BENIN JOURNAL OF STATISTICS
ISSN 2682-5767
Vol. 8, pp. 13– 35 (2025)

On Zero Truncated Under-dispersed Count Data

H. B. Lawal
College of Postgraduate Studies Summit University, Offa, Kwara State, Nigeria

(*Received: 14 June 2024; Accepted: 27 September 2024*)

Abstract. In this paper, we employ two and three parameter distributions to model the under-dispersed zero truncated data presented in a recent article. We illustrate with a large frequency data having many categories that, the zero-truncated negative binomial and its mixture distributions always fail to converge and consequently producing estimated probabilities that sum more than 1 within the range of the data. The two-parameter zero-truncated versions of the following distributions are considered here: the generalized Poisson (ZTGP), the new logarithmic distribution (ZTNLD), the new geometric discrete Pareto distribution (ZTNGDP), the generalized Poisson-Lindley (ZTGPLD), the zero-truncated Poisson-Exponential-Gamma (ZTPEGD) and the Quasi-Negative Binomial (ZTQNBD2). The three-parameter distributions similarly considered include the zero-truncated Quasi-negative binomial (ZTQNBD), the zero-truncated Inverse Trinomial (ZTIT), the zero-truncated Delaporte (ZTDLPD), the zero-truncated Negative binomial-Erlang distribution (ZTNB-ELD), the New three parameter Poisson-Lindley distribution (ZTNTPLD) and the New three-parameter size-biased Poisson- Lindley distribution (NTPSBPLD) zero-truncated versions. Three frequency data sets were employed and we further extend our analyses to zero-truncated count data having co-variates (GLM) by utilizing the Nigerian Health Insurance Survey (NHIS) data. Among the models, the ZTPEGD and the ZTDLPD perform better than all the other models for the example data sets employed in this study. The Delaporte is particularly attractive and easier to program in R since there is a Delaporte package in R that can be appropriately used. All models in this study were implemented with SAS NLMIXED and corresponding written R codes. Two of the R codes are presented in Appendices I and II.

Keywords: Underdispersion, zero-truncated, Delaporte, three-parameter distributions.

Published by: Department of Statistics, University of Benin, Nigeria

1. Introduction

Umar et al. (2019) fitted a series of models to three frequency data sets that all exhibit underdispersion, where the dispersion index DI $= s^2/\bar{y} < 1$ for each data set. Of concern is the use of negative binomial mixture distributions, such as the three-parameter zero-truncated negative binomial-Erlang (ZTNB-ELD),

Corresponding author Email: blawal66@gmail.com

the four parameter zero-truncated negative binomial beta exponential (ZTNB-BED) distribution as well as the zero-truncated Negative binomial (ZTNB). The other employed distributions are all one-parameter type including the zero-truncated Poisson (ZTPD) distribution. These negative binomial mixture distributions were shown to fit the data sets better than the other models, most especially for the metaphitamine data set in Table 2 and the accident data in Table 3 of Umar et al. (2019) based on goodness-of-fit statistics -2LL and AIC. It is well known that for under-dispersed data, the negative binomial or its mixture distributions have convergence problems for any fitting algorithm. Thus attempt to fit the ZTNBD or ZTNB-BED will lead to convergence problems, manifesting into estimated probabilities sometimes greater than 1. The results of these models as presented in Table 2 of Umar et al. (2019) buttress this problem with sums of expected frequencies being greater than the sample size of $(n = 3345.0)$. Table 2 data set is far more intractable than data sets in Tables 1 and 3 in Umar et al. (2019) because they have smaller observed sample sizes. In actuality, for all discrete distributions, the sum of estimated probabilities are always less than 1 (Lawal, 2018)-the exception sometimes being the Poisson. Therefore the sum of expected frequencies of 2245.3 and 3345.1 in ZTNBD and ZTNB-ELD respectively in Umar et al. (2019) should have been a warning sign.

In this paper, we shall consider only two or more parameter zero-truncated distributions as it has been established that none of the one-parameter ZT distributions fit the data. Among the two-parameter distributions that we will consider are the generalized Poisson (ZTGP), the zero-truncated new Geometric Discrete Pareto Distribution (ZTNGDP), the zero-truncated Poisson Exponential Gamma distribution (ZTPEGD); the zero-truncated two-parameter quasi-negative binomial distribution designated here as (ZTQNBD_2); the zero-truncated generalized Poisson-Lindley (ZTGPLD) and the zero-truncated new logarithmic distribution (ZTNLD).

Also considered in this paper are the three parameter zero-truncated distributions Quasi-negative binomial (ZTQNBD), the zero-truncated Inverse-trinomial (ZIT), the zero- truncated Delaporte (ZTDLD); the zero-truncated new three-parameter Poisson-Lindley distribution (NTPLD); the zero-truncated new three-parameter size-biased Poisson Lindley distribution (NTPSBPLD); the zero-truncated negative binomial-Erlang (ZTNB-ELD) distribution as well as the four parameter zero-truncated extended Com Poisson (ZTECOM) distribution. These distributions are briefly described in the next sections.

Most of these distributions belong to the ABM class (Awad et al., 2016), named after the authors, whose variance functions are of the form:

$$V_p(\mu) = \mu \left(1 + \frac{\mu}{p}\right)^r, \quad p > 0, \; r = 1, 2, \ldots \tag{1}$$

Belonging to this class is the Poisson $(r = 0)$, the Negative binomial-NB $(r = 1)$ and the generalized Poisson- GP $(r = 2)$. The variance of the NB is $\sigma^2 = \mu(1 + k\mu) > \mu$, the NB therefore is mostly appropriate for over-dispersed count data. An attempt to implement even the ZTNB in R package *glmmTMB()*

for the emphipetamine data will return a zero for the NB dispersion parameter k. Attempt to implement ZTNB-ELD for instance in R using optim optimizer returns convergence issues-simply because the data in Umar's Table 2 is strongly under-dispersed.

2. Materials and Methods

2.1 *Zero-Truncated Discrete Distributions*

For a response variable Y with a probability density function $f(y|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of its parameters. a zero-truncated model at $y = 0$ has the distribution

$$f(y_i, \boldsymbol{\theta}) = \frac{f(y_i, \boldsymbol{\theta})}{\Pr(Y_i > 0)} = \frac{f(y_i, \boldsymbol{\theta})}{1 - \Pr(Y_i = 0)} \quad y_i = 1, 2, \ldots, \tag{2}$$

where $\Pr(Y_i = 0)$ is designated here as $f(0)$.
We present in the following sections, brief descriptions of the zero-truncated two and three-parameter discrete distributions employed in this study.
Employing (2), it is not too difficult to show that the zero-truncated Negative Binomial has the pmf displayed in (3).

$$f_{rzt}(y|r, p) = \frac{\Gamma(r + y)p^y(1 - p)^r}{y!\Gamma(r)[(1 - f(0)]}, \quad y = 1, 2, \ldots \tag{3}$$

with $f(0) = (1 - p)^r$

2.1.1 The Zero-Truncated Generalized Poisson Distribution-ZTGP:

The type I generalized Poisson distribution (Consul and Famoye, 1992) has the following pmf:

$$\Pr(y_i, \mu_i, \alpha) = \left(\frac{\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \frac{(1 + \alpha y_i)^{y_i - 1}}{y_i!} \exp\left\{-\frac{\mu_i(1 + \alpha y_i)}{(1 + \alpha\mu_i)}\right\}, \quad y_i = 0, 1, \ldots \tag{4}$$

with mean $E(Y_i) = \mu_i$ and $\text{Var}(Y_i) = \mu_i(1 + \alpha\mu_i)^2$. Again, employing (2), its corresponding zero-truncated pmf is:

$$f_{zt}(y_i; \mu, \alpha) = \frac{\left(\frac{\mu_i}{1 + \alpha\mu_i}\right)^{y_i} \frac{(1 + \alpha y_i)^{y_i - 1}}{y_i!} \exp\left(-\frac{\mu_i(1 + \alpha y_i)}{(1 + \alpha\mu_i)}\right)}{1 - \exp\left[-\frac{\mu_i}{(1 + \alpha\mu_i)}\right]}, \quad y_i = 1, 2, \ldots \tag{5}$$

2.1.2 The Generalized Poisson-Lindley

The Poisson generalized Lindley (PGL) proposed in Atikankul (2023) is a mixture of Poisson and generalized Lindley (GL) distributions and has the pmf:

$$f(y,\alpha,\theta)=\frac{\theta^{\alpha}\Gamma(\alpha+y-1)[\alpha(\theta+2)-\theta+y-2]}{\Gamma(\alpha)\Gamma(y+1)(\theta+1)^{\alpha+y+1}};\ y=0,1,\ldots \tag{6}$$

with $\alpha>1$ and $\theta>0$. $f(0)=\dfrac{\theta^{\alpha}(\theta+2)}{(1+\theta)^{\alpha+1}}$.
Hence, the zero-truncated pmf becomes:

$$f_{zt}(y,\alpha,\theta)=\frac{\theta^{\alpha}\Gamma(\alpha+y-1)[\alpha(\theta+2)-\theta+y-2]}{\Gamma(\alpha)\Gamma(y+1)(\theta+1)^{\alpha+y+1}[(1-f(0)]};\quad y=1,2,\ldots \tag{7}$$

