Journal of Modern Applied Statistical Methods

Volume 8 | Issue 2 Article 8

11-1-2009

On Some Discrete Distributions and their Applications with Real Life Data

Shipra Banik
Independent University, Dhaka, Bangladesh, shiprabanik@yahoo.com.au

B. M. Golam Kibria
Florida International University, kibriag@fiu.edu

Journal of Modern Applied Statistical Methods
November 2009, Vol. 8, No. 2, 423-447

1538 – 9472/09/$95.00

On Some Discrete Distributions and their Applications with Real Life Data

Shipra Banik
Independent University, Bangladesh

B. M. Golam Kibria
Florida International University

This article reviews some useful discrete models and compares their performance in terms of the high frequency of zeroes, which is observed in many discrete data (e.g., motor crash, earthquake, strike data, etc.). A simulation study is conducted to determine how commonly used discrete models (such as the binomial, Poisson, negative binomial, zero-inflated and zero-truncated models) behave if excess zeroes are present in the data. Results indicate that the negative binomial model and the ZIP model are better able to capture the effect of excess zeroes. Some real-life environmental data are used to illustrate the performance of the proposed models.

Key words: Binomial Distribution; Poisson distribution; Negative Binomial; ZIP; ZINB.

Introduction

Statistical discrete processes – for example, the number of accidents per driver, the number of insects per leaf in an orchard, the number of thunderstorms per year, the number of earthquakes per year, the number of patients visit emergency room in a certain hospital per day - often occur in real life. To approximate (or fit) a process, statistical probabilistic distributions are often used. Thus, fitting a process has been drawn considerable attention in the literature of many fields, for example, engineering (Lord, et al., 2005), ecology (Warton, 2005), biological science (Lloyd-Smith, 2007; Bliss & Fisher, 1953), epidemiology (Bohning, 1998), entomology (Taylor, 1961), zoology (Fisher, 1941), bacteriology (Neyman, 1939).

A broad range of probability models are commonly used in applied literature to fit discrete processes. These include: binomial model, Poisson model, negative binomial model, zero-inflated models and zero-truncated models. Binomial distribution models represent the total number of successes in a fixed number of repeated trials when only two outcomes are possible on each trial. Poisson distributions approximate rare-event processes (e.g., accident occurrences, failures in manufacturing or processing, etc.). An important restriction of the Poisson distribution is that its mean and variance are equal.

In reality, discrete processes often exhibit a large variance and a small mean and thus, display over-dispersion with a variance-to-mean value greater than 1 (Bliss & Fisher, 1953; Warton, 2005; Ross & Preece, 1985; White & Bennetts, 1996). Therefore, in real life, the Poisson assumption is often violated. A negative binomial distribution may be used for modeling purposes because it uses an additional parameter to describe the variance of a variable. Hence, the negative binomial distribution is considered as the first alternative to the Poisson distribution when the process is over-dispersed.

However, in many situations (e.g., road crash data), the chance of observing zero is

Shipra Banik is an Assistant Professor in the School of Engineering and Computer Science, Independent University, Bangladesh. Email: banik@secs.iub.edu.bd. B. M. Golam Kibria is an Associate Professor in the Department of Mathematics and Statistics at the Florida International University. He is the overseas managing editor of the *Journal of Statistical Research*, coordinating editor for the *Journal of Probability and Statistical Science*. He is an elected fellow of the *Royal Statistical Society* and the *International Statistical Institute.* Email: kibriag@fiu.edu.

greater than expected. Reasons may include failing to observe an event during the observational period and an inability to ever experience an event. Some researchers (Warton, 2005; Shankar, et al., 2003; Kibria, 2006) have applied zero-inflated models to model this type of process (known as a dual-states process: one zero-count state and one other normal-count state). These models generally capture apparent excess zeroes that commonly arise in some discrete processes, such as road crash data, and improve statistical fit when compared to the Poisson and the negative binomial model. The reason is that data obtained from a dual-state process often suffer from over-dispersion because the number of zeroes is inflated by the zero-count state. A zero-inflated model (introduced by Rider, 1961) is defined by

$$P(X=k)=\begin{cases}1-\theta & \text{if } X=0\\ \theta P(X;\mu_i) & \text{if } X>0\end{cases}$$

where θ is the proportion of non-zero values of X and $P(X;\mu_i)$ is a zero-truncated probability model fitted to normal-count states. To address phenomena with zero-inflated counting processes, the zero-inflated Poisson (ZIP) model and the zero-inflated negative binomial (ZINB) model have been developed. A ZIP model is a mix of a distribution that is degenerate at zero and a variant of the Poisson model. Conversely, the ZINB model is a mix of zero and a variant of negative binomial model.

Opposite situations from the zero-inflated models are also encountered; this article examines processes that have no zeroes: the zero-truncated models. If the Poisson or the negative binomial model is used with these processes, the procedure tries to fit the model by including probabilities for zero values. More accurate models that do not include zero values should be able to be produced. If the value of zero cannot be observed in any random experiment, then these models may be used. Two cases are considered: (1) the zero-truncated Poisson model, and (2) the zero-truncated negative binomial model.

Given a range of possible models, it is difficult to fit an appropriate discrete model. The main purpose of this article is to provide guidelines to fit a discrete process appropriately. First, a simulation study was conducted to determine the performance of the considered models when excess zeroes are present in a dataset. Second, the following real-life data (For details, see Table 4.1) were analyzed, the numbers of:

1. Road accidents per month in the Dhaka district,
2. People visiting the Dhaka medical hospital (BMSSU) per day,
3. Earthquakes in Bangladesh per year, and
4. Strikes (hartals) per month in Dhaka.

Statistical Distribution: The Binomial Distribution

If $X \sim B(n, p)$, then the probability mass function (pmf) of X is defined by

$$P(X=k;n,p)=n_{C_K}\,p^k(1-p)^{n-k},\quad k=0,1,2,\ldots,n. \qquad (2.1)$$

where n is the total number of trials and p is the probability of success of each trial. The moment generating function (mgf) of (2.1) is

$$M_X(t)=(p+qe^t)^n,$$

thus, E(X), V(X) and skewness (S_k) of (2.1) are np, npq and $[(1-2p)^2/npq]$ respectively.

Statistical Distribution: The Poisson Distribution

In (2.1) if $n\to\infty$ and $p\to 0$, then X follows the Poisson distribution with parameter $\lambda(>0)$ (denoted $X \sim P(\lambda)$). The pmf is defined as

$$P(X=k;\lambda)=\frac{e^{-\lambda}\lambda^k}{k!},\ k=0,1,2,\ldots \qquad (2.2)$$

where λ denotes expected number of occurrences. The mgf of (2.2) is

$$M_X(t)=e^{\lambda(e^t-1)},$$

thus, E(X) and V(X) of (2.2) are the same, which is λ and S_k is equal to $1/\lambda$.

Statistical Distribution: The Negative Binomial Distribution

If X~NB(k,p), then the pmf of X is given by

$$P(X;k,p)=\frac{\Gamma(k+X)}{X!\Gamma k}p^{X}q^{-(k+X)},\quad p>0,\ k>0,\ X=0,\ 1,\ 2,... \tag{2.3}$$

where p is the chance of success in a single trial and k are the number of failures of repeated identical trials. If $k\to\infty$, then X~P(λ), where $\lambda = kp$. The mgf of (2.3) is

$$M_X(t)=(q-pe^{t})^{-k},$$

thus, E(X), V(X) and S_k of (2.3) are kp, kpq and $[(1+2p)^2/kpq]$ respectively.

Statistical Distribution: The Zero-Inflated Poisson (ZIP) Distribution

If X~ZIP(θ, λ) with parameters θ and λ, then the pmf is defined by

$$P(X;\theta,\lambda)=\begin{cases}1-\theta & if\ X=0\\ \theta\dfrac{P(X;\lambda)}{1-e^{-\lambda}} & if\ X>0\end{cases} \tag{2.4}$$

where P(X; λ) is defined in (2.2) and θ is the proportion of non-zero values of X. The mgf of (2.4) is

$$M_X(t)=(1-\theta)+\frac{\theta e^{-\lambda}}{1-e^{-\lambda}}(e^{\lambda e^{t}}-1),$$

thus, E(X), V(X) and S_k of (2.4) are

$$\frac{\theta\lambda}{1-e^{-\lambda}},\ \frac{\theta\lambda}{1-e^{-\lambda}}\left[\lambda+1-\frac{\theta\lambda}{1-e^{-\lambda}}\right]$$

and

$$\frac{\lambda^2+3\lambda+1-\left(\frac{\theta\lambda}{1-e^{-\lambda}}\right)\left[3\lambda+3-\left(\frac{\theta\lambda}{1-e^{-\lambda}}\right)\right]^2}{\left(\frac{\theta\lambda}{1-e^{-\lambda}}\right)\left[\lambda+1-\left(\frac{\theta\lambda}{1-e^{-\lambda}}\right)\right]^3}$$

respectively.

Statistical Distribution: Zero-Inflated Negative Binomial (ZINB) Distribution

If X~ZINB(θ, k, p), then the pmf of X is defined by

$$P(X;\theta,k,p)=\begin{cases}1-\theta & if\ X=0\\ \theta\dfrac{P(X;k,p)}{1-q^{-k}} & if\ X>0\end{cases} \tag{2.5}$$

where P(X; k, p) is defined in (2.3) and θ is the proportion of non-zero values of X. The mgf of (2.5) is

$$M_X(t)=(1-\theta)+\frac{\theta}{1-q^{-k}}[(q-pe^{t})^{-k}-1],$$

thus, E(X), V(X) and S_k of (2.5) are

$$\frac{\theta kp}{1-q^{-k}},\ \frac{\theta kp}{1-q^{-k}}\left[p(k+1)+1-\left(\frac{\theta kp}{1-q^{-k}}\right)\right]$$

and

$$\frac{(kp)^2+3kp+1-\left(\frac{\theta kp}{1-q^{-k}}\right)\left[3kp+3-\left(\frac{\theta kp}{1-q^{-k}}\right)\right]^2}{\left(\frac{\theta kp}{1-q^{-k}}\right)\left[kp+1-\left(\frac{\theta kp}{1-q^{-k}}\right)\right]^3}$$

respectively. As $k\to\infty$, ZINB (θ, k, p) ~ZIP(θ, λ), where $\lambda = kp$.

Statistical Distribution: Zero-Truncated Poisson (ZTP) Distribution

If X~ZTP(λ), then the pmf of X is given by

$$P(X=x\mid X>0)=\frac{P(X=x)}{P(X>0)}=\frac{P(X;\lambda)}{1-P(X=0)}=\frac{P(X;\lambda)}{(1-e^{-\lambda})}$$

for x = 1, 2, 3, …, where P(X; λ) is defined in (2.2). The mgf of this distribution is

$$M_X(t) = \frac{e^{-\lambda}}{1-e^{-\lambda}}(e^{\lambda e^t} - 1),$$

thus, E(X) and V(X) are $\lambda(1-e^{-\lambda})^{-1}$ and S_k is $(1-e^{-\lambda})(\lambda+3+\lambda^{-1}) - 3(\lambda+1) - (1-e^{-\lambda})^{-1}$.

