% and 7.78 for a 10% significance level.

Table 4 : Estimated Consumer surplus ($CDN) per moose hunting trip

	CS/trip ($CDN)	Standard error	95% confidence interval
Conventional count data models			
Poisson	240.39	26.98	[199.3 301.4]
Negbin II	180.40	15.82	[156.0 215.8]
Alternative count data models			
Geometric	187.60	14.84	[163.4 221.8]
Modified logarithmic	145.53	11.77	[128.0 175.1]
Modified Borel	116.11	172.73	[67.7 383.7]
Modified Yule	77.98	48.34	[48.9 163.3]
Generalised negbin	171.30	14.90	[145.9 204.2]
HGP1	167.87	17.11	[142.2 207.1]
HGPII	181.33	38.10	[134.3 271.9]
HGPIII	202.62	22.21	[169.1 254.0]

Estimated standard errors and 95% confidence intervals are based on the results of a Monte Carlo simulation involving 1000 replications.

Figure 1: probability distributions of four alternative single-parameter count data models

Geometric distribution	Modified Borel distribution	Modified logarithmic distribution	Modified Yule distribution

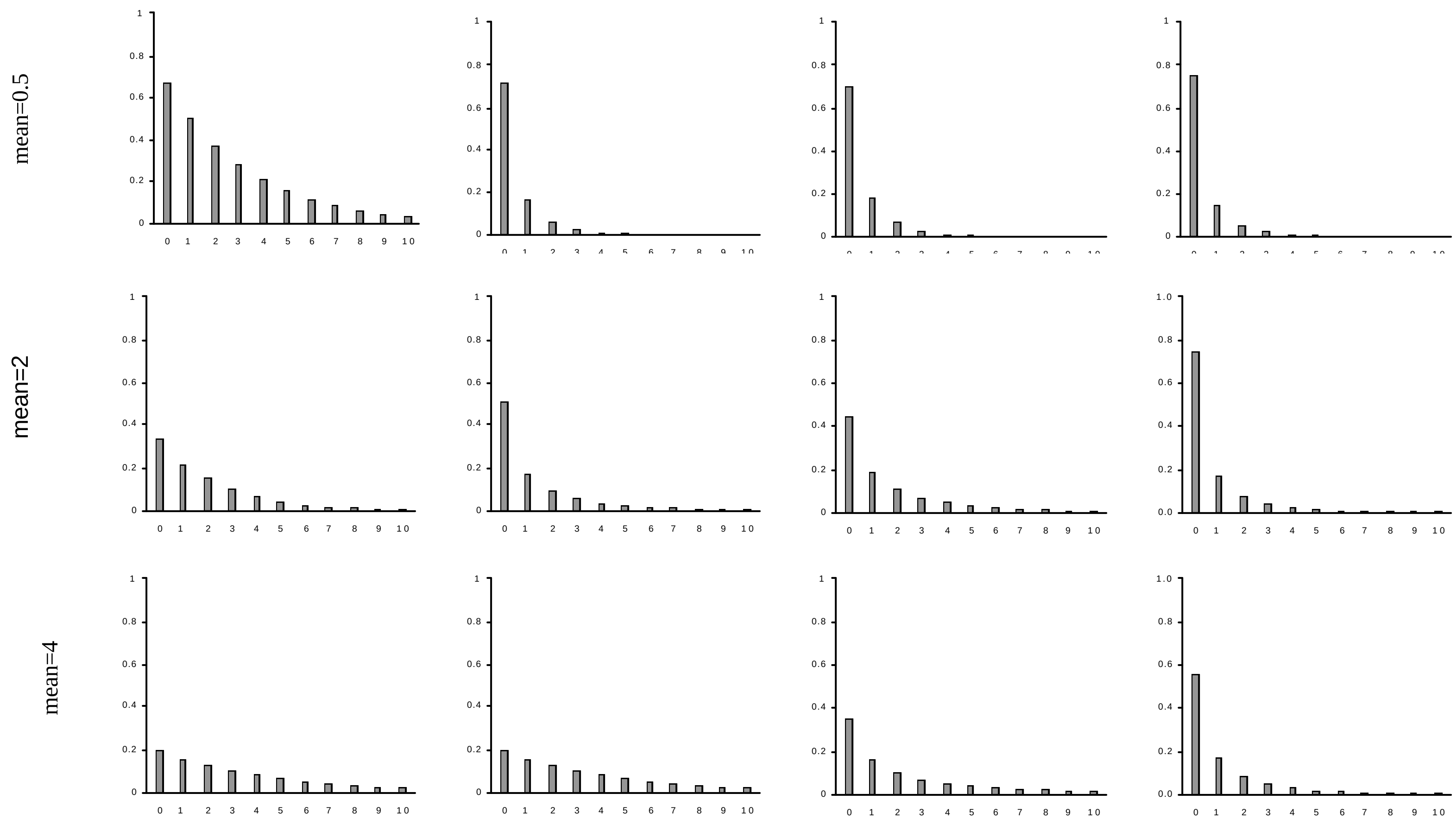

Figure 2 : Probability distributions of the generalised Poisson distribution

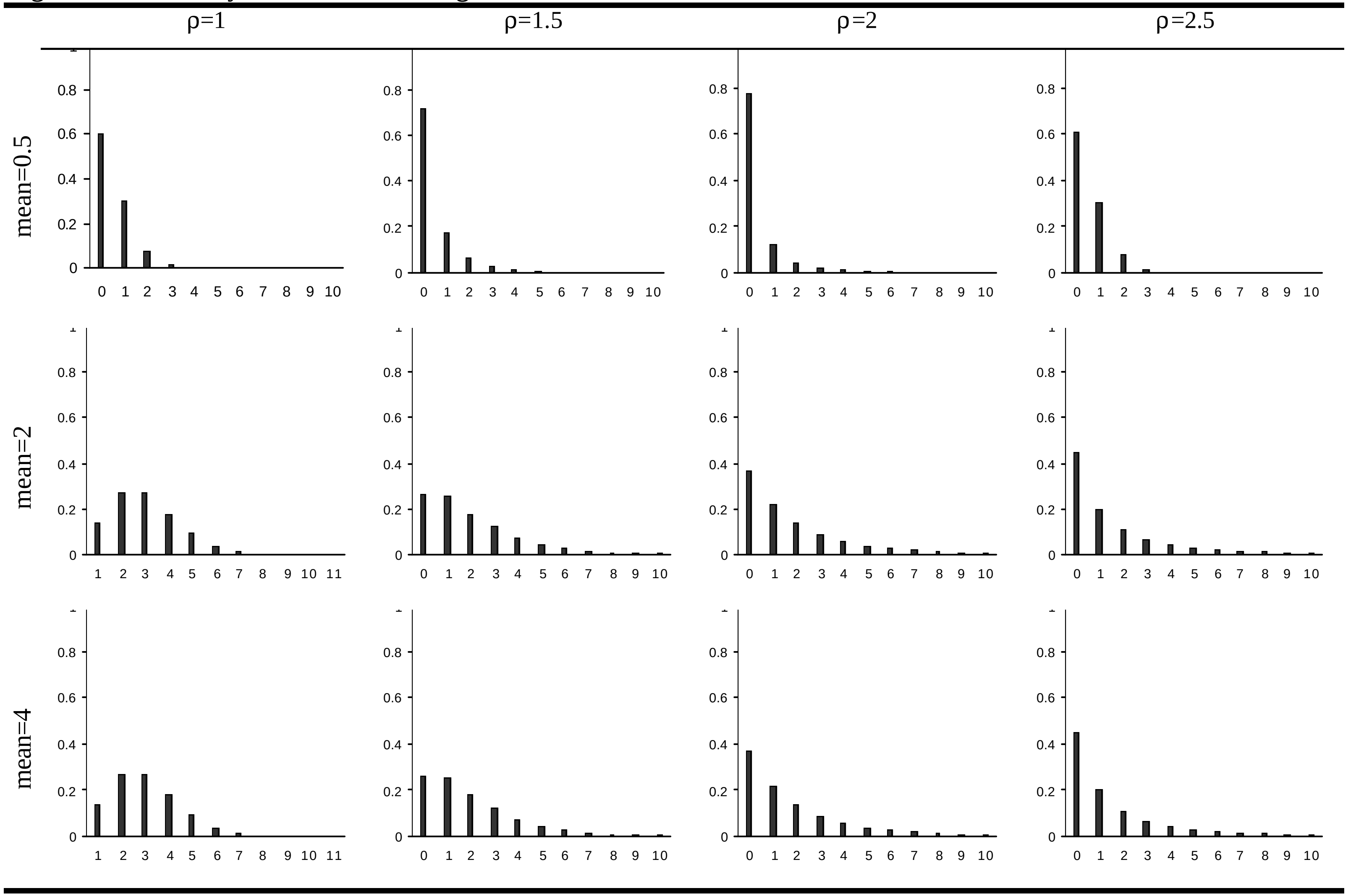

Figure 3: Probability distributions of the restricted generalized Poisson model