2.1.3 The New Geometric Discrete Pareto Distribution-NGDP

The NGDP proposed in Bhati and Bakouch (2019) has the pmf:

$$f(y|q,\alpha)=\frac{q^{y}}{(y+1)^{\alpha}}-\frac{q^{(y+1)}}{(y+2)^{\alpha}},\quad y=0,1,2,\ldots,\quad 0<q<1,\ \alpha\geq 0. \tag{8}$$

Its mean and variance can be computed from expressions in (9a) and (9b) respectively,

$$\mu_y=q\Phi(q,\alpha,2) \tag{9a}$$
$$\sigma_y^2=2q\Phi(q,\alpha-1,2)-q\Phi(q,\alpha,2)[3+q\Phi(q,\alpha,2)] \tag{9b}$$

where $\Phi(z,s,a)=\sum_{k=0}^{\infty}\dfrac{z^k}{(a+k)^s}$. Its zero-truncated has the pmf:

$$f_{zt}(y|q,\alpha)=\left(\frac{q^{y}}{(y+1)^{\alpha}}-\frac{q^{(y+1)}}{(y+2)^{\alpha}}\right)/[1-f(0)],\quad y=1,2,\ldots,\quad 0<q<1,\ \alpha\geq 0. \tag{10}$$

where $f(0)=1-\dfrac{q}{2^{\alpha}}$

2.1.4 The New logarithmic Distribution-ZTNLD

Gómez-Déniz, Sarabia and Calderin-Ojeda (2011) proposed the new logarithmic distribution (NLD) whose pmf has the form:

$$f(y|\alpha,\theta)=\frac{\log(1-\alpha\theta^{y})-\log(1-\alpha\theta^{y+1})}{\log(1-\alpha)};\quad y=0,1,\ldots,\ 0<\theta<1;\ \alpha<1(\alpha\neq 0) \tag{11}$$

Its mean and variance can be computed from expressions in (12a) and (12b) respectively,

$$\mu_y = \frac{1}{\log(1-\alpha)}\sum_{y=1}^{\infty}\log(1-\alpha\theta^y) \tag{12a}$$

$$\sigma_y^2 = \frac{1}{\log(1-\alpha)}\sum_{y=1}^{\infty}(2y-1)\log(1-\alpha\theta^y) - \mu_y^2 \tag{12b}$$

Consequently, its zero-truncated pmf is of the form:

$$f_{zt}(y|\alpha,\theta) = \left(\frac{\log(1-\alpha\theta^y) - \log(1-\alpha\theta^{y+1})}{\log(1-\alpha\theta)}\right); \quad y = 1, 2, \ldots, \tag{13}$$

with $0 < \theta < 1;\ \alpha < 1(\alpha \neq 0)$

2.1.5 The Two-Parameter Quasi Negative Binomial: QNBD$_2$

Shoukri and Aleid (2022) proposed the two-parameter quasi-negative binomial distribution, whose pmf is given in (14)

$$f(y;\theta,\beta) = \frac{\beta-1}{(\beta-1)+\beta y}.\frac{\Gamma(\beta+\beta y)\theta^y(1-\theta)^{\beta+\beta y-y-1}}{y!\Gamma(\beta+\beta y-y)}, \quad y = 0, 1, 2, \ldots \tag{14}$$

with $0 < \theta < 1$ and $0 < \theta\beta < 1$. Shoukri and Aleid (2022) give the properties of this distribution and its moments. Thus, the zero-truncated version of this distribution has the pmf:

$$f_{zt}(y;\theta,\beta) = \frac{\beta-1}{(\beta-1)+\beta y}.\frac{\Gamma(\beta+\beta y)\theta^y(1-\theta)^{\beta+\beta y-y-1}}{y!\Gamma(\beta+\beta y-y)[1-(1-\theta)^{\beta-1}]}, \quad y = 1, 2, \ldots \tag{15}$$

We will designate this distribution as ZTQNBD$_2$ with $f(0) = (1-\theta)^{\beta-1}$.

2.1.6 The Zero-Truncated Poisson-Exponential Gamma Distribution-ZTPEGD

The zero-truncated Poisson-Exponential Gamma distribution Umar (2019) has the pmf:

$$f_{ztpegd}(y;\theta,\alpha) = \frac{\theta^2(\theta+1)^\alpha y! + \theta^\alpha(\theta+1)\Gamma(\alpha+y)}{\{\Gamma(\alpha)(\theta+1)[(\theta+1)^\alpha - \theta^\alpha] + \theta(\theta+1)^\alpha\}(\theta+1)^y y!}, \quad y = 1, 2, \ldots \tag{16}$$

with $\alpha, \theta > 0$.

2.1.7 Log-likelihood functions for two-parameter Distributions

For a single observation i, the log-likelihood for the zero-truncated two-parameter distributions, ZTNB, ZTGP, ZTPLD, ZTNGBD, ZTNLD, ZTQNBD_2, and ZTPEGD are presented respectively in LL1 to LL7 in (17).

$$\text{LL1} = \log[\Gamma(r+y_i)] + y_i \log(p) + r\log(1-p) - \log y_i! - \log[\Gamma(r)] - \log[1-(1-p)^r] \tag{17a}$$

$$\text{LL2} = y_i \log\left(\frac{\mu_i}{1+\alpha\mu_i}\right) + (y_i - 1)\log(1+\alpha y_i) - \frac{\mu_i(1+\alpha y_i)}{1+\alpha\mu_i} - \log(y_i!) - \log\left\{1-\exp\left[\frac{-\mu_i}{(1+\alpha\,\mu_i)}\right]\right\} \tag{17b}$$

$$\text{LL3} = \log\left[\frac{\theta^\alpha\Gamma(\alpha+y_i-1)[\alpha(\theta+2)-\theta+y_i-2]}{\Gamma(\alpha)\Gamma(y_i+1)(\theta+1)^{\alpha+y_i+1}[(1-f(0)]}\right] \tag{17c}$$

$$\text{LL4} = \log\left[\frac{q_i^y}{(y_i+1)^\alpha} - \frac{q^{(y_i+1)}}{(y_i+2)^\alpha}\right] - \log(q) + \alpha\log(2) \tag{17d}$$

$$\text{LL5} = \log\left[\frac{\log(1-\alpha\theta_i^y) - \log(1-\alpha\theta^{y_i+1})}{\log(1-\alpha\theta)}\right] \tag{17e}$$

$$\text{LL6} = \log\left[\frac{\beta-1}{(\beta-1)+\beta y_i}\cdot\frac{\Gamma(\beta+\beta y_i)\theta_i^y(1-\theta)^{\beta+\beta y_i-y_i-1}}{y_i!\Gamma(\beta+\beta y_i-y_i)[1-(1-\theta)^{\beta-1}]}\right] \tag{17f}$$

$$\text{LL7} = \log\left[\frac{\theta^2(\theta+1)^\alpha y_I! + \theta^\alpha(\theta+1)\Gamma(\alpha+y_i)}{\{\Gamma(\alpha)(\theta+1)[(\theta+1)^\alpha-\theta^\alpha]+\theta(\theta+1)^\alpha\}(\theta+1)^{y_i}y_i!}\right] \tag{17g}$$

2.2 *Zero-truncated three parameter discrete distributions*

We also consider some three-parameter distributions that have received considerable attention for implementing zero-truncated data in the literature. This includes the quasi-negative binomial (Lawal, 2019), Li et al., 2011), The Inverse trinomial (Lawal, 2019), The Delaporte distribution, the Negative binomial-Erlang distribution (ZTNB-ED) Umar et al., 2019), the New three-parameter Poisson-Lindley distribution (NTPLD) proposed in Das et al. (2018), and the New three-parameter size-biased Poisson-Lindley distribution (NTPSBPLD) proposed in Shanker and Shukla (2020). We briefly present the zero-truncated pmfs of these distributions in the following sections.

2.2.1 The Quasi-Negative Binomial-QNBD

The quasi-negative binomial distribution recently employed in Li et al. (2011) has the pmf (Hassan and Bilal, 2008) given by:

$$P(Y=y) = \begin{cases} \dfrac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)}\left(\dfrac{1}{1+cy}\right)\left(\dfrac{1+cy}{1+b+cy}\right)^y\left(\dfrac{b}{1+b+cy}\right)^\alpha, & y = 0, 1, \ldots \\ 0 & \text{for } y > m \text{ if } c < 0 \end{cases} \tag{18}$$

where $\alpha > 0$, $b > 0$ and $m = [-1/c]$ with m being the largest positive integer for which $1+mc = 0$. The QNBD in (18) reduces to the negative binomial (NB) distribution when the parameter $c = 0$.
An equivalent form of the QNBD model is proposed in Janardan (1975); Hassan and Bilal (2008) and has the pmf of the form:

$$P(Y=y)=\binom{\alpha+y-1}{y}\frac{\theta_1(\theta_1+\theta_2 y)^{y-1}}{(1+\theta_1+\theta_2 y)^{\alpha+y}},\quad y=0,1,\ldots, \tag{19}$$

The two models are equivalent with $c=\frac{\theta_2}{\theta_1},\ b=\frac{1}{\theta_1}$, and $a=\alpha$. The mean and variance of the QNBD can only be obtained via a series of recurrence relations. Thus the moments do not have closed form expressions and are consequently almost intractably difficult to compute. We propose a method for computing these later in this paper. Its zero truncated pmf is:

$$f_{zt}(\alpha,b,c)=\frac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)[1-f(0)]}\left(\frac{1}{1+cy}\right)\left(\frac{1+cy}{1+b+cy}\right)^{y}\left(\frac{b}{1+b+cy}\right)^{\alpha},y=1,2,\ldots \tag{20}$$

where $f(0)=\left(\frac{b}{1+b}\right)^{\alpha}$.