Statistical Distribution: Zero-Truncated Negative Binomial (ZTNB) Distribution

If X~ZTNB(k, p), then the pmf of X is given by

$$P(X = x \mid X > 0) = \frac{P(X; k, p)}{1 - P(X = 0)} = \frac{P(X; k, p)}{1 - q^{-k}}$$

for x = 1, 2, 3, …, where P(X; k, p) is defined in (2.3). The mgf of this distribution is

$$M_X(t) = \frac{1}{1-q^{-k}}[(q - pe^t)^{-k} - 1],$$

thus, E(X), V(X) and S_k are

$$\frac{kp}{1-q^{-k}}, \frac{kp}{1-q^{-k}}\left[p(k+1) + 1 - \left(\frac{kp}{1-q^{-k}}\right)\right]$$

and

$$\frac{(kp)^2 + 3kp + 1 - \left(\frac{kp}{1-q^{-k}}\right)\left[3kp + 3 - \left(\frac{kp}{1-q^{-k}}\right)\right]^2}{\left(\frac{kp}{1-q^{-k}}\right)\left[kp + 1 - \left(\frac{kp}{1-q^{-k}}\right)\right]^3}$$

respectively.

Parameter Estimation

To estimate the parameters of the considered models, the most common methods are the method of moment estimation (MME) (Pearson, 1894) and the maximum likelihood estimation (MLE) method (Fisher, 1922). The latter method has been used extensively since in the early 1900s, due to its properties of being consistent, asymptotically normal and having minimum variances for large samples.

The Moment Estimation Method (MME)

Consider the k^{th} moments of a random variable X. By notation,

$$M_k = \sum_{i=1}^{n} \frac{X_i^k}{n} = E(X^k),\ k = 1, 2, 3, \ldots,$$

thus,

$$M_1 = E(X),\ M_2 = \sum_{i=1}^{n} \frac{X_i^2}{n}, M_3 = \sum_{i=1}^{n} \frac{X_i^3}{n}.$$

The Maximum Likelihood Estimation Method (MLE)

Find the log-likelihood function for a given distribution and take a partial derivative of this function with respect to each parameter and set it equal to 0; solve it to find the parameters estimate.

Binomial Distribution: Moment Estimator of p

Based on (2.1), it is known that E(X) = np, therefore, $M_1 = np$. Simplification results in:

$$B_{\hat{p}(mom)} = E(X)/n = M_1/n.$$

Binomial Distribution: Maximum Likelihood Estimator of p

The log-likelihood expression of (2.1) is

$$LogL(X;n,p) = Constant + \sum_{i=1}^{n} X_i \log p + n - \sum_{i=1}^{n} X_i \log(1-p);$$

differentiating the above expression with respect to p, the following equation is obtained

$$\frac{\partial \log L(X; n, p)}{\partial p} = \frac{\sum_{i=1}^{n} X_i}{p} - \frac{n - \sum_{i=1}^{p} X_i}{1-p}.$$

Simplifying results in:

$$B_{\hat{p}(ml)} = \sum_{i=1}^{n} X_i / n$$

.

Poisson Distribution: Moment Estimator of λ

Based on (2.2), it is known that E(X) = λ, thus,

$$P_{\hat{\lambda}(mom)} = M_1.$$

Poisson Distribution: Maximum Likelihood Estimator of λ

The log-likelihood expression of (2.2) is

$$LogL(X;\lambda) = -n\lambda + \sum_{i=1}^{n} X_i \log \lambda - \sum_{i=1}^{n} (\log X_i)!$$

Differentiating the above expression with respect to λ, results in

$$\frac{\partial \log L(X;\lambda)}{\partial \lambda} = -n + \frac{\sum_{i=1}^{p} X_i}{\lambda}$$

and, after simplification,

$$P_{\hat{\lambda}(ml)} = M_1$$

thus

$$P_{\hat{\lambda}(mom)} = P_{\hat{\lambda}(ml)} = M_1.$$

Negative Binomial Distribution: Moment Estimators of p and k

Based on (2.3), it is known that E(X) = kp and V(X) = kpq, thus

$$M_1 = kp \qquad (2.6)$$

and

$$M_2 - M_1^2 = kpq \qquad (2.7)$$

Solving (2.7) for q results in $\hat{q} = \frac{M_2 - M_1^2}{M_1}$, and because it is known (based on 2.3) that $q - p = 1$,

$$NB_{\hat{p}(mom)} = (s^2 / M_1) - 1.$$

Solving (2.6), results in

$$NB_{\hat{k}(mom)} = \frac{M_1^2}{s^2 - M_1}$$

where

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - M_1)^2 .$$

Negative Binomial Distribution: Maximum Likelihood Estimators of p and k

The log-likelihood expression of (2.3) is

$$LogL(X; k, p) = \sum_{i=1}^{n} \log\left(\frac{\Gamma(X_i + k)}{X_i!\,\Gamma k}\right) + \sum_{i=1}^{n} X_i \log p - \sum_{i=1}^{n} (k + X_i) \log q$$

Differentiating the above expression with respect to p and k, the following equations result:

$$\frac{\partial LogL(X;k,p)}{\partial p} = \frac{\sum_{i=1}^{n} X_i}{p} - \frac{\sum_{i=1}^{n} (k + X_i)}{1+p} \qquad (2.8)$$

and

$$\frac{\partial LogL(X;k,p)}{\partial k} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=0}^{X_i - 1} \frac{j}{1 + jk^{-1}} + k^2 \ln(1+p) - \frac{kp(E(X)+k)}{1+p} \qquad (2.9)$$

Solving (2.8), results in $NB_{p(ml)} = M_1 / \hat{k}$. It was observed that $NB_{\hat{k}(ml)}$ does not exist in closed form, thus, $NB_{\hat{k}(ml)}$ was obtained by optimizing numerically (2.9) using the Newton-Raphson optimization technique where $p = \hat{p}$.

ZIP Distribution: Moment Estimators of θ and λ

It is known for (2.4) that

$$E(X) = \frac{\theta\lambda}{1 - e^{-\lambda}}$$

and

$$V(X) = \frac{\theta\lambda}{1 - e^{-\lambda}} \left[\lambda + 1 - \frac{\theta\lambda}{1 - e^{-\lambda}} \right],$$

thus,

$$M_1 = \frac{\theta\lambda}{1 - e^{-\lambda}},$$

and

$$M_2 - M_1^2 = \frac{\theta\lambda}{1 - e^{-\lambda}} \left[\lambda + 1 - \frac{\theta\lambda}{1 - e^{-\lambda}} \right].$$

Simplifying the above equations results in

$$ZIP_{\hat{\lambda}(mom)} = \frac{M_2}{M_1} - 1$$

and

$$ZIP_{\hat{\theta}(ml)} = \frac{M_1(1-e^{-\hat{\lambda}})}{\hat{\lambda}}.$$

ZIP Distribution: Maximum Likelihood Estimators of θ and λ

The log-likelihood expression of (2.4) is

$$LogL(X;\ \theta,\ \lambda) = \sum_{i=1}^{n} I(X_i = 0)\log(1-\theta) + \sum_{i=1}^{n} I(X_i > 0)\{\log\theta + X_i \log\lambda - \log(e^{\lambda}-1) - \log X_i!\}$$

Differentiating the above expression with respect to θ and λ, results in

$$\frac{\partial\, log\, L(X;\theta,\lambda)}{\partial\theta} = -\frac{\sum_{i=1}^{n} I(X_i = 0)}{1-\theta} + \frac{\sum_{i=1}^{n} I(X_i > 0)}{\theta}$$

$$\frac{\partial \log \mathrm{L}(\mathrm{X};\theta,\lambda)}{\partial\lambda} = \sum_{i=1}^{n} \mathrm{X}_i / \lambda - \sum_{i=1}^{n} \mathrm{I}(\mathrm{X}_i > 0) - \sum_{i=1}^{n} \mathrm{I}(\mathrm{X}_i > 0)\frac{e^{-\lambda}}{e^{\lambda}-1}$$

After the above equations are simplified for θ and λ, the following are obtained:

$$ZIP_{\hat{\theta}(mom)} = \sum_{i=1}^{n} I(X_i > 0)/n$$

and

$$ZIP_{\hat{\lambda}(ml)} = E(X)(1-e^{-\lambda}).$$

where E(X) is the expected value of the non-zero occurrences of X (λ does not have a closed form solution, hence the Newton-Raphson algorithm was used to find $\hat{\lambda}$ iteratively.)

ZINB Distribution: Moment Estimators of θ, k, p

Moment estimators of θ, k, p do not exist.

ZINB Distribution: Maximum Likelihood Estimators of θ, k, p

The log-likelihood expression of (2.5) is

$$LogL(X;\theta,k,p) = \sum_{i=1}^{n} I(X_i = 0)\log(1-\theta) + \sum_{i=1}^{n} I(X_i > 0)\log\theta + \sum_{i=1}^{n} \log\left(\frac{\Gamma(X_i+K)}{X_i!\Gamma K}\right) - \sum_{i=1}^{n} (k+X_i)\log q + \sum_{i=1}^{n} X_i \log p - \sum_{i=1}^{n} \log(1-q^{-k})$$

Differentiating the above with respect to each of parameters, results in the following estimators for θ, k, and p:

$$ZINB_{\hat{\theta}(ml)} = \sum_{i=1}^{n} I(X_i > 0)/n.$$

Other estimates $\hat{p}$ *and* $\hat{k}$ were found iteratively: k, p is given by

$$ZINB_{\hat{p}(ml)} = E(X)\{1-(1+kp)^{-k^{-1}}\} \quad (2.10)$$

thus, the solution of $\hat{p}$ has the same properties as described above. Because the score equation for $\hat{k}$ does not have a simple form, k was estimated numerically given the current estimate of $\hat{p}$ from (2.10) (for details, see Warton, 2005).

ZTP Distribution

The estimated parameters $ZTP_{\hat{\lambda}(mom)}$ and $ZTP_{\hat{\lambda}(ml)}$ are similar to $ZIP_{\hat{\lambda}(mom)}$ and $ZIP_{\hat{\lambda}(ml)}$, where the log-likelihood expression for this distribution is given by

$$LogL(X;\ \lambda) = \sum_{i=1}^{n} I(X_i > 0)\{X_i\ log\ \lambda - log(e^{\lambda}-1) - log\ X_i!\}$$

.

ZTNB Distribution

The estimated parameters $ZTNB_{\hat{p}(ml)}$ and $ZTNB_{\hat{k}(ml)}$ are similar to $ZINB_{\hat{p}(ml)}$ and $ZINB_{\hat{k}(ml)}$, where the log-likelihood expression for this distribution is given by

$$LogL(X;\ k,\ p) =$$

$$\sum_{i=1}^{n}\log(\frac{\Gamma(X_i+k)}{X_i!\Gamma k})+\sum_{i=1}^{n}X_i\log p$$
$$-\sum_{i=1}^{n}(k+X_i)\log q-\sum_{i=1}^{n}\log(1-q^{-k})$$

Methods for Comparison of the Distributions:
Goodness of Fit (GOF) Test

The GOF test determines whether a hypothesized distribution can be used as a model for a particular population of interest. Common tests include the χ^2, the Kolmogorov-Smirnov and the Anderson-Darling tests. The χ^2 test can be applied for discrete models; other tests tend to be restricted to continuous models. The test procedure for the χ^2 GOF is simple: divide a set of data into a number of bins and the number of points that fall into each bin is compared to the expected number of points for those bins (if the data are obtained from the hypothesized distribution). More formally, suppose:

H_0: Data follows the specified population distribution.
H_1: Data does not follow the specified population distribution.

If the data is divided into bins, then the test statistic is:

$$\chi^2_{cal}=\sum_{i=1}^{s}\frac{(O_i-E_i)^2}{E_i} \qquad (2.11)$$

where O_i and E_i are the observed and expected frequencies for bin i. The null, H_o, is rejected if $\chi^2_{cal}>\chi^2_{df,\alpha}$, where degrees of freedom (df) is calculated as (s−1− # of parameters estimated) and α is the significance level.

Methodology

Simulation Study

Because the outcome of interest in many fields is discrete in nature and generally follows the binomial, Poisson or the negative binomial distribution. It is evident from the literature that these types of variables often contain a high proportion of zeroes. These zeroes may be due to either the presence of a population with only zero counts and/or over-dispersion. Hence, it may be stated that - to capture the effect of excess zeroes - it is necessary to investigate which model would best fit a discrete process. Thus, a series of simulation experiments was conducted to determine the effect of excess zeroes on selected models. These simulation studies reflect how commonly used discrete models behave if excess zeroes are present in a set of data.

Simulation Experiment Design

A sample, $X = \{X_1, X_2, \ldots, X_n\}$, was obtained where data were generated from a Poisson model with:

λ: 1.0, 1.5, 2.0 and 2.5;
n: 10, 20, 30, 50, 100, 150, 200; and
10%, 20%, 80% zeroes.

Different data sets for different sample sizes and λs were generated to determine which model performs best if zeroes (10% to 80%) are present in a dataset. To select the possible best model, the Chi-square GOF statistic defined in (2.11) for all models were calculated. If the test was not statistically significant, then the data follows the specified (e.g., binomial, Poisson, or other) population distribution. Tables 3.1-3.8 show the GOF statistic values for all proposed distributions. Both small and large sample behaviors were investigated for all models, and all calculations were carried out using the programming code MATLAB (Version 7.0).

Results

Tables 3.1 to 3.8 show that the performance of the models depends on the sample size (n), λ and the percentage of zeroes included in the sample. It was observed that, as the percentage of zeroes increases in the sample, the proportion of over dispersion decreases and performance of the binomial and Poisson distributions decrease. For small sample sizes, most of the models fit well; for large sample sizes, however, both the binomial and Poisson performed poorly compared to others. For samples containing a moderate to high percentage of zeroes, the negative binomial performed best followed by ZIP and ZINB. Based on simulations, therefore, in the presence of excess zeroes the negative binomial model and the ZIP (moment estimator of parameters) model to approximate a real discrete process are recommended.

Table 3.1: Simulation Results for 10% Zero-Inflation

n	λ	The Chi-Square GOF Statistic								
		(Mean, Var)	B_{mom}	B_{ml}	Poisson	NB_{mom}	NB_{ml}	ZIP_{mom}	ZIP_{ml}	$ZINB_{ml}$
10	1.0	(1.87, 1.55)	4.86	1.53	0.71	-	-	1.45	0.93	1.08
	1.5	(1.88, 1.61)	4.59	1.45	0.49	-	-	0.80	0.61	0.48
	2.0	(2.11, 1.11)	6.99	3.24	3.94	-	-	1.26	4.54	6.34
	2.5	(2.66, 2.00)	2.41	14.34	2.11	-	-	-	1.32	1.96
20	1.0	(1.29, 0.59)	3.14	1.03	3.56	-	-	0.87	0.22	57.69*
	1.5	(1.55, 1.67)	50.39*	22.04*	6.80	6.57	6.74	6.95	6.00	4.35
	2.0	(1.89, 1.43)	8.03	2.07	0.64	-	-	1.70	0.75	0.70
	2.5	(2.42, 2.47)	33.68*	15.00	4.53	4.53	4.43	4.68	5.15	6.19
30	1.0	(1.28, 0.71)	4.74	1.22	3.04	-	-	0.85	0.95	111.05*
	1.5	(1.78, 1.45)	10.13	3.22	1.16	-	-	1.67	1.55	1.03
	2.0	(2.11, 2.41)	8.48	31.46*	11.36	9.80	10.08	11.81	7.60	8.91
	2.5	(2.25, 2.66)	129.94*	51.96*	6.42	5.15	5.33	6.01	8.45	4.31
50	1.0	(1.50, 1.00)	21.9	6.37	5.58	-	-	5.97	5.03	4.76
	1.5	(1.82, 1.25)	10.75	2.09	1.73	-	-	6.93	2.28	4.78
	2.0	(2.33, 2.04)	57.09*	24.38*	6.99	-	-	10.67	8.55	9.52
	2.5	(2.68, 2.21)	34.43*	26.04*	10.41	-	-	50.46*	11.36	16.67
100	1.0	(1.24, 0.54)	13.33	4.32	15.87	-	-	5.06	2.04	217.25*
	1.5	(1.67, 1.29)	32.99*	11.59	6.36	-	-	2.50	2.63	7.72
	2.0	(1.93, 1.74)	73.11*	27.26*	4.50	-	-	3.37	4.37	1.79
	2.5	(2.50, 3.27)	379.60*	159.59*	9.64	4.74	4.78	11.31	12.02	7.29
150	1.0	(1.21, 0.64)	27.15*	18.09*	28.80*	-	-	1.10	2.05	630.30*
	1.5	(1.68, 1.35)	33.19*	11.91	6.94	-	-	1.51	1.83	9.04
	2.0	(2.50, 2.49)	108.42*	51.92*	5.80	-	-	6.54	7.28	17.72*
	2.5	(2.05, 1.80)	54.72*	19.62*	2.98	-	-	3.44	3.86	9.24
200	1.0	(1.31, 0.87)	33.01*	20.37*	25.95*	-	-	3.81	3.53	48.64*
	1.5	(1.76, 1.58)	281.53*	111.60*	11.22	-	-	8.43	10.59	18.00*
	2.0	(2.16, 2.16)	98.78*	39.66*	0.52	-	-	0.51	2.67	9.86
	2.5	(2.52, 2.49)	108.42*	51.92*	5.80	-	-	6.54	7.28	17.72*

Notes: $\chi^2_{9,0.05}$ =16.91; *indicates significance (data do not follow the mentioned distribution) at α = 5%; - indicates over-dispersion; mom denotes moment estimator, and ml denotes MLE of model parameters.

Table 3.2: Simulation Results for 20% Zero-Inflation

n	λ	The Chi-Square GOF Statistic								
		(Mean, Var)	B_{mom}	B_{ml}	Poisson	NB_{mom}	NB_{ml}	ZIP_{mom}	ZIP_{ml}	$ZINB_{ml}$
10	1.0	(0.75, 0.91)	2.44	0.87	0.54	0.56	0.54	0.67	1.49	19.74*
	1.5	(1.88, 1.61)	1.90	1.36	0.21	0.33	0.23	0.11	0.65	0.44
	2.0	(2.11, 1.11)	5.13	1.97	1.07	-	-	2.51	2.08	1.78
	2.5	(2.40, 2.26)	4.46	52.17*	2.16	-	-	5.88	1.24	2.33
20	1.0	(1.14, 0.74)	9.12	2.43	3.54	-	-	5.78	5.42	83.15*
	1.5	(1.76, 1.69)	9.47	6.16	3.42	-	-	4.59	5.00	3.93
	2.0	(2.05, 2.26)	43.37*	21.00*	6.54	6.17	6.30	6.01	7.87	7.19
	2.5	(2.25, 2.72)	22.62*	17.07*	7.25	6.46	6.41	5.08	7.77	7.53
30	1.0	(1.44, 1.11)	16.47	4.57	1.98	4.14	-	-	4.42	1.94
	1.5	(1.51, 1.87)	15.05	6.69	2.80	3.36	3.13	3.70	3.85	3.97
	2.0	(1.71, 3.02)	183.02*	79.97*	10.86	2.53	2.23	3.38	17.55*	1.50
	2.5	(1.89, 1.87)	22.97*	14.24	6.21	-	-	7.73	9.48	8.64
50	1.0	(1.07, 0.73)	8.91	1.90	3.01	-	-	0.72	0.87	226.10*
	1.5	(1.27, 1.12)	9.23	2.64	0.84	-	-	0.82	2.45	2.57
	2.0	(1.42, 1.58)	36.468	14.27	0.83	0.40	0.45	0.74	4.39	0.90
	2.5	(2.31, 4.26)	1.16e+003*	473.03*	25.04*	9.44	9.44	17.75*	29.23*	21.43*
100	1.0	(1.27, 1.11)	12.51	4.70	2.93	-	-	5.07	8.63	6.35
	1.5	(1.54, 1.98)	111.79*	44.34*	1.81	1.08	1.13	1.68	8.20	3.94
	2.0	(1.90, 2.02)	49.98*	29.58*	10.09	9.46	9.43	7.80	13.15	12.44
	2.5	(2.21, 2.95)	129.95*	73.67*	17.07*	10.76	10.42	3.83	13.62	11.96
150	1.0	(1.15, 0.92)	26.42*	6.98	4.04	-	-	0.79	3.19	15.99
	1.5	(1.38, 1.45)	521.78*	206.30*	13.68	11.50	11.97	13.60	37.08*	11.87
	2.0	(2.09, 2.85)	643.42*	284.48*	27.75*	9.65	8.65	4.70	31.70*	9.87
	2.5	(1.81, 2.35)	163.00*	77.96*	10.29	2.76	2.33	2.08	17.60*	2.32
200	1.0	(1.07, 0.80)	16.68	2.76	5.02	-	-	0.60	2.64	931.68*
	1.5	(1.43, 1.39)	31.52*	12.61	2.07	-	-	2.76	12.21	7.39
	2.0	(1.78, 2.20)	269.23*	119.66*	12.80	7.57	7.52	6.05	23.29*	12.43
	2.5	(2.09, 2.85)	643.42*	284.48*	27.75*	9.65	8.65	4.70	31.71*	9.87

Notes: $\chi^2_{9,0.05}$ =16.91; *indicates significance (data do not follow the mentioned distribution) at α = 5%; - indicates over-dispersion; mom denotes moment estimator, and ml denotes MLE of model parameters.