2.2.2 The Inverse-Trinomial Distribution-IT

The inverse trinomial distribution Shimizu and Yanagimoto (1991) which is derived from the Lagrangian expression has the probability mass function of the form :

$$P(Y=y)=\frac{\lambda p^{\lambda}q^{y}}{y+\lambda}\sum_{t=0}^{|y/2|}\frac{(y+\lambda)!}{t!(t+\lambda)!(y-2t)!}.\left(\frac{pr}{q^2}\right)^{t} \tag{21}$$

$y=0,1,\ldots;\lambda>0,p\geq r$ and $p+q+r=1$. It is so named because its cumulant generating function is the inverse of that for the trinomial distribution. The IT model was employed for over-dispersed medical count data by Phang and Ong (2014). It is a member of the Takac family distribution with a cubic variance function of the mean.

For the zero-truncated ITD, with parameter $\Pr(Y>0)=1-\Pr(Y=0)=1-p^{\lambda}$. Hence, the pmf of zero-truncated ITD random variable Y_t becomes

$$f_{zt}(y)=\frac{\lambda p^{\lambda}q^{y}}{(y+\lambda)}.\frac{1}{(1-p^{\lambda})}\sum_{t=0}^{|y/2|}\frac{(y+\lambda)!}{t!(t+\lambda)!(y-2t)!}.\left(\frac{pr}{q^2}\right)^{t} \tag{22}$$

for $y=1,2,\ldots$. Its means and variances will be computationally obtained.

2.2.3 The Delaporte Distribution-ZTDLPD

The discrete Delaporte probability distribution with parameters α,β, and λ is a convolution of the negative binomial distribution with a Poisson distribution and has the pmf,

$$f(y|\alpha,\beta,\lambda)=\sum_{i=0}^{y}\frac{\Gamma(\alpha+i)\beta^{i}\lambda^{y-i}e^{-\lambda}}{i!\Gamma(\alpha)(1+\beta)^{\alpha+i}(y-i)!} \tag{23}$$

$y=0,1,2,\ldots$ and $\alpha,\beta,\lambda>0$. The distribution has also been described as a compound Poisson distribution with two components. The first component being a fixed λ parameter and second component which has gamma-distribution with parameters α and β. The distribution is named after Pierre Delaporte and is designated here as the DLPD. Its zero-truncated pmf using (2) is displayed in (24).

$$f_{zt}=\left(\frac{1}{1-f(0)}\right)\sum_{i=0}^{y}\frac{\Gamma(\alpha+i)\beta^{i}\lambda^{y-i}e^{-\lambda}}{i!\Gamma(\alpha)(1+\beta)^{\alpha+i}(y-i)!},\quad y=1,2,\ldots \tag{24}$$

where $f(0)=\dfrac{e^{-\lambda}}{(1+\beta)^{\alpha}}$.

2.2.4 *The New 3-parameter Poisson-Lindley: NTPPLD*

The New three-parameter Poisson-Lindley distribution (NTPLD) proposed in Das et al. (2018) has the pmf:

$$f(y;\alpha,\beta,\theta)=\frac{\theta^{2}}{(\theta+1)^{y+2}}.\left(1+\frac{\alpha+\beta y}{\theta\alpha+\beta}\right),\ y=0,1,2,\ldots \tag{25}$$

with $\theta,\beta>0$ and $\theta\alpha+\beta>0$. $f(0)=\dfrac{\theta^{2}(\theta\alpha+\alpha+\beta)}{(\theta+1)^{2}(\theta\alpha+\beta)}$. Hence, its zero-truncated pmf is given in (26).

$$f_{zt}(y;\alpha,\beta,\theta)=\frac{\theta^{2}}{(\theta+1)^{y+2}}.\left(\frac{\theta\alpha+\alpha+\beta y+\beta}{(\theta\alpha+\beta)[1-f(0)]}\right),\ y=1,2,\ldots \tag{26}$$

if $\beta=\theta$ in (26), we have:

$$f_{zt}(y;\alpha,\theta)=\frac{\theta^{2}}{(\theta+1)^{y+2}}.\left(\frac{\theta\alpha+\alpha+\theta y+\theta}{(\theta\alpha+\theta)[1-f_{a}(0)]}\right),\ y=1,2,\ldots \tag{27}$$

where $f_{a}(0)=\dfrac{\theta^{2}(\theta\alpha+\alpha+\theta)}{(\theta+1)^{2}(\theta\alpha+\theta)}$

2.2.5 *The New 3-parameter sized-biased Poisson-Lindley: NTPSBPLD*

The New three-parameter size-biased Poisson- Lindley distribution (NTPSB-PLD) proposed in Shanker and Shukla (2020) has the zero-truncated pmf:

$$f_{zt}(y;\theta,\alpha,\beta) = \frac{\theta^3}{\theta\alpha+2\beta}\frac{y(\beta y+\theta\alpha+\alpha+\beta)}{(\theta+1)^{y+2}}, \quad y=1,2,\ldots \tag{28}$$

with $\alpha, \beta > 0$. Its mean and variance are respectively,

$$\mu = \frac{\alpha\theta^2+2(\alpha+\beta)\theta+6\beta}{\theta(\theta\alpha+2\beta)} \tag{29a}$$

$$\sigma^2 = \frac{2[\alpha^2\theta^3+(\alpha^2+5\alpha\beta)\theta^2+(6\beta^2+6\alpha\beta)\theta+6\beta^2]}{\theta^2(\theta\alpha+2\beta)^2} \tag{29b}$$

The two-parameter variants of the model in (28) are the SBQPLD (Size-based quasi Poisson-Lindley distribution when $\beta = \theta$) and SBNQPLD (Size-based new quasi Poisson-Lindley distribution when $\alpha = \theta$). Both were suggested by Shanker and Mishra (2013, 2017) respectively.

2.2.6 *The Zero-truncated Negative Binomial-Erlang Distribution: ZTNB-ELD*

The Zero-truncated negative Binomial-Erlang distribution has the pmf (Bodhisuwan et al., 2017),

$$f_{zt}(y;r,c,k) = \frac{\binom{y+r-1}{y}\sum_{j=0}^{y}\binom{y}{j}(-1)^j\left(\frac{c}{c+r+j}\right)^k}{1-\left(\frac{c}{c+r}\right)^k}, \quad y=1,2,\ldots \tag{30}$$

with $r, c, k > 0$.

2.2.7 *The Negative binomial COM-Poisson Distribution: NBCOM*

The Negative binomial Com-Poisson is provided in Zhang et al. (2018) and has parameters (r, ν, p) with the pmf given by

$$f(y;r,\nu,p) = \left[\frac{\Gamma(r+y)}{y!\Gamma(r)}\right]^\nu p^y(1-p)^r.\frac{1}{C(r,\nu,p)}; \quad y=0,1,\ldots \tag{31}$$

where $r, v \in (0,\infty)$ and $p \in (0,1)$ with

$$C(r,\nu,p) = \sum_{j=0}^{\infty}\left[\frac{\Gamma(r+j)}{j!\Gamma(r)}\right]^\nu p^j(1-p)^r \tag{32}$$

being the normalizing constant. $f(0) = (1-p)^r C^{-1}$. Consequently, using (2), its zero-truncated pmf becomes,

$$f_{zt}(y; r, \nu, p) = \left[\frac{\Gamma(r+y)}{y!\Gamma(r)}\right]^{\nu} \frac{p^y (1-p)^r}{[C(r,\nu,p) - (1-p)^r]}; \quad y = 1, 2, \ldots \tag{33}$$