Table 3.3: Simulation Results for 30% Zero-Inflation

n	λ	The Chi-Square GOF Statistic								
		(Mean, Var)	B_{mom}	B_{ml}	Poisson	NB_{mom}	NB_{ml}	ZIP_{mom}	ZIP_{ml}	$ZINB_{ml}$
10	1.0	(0.87, 0.69)	2.84	0.45	0.44	-	-	0.94	1.16	27.88*
	1.5	(1.12, 1.83)	3.96	2.01	0.55	0.55	0.44	0.91	1.43	0.67
	2.0	(2.10, 2.98)	5.61	14.43	3.24	1.23	1.09	1.51	0.49	0.36
	2.5	(2.30, 4.90)	18.52*	19.26*	6.62	3.92	4.02	1.51	3.76	2.96
20	1.0	(0.94, 0.71)	3.21	0.51	0.86	-	-	0.15	0.39	92.09*
	1.5	(1.38, 1.78)	35.44*	15.61	4.36	3.13	3.13	2.60	9.45	2.50
	2.0	(1.70, 2.09)	20.71*	17.90*	11.35	10.73	10.68	8.74	13.98	11.83
	2.5	(1.65, 2.39)	17.36*	11.52	4.39	3.82	3.82	2.42	2.03	4.38
30	1.0	(1.0, 0.88)	7.18	1.64	0.16	-	-	0.50	2.29	190.49*
	1.5	(1.0, 0.91)	7.71	2.04	0.18	-	-	0.59	2.85	197.58*
	2.0	(2.03, 3.89)	32.69*	52.35*	21.40*	6.51	3.97	6.44	4.58	3.54
	2.5	(1.29, 2.21)	80.02*	33.24*	4.13	1.99	1.78	5.92	12.27	4.58
50	1.0	(0.84, 0.97)	6.60	3.12	1.70	1.69	1.67	1.23	5.90	3.52
	1.5	(1.48, 2.30)	283.18*	120.03*	13.83	5.50	5.26	2.89	29.87*	4.96
	2.0	(1.67, 2.09)	64.25*	35.99*	11.66	9.08	8.83	6.04	14.96	9.07
	2.5	(2.02, 3.32)	551.89*	241.91*	37.23*	8.70	7.04	1.51	104.44*	7.85
100	1.0	(0.92, 0.84)	12.54	3.18	0.59	-	-	0.59	5.54	663.75*
	1.5	(1.23, 1.72)	156.60*	67.78*	9.01	1.21	0.95	1.41	34.08*	0.48
	2.0	(1.47, 2.04)	146.74*	69.81*	13.45	4.26	3.48	1.25	28.87*	1.80
	2.5	(1.92, 3.14)	2.62e+003*	1.06e+003*	51.79*	21.29*	20.20*	7.95	66.55*	20.26*
150	1.0	(0.93, 1.00)	37.20*	13.93	0.89	0.28	0.29	0.25	17.34*	11.65
	1.5	(1.27, 1.50)	46.01*	26.46*	10.28	8.17	8.05	4.44	27.61*	9.47
	2.0	(1.82, 2.79)	532.08*	273.65*	61.18*	26.09*	22.16*	4.35	64.56*	20.42*
	2.5	(1.44, 2.07)	281.77*	127.63*	22.52*	5.78	4.76	6.99	54.01*	2.57
200	1.0	(0.95, 0.99)	72.01*	27.23*	1.18	0.83	0.86	1.22	22.25*	16.74
	1.5	(1.15, 1.47)	88.14*	44.03*	10.55	4.57	4.41	2.79	44.02*	5.86
	2.0	(1.61, 2.49)	1.75e+003*	723.43*	48.01*	12.80	11.61	3.90	92.98*	10.08
	2.5	(1.82, 2.79)	532.08*	273.65*	61.18*	26.098*	22.16*	4.35	64.56*	20.42*

Notes: $\chi^2_{9,0.05}$ =16.91; *indicates significance (data do not follow the mentioned distribution) at α = 5%; - indicates over-dispersion; mom denotes moment estimator, and ml denotes MLE of model parameters.

Table 3.4: Simulation Results for 40% Zero-Inflation

n	λ	The Chi-Square GOF Statistic								
		(Mean, Var)	B_{mom}	B_{ml}	Poisson	NB_{mom}	NB_{ml}	ZIP_{mom}	ZIP_{ml}	$ZINB_{ml}$
10	1.0	(0.66, 1.46)	2.26	1.87	0.96	0.53	0.58	0.17	1.86	0.66
	1.5	(1.20, 1.73)	4.03	3.02	1.38	1.31	1.28	0.67	2.61	1.46
	2.0	(1.40, 2.04)	17.73*	9.68	3.77	2.71	2.68	1.85	6.30	2.68
	2.5	(1.90, 3.21)	10.75	27.79*	6.48	3.99	3.90	2.16	3.84	4.04
20	1.0	(0.83, 0.97)	21.23*	8.40	3.17	2.69	2.79	2.85	9.66	2.70
	1.5	(0.84, 0.69)	8.90	2.35	1.63	-	-	2.89	4.66	79.64*
	2.0	(1.07, 2.07)	42.96*	22.27*	8.79	3.94	3.62	1.38	24.15*	2.86
	2.5	(1.70, 3.58)	40.06*	26.74*	9.49	3.32	3.24	0.34	8.61	1.42
30	1.0	(0.76, 1.02)	75.62*	31.05*	6.16	3.57	3.64	5.07	28.62*	4.84
	1.5	(0.92, 1.27)	35.83*	15.13	2.78	0.62	0.62	0.95	14.79	0.71
	2.0	(1.41, 2.89)	267.73*	126.08*	47.17*	27.79*	28.18*	13.98	103.35*	35.84*
	2.5	(1.36, 2.30)	81.23*	39.25*	9.82	2.59	2.16	1.75	22.33*	1.91
50	1.0	(1.04, 1.57)	150.54*	63.48*	8.05	1.16	0.99	0.98	39.99*	0.59
	1.5	(1.17, 1.65)	70.74*	33.93*	9.43	4.40	3.46	1.46	26.17*	1.39
	2.0	(1.51, 2.50)	200.58*	95.75*	22.46*	8.65	7.25	2.60	38.05*	5.88
	2.5	(1.34, 2.36)	115.07*	62.35*	23.13*	8.98	7.79	4.56	50.28*	8.02
100	1.0	(0.87, 1.29)	232.718	99.50*	14.12	3.43	3.24	2.03	88.05*	3.19
	1.5	(1.06, 1.70)	303.13*	134.46*	35.92*	14.85	12.96	7.92	119.83*	10.78
	2.0	(1.26, 1.87)	144.32*	74.04*	21.62*	9.68	8.51	2.94	50.29*	6.50
	2.5	(1.51, 3.11)	1.78e+003*	742.21*	77.04*	11.73	8.38	4.45	181.85*	4.43
150	1.0	(0.79, 0.91)	33.33*	14.98	3.49	1.57	1.44	0.68	28.07*	1.4e+003*
	1.5	(1.03, 1.62)	795.95*	330.68*	27.73*	4.15	3.65	1.85	160.24*	3.04
	2.0	(1.60, 3.65)	2.18e+004*	8.6e+003*	221.00*	20.44*	15.72	9.98	623.30*	5.46
	2.5	(1.31, 2.38)	1.2e+003*	571.42*	81.64*	18.53*	14.69	12.03	232.37*	16.67
200	1.0	(0.75, 1.00)	206.88*	87.74*	11.31	2.48	2.36	2.94	91.17*	7.66
	1.5	(1.06, 1.55)	210.87*	103.49*	25.11*	6.13	4.77	0.47	95.82*	2.48
	2.0	(1.24, 2.04)	1.68e+003*	699.68*	62.48*	13.81	12.28	5.15	225.91*	10.27
	2.5	(1.60, 3.65)	2.18e+004*	8.69e+003*	221.00*	20.44*	15.72	9.98	623.30*	5.46

Notes: $\chi^2_{9,0.05}$ =16.91; *indicates significance (data do not follow the mentioned distribution) at α = 5%; - indicates over-dispersion; mom denotes moment estimator, and ml denotes MLE of model parameters.

Table 3.5: Simulation Results for 50% Zero-Inflation

n	λ	The Chi-Square GOF Statistic								
		(Mean, Var)	B_{mom}	B_{ml}	Poisson	NB_{mom}	NB_{ml}	ZIP_{mom}	ZIP_{ml}	$ZINB_{ml}$
10	1.0	(0.60, 0.48)	1.40	0.02	0.19	-	-	0.13	0.16	17.86*
	1.5	(1.10, 2.10)	43.97*	19.56*	5.32	1.60	1.47	0.94	17.22*	1.33
	2.0	(1.00, 1.75)	27.08*	13.12	6.53	4.48	4.46	2.63	17.72*	4.57
	2.5	(0.62, 0.83)	8.88	4.02	2.78	2.54	2.54	2.08	6.72	62.37*
20	1.0	(0.66, 0.82)	10.80	4.24	0.93	0.29	0.32	0.28	5.81	241.12*
	1.5	(0.56, 0.79)	20.75*	8.55	2.43	0.77	0.80	0.81	11.44	0.66
	2.0	(0.85, 1.29)	59.76*	24.70*	4.43	0.95	0.99	1.61	24.03*	0.93
	2.5	(1.21, 2.50)	169.46*	74.33*	13.89	3.18	2.60	0.72	44.27*	3.36
30	1.0	(0.42, 0.55)	16.33	6.96	3.94	2.75	2.59	1.97	10.81	121.28*
	1.5	(0.85, 1.51)	74.89*	73.14*	14.97	3.17	2.85	2.90	78.40*	2.11
	2.0	(1.03, 2.24)	232.29*	113.74*	26.09*	9.27	8.85	3.24	106.43*	9.77
	2.5	(1.41, 3.10)	225.50*	107.76*	28.80*	9.25	7.93	2.21	60.34*	11.61
50	1.0	(0.64, 0.91)	89.04*	37.10*	6.18	0.95	0.89	1.01	42.96*	1.12
	1.5	(0.89, 1.61)	42.06*	22.55*	8.57	2.36	2.33	2.33	30.61*	1.65
	2.0	(0.92, 2.06)	1.09e+003*	448.56*	51.15*	5.58	4.49	4.26	371.95*	2.83
	2.5	(1.30, 2.75)	551.89*	241.91*	37.23*	8.70	7.04	1.51	104.44*	7.85
100	1.0	(0.57, 0.82)	107.32*	46.20*	14.72	4.27	3.34	2.21	66.71*	0.87
	1.5	(0.78, 1.17)	284.75*	118.19*	20.75*	3.50	3.12	5.04	124.37*	2.16
	2.0	(1.10, 2.27)	822.90*	358.46*	49.51*	10.05	8.64	0.52	204.08*	7.83
	2.5	(1.25, 2.68)	577.43*	281.76*	68.86*	15.87*	11.79	1.33	195.13*	15.77
150	1.0	(0.61, 0.83)	94.65*	44.15*	17.37*	8.16	6.78	3.50	68.03*	1.8e+003*
	1.5	(0.91, 1.55)	877.63*	371.87*	52.36*	11.54	9.98	3.97	305.59*	7.01
	2.0	(1.28, 2.68)	8.3e+003*	3.3e+003*	179.29*	38.23*	35.12*	8.73	758.37*	38.32*
	2.5	(1.16, 2.43)	1.6e+003*	713.45*	94.96*	12.30	8.31	1.29	386.70*	6.23
200	1.0	(0.65, 0.90)	116.46*	55.23*	16.14	4.99	4.20	1.02	85.16*	7.88
	1.5	(0.82, 1.30)	210.40*	104.89*	29.59*	6.11	5.12	0.14	147.99*	2.77
	2.0	(1.23, 2.95)	4.4e+004*	1.7e+004*	312.14*	13.44	10.50	8.88	2.4e+003*	5.87
	2.5	(1.28, 2.68)	8.3e+003*	3.3e+003*	179.29*	38.23*	35.12*	8.73	758.37*	38.32*

Notes: $\chi^2_{9,0.05}$ =16.91; *indicates significance (data do not follow the mentioned distribution) at α = 5%; - indicates over-dispersion; mom denotes moment estimator, and ml denotes MLE of model parameters.