2.2.8 *Log-likelihood functions for three-Parameter Distributions*

The log likelihood of a single observation i from **ZTQNBD, ZTIT, ZTNBCOM, NTPSBPLD, NTPPLD, ZTNB-ELD, and ZTDLPD** are given in expressions (34a) to (34g) respectively:

$$\text{LL1} = \log\Gamma(y_i+\alpha) - \log(y_i!) - \log[\Gamma(\alpha)] + \log\left(\frac{1}{1+cy_i}\right) + y_i \log\left(\frac{1+cy_i}{1+b+cy_i}\right) + \alpha\log\left(\frac{b}{1+b+cy_i}\right) - \log[1-f(0)] \tag{34a}$$

$$\text{LL2} = \log(\lambda) + \lambda\log(p) + y_i\log(q) - \log(y_i+\lambda) + \log\left[\sum_{t=0}^{|y_i/2|} \frac{(y_i+\lambda)!}{t!(t+\lambda)!(y_i-2t)!} \cdot \left(\frac{pr}{q^2}\right)^t\right] - \log[1-f(0)] \tag{34b}$$

$$\text{LL3} = y_i\log(p) + r\log(1-p) + \nu\log\Gamma(y_i+r) - \nu\log(y_i!) - \nu\log\Gamma(r) - \log C(r,\nu,p) - \log[1-f(0)] \tag{34c}$$

$$\text{LL4} = 3\log(\theta) - \log(\theta\alpha+2\beta) + \log(y_i) + \log(\beta y_i + \theta\alpha + \alpha + \beta) - (y_i+2)\log(\theta+1) \tag{34d}$$

$$\text{LL5} = 2\log(\theta) - (y_i+2)\log(\theta+1) + \log(\theta\alpha+\alpha+\beta y_i+\beta) - \log(\theta\alpha+\beta) - \log[1-f(0)] \tag{34e}$$

$$\text{LL6} = \log(\binom{y_i+r-1}{y_i}) + \log\left[\sum_{j=0}^{y_i}\binom{y_i}{j}(-1)^j\left(\frac{c}{c+r+j}\right)^k\right] - \log[1-f(0)] \tag{34f}$$

$$\text{LL7} = \log\left[\sum_{i=0}^{y_i}\frac{\Gamma(\alpha+i)\beta^i\lambda^{y_i-i}e^{-\lambda}}{i!\Gamma(\alpha)(1+\beta)^{\alpha+i}(y_i-i)!}\right] - \log\left[1-f(0)\right] \tag{34g}$$

2.2.9 *Estimation*

Maximum-likelihood estimation of all log-likelihoods in (17) and (34) is carried out with PROC NLMIXED in SAS, which minimizes the function $-LL(y,\Theta)$ over the parameter space Θ numerically. The integral approximations in PROC NLMIXED is the Adaptive Gaussian Quadrature Pinheiro and Bates (20??) and our choice optimization algorithm here is the Newton-Raphson techniques.

3. Results and Discussion

The above two-parameter and three-parameter truncated models are applied to three frequency distributed example data sets presented below.

3.1 *Two-Parameter Applications*

Firstly, we apply the two-parameter zero-truncated models to these three data examples and compare our results with the three-parameter ZT models.

3.1.1 *Example I-Distribution of Stillbirths*

The data in this example presented in Table 1 is the frequency distribution of mothers with at least one stillbirth in the urban areas of Nigeria (NDHS, 2013). The data is also recently analyzed in Umar et al. (2019). It has an observed mean and variance 1.44453 and 0.5975 respectively. Consequently, the dispersion index is DI $= 0.4134 \ll 1$ indicating a strong under-dispersion. The data has a sample size of $n = 658$.

Table 1: Distribution of stillbirths in Urban areas of Nigeria

Y	Count	ZTNB	ZTGP	ZTNGDP	ZTPEGD	$ZTQNBD_2$	ZTGPLD	ZTNLD
1	453	450.5525	450.2318	450.7300	450.4029	467.5226	452.1750	451.5532
2	139	146.1122	146.7236	145.9022	145.8231	126.4600	144.4108	145.0581
3	49	43.8356	43.6470	43.8047	44.4271	40.6566	43.2590	43.4491
4	12	12.6190	12.5057	12.6434	12.7086	14.3766	12.8039	12.7244
5	5	3.5407	3.5241	3.5619	3.4467	5.4003	3.7757	3.7014
					(4.6383)			
Total	658	656.6601	656.6322	656.6422	656.8084	654.4161	656.4244	656.4862
y_a		20	17	17	16	25	18	25
MLE		$\hat{p}$=0.2514	$\hat{\mu}$=0.5577	$\hat{q}$=0.2545	$\hat{\alpha}$=3.4118	$\hat{\beta}$=29.6127	$\hat{\alpha}$=0.9507	$\hat{\alpha}$= -0.8619
		$\hat{r}$=1.5794	$\hat{\tau}$=0.2579	$\hat{\alpha}$= -0.5260	$\hat{\theta}$= 4.2920	$\hat{\theta}$=0.0078	$\hat{\theta}$=2.3876	$\hat{\theta}$=0.2901
μ	1.4453							
σ^2	0.5975							
$\bar{y}$		1.4453	1.4453	1.4453	1.4452	1.4452	1.4453	1.4455
s^2		0.6087	0.6082	0.6070	0.6033	0.7536	0.6228	0.6188
X_g^2		1.0012	1.1030	0.9882	0.8726	5.5666	1.0397	1.0169
d.f.		2	2	2	2	2	2	2
p-value		0.6062	0.5761	0.6101	0.6464*	0.0618	0.5946	0.6014
X_W^2		644.8668	645.3370	643.5376	650.6427	520.9050	630.2672	634.3413
d.f		655	655	655	655	655	655	655
-2LL		1174.1	1174.2	1174.1	1173.7	1180.8	1174.4	1174.3
AIC		1178.1	1178.2	1178.1	1177.7	1184.8	1178.4	1178.3
BIC		1187.1	1187.2	1187.1	1186.7	1193.7	1187.4	1187.3

(1) As observed in Lawal (2018) all the models (as typical of discrete distributions) produce cumulative sum of expected values not summing to the sample size $n = 658$. The estimated probabilities do not sum to 1.000 (and consequently, the expected values summing to $n = 658$ until $y_a = 16$ for the ZTPEGD for example, which is outside the observed range $1 \leq Y \leq 5$ for the data. The mean and variance of this distribution can only be estimated at this point. The values of y_a at which the estimated probabilities for each of these distributions sum to 1 are given by y_a. In appendix I (providing theR codes and outputs) for the ZTPEGD for example, the empirical mean and variance at $Y = 5$ are 1.4337 and 0.5630 respectively. Whereas the theoretical mean and variance of the ZTPEGD are 1.4452 and 0.6033 attained at Y=16 respectively when the estimated probabilities actually sum to 1.000000 (using six digits). Note that the means and variance remain constant at $Y > 16$

(2) For the ZTPEGD again as an example, $\sum_{i=0}^{5} \hat{m}_i = 656.8084$ with actual estimated expected value being $\hat{m}_5 = 3.4467$. However, in computing the grouped Pearson's X^2, we had used 3.4467 + (658-656.8084) = 4.6383. This is similarly applied to all the models in the computations of X_g^2.

(3) In computing Pearson's group X_g^2, we have employed the Lawal's (1980) rule which allows expected values to be as small as $rd^{-3/2}$, where r is the number of expected values less than 3, and d is the corresponding d.f. All the two-parameter models satisfy the criterion hence, we do not need

to collapse any categories. The resulting Pearson's X_g^2 are displayed with the corresponding d.f. based on $5-2-1=2$ d.f., since two parameters are estimated from the model. Based on the grouped Pearson's X^2, model ZTPEGD is the most parsimonious with a p-value of 0.6464 and is closely followed by the ZTNGDP. ZTPEGD also has the lowest -2LL or AIC.

(4) We observe that the computed means $\bar{y}$ for all the models are all very close to the observed mean of 1.4453. The closest estimated variance to the observed variance is 0.6033, that of ZTPEGD. ZTNB, ZTGP and ZTNGDP also have estimated variances that are very close to the observed. The other distributions over-estimated the observed variance in the data.

(5) The Wald's goodness-of-fit test statistic $X_W^2 = \sum_{i=1}^{N} \frac{(y_i - \hat{m}_i)^2}{\hat{\sigma}_i^2}$ are similarly presented. The most parsimonious with this GOF is the ZTQNBD_2. This is not surprising as it gives the highest estimate of the observed variance and since X_W^2 is a function of the reciprocal of the variance, we expect it to lowest for all the models.

3.1.2 Example II: The Methamphetamine Data

The frequency data set in Table 2 gives the number of Methamphetamine from the Office of the Narcotics Control Board (ONCB), Thailand in a Bangkok metropolitan region (Bodhisuwan et al., 2017). The data were presented in Umar et al. (2019). The observed mean and variance for this data are $\mu = 1.1250$ and $\sigma^2 = 0.4054$. Consequently, the dispersion index DI $= 0.36 \ll 1$. Hence, this data is strongly under-dispersed.