Table 3.6: Simulation Results for 60% Zero-Inflation

n	λ	The Chi-Square GOF Statistic								
		(Mean, Var)	B_{mom}	B_{ml}	Poisson	NB_{mom}	NB_{ml}	ZIP_{mom}	ZIP_{ml}	$ZINB_{ml}$
10	1.0	(0.44, 0.52)	3.33	1.14	0.45	0.29	0.36	0.33	1.73	31.31*
	1.5	(0.60, 0.93)	0.93	0.60	0.07	0.17	0.08	0.30	0.42	0.70
	2.0	(0.80, 1.73)	11.61	5.97	2.33	0.59	0.59	0.43	7.49	0.50
	2.5	(0.77, 1.69)	5.19	4.88	2.81	2.01	2.48	0.85	4.85	2.08
20	1.0	(0.35, 0.36)	2.66	0.61	0.09	0.06	0.09	0.13	0.91	36.26*
	1.5	(0.57, 0.81)	17.04*	7.31	2.20	0.56	0.50	0.29	10.51	0.13
	2.0	(0.95, 2.26)	522.66*	215.30*	25.74*	2.22	1.70	1.57	181.72*	0.61
	2.5	(1.36, 4.13)	102.70*	58.03*	20.07*	3.54	2.50	1.39	29.57*	4.79
30	1.0	(0.48, 0.56)	16.46	6.57	3.45	2.69	2.58	2.17	10.16	129.52*
	1.5	(0.79, 1.31)	70.10*	33.22*	14.84	8.04	7.31	3.93	50.71*	6.55
	2.0	(0.99, 1.98)	550.56*	116.98*	30.89*	9.87	8.45	2.89	98.09*	9.03
	2.5	(1.00, 2.59)	100.29*	464.03*	56.83*	7.57	6.21	2.37	407.44*	5.59
50	1.0	(0.37, 0.38)	8.58	2.31	0.53	0.44	0.52	0.52	3.41	104.71*
	1.5	(0.55, 1.30)	155.03*	66.52*	12.75	0.66	0.65	3.41	89.77*	2.93
	2.0	(0.91, 2.12)	370.66*	171.86*	44.30*	12.42	11.59	5.35	230.82*	12.37
	2.5	(1.12, 2.94)	4.8e+003*	1.95e+003*	118.24*	10.88	8.87	3.87	893.32*	5.71
100	1.0	(0.52, 0.87)	2.1e+003*	862.93*	59.97*	1.48	1.60	6.54	796.85*	4.73
	1.5	(0.57, 1.27)	1.7e+003*	729.47*	89.54*	4.64	3.85	7.12	1.01e+003*	3.43
	2.0	(0.82, 1.96)	4.2e+003*	1.75e+003*	121.95*	9.62	7.23	4.40	1.47e+003*	3.86
	2.5	(1.09, 2.79)	1.5e+003*	718.16*	136.83*	21.88*	18.27*	5.12	747.44*	14.69
150	1.0	(0.41, 0.50)	64.26*	25.91*	5.86	0.79	0.63	0.71	35.15*	2.02e+003*
	1.5	(0.68, 1.25)	426.14*	196.29*	58.20*	12.82	9.72	3.99	306.42*	7.49
	2.0	(1.07, 2.72)	9.1e+003*	3.87e+003*	298.17*	34.00*	27.98*	4.46	3.27e+003*	8.91
	2.5	(0.81, 1.69)	2.5e+003*	1.06e+003*	120.97*	8.71	5.49	3.67	1.05e+003*	5.61
200	1.0	(0.51, 0.87)	882.24*	369.87*	63.28*	4.00	2.63	3.03	482.26*	1.13
	1.5	(0.55. 0.93)	965.77*	409.93*	55.95*	6.49	5.24	2.93	498.47*	3.36
	2.0	(0.86, 1.74)	650.33*	329.36*	110.74*	30.81*	25.08*	5.69	573.57*	21.21*
	2.5	(1.07, 2.72)	9.1e+003*	3.87e+003*	298.17*	34.20*	27.98*	4.46	2.37e+003*	18.91*

Notes: $\chi^2_{9,0.05}$ =16.91; *indicates significance (data do not follow the mentioned distribution) at α = 5%; - indicates over-dispersion; mom denotes moment estimator, and ml denotes MLE of model parameters.

Table 3.7: Simulation Results for 70% Zero-Inflation

n	λ	The Chi-Square GOF Statistic								
		(Mean, Var)	B_{mom}	B_{ml}	Poisson	NB_{mom}	NB_{ml}	ZIP_{mom}	ZIP_{ml}	$ZINB_{ml}$
10	1.0	(0.60, 1.15)	29.65*	13.37	5.27	1.86	1.71	0.83	22.54*	1.39
	1.5	(0.50, 0.72)	11.8	5.37	3.54	2.84	2.83	2.07	8.99	82.23*
	2.0	(0.90, 2.32)	10.40	8.74	4.82	1.22	1.18	0.11	7.89	1.36
	2.5	(0.80, 1.73)	4.99	6.31	3.49	1.88	2.47	0.71	4.64	51.86*
20	1.0	(0.36, 0.57)	1.80	0.85	0.21	0.13	0.07	0.42	0.82	1.68
	1.5	(0.27, 0.33)	4.68	1.65	0.60	0.22	0.24	0.20	2.22	52.16*
	2.0	(0.57, 1.25)	23.82*	13.34	7.45	3.91	4.25	1.51	23.59*	3.71
	2.5	(1.10, 3.98)	293.96*	133.96*	31.52*	6.20	6.52	0.91	149.41*	1.67
30	1.0	(0.39, 0.61)	39.03*	17.02*	5.44	0.83	0.62	0.29	25.18*	0.13
	1.5	(0.72, 2.06)	109.55*	54.26*	20.14*	5.12	5.25	0.85	101.27*	1.79
	2.0	(0.82, 2.29)	790.34*	335.52*	50.67*	7.50	6.50	1.20	490.86*	14.03
	2.5	(0.50, 0.74)	23.87*	11.81	5.61	2.66	2.38	1.22	19.68*	442.47*
50	1.0	(0.41, 0.78)	477.66*	198.32*	30.46*	1.24	0.87	0.91	267.62*	0.33
	1.5	(0.47, 0.92)	202.94*	88.76*	31.60*	8.81	7.51	3.60	157.73*	5.92
	2.0	(0.56, 1.14)	159.68*	71.58*	18.19	1.70	1.46	1.53	112.87*	0.87
	2.5	(0.87, 3.55)	1.29e+003*	542.57*	70.87*	6.72	7.13	8.29	675.26*	104.59*
100	1.0	(0.29, 0.39)	61.89*	26.49*	14.07	7.28	6.18	4.32	39.15*	486.00*
	1.5	(0.43, 0.77)	242.08*	110.12*	30.088	5.98	4.92	1.25	181.32*	2.70
	2.0	(0.71, 2.01)	1.1e+003*	4.62e+003*	290.16*	13.95	11.20	2.57	6.0e+003*	8.88
	2.5	(0.81, 2.12)	4.8e+003*	2.01e+003*	174.84*	23.30*	21.50*	5.88	2.3e+003*	29.29*
150	1.0	(0.40, 0.71)	311.81*	146.78*	46.22*	5.37	5.81	1.57	270.69*	1.45
	1.5	(0.49, 0.93)	396.50*	189.83*	60.48*	19.02*	17.47*	6.49	335.32*	13.90
	2.0	(0.73, 1.86)	7.0e+003*	2.92e+003*	327.54*	29.06*	21.54*	5.85	3.8e+003*	24.25*
	2.5	(0.57, 1.18)	749.89*	336.53*	89.94*	20.35*	16.90	4.70	589.06*	12.55
200	1.0	(0.29, 0.38)	75.97*	33.90*	11.71	1.88	1.36	0.40	50.56*	2.3e+003*
	1.5	(0.56, 1.34)	1.9e+004*	7.99e+003*	304.89*	13.25	11.95	2.23	8.0e+003*	6.22
	2.0	(0.73, 1.94)	4.6e+003*	1.98e+003*	227.50*	30.41*	27.78*	5.59	2.7e+003*	15.97
	2.5	(0.73, 1.86)	7.0e+003*	2.92e+003*	327.54*	29.06*	21.54*	5.85	3.8e+003*	24.35*

Notes: $\chi^2_{9,0.05}$ =16.91; *indicates significance (data do not follow the mentioned distribution) at α = 5%; - indicates over-dispersion; mom denotes moment estimator, and ml denotes MLE of model parameters.

Table 3.8: Simulation Results for 80% Zero-Inflation

n	λ	The Chi-Square GOF Statistic								
		(Mean, Var)	B_{mom}	B_{ml}	Poisson	NB_{mom}	NB_{ml}	ZIP_{mom}	ZIP_{ml}	$ZINB_{ml}$
10	1.0	(0.22, 0.44)	0	0.64	0.31	0.03	0.28	0.10	0.0032	0
	1.5	(0.50, 1.16)	3.17	3.24	2.05	0.51	0.67	0.07	2.91	31.48*
	2.0	(0.60, 1.82)	2.81	4.41	2.74	0.44	0.79	0.46	2.21	0.40
	2.5	(0.50, 1.16)	3.17	3.24	2.05	0.51	0.67	0.07	2.91	31.48*
20	1.0	(0.31, 0.67)	108.52*	46.91*	14.89	2.32	1.98	0.82	86.19*	1.54
	1.5	(0.40, 1.30)	1.80	2.65	1.54	0.56	0.39	8.46	1.17	64.35*
	2.0	(0.70, 2.32)	139.56*	64.29*	18.14*	2.61	2.33	0.16	134.54*	6.49
	2.5	(0.65, 2.55)	184.74*	81.78*	20.07*	2.28	3.38	0.94	173.58*	10.92
30	1.0	(0.11, 0.18)	24.48*	10.85	5.71	1.27	1.09	0.65	15.83	208.91*
	1.5	(0.34, 1.01)	53.98*	24.48*	8.24	0.46	0.47	3.31	41.92*	9.42
	2.0	(0.46, 1.29)	452.56*	189.90*	33.05*	2.09	1.79	0.32	337.42*	0.36
	2.5	(0.31, 0.57)	79.11*	35.05*	12.03	2.12	1.70	0.64	59.23*	0.94
50	1.0	(0.19, 0.29)	41.41*	18.28*	9.89	3.64	3.10	1.96	27.61*	328.00*
	1.5	(0.39, 1.05)	580.55*	247.89*	47.62*	4.17	3.73	0.27	481.91*	2.36
	2.0	(0.42, 0.82)	167.33*	76.05*	29.76*	9.56	8.86	3.91	144.92*	4.1e+003*
	2.5	(0.38, 0.85)	38.27*	20.88*	11.31	3.89	4.04	1.42	37.61*	2.45
100	1.0	(0.22, 0.36)	247.15	105.64	29.37*	1.74	1.19	0.93	147.52*	7.6e+003*
	1.5	(0.29, 0.56)	863.01*	362.55*	57.75*	1.83	1.08	0.27	513.18*	0.10
	2.0	(0.44, 1.30)	1.86e+004*	7.6e+003*	540.37*	12.36	10.00	2.66	1.2e+004*	11.11
	2.5	(0.54, 1.83)	1.59e+006*	6.2e+005*	7.1e+003*	9.56	7.22	2.52	6.2e+005*	1.50
150	1.0	(0.23, 0.41)	365.82*	160.49*	51.41*	5.92	4.22	1.42	256.57*	9.9e+003*
	1.5	(0.29, 0.55)	755.46*	331.02*	71.96*	4.16	2.68	0.42	543.71*	0.52
	2.0	(0.45, 1.12)	9.3e+003*	3.8e+003*	369.57*	20.28*	16.55	3.03	6.4e+003*	26.67*
	2.5	(0.48, 1.32)	1.7e+004*	7.2e+003*	530.55*	17.27*	13.11	2.32	1.1e+004*	17.63*
200	1.0	(0.22, 0.41)	1.46e+003*	611.40*	86.64*	3.60	3.36	1.53	776.13*	2.75
	1.5	(0.36, 0.84)	6.9e+003*	2.8e+003*	280.71*	9.29	6.20	0.34	4.5e+003*	7.44
	2.0	(0.42, 1.30)	4.1e+005*	1.6e+005*	4.4e+003*	12.41	8.16	4.40	2.1e+005*	1.73
	2.5	(0.45, 1.12)	9.3e+003*	3.8e+003*	369.57*	20.28*	16.55	3.03	6.4e+003*	26.57*

Notes: $\chi^2_{9,0.05}$ =16.91; *indicates significance (data do not follow the mentioned distribution) at α = 5%; - indicates over-dispersion; mom denotes moment estimator, and ml denotes MLE of model parameters.

Applications to Real Data Sets

The selected processes were fitted using theoretical principles and by understanding the simulation outcomes. Theoretical explanations of a discrete process are reviewed as follows: Generally, a process with two outcomes (see Lord, et al., 2005, for details) follows a Bernoulli distribution. To be more specific, consider a random variable, which is NOA. Each time a vehicle enters any type of entity (a trial) on a given transportation network, it will either be involved in an accident or it will not.

Thus, $X \sim B(1,p)$, where p is the probability of an accident when a vehicle enters any transportation network. In general, if n vehicles are passing through the transportation network (n trials) they are considered records of NOA in n trials, thus, $X \sim B(n,p)$. However, it was observed that the chance that a typical vehicle will cause an accident is very small out when considering the millions of vehicles that enter a transportation network (large number of n trials). Therefore, a $B(n,p)$ model for X is approximated by a $P(\lambda)$ model, where λ represents expected number of accidents. This approximation works well when λs are constant, but it is not reasonable to assume that λ across drivers and road segments are constant; in reality, this varies with each driver-vehicle combination. Considering NOA from different roads with different probabilities of accidents for drivers, the distribution of accidents have often been observed over-dispersed: if this occurs, $P(\lambda)$ is unlikely to show a good fit. In these cases, the negative binomial model can improve statistical fit to the process. In the literature, it has been suggested that over-dispersed processes may also be characterized by excess zeroes (more zeroes than expected under the $P(\lambda)$ process) and zero-inflated models can be a statistical solution to fit these types of processes.

Traffic Accident Data

The first data set analyzed was the number of accidents (NOA) causing death in the Dhaka district per month; NOA were counted for each of 64 months for the period of January 2003 to April 2008 and the data are presented in Table 4.1.1 and in Figure 4.1.1.

Table 4.1.1: Probability Distribution of NOA

NOA	Observed Months
0	0.12
1	0.24
2	0.17
3	0.22
4	0.17
5	0.05
6	0
7	0.02
Total	64 months

Table 4.1: Dataset Properties

Type	Time Period	n	Data Source
The number of traffic accidents (NOA) in the Dhaka district per month	Jan-2003 to April-2008	64 months	The Daily Star Newspaper
The number of peoples visiting (NOPV) Dhaka BMSSU per day	April-2007 to July-2007	74 days	BMSSU, Dhaka.
The number of earthquakes (NEQ) in Bangladesh per year	1973 to 2008	37 years	http://neic.usgs.gov/cgi-bin/epic/epic.cgi
The number of hartals (NOH) in the city of Dhaka per month	Jan-1972 to Dec-2007	432 months	Dasgupta (2001) and the Daily Star Newspaper

A total of 141 accidents occurred during the considered periods (see Figure 4.1.1) and the rate of accidents per month was 2.2 (see Table 4.1.2). Figure 4.1.1 also shows that since 2004 the rates have decreased. Main causes of road accidents identified according to Haque, 2003 include: rapid increase in the number of vehicles, more paved roads leading to higher speeds, poor driving and road use knowledge, skill and awareness and poor traffic management. The observed and expected frequencies with GOF statistic values are tabulated and shown in Table 4.1.2. The sample mean and variance indicate that the data shows over dispersion; about 16% of zeroes are present in the NOA data set. According to the GOF statistic, the binomial model fits poorly, whereas the Poisson and the negative binomial appear to fit. The excellent fits of different models are illustrated in Figure 4.1.2; based on the figure and the GOF statistic, the negative binomial model was shown to best fit NOA.

Figure 4.1.1: NOA by Year

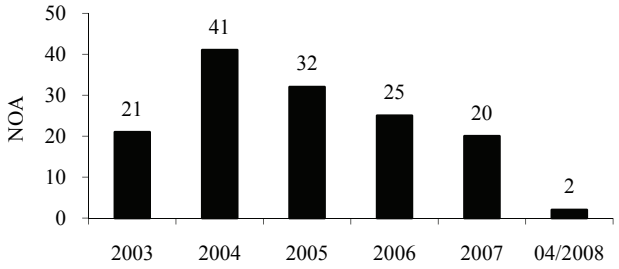


Table 4.1.2: Observed and Fitted Frequencies of NOA

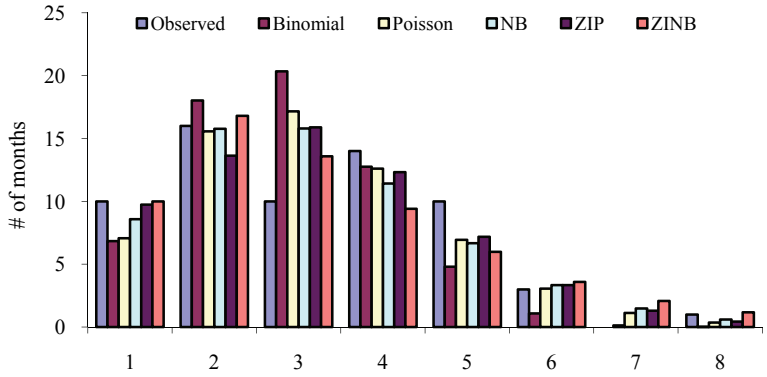
NOA	0	1	2	3	4	5	6	7	Chi-Square GOF Statistic
Observed Months	10	16	10	14	10	3	0	1	
B_{mom}	6.8	18.2	20.3	12.7	4.8	1.0	0.1	0	150.93*
B_{ml}	4.5	14.6	20.1	15.3	7.0	1.9	0.2	0	63.08*
Poisson	7.0	15.5	17.1	12.5	6.9	3.0	1.1	0.3	8.02
NB_{mom}	8.4	15.7	15.8	11.5	6.9	3.3	1.4	0.5	6.43
NB_{ml}	8.5	15.7	15.8	11.4	6.6	3.3	1.4	0.6	6.39
ZIP_{mom}	9.7	13.6	15.8	12.3	7.1	3.3	1.3	0.4	6.01
ZIP_{ml}	10.0	16.5	16.8	11.4	5.8	2.3	0.8	0.2	9.90
$ZINB_{ml}$	10.0	16.8	13.5	9.4	6.0	3.6	2.0	1.1	8.11

Mean = 2.2 and Variance = 2.6

Parameter Estimates: $B_{\hat{p}(mom)} = 0.27$, $B_{\hat{p}(ml)} = 0.31$, $P_{\hat{\lambda}(mom/ml)} = 2.2$, $NB_{\hat{p}(mom)} = 0.84$, $NB_{\hat{k}(ml)} = 11.96$, $NB_{\hat{p}(mom)} = 0.83$, $NB_{\hat{k}(ml)} = 11.09$, $ZIP_{\hat{p}(mom)} = 0.84$, $ZIP_{\hat{\lambda}(mom)} = 2.32$, $ZIP_{\hat{p}(ml)} = 0.84$, $ZIP_{\hat{\lambda}(ml)} = 2.03$, $ZINB_{\hat{\theta}(ml)} = 0.84$, $ZINB_{\hat{p}(ml)} = 0.53$, $ZINB_{\hat{k}(ml)} = 2.49$

Note: See footnotes, Table 3.1

Figure 4.1.2: Distribution of NOA for Different Models



Number of Patient Visits (NOPV) at Hospital

The number of patient visits (NOPV) data were collected from the medical unit of the Dhaka BMSSU medical hospital for the period of 26 April 2007 to 23 July 2007, where the variable of interest is the total number of patients visit in BMSSU per day. The frequency distribution for NOPV is reported in Table 4.2.1, which shows that the patients visiting rate per day is 142.36; this equates to a rate of 14.23 per working hour (see Table 4.2.2). Expected frequencies and GOF statistic values were tabulated and are shown in Table 4.2.2 and Figure 4.2.2 shows a bar chart of observed vs. expected frequencies. Tabulated results and the chart show that the negative binomial model and the ZTNB model (ZTNB model had the best fit) fit NOPV data well compared to other models. Based on this analysis, the ZTNB model is recommended to accurately fit NOPV per day.