Table 2: Observed and Expected frequencies for the Methamphetamine data for the Models

Y	Count	ZTNB	ZTGP	ZTNGDP	ZTPEGD	ZTQNBD_2	ZTGPLD	ZTNLD
1	3114	2982.8529	2993.3116	3059.2187	2976.6175	2986.6343	2988.1818	3114.0001
2	163	310.9127	297.7201	235.8907	327.9534	308.2312	314.0136	103.9429
3	23	43.1453	44.2705	37.0064	36.0160	42.2477	37.4655	49.0458
4	20	6.7323	7.7955	8.6217	3.9343	6.5594	4.6617	27.8294
5	9	1.1203	1.5076	2.5895	0.4272	1.0938	0.5907	17.0923
6	3	0.1941	0.3095	0.9288	0.0461	0.1911	0.0755	10.9498
7	3	0.0346	0.0662	0.3800	0.0049	0.0345	0.0097	7.1909
8	3	0.0063	0.0146	0.1720	0.0005	0.0064	0.0013	4.7959
9	4	0.0012	0.0033	0.0844	0.0001	0.0012	0.0002	3.2305
10	3	0.0002	0.0008	0.0442	0.0000	0.0002	0.0000	2.1904
Total	3045	3344.9999	3344.9998	3344.9364	3344.9999	3344.9999	3344.9999	3340.2679
y_a		14	12	82	11	13	11	55
		$\hat{p}$=0.0030*	$\hat{\mu}$=0.0012	$\hat{\alpha}$=6.0671	$\hat{\alpha}$=2.2993	$\hat{\theta}$=0.1701	$\hat{\alpha}$=1.2065	$\hat{\alpha}$=1.4549
		$\hat{r}$=0.20778*	$\hat{\tau}$=100.06	$\hat{q}$=1.0000*	$\hat{\theta}$=9.2679	$\hat{\beta}$=1.0771	$\hat{\theta}$=6.7811	$\hat{\theta}$=0.6873
μ	1.1250							
σ^2	0.4054							
$\bar{y}$		1.1265	1.1249	1.1065	1.1237	1.1250	1.1213	1.1760
s^2		0.1540	0.1579	0.1581	0.1388	0.1519	0.1400	0.7796
X_W^2		8803.7247	8588.3845	8581.9372	9764.5230	8927.9409	9682.7013	1750.2414
d.f.		3342	3342	3342	3342	3342	3342	3342
-2LL		2505.4	2457.8	2310.3	2629.2	2502.1	2576.6	2268.1
AIC		2509.4	2461.8	2314.3	2633.2	2506.1	2580.6	2272.1
BIC		2521.6	2474.0	2326.5	2645.4	2518.3	2592.8	2284.3

(1) Again all the models have sum of expected values less than $n = 3345$ within the observed range $1 \leq Y \leq 10$ of the observed data. The estimated probabilities only sum to 1 at $Y = y_a$ which is outside the observed range of the data. These y_a are displayed for each model.. For the ZTNBD, because the data is strongly under-dispersed, the ZTNB model fails to converge for various combinations of r and p. Thus, SAS PROC NLMIXED gives a warning that the variance-covariane was computed using the Moore-Penrose approach. The consequent of this, is that we may have sum of estimated probabilities being greater than 1 before $\sum \hat{m}_i = n$.

(2) None of the models fits the data. While QNBD_2, ZTGP, and ZTPEGD give estimated means that are either equal or slightly less than the observed $\mu = 1.1250$. All the models underestimated the true variance of the data which is 0.4054 (the exception being the ZTNLD which overestimates the variance). This under estimation of the observed variance in the data accounts for the lack of fit of all these models.

(3) Based on the -2LL and AIC, the most parsimonious model would be ZTNLD but produces estimated moments that grossly over-estimated the observed moments.

3.2 *Three Parameter Zero-Truncated Models' Applications*

We present the results of applying some of the three-parameter zero truncated models to the following data:

3.2.1 Example III: Household migrant Data

The data in this example in Table 3 gives the distribution of the number of households having at least one migrant (Y), according to the number of observed migrants as reported in Singh and Yadav (1971) and analyzed in Shanker and Shukla (2020).

Results of the application of zero-truncated three-parameter models to the household migrant data in Table 3 indicate that all provide estimated means that are very or exactly close to the observed mean of 1.5475. However, of the models, the zero-truncated Delaporte provides the most parsimonious model. In general we observe the following:

- The probabilities of each distributions do not sum to 1.000 until y_a displayed in the table, which is outside the range $1 \leq Y \leq 8$.
- Based on -2LL and AIC, the ZTDLPD is the most parsimonious.
- However, based on Pearson's grouped GOF, the most parsimonious model would be the zero-truncated quasi-negative binomial with $X_g^2 = 2.6519$ on 4 d.f. (pvalue=0.6177)
- Based on Wald's GOF, the most parsimonious is the ZTDLPD because it has a larger estimated variance $s^2 = 0.8128$, which happens to be the closest to the observed variance of 0.8186 in the data.
- All the three models ZTQNBD, ZTIT and ZTDLPD outperform the others and would each be suitable for this data set. They outperform the ZTNB-ELD.

Table 3: Zero-truncated Three-parameters Applied to the Household Migrant Data

Y	count	ZTQNBD	ZTIT	ZTDLPD	NBCOM	NTPSBPLD	ZTNTPLD	ZTNB-ELD
1	375	376.0993	375.4345	373.4299	376.3302	363.6621	376.3375	376.1869
2	143	141.6323	143.3341	146.9495	140.4413	156.3045	140.2411	141.3854
3	49	48.1418	46.7632	46.6549	48.8596	50.3856	49.0492	48.2401
4	17	15.9564	16.1342	14.6733	16.4069	14.4374	16.4780	16.0364
5	2	5.3287	5.4609	5.0534	5.3930	3.8783	5.3837	5.3487
6	2	1.8207	1.8789	1.9116	1.7476	1.0002	1.7234	1.8144
7	1	0.6411	0.6477	0.7682	0.5605	0.2508	0.5431	0.6300
8	1	0.2334	0.2251	0.3195	0.1784	0.0616	0.1691	0.2246
Total	590	589.8537	589.8787	589.7603	589.9177	589.9805	589.9250	589.8666
y_a		27	21	25	19	17	18	18
		$\hat{a}$=5.4112	$\hat{p}$=0.8728	$\hat{\alpha}$=0.2627	$\hat{\nu}$=5.7330	$\hat{\alpha}$=12.7568	$\hat{\alpha}$=261.40	$\hat{r}$=4.5305
		$\hat{b}$=8.4846	$\hat{r}$=0.0224	$\hat{\beta}$=0.8944	$\hat{p}$=0.3016	$\hat{\theta}$=3.6532	$\hat{\beta}$=901.91	$\hat{c}$=32.1112
		$\hat{c}$=0.1169	$\hat{\lambda}$=2.7219	$\hat{\lambda}$=0.5798	$\hat{r}$=1.0757	$\hat{\beta}$= -0.0002	$\hat{\theta}$=2.5680	$\hat{k}$=4.4964
μ	1.5475							
σ^2	0.8186							
$\bar{y}$		1.5474	1.5475	1.5475	1.5475	1.5475	1.5475	1.5474
s^2		0.8094	0.8075	0.8128	0.7993	0.6973	0.7972	0.8085
X_g^2		2.6519	2.7902	2.8661	4.6854	8.2326	4.1152	3.5761
d.f		4	4	4	3	2	3	4
pvalue		0.6177	0.5935	0.5805	0.1963	0.0163	0.2493	0.4664
X_W^2		595.7063	597.1184	593.2251	603.2790	691.4707	604.8212	596.3681
d.f.		587	587	587	587	587	587	587
-2LL		1185.4	1185.6	1185.2	1185.9	1189.6	1186.0	1185.5
AIC		1191.4	1191.6	1191.2	1191.9	1195.6	1192.0	1191.5
BIC		1204.5	1204.7	1204.3	1205.0	1208.7	1205.1	1204.5

3.2.2 Example IV: The Methamphetamine Data

In this section, we apply the three parameter zero-truncated models to the Methamphetamine data previously presented in Table 2. The results of implementing some of these models are presented in Table 4.

The zero-truncated Delarporte distribution (ZTDLPD) is the most parsimonious amongst the distributions considered here. It has a grouped X^2 of 8.2587 on 6 d.f. (p-value=0.2198). A good fit. Unfortunately, because of the strong under-dispersion of this data set, all the other three parameter zero-truncated distributions considered in Table 3 fail to converge. This is not unusual especially, for negative binomial mixture models. We present in Appendix II the R codes and output in the implementation of ZTDLPD to this data set. Similar to the data in Table 1, the sum of estimated probabilities in the observed range $1 \leq Y \leq 10$ is $3342.4870 < n = 3345$. The $\hat{\pi}_i$ do not sum to 1.0000 until Y=62. At $Y = 62$ the estimated theoretical mean and variance are 1.1250 and 0.4435 respectively- both of which are very close to the observed moments in the data. We observed here that the estimated variances from the ZTNTPLD and NSBPLD models grossly under estimated the observed variance in the data-hence, their inability to fit the data.The ZTNB-ELD does not behave as indicated in Umar et al. (2019). Like all mixed negative binomial models, it suffers from convergence with under-dispersed count data. Thus, its parameters exist here for $1 \leq Y \leq 32$. For values of $Y > 32$, the sum of expected values will be greater than the sample size $n = 3345$. This explains the sum of expected values being 3345.1 in Umar et al. (2019) and it should not be. Under this condition therefore, its estimated variance is less than the observed variance and it therefore does not fit the data but performs better than the ZTNTPLD and NSBPLD.

Table 4: Distribution of the number of Methamphetamine

Y	Count	ZTNTPLD	NSBPLD	ZTDLPD	ZTNB-ELD
1	3114	2973.4321	2963.1491	3114.5778	2907.0197
2	163	330.2936	348.5078	161.4043	325.2221
3	23	36.6895	30.7420	28.8169	72.5858
4	20	4.0755	2.4105	14.7027	22.7673
5	9	0.4527	0.1772	8.8126	8.8004
6	3	0.0503	0.0125	5.5375	3.9247
7	3	0.0056	0.0009	3.5848	1.9435
8	3	0.0006	0.0001	2.3711	1.0427
9	4	0.0001	0.0000	1.5940	0.5961
10	3	0.0000	0.0000	1.0854	0.3589
				(3.5983)	(1.0995)
Total	3345	3345.0000	3345.0000	3342.4870	3344.2614
y_a		11	10	62	32
		$\hat{\alpha}$=414.00	$\hat{\alpha}$=0.3660	$\hat{\alpha}$=0.0038	$\hat{r} \approx 0.000$
		$\hat{\beta}$=0.0214	$\hat{\beta}$=0.0000001	$\hat{\beta}$=3.1196	$\hat{c}$=4.3655
		$\hat{\theta}$=8.0024	$\hat{\theta}$=16.0048	$\hat{\lambda}$=0.0721	$\hat{k}$=0.3457
μ	1.1250				
σ^2	0.4054				
$\bar{y}$		1.1250	1.1250	1.1250	1.1880
s^2		0.1406	0.1328	0.4435	0.3765
X_W^2		9644.2029	10,211.347	3057.052	3619.972
d.f.		3342	3342	3342	3342
-2LL		2624.8	2752.5	2207.8	2374.1
AIC		2630.8	2758.2	2213.8	2380.1
BIC		2649.2	2776.8	2232.1	2398.4
X_g^2		> 158.0	> 676.0	8.2587	> 134.0
d.f.		1	1	6	5
p-value		0.0000	0.0000	0.2198	0.0000

3.3 *GLM Applications-The NHIS Data*

Adesina et al. (2021) re-analyzed the National Health Insurance Scheme (NHIS) data that is fully described in Mendeley Data website, *https:/data.mendeley/z7wznk53cf/8*. The data, obtained from health facilities in Ota, Ogun State, Nigeria has 1647 patients. The response variable of interest is Y-the number of encounter visits to the doctors The predictors in the data set are:

- sex- gender of patients (male=1, female=0)
- age- age of patients
- fup- (follow-up=1, no-follow-up=0)
- ecs- nature of admission (1=for in patients, 0=for outpatients)

The first and last five observations of the data are presented below.

ID	sex	age	fup	ecs	Y
1	1	27	1	0	1
2	0	38	1	0	3
3	0	55	1	0	1
4	0	41	1	0	1
5	0	34	1	0	4
...	...	...	...	...	...
1643	1	17	0	0	1
1644	1	51	0	0	2
1645	0	10	0	0	1
1646	0	10	0	0	1
1647	1	48	0	0	1

The variable Y has $\bar{y} = 3.3892$; $s^2 = 11.5987$, thus giving a dispersion index of $3.4223 > 1$ thus indicating strong over-dispersion. Further, Y has the range $[1, 27]$, thus it is truncated at $Y = 0$, with 98.54% of the data, that is, 1623 observations in the range [1,15], indicating extreme values in the range [16,27] with just 1.46% of the data. In figure 1 is the frequency distribution of Y-the number of encounter visits to the doctor. Clearly, the bulk of the data is between 1 and 9.

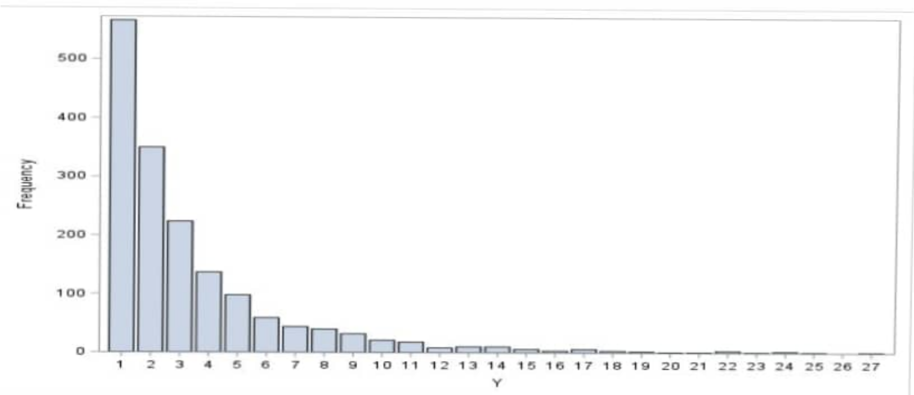

Figure 1: Frequency Distribution of Y-the number of Visits

We present in Table 5, the results of applying some zero-truncated two and three parameter models to the National Health Insurance Scheme (NHIS) data with the linear predictor being

$$\boldsymbol{x'\beta} = \beta_0 + \beta_1\text{sex} + \beta_2\text{age} + \beta_3\text{fup} + \beta_4\text{eclass}$$

The ZTNB and ZTGP are modeled as $\mu_i = \exp(x_i^{'}\beta)$. Parameters q, θ and r in ZTNGDP, ZTNLD and ZTNB-ELD respectively are modeled in the logit form. That is, for example, for the ZTNLD, $\theta_i = 1/[1 + \exp(-x_i^{'}\beta)]$. The parameter λ in ZTIT and ZTDLPD are in the exponential form $\lambda_i = \exp(x_i^{'}\beta)$, while parameter b in the ZTQNBD is modeled as $1/\exp(x_i^{'}\beta)$.
Results in Table 5 indicate that among the two-parameter models employed here, the most parsimonious model is the zero-truncated generalized Poisson with a p-value of 0.6878 and also gives the lowest AIC and BIC GOFs. The

Table 5: Parameter estimates and GOF statistics for ZT models applied to the NHIS Data

Effect	ZTNB	ZTGP	ZTNGDP	ZTNLD	ZTQNBD	ZTIT	ZTDLPD	ZTNB-ELD
Intercept	0.3267	0.6085	1.2276	1.0617	1.0451	0.0092	-0.5616	15.8794
sex	-0.0144	-0.0122	-0.0205	-0.0169	-0.0112	0.0270	0.0465	-15.4953
age	0.0043	0.0038	0.0050	0.0043	0.0040	0.0076	0.0045	1.0154
fup	0.1046	0.0945	0.1369	0.1135	0.0901	0.0607	0.0919	-5.0229
ecs	0.3323	0.3065	0.4325	0.3544	0.2919	0.2509	0.2405	9.7604
MLE	$\hat{r}$=0.3952	$\hat{\tau}$=0.5625	$\hat{\alpha}$=0.5527	$\hat{\alpha}$=0.8937	$\hat{\alpha}$=0.5211	$\hat{p}$=0.5694	$\hat{\alpha}$=0.4111	$\hat{c}$=7.6485
					$\hat{c}$=0.0107	$\hat{r}$=0.1438	$\hat{\beta}$=4.5160	$\hat{k}$=8.2352
-2LL	6709.3	6705.8	6707.5	6709.6	6708.1	6707.1	6710.5	6713.5
AIC	6721.3	6717.8	6719.5	6721.6	6722.1	6721.1	6724.5	6727.5
X^2	1699.83	1612.4480	1662.3487	1719.6386	1616.6445	1598.9972	1631.3057	1596.7487
d.f.	1641	1641	1641	1641	1640	1640	1640	1640
p-value	0.1523	0.6878	0.3510	0.0865	0.6547	0.7613	0.5558	0.7735

three-parameter models considered all fit the data very well with the zero-truncated NB-Erlang distribution being the most parsimonious with a p-value of 0.7735 and is closely followed by the ZIIT. Over all, the three-parameter models provide a better fit to the data than the two-parameter models. The ZTPEGD performs poorly here with AIC and BIC being 6724.2 and 6756.7 respectively with a Wald's GOF of 1710.0594 on 1641 d.f.

4. Conclusion

The negative binomial and its mixture models are unsuitable for modeling under-dispersed count data because they usually have convergence problems. However, we have demonstrated that alternative two and three parameter distributions can be employed. Among the two-parameter distributions, the ZTPEGD (zero truncated Poisson -Exponential Gamma distribution) seems to fair better for frequency under-dispersed count data. However, the zero-truncated Delaporte (ZTDLPD) by far performs much better than all the others. In particular, it fits the intractable metaphitamine data in Table 4. All the models are implemented in SAS PROC NLMIXED and R using *Optim* as our optimizer. SAS does not have a routine for the Delaporte distribution, so this was programmed. The results from both SAS and R are very close, but SAS gives better results because of its array of optimization algorithms.

References

Abouammoh, A.M, Alshangiti, A.M. and Ragab IE (2015). A new Generalized Lindley Distribution. Journal of Statistical Computation and Simulation, 85(10),3662-3678.