Table 4.2.1: Frequency Distribution of NOPV

NOPV	Observed Days
51-83	1
84-116	12
117-149	32
150-182	23
183-215	6

Figure 4.2.1: Trend to Visits in BMSSU per Day

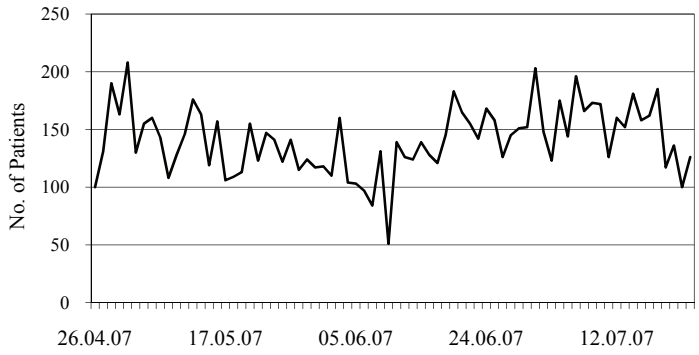


Table 4.2.2: Observed and Fitted Frequencies of NOPV

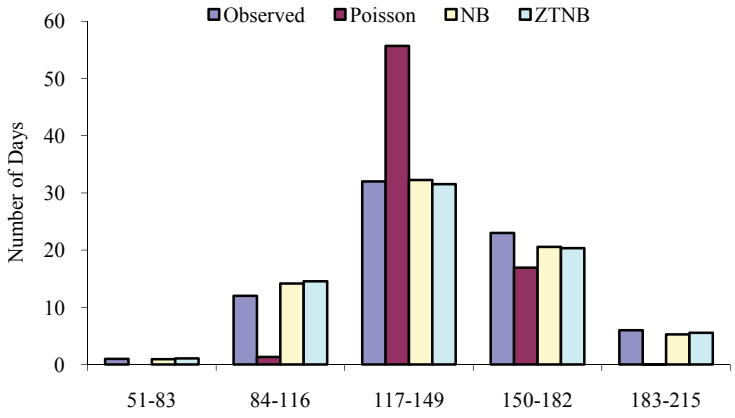
NOPV	51-83	84-116	117-149	150-182	183-215	Chi-Square GOF Statistic
Observed Days	1	12	32	23	6	
B_{mom}	0	0.02	66.22	7.75	0	1.5e+013*
B_{ml}	0	0.02	66.22	7.75	0	1.5e+013*
Poisson	0	1.33	55.69	16.94	0.02	1.46e+005*
NB_{mom}	0.92	14.17	32.27	20.58	5.28	0.7196
NB_{ml}	1.08	14.56	31.54	20.35	5.55	0.8453
ZIP_{mom}	73.03	0	0	0	0	-
ZIP_{ml}	0	1.33	55.69	16.94	0.02	1.46e+005*
$ZINB_{ml}$	0	0	0	0	0	-
ZTP_{ml}	0	1.33	55.69	16.94	0.02	1.46e+005
$ZTNB_{ml}$	1.08	14.56	31.54	20.35	5.55	0.84

Mean =142.36 and Variance =831.87

Parameter estimates: $B_{\hat{p}(mom)} = 0.65$, $B_{\hat{P}(ml)} = 0.65$, $P_{\hat{\lambda}(mom/ml)} = 140.78$, $NB_{\hat{p}(mom)} = 0.16$, $NB_{\hat{k}(mom)} = 26.82$, $NB_{\hat{p}(ml)} = 0.16$, $NB_{\hat{k}(ml)} = 26.82$, $ZIP_{\hat{p}(mom)} = 0.01$, $ZIP_{\hat{\lambda}(mom)} = 1.08e+004$, $ZIP_{\hat{p}(ml)} = 1.0$, $ZIP_{\hat{\lambda}(ml)} = 0.14$, $ZINB_{\hat{\theta}(ml)} = 1.0$, $ZINB_{\hat{p}(ml)} = 0.14$, $ZINB_{\hat{k}(ml)} = 831.87$, $ZTP_{\hat{\lambda}(ml)} = 140.78$, $ZTNB_{\hat{p}(ml)} = 0.14$ and $ZTNB_{\hat{k}(ml)} = 831.87$

Note: See footnotes, Table 3.1

Figure 4.2.2: Distribution of NOPV for Different Models



Earthquake Data

The third variable of interest is the number of earthquakes (NEQ) that occurred in Bangladesh from 1973 to January 2008 (based on available data). This data set was extracted from the http://earthquake.usgs.gov site and is presented in Table 4.3.1. The number of earthquakes per year is presented in Figure 4.3.1 and their magnitudes are displayed in Figure 4.3.2. The frequency distribution of earthquakes in Bangladesh is shown in Table 4.3.1. Table 4.3.2 shows a total of 127 earthquakes occurred in Bangladesh during the selected time period and that the average yearly earthquake rate is 3.43. The observed frequencies, the expected frequencies and the GOF statistic values for NEQ data are reported in Table 4.3.2. Sample mean and variance equal 3.43 and 10.19 respectively (shows over dispersion). It was found that the negative binomial model fits this data well (see Figure 4.3.3), whereas other models indicate lack of fit. Thus, based on this study, the distribution of NEQ follows the negative binomial distribution with a proportion of earthquakes equaling 0.29 per year.

Figure 4.3.1: Number of Earthquakes in Bangladesh per Year

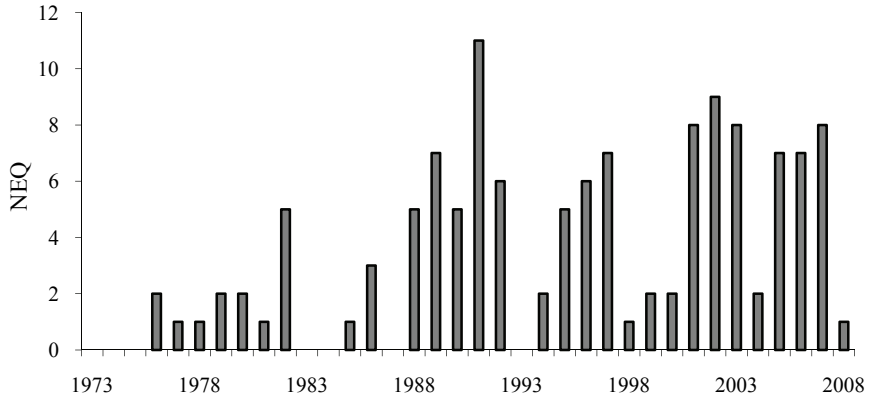


Figure 4.3.2: Earthquake Magnitudes

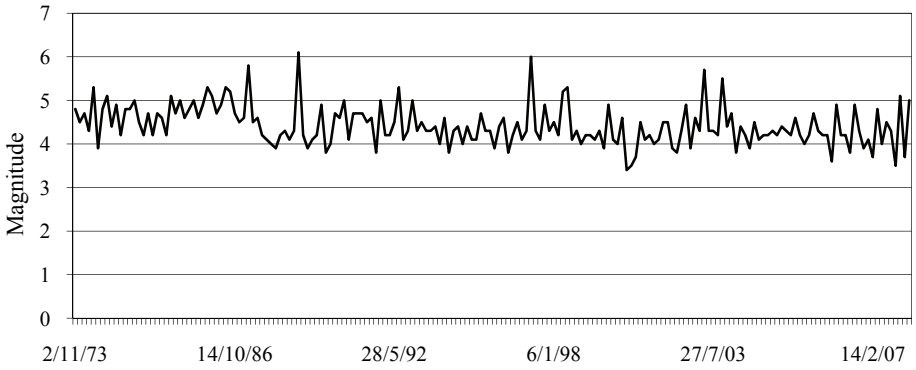


Table 4.3.1: Frequency Distribution of NEQ

NEQ	0	1	2	3	4	5	6	7 or More
Number of Years	7	6	7	1	0	4	2	9

Table 4.2.2: Observed and Fitted Frequencies of NOPV

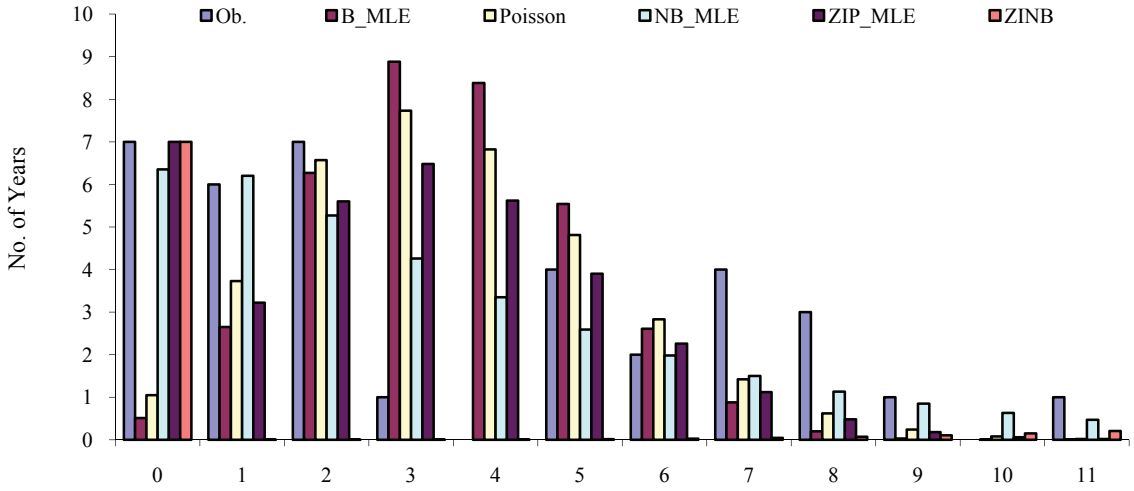
NEQ	Ob.	B_{mom}	B_{ml}	Poisson	NB_{mom}	NB_{ml}	ZIP_{ml}	ZIP_{ml}	$ZINB_{ml}$
0	7	0.78	0.51	1.05	4.96	6.35	12.32	7.00	7.00
1	6	3.58	2.65	3.73	6.26	6.20	0.60	3.22	0.0002
2	7	7.45	6.27	6.57	5.68	5.27	1.62	5.60	0.0009
3	1	9.31	8.88	7.73	4.78	4.26	2.89	6.48	0.0026
4	0	7.75	8.38	6.82	3.80	3.35	3.86	5.62	0.0065
5	4	4.52	5.54	4.81	2.92	2.59	4.12	3.90	0.0137
6	2	1.88	2.61	2.83	2.18	1.98	3.67	2.26	0.0258
7	4	0.56	0.88	1.42	1.60	1.50	2.80	1.12	0.0443
8	3	0.11	0.20	0.62	1.16	1.13	1.86	0.48	0.0708
9	1	0.01	0.03	0.24	0.83	0.85	1.10	0.18	0.1064
10	0	0	0	0.08	0.59	0.63	0.59	0.06	0.1518
11	1	0	0	0.02	0.41	0.47	0.28	0.02	0.2071
GOF Statistic		1.9e+004*	7.7e+003*	97.72*	16.23	15.25	77.09*	83.67*	2.28e+005

Mean =3.43 and Variance =10.19

Parameter estimates: $B_{\hat{p}(mom)} = 0.29$, $B_{\hat{p}(ml)} = 0.32$, $P_{\hat{\lambda}(mom/ml)} = 3.52$, $NB_{\hat{p}(mom)} = 0.34$, $NB_{\hat{k}(mom)} = 1.86$, $NB_{\hat{p}(ml)} = 0.27$, $NB_{\hat{k}(ml)} = 1.35$, $ZIP_{\hat{p}(mom)} = 0.65$, $ZIP_{\hat{\lambda}(mom)} = 5.33$, $ZIP_{\hat{p}(ml)} = 0.80$, $ZIP_{\hat{\lambda}(ml)} = 3.47$, $ZINB_{\hat{\theta}(ml)} = 0.80$, $ZINB_{\hat{p}(ml)} = 0.25$, $ZINB_{\hat{k}(ml)} = 10.19$.

Note: See footnotes, Table 3.1

Figure 4.3.3: Distribution of NEQ for Different Models



Hartal (Strike) Data

The fourth variable is the number of hartals (NOH) per month observed in Dhaka city from 1972 to 2007. Data from 1972 to 2000 was collected from Dasgupta (2001) and from 2001-2007 was collected from the daily newspaper, the Daily Star. Historically, the hartal phenomenon has respectable roots in Ghandi's civil disobedience against British colonialism (the word hartal, derived from Gujarati, is closing down shops or locking doors). In Bangladesh today, hartals are usually associated with the stoppage of vehicular traffic, closure of markets, shops, educational institutions and offices for a specific period of time to articulate agitation (Huq, 1992). When collecting monthly NOH data, care was taken to include all events that were consistent with the above definition of hartal (e.g., a hartal lasting 4 to 8 hours was treated as a half-day hartal, 9 to 12 hours as a full-day hartal; for longer hartals, each 12 hour period was treated as a full-day hartal). Historical patterns of hartals in Dhaka city, NOH with respect to time are plotted in Figure 4.4.1, and the frequency distribution of NOH is shown in Table 4.4.1. Between 1972 and 2007, 413 hartals were observed and the monthly hartal rate is 0.96 per month (see Table 4.4.2). Figure 4.4.1 shows the NOH for two periods: 1972-1990 (post-independence) and 1991-2007 (parliamentary democracy). It has been observed that the NOH have not decreased since the Independence in 1971. Although there were relatively few hartals in the early years following independence, the NOH began to rise sharply after 1981, with 101 hartals between 1982 and 1990. Since 1991(during the parliamentary democracy), the NOH have continued to rise with 125 hartals occurring from 1991-1996. Thus, the democratic periods (1991-1996 and 2003-2007) have experienced by far the largest number of hartals. Lack of political stability was found to be the main cause for this higher frequency of hartals (for details, see Beyond Hartals, 2005, p. 11). From Table 4.4.1, it may be observed that the hartal data contains about 60% of zeroes. Table 4.4.2 indicates that NOH process displays over-dispersion with a variance to mean > 1. According to data in this study (Table 4.4.2), the negative binomial distribution to model NOH with 31% chance of hartal per month is recommended.

Figure 4.4.1: Total Hartals in Dhaka City: 1972-2007

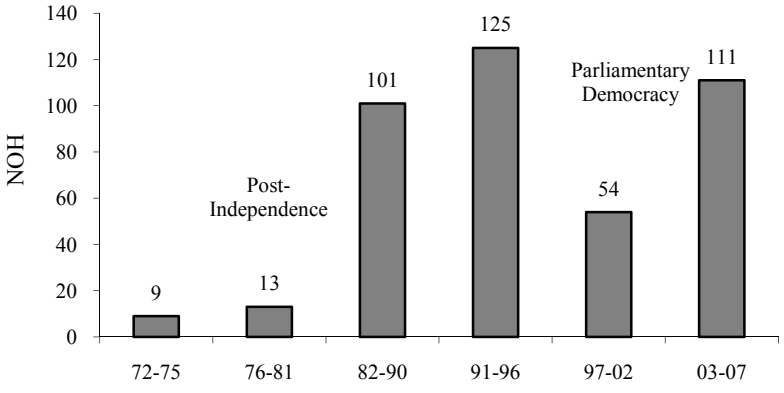


Table 4.4.1: Frequency Distribution of NOH

NOH	Number of Months
0	257
1	86
2	41
3	14
4	9
5	8
6	10
7 or More	7

Table 4.4.2: Observed and Fitted Frequencies of NOH

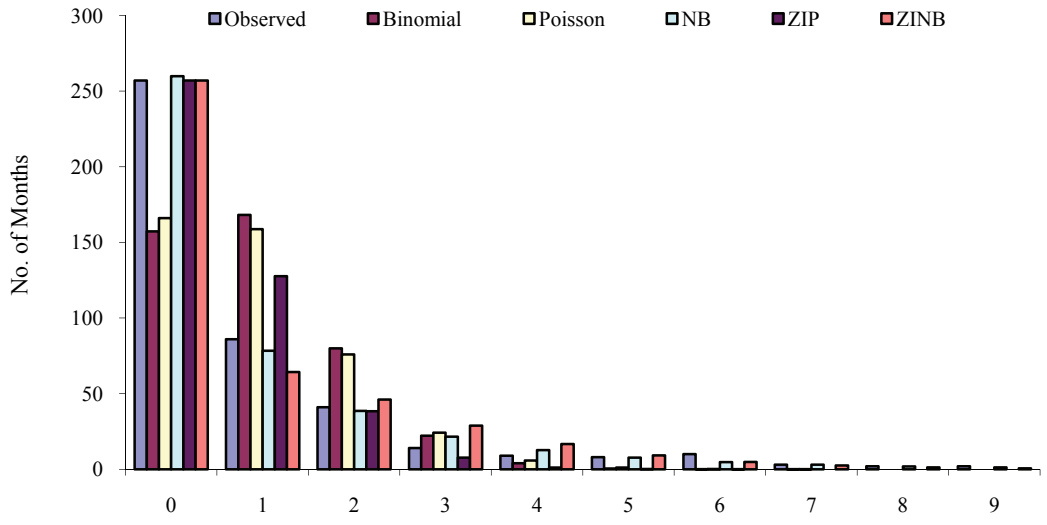
NOH	0	1	2	3	4	5	6	7	8	9	Chi-Square GOF Statistic
Observed Months	257	86	41	14	9	8	10	3	2	2	
B_{mom}	179.76	165.54	67.75	16.17	2.48	0.25	0.01	0.008	0	0	1.8e+007*
B_{ml}	157.23	168.18	79.95	22.17	3.95	0.47	0.03	0.001	0	0	5.4e+006*
Poisson	166.06	158.76	75.89	24.18	5.78	1.10	0.17	0.02	0	0	1.5e+004*
NB_{mom}	267.77	72.94	36.01	20.43	12.35	7.73	4.96	3.23	2.13	1.42	11.78
NB_{ml}	259.92	78.32	38.61	21.50	12.66	7.70	4.78	3.01	1.91	1.23	10.79
ZIP_{mom}	296.82	24.80	34.51	32.01	22.27	12.39	5.75	2.28	0.79	0.24	194.87*
ZIP_{ml}	257	127.66	38.34	7.67	1.15	0.13	0.01	0.00	0.00	0.00	7.3e+005*
$ZINB_{ml}$	257.00	64.28	46.11	28.80	16.65	9.17	4.87	2.53	1.28	0.64	27.88*

Mean =0.96 and Variance =3.35

Parameter estimates: $B_{\hat{P}(mom)} = 0.09$, $B_{\hat{P}(ml)} = 0.10$, $P_{\hat{\lambda}(mom/ml)} = 0.95$, $NB_{\hat{p}(mom)} = 0.28$, $NB_{\hat{k}(mom)} = 0.38$, $NB_{\hat{p}(ml)} = 0.31$, $NB_{\hat{k}(ml)} = 0.44$, $ZIP_{\hat{p}(mom)} = 0.31$, $ZIP_{\hat{\lambda}(mom)} = 2.78$, $ZIP_{\hat{p}(ml)} = 0.40$, $ZIP_{\hat{\lambda}(ml)} = 0.60$, $ZINB_{\hat{\theta}(ml)} = 0.40$, $ZINB_{\hat{p}(ml)} = 0.56$ and $ZINB_{\hat{k}(ml)} = 2.26$

Note: See footnotes, Table 3.1

Figure 4.4.2: Distribution of NOH for Different Models



Conclusion

This study reviewed some discrete models and compared them by assuming different amounts of zeroes in a sample. The following models were considered: the binomial model, the Poisson model, the negative binomial model and the zero-inflated and truncated models. A simulation study was conducted to observe the effects of excess zeroes on selected models, where data was generated from the Poisson model. This simulation study indicated that both the negative binomial and the ZIP models were useful to model discrete data with excess zeroes in the sample. Other models fit data containing excess zeroes poorly. Real-life examples were also used to illustrate the performance of the proposed models. All processes exhibited over-dispersion characteristic and could be fit well by the negative binomial model, with the exception of number of patients per day visiting a medical hospital, this data was better fit by ZTNB.

Acknowledgements

The authors are grateful to the management of the BMSSU medical college and hospital for providing records on the number of outpatient visits per day for the period of 26 April 2007 to 23 July 2007. We gratefully acknowledge contributions of the USGS staff for providing us with information about earthquakes in Bangladesh. Further we wish to thank the library staff of the Press Institute of Bangladesh for providing us information about hartal statistics. The article was partially completed while the second author was visited ISRT, Dhaka University, Bangladesh and ISI, Calcutta, India during July 2006.

References

Bliss, C. I., & Fisher, R. A. (1953). Fitting the negative binomial distribution to biological data. *Biometrics*, *9*, 176-200.

Bohning, D. (1998). Zero-inflated Poisson models and C.A.M.A.N: A tutorial collection of evidence. *Biometrical Journal*, *40*, 833-843.

Dasgupta, A. (2001). *Sangbadpatrey Hartalchitra*. Press Institute of Bangladesh, Dhaka.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series* A, *222*, 309-368.

Fisher, R. A. (1941). The negative binomial distribution. *Annals of Eugenics*, *11*, 182-187.

Hoque, M. M. (2003). *Injuries from road traffic accidents: A serious health threat to the children*. Proceedings published on the World Health Day, 2003.

Huq, E. (1992). *Bangla academy byabaharik bangla abhidhan*. Dhaka: Bangla Academy.

Kibria, B. M. G. (2006). Applications of some discrete regression models for count data. *Pakistan Journal of Statistics and Operation Research*, *2*(*1*), 1-16.

Lloyd-Smith, J. D. (2007). Maximum likelihood estimation of the negative binomial dispersion parameter for highly over-dispersed data with applications to infectious diseases. *pLos one*, *2*, 1-8.

Lord, D., Washington, S. P., & Ivan, J. N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis & Prevention*, *37*, 35-46.

MATLAB (version 7.0), MathWorks, New York, 2004-2007.

Neyman, J. (1939). On a new class of contagious distribution, applicable in entomology and bacteriology. *Annals of Mathematical Statistics*, *10*, 35-57.

Pearson, K. (1894). Contribution to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London*, *Series A*, 71-110.

Rider, P. R. (1961). Estimating the parameters of mixed Poisson, binomial and Weibul distributions by method of moments. *Bulletin de l'Institut International de Statistiques*, *38*, Part 2.

Ross, G. J. S., & Preece, D. A. (1985). The negative binomial distribution. *The Statistician*, *34*, 323-336.

Shankar, V. N., Ulfarsson, G. F., Pendyala, R. M., & Nebergal, M. B. (2003). Modeling crashes involving pedestrians and motorized traffic. *Safety Science*, *41*, 627-640.

Taylor, L. R. (1961). Aggregation, variance and the mean. *Nature*, *189*, 732-735.

United Nations. (2005). *Beyond Hartals: Towards Democratic Dialogue in Bangladesh*. United Nations Development Program Bangladesh, ISBN 984-32-1424-2, March 2005.

Warton, D. I. (2005). Many zeroes do not mean zero inflation: comparing the goodness of-fit of parametric models to multivariate abundance data. *Environmetris*, *16*, 275-289.

White, G. C., & Bennetts, R. E. (1996). Analysis of frequency count data using the negative binomial distribution. *Ecology*, *77*, 2549-2557.

Zhang, Y. Z., & Lord, D. (2007). Estimating dispersion parameter of negative binomial distribution for analyzing crash data using bootstrapped maximum likelihood method, *Journal of Transportation Research Board,* Issue Number: *2019*, 15-21.