Adesina, O.S., Agunbiade, D.A. and Oguntunde, P. E. (2021). Flexible Bayesian Dirichlet mixtures of generalized linear mixed models for count data. Scientific African, 13, 1–11.

Atikankul, Y. (2023). A New Generalized Lindley Regression Model. Austrian Journal of Statistics, 52, 39-50.

Awad, Y.S., Bar-Lev,S.K. and Makov, U. (2016). A new class counting distributions embedded in the Lee-Carter model for mortality projections. A Bayesian approach. Technical report, No 146. Actuarial Research Center, University of Haifa. Israel.

Bhati,D. and H.S. Bakouch (2019). A new infinitelydivisible discrete distribution with applications to count data modelling. Communications in Statistics-Theory and Methods, 48 (6), 140–16.

Bodhisuwan,W., Pudprommarat, C., Bodhisuwan, R. and Saothayanun, L. (2017). Zero-truncated Negative Binomial-Erlang Distribution. AIP conference Proceedings,

1905, 050011, American Institute of Physics.
Consul, P.C, and Famoye, F. (1992). Generalized Poisson Regression Model. Comm. Statist. Theory Methods, 21(1), 89–109.
Das, K.K., Ahmad, J, and Bhattachajee, S. (2018). A new three-parameter Poisson-Lindley distribution for modeling over dispersed count data. Internal Journal of Applied Engineering Research, 13(25), 16468–16477.
Delaporte, Pierre J. (1960). "Quelques problmes de statistiques mathmatiques poses par l'Assurance Automobile et le Bonus pour non sinistre" [Some problems of mathematical statistics as related to automobile insurance and no-claims bonus]. Bulletin Trimestriel de l'Institut des Actuaires Franais (in French). 227: 87102.
El-Morshedy,M, Eliwa,M.S., and Nagy, H. (2020). A new two-parameter exponentiated discrete Lindley distribution: Properties, estimation and applications. Journal of Applied Statistics, 47 (2): 354–75.
G´omez-D´eniz, Sarabia, J.M, and Calderin-Ojeda,E. (2011). A new discrete distribution with actuarial applications. Insurance: Mathematics and Economics, 48 (3), 406–12.
G´omez-D´eniz, E., Hern´andez-Bastida, A., and Fern´andez-S´anchez, M.P.A. (2016). A Suitable Discrete Distribution for Modelling Automobile Claim Frequencies. Bulletin of the Malaysian Mathe- matical Sciences Society, 39, 633–647.
Hassan, A. and Bilal, S. (2008). On some properties of Quasi-Binomial Distribution and Its Applications. Journal of Modern Applied Statistical Methods, 7(2), 616–63
Janardan, K. G. (1975). Markov-Polya urn-model with predetermined strategies. Gujarat Statist. Rev., 2(1),17–32.
Konggrod,S., Bodhisuwan,W. and Payakkapong, P. (2014). The Negative binomial-Erlang distribution with Applications. International Journal of Pure and Applied Mathematics, 92(3), 380–401.
Lawal, H.B. (1980). Tables of percentage points of Pearson's goodness-of-fit statistic for use with small expectations. Applied Statistics, 29, 292–298.
Lawal, H.B. (2018). Correcting for Non-Sum to 1 Estimated Probabilities in Applications of Discrete Probability Models to Count Data. International Journal of Statistics and Probability, 6, 119–131.
Lawal, H. B. (2019). Quasi-Negative Binomial, Inverse Trinomial and Negative Binomial Generalized Exponential Distributions and their Applications. Journal of Probability and Statistical Science- JPSS, 17(2), 205–222.
Li, S., Yang,F., Famoye,F., Lee, C., Black, D. (2011). Quasi-negative binomial distribution: Properties and applications. Computational Statistics and Data Analysis, 55(20), 2363-2371.
Mahmoudi, E. and Zakerzadeh,H. (2010). Generalized Poisson-Lindley Distribution. Communication in Statistics: Theory and Methods, 39(10), 1785–1798.
Phang, Y.N. and Loh,E.F. (2014). Statistical Analysis for overdispersed medical Count Data.International Scholarly and Scientific Research, 8(2), 292–294.
Pinheiro, J.C. and Bates,D.M.(1995). Approximations to the Log-Likelihood Function in the Non-linear Mixed-Effects Model Journal of Computational and Graomegacal Statistics, 4, 12–35.
Sankaran, M. (1970). The discrete Poisson-Lindley distribution, Biometrics, 26, 145–149.
Shanker, R., and Shukla, K. K.(2013).On a size-based quasi poisson-lindley distribution. International Journal of probability and Statistics, 2(21), 28–34.
Shanker, R., Shukla, K. K.(2020). A new three-parameter sized-biased Poisson-Lindley distribution with properties and applications. Biometrics and Biostatistics International Journal, 9(1), 1–14.
Shanker,R., and Mishra,A. (2017). On size-based two-parameter poisson-lindley distribution and its applications. American Journal of Mathematics and Statistics, 7(3), 99–107.
Shimizu, K. and Yanagimoto, T. (1991). The inverse trinomial Distribution. Japan Journal of Applied Statistics, 20, 89–96.
Shmueli, G., Minka, T., Borle, J., and Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. J. R. Stat. Soc., Series C, 54:127–142.
Shoukri, M. M. and Aleid, M.M. (2022). Quasi-Negative Binomial: Properties, Parametric

Estimation, Regression Model and Application to RNA-SEQ Data. Open Journal of Statistics, 12. 216–237. DOI:10.4236/ojs.2022.122016

Singh, S.N. and Yadav, R.c. (1971). trends in rural out-migration at household level. Rural Demography, 8(1), 53–61.

Umar, M. A. (2019). A zero-truncated Poisson-Exponential-Gamma Distribution and its Applications. M.Sc. Dissertation, University of Ilorin, Ilorin. Nigeria.

Umar, M.A., Jimoh,K., and Yahya, W.B. (2019). A note on the applications of some zero truncated distributions. Edited Proceedings of the 3rd International Conference of the PSSN, 3, 499–505.

Zakerzadeh, H. and Dolati, A. (2009). Generalized Lindley Distribution. Journal of Mathematical Extension, 3, 13–25.

Zhang, H., Tan, K. and Li, B. (2018). COM-negative binomial distribution:modeling overdispersion and utrahigh zero-inflated count data. Front. Math. China, 13(4), 967–998

Appendix I: ZTPEGD R output for the Stillbirth data in Table 1

```
> y <- c(rep(1,453), rep(2,139), rep(3,49), rep(4,12), rep(5,5))
>
> ztpegd=function(gam,y){
+   a=gam[1]
+   theta=gam[2]
+   u1=(theta+1)^a
+   z1=(theta^2)*u1*exp(lgamma(y+1))
+   z2=(theta^a)*(theta+1)*exp(lgamma(a+y))
+   z3=exp(lgamma(a))*(theta+1)*(u1-theta^a)+(theta*u1)
+   LL=log(z1+z2)-log(z3)-y*log(theta+1)-lgamma(y+1)
+   return(Re(-sum(LL)))
+ }
>
> fit= optim(fn=ztpegd, c(a=2.5, theta=4), method="BFGS", y=y, hessian = TRUE)

> fit
$value
[1] 586.8564

$counts
function gradient
      27       14

$convergence
[1] 0

$message
NULL

$hessian
              a     theta
a      39.37807 -25.29966
theta -25.29966  16.58869

>
> ##Nelder-Mead
> #est=fit$estimate
> est=fit$par
>
> #Parameter Standard Errors
> covm <- solve(fit$hessian)
> se1=sqrt(diag(covm))
> Z.value=est/se1
> pvalue=2*(1-pnorm(abs(Z.value)))
> ra=round(cbind(est, se1,Z.value, pvalue),digits=4)
> ra
         est    se1 Z.value pvalue
a     3.4113 1.1228  3.0381 0.0024
theta 4.2912 1.7300  2.4805 0.0131
>
> k=2
> AIC=2*(fit$value)+2*k
> n=length(y)
> BIC=2*(fit$value)+k*log(n)
> VIC=cbind(AIC,BIC)
> VIC
          AIC      BIC
[1,] 1177.713 1186.691
>
> ##Generate estimated probabilities, mean and variance##
> aa <- fit$par[1]
> theta <- fit$par[2]
> DF=658
>
> # Initialize result storage
> results <- data.frame(
+   i = integer(),
+   prob = numeric(),
+   cum = numeric(),
+   fit1 = numeric(),
+   ss1 = numeric(),
+   mean = numeric(),
+   v1 = numeric(),
+   var = numeric()
+ )
>
> # Initialize cumulative variables
> ss1 <- 0
> cum <- 0
> mean <- 0
> v1 <- 0
>
> # Loop through s from 1 to 5
> for (i in 1:20) {
+   u11=(theta+1)^aa
+   z11=(theta^2)*u11*exp(lgamma(i+1))
+   z22=(theta^aa)*(theta+1)*exp(lgamma(aa+i))
```

```
+   z31=exp(lgamma(aa))*(theta+1)*(u11-theta^aa)+(theta*u11)
+   LL1=log(z11+z22)-log(z31)-i*log(theta+1)-lgamma(i+1)
+   prob=exp(LL1)
+   fit1 <- DF * prob
+   ss1 <- ss1+fit1
+   cum <- cum + prob
+   mean <- mean+(prob * i)
+   v1 <- v1 + (i * i * prob)
+   var <- v1 - (mean * mean)
+
+   # Store results
+   results <- rbind(results, data.frame(
+     i = i,
+     prob = prob,
+     cum=cum,
+     fit1 = fit1,
+     ss1 = ss1,
+     mean = mean,
+     v1 = v1,
+     var = var
+   ))
+ }
>
> # Print final results
> round(results,4)
         i   prob    cum     fit1      ss1   mean     v1    var
         ======================================================
         1 0.6845 0.6845 450.4080 450.4080 0.6845 0.6845 0.2160
         2 0.2216 0.9061 145.8189 596.2269 1.1277 1.5709 0.2992
         3 0.0675 0.9736  44.4258 640.6527 1.3303 2.1786 0.4090
         4 0.0193 0.9930  12.7086 653.3613 1.4075 2.4876 0.5065
         5 0.0052 0.9982   3.4469 656.8082 1.4337 2.6186 0.5630 **
         6 0.0014 0.9995   0.8954 657.7036 1.4419 2.6676 0.5885
         7 0.0003 0.9999   0.2246 657.9282 1.4443 2.6843 0.5983
         8 0.0001 1.0000   0.0548 657.9830 1.4449 2.6896 0.6018
         9 0.0000 1.0000   0.0130 657.9961 1.4451 2.6912 0.6028
        10 0.0000 1.0000   0.0030 657.9991 1.4452 2.6917 0.6032
        11 0.0000 1.0000   0.0007 657.9998 1.4452 2.6918 0.6033
        12 0.0000 1.0000   0.0002 658.0000 1.4452 2.6919 0.6033
        13 0.0000 1.0000   0.0000 658.0000 1.4452 2.6919 0.6033
        14 0.0000 1.0000   0.0000 658.0000 1.4452 2.6919 0.6033
        15 0.0000 1.0000   0.0000 658.0000 1.4452 2.6919 0.6033
        16 0.0000 1.0000   0.0000 658.0000 1.4452 2.6919 0.6033 ***
        17 0.0000 1.0000   0.0000 658.0000 1.4452 2.6919 0.6033
        18 0.0000 1.0000   0.0000 658.0000 1.4452 2.6919 0.6033
        19 0.0000 1.0000   0.0000 658.0000 1.4452 2.6919 0.6033
        20 0.0000 1.0000   0.0000 658.0000 1.4452 2.6919 0.6033
>
> ##Compute mean and variance using the method of moments##
> suma=0
> sumb=0
> for (k in 1:100){
+   u21=(theta+1)^aa
+   z21=(theta^2)*u21*exp(lgamma(k+1))
+   z23=(theta^aa)*(theta+1)*exp(lgamma(aa+k))
+   z32=exp(lgamma(aa))*(theta+1)*(u11-theta^aa)+(theta*u21)
+   LL2=log(k)+log(z21+z23)-log(z32)-k*log(theta+1)-lgamma(k+1)
+   LL3=2 * log(k)+log(z21+z23)-log(z32)-k*log(theta+1)-lgamma(k+1)
+   suma=suma+exp(LL2)
+   sumb=sumb+exp(LL3)
+ }
> mu=suma
> sig=sumb-(mu*mu)
> xx=((y-mu)**2)/sig
> yq=cbind(mu,sig)
> head(yq,10)
            mu       sig
theta 1.445187 0.6033023
>
> X2=sum(((y-mu)^2)/sig)
> X2
[1] 650.6363
>
```

Appendix II: ZTDLPD R codes and output for the data in Table 4

```
> y <- c(rep(1,3114), rep(2,163), rep(3,23), rep(4,20), rep(5,9),
+        rep(6,3), rep(7,3), rep(8,3), rep(9,4), rep(10,3))
>
> library(Delaporte)
> dlpd=function(theta,y){
+   aa=theta[1]
+   beta=theta[2]
+   lambda=theta[3]
+ f0=exp(-lambda-aa*log(1+beta))
+   z=ddelap(y,aa,beta,lambda, log=FALSE)
+ LL=log(z)-log(1-f0)
```

```
+  return(Re(-sum(LL)))
+ }
> fit= optim(fn =dlpd, c(aa=0.003, beta=3.11, lambda=0.06), method="BFGS",y=y, hessian = TRUE)
> fit
$par
         aa        beta      lambda
0.003831181 3.054836649 0.071978555

$value
[1] 1103.886

$counts
function gradient
     104        9

$convergence
[1] 0

$message
NULL

$hessian
                aa       beta      lambda
aa     5500689.216 6368.21580 -97856.9534
beta      6368.216   11.74804   -225.0077
lambda  -97856.953 -225.00770  13261.4181

>
> #, method="BFGS"
> ##Nelder-Mead
> #est=fit$estimate
> est=fit$par
>
> #Parameter Standard Errors
> covm <- solve(fit$hessian)
> se1=sqrt(diag(covm))
> Z.value=est/se1
> pvalue=2*(1-pnorm(abs(Z.value)))
> ra=round(cbind(est, se1,Z.value, pvalue),digits=4)
> ra
          est    se1 Z.value pvalue
aa     0.0038 0.0007  5.3960      0
beta   3.0548 0.5511  5.5428      0
lambda 0.0720 0.0107  6.7014      0
>
>
> k=3                              ##Number of estimated parameters
> AIC_COMu = 2*(fit$value)+ 2*k
> AIC_COMu
[1] 2213.772
> n=length(y)
> BIC_COMu = 2*(fit$value)+ k*log(n)
> YIC=cbind(AIC_COMu,BIC_COMu)
> YIC
     AIC_COMu BIC_COMu
[1,] 2213.772 2232.118
>
>
>
> ##Generate estimated probabilities, mean and variance##
>
>
> aa <- fit$par[1]
> beta <- fit$par[2]
> lambda <- fit$par[3]
> DF <- 3345
> f00=exp(-lambda-aa*log(1+beta))
>
> # Initialize result storage
> results <- data.frame(
+   i = integer(),
+   prob = numeric(),
+   cum = numeric(),
+   fit1 = numeric(),
+   ss1 = numeric(),
+   mean = numeric(),
+   v1 = numeric(),
+   var = numeric()
+ )
>
> # Initialize cumulative variables
> ss1 <- 0
> cum <- 0
> mean <- 0
> v1 <- 0
>
> # Loop through s from 1 to 10
> for (i in 1:10) {
+ zz=ddelap(i, aa, beta, lambda, log=FALSE)
+ LL3=log(zz)-log(1-f00)
+ prob=exp(LL3)
```

```
+   fit1 <- DF * prob
+   ss1 <- ss1+fit1
+   cum <- cum + prob
+   mean <- mean+(prob * i)
+   v1 <- v1 + (i * i * prob)
+   var <- v1 - (mean * mean)
+
+   # Store results
+   results <- rbind(results, data.frame(
+     i = i,
+     prob = prob,
+     cum=cum,
+     fit1 = fit1,
+     ss1 = ss1,
+     mean = mean,
+     v1 = v1,
+     var = var
+   ))
+ }
>
> # Print final results
> round(results,4)
         i   prob    cum      fit1      ss1   mean     v1    var
         =======================================================
         1 0.9310 0.9310 3114.2745 3114.274 0.9310 0.9310 0.0642
         2 0.0484 0.9794  161.8032 3276.078 1.0278 1.1245 0.0682
         3 0.0087 0.9881   29.0112 3305.089 1.0538 1.2026 0.0921
         4 0.0044 0.9925   14.7418 3319.831 1.0714 1.2731 0.1252
         5 0.0026 0.9951    8.7911 3328.622 1.0846 1.3388 0.1625
         6 0.0016 0.9967    5.4956 3334.117 1.0944 1.3979 0.2002
         7 0.0011 0.9978    3.5395 3337.657 1.1018 1.4498 0.2358
         8 0.0007 0.9985    2.3291 3339.986 1.1074 1.4943 0.2680
         9 0.0005 0.9990    1.5578 3341.544 1.1116 1.5321 0.2965
        10 0.0003 0.9993    1.0553 3342.599 1.1147 1.5636 0.3210
> # Print final results
> #round(print(results),4)
>
>
> ##Compute mean and variance using the method of moments##
> suma=0
> sumb=0
>  for (k in 1:1000){
+ f01=exp(-lambda-aa*log(1+beta))
+   zf=ddelap(k,aa,beta,lambda, log=FALSE)
+ LL2=log(k)+log(zf)-log(1-f01)
+ LL3=2*log(k)+log(zf)-log(1-f01)
+ suma=suma+exp(LL2)
+ sumb=sumb+exp(LL3)
+ }
> mu=suma
> sig=sumb-(mu*mu)
> xx=((y-mu)**2)/sig
> yq=cbind(mu,sig)
> head(yq,10)
             mu       sig
lambda 1.124358 0.4346404
>
> X2=sum(((y-mu)^2)/sig)
> X2
[1] 3119.284
>
```