International Journal of Statistics and Probability; Vol. 7, No. 6; November 2018
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

GLM for Some Class of Com-Poisson Distributions with Applications

Bayo H. Lawal[1]

[1] Department of Statistics & Mathematical Sciences, Kwara State University, Nigeria.

Correspondence: Bayo H. Lawal, Department of Statistics & Mathematical Sciences, Kwara State University, Malete. Nigeria.

Received: July 2, 2018 Accepted: July 20, 2018 Online Published: August 17, 2018

doi:10.5539/ijsp.v7n6p1 URL: https://doi.org/10.5539/ijsp.v7n6p1

Abstract

In this paper, we present regression models (GLM) for the class of Conway-Maxwell-Poisson (Com-Poisson) distributions. This class of models include the Com-Poisson, the Com-Poisson negative binomial, the Generalized Com-Poisson and the Extended Com-Poisson distributions, all of which have been presented in various literatures within the last five years. While these distributions have been applied most especially to frequency count data exhibiting over or under dispersion, not much has been presented in the application of this class of models to data having several covariates (the exception being the Com-Poisson itself). Thus in this paper, we present the generalized linear model formulation for these distributions and compare our results with the baseline Com-Poisson and Poisson models. Two data sets are employed in this application. We further extended our discussion to the zero-inflated versions of these distributions and applying same to a well established data with having 64% zero observations. All the models are fitted using SAS PROC NLMIXED. In all cases, empirical means and variances are generated which leads to our ability to compute the Wald's goodness-of-fit test statistic for all the models employed in this paper.

Keywords: Com-Poisson, NDHS, under-dispersion, empirical mean, zero-inflated models

1. Introduction

Most often, discussions are based on fitting distributions for over-dispersed data. Not much is focussed on under-dispersed count data. The Com-Poisson distribution (Shmueli *et al.*) however can be used to model both under and over dispersed count data. Consequently, we have applied the Com-Poisson model to three data sets having covariates in this study. Thus, our goal here is to fit the generalized linear model versions to the class of Com-Poisson to these data sets. We note here that the Com-Poisson distribution and its generalized linear model (GLM) has been explored and found very flexible in handling count data. See results in Lord *et al.* (2008), Sellers *et al.* (2012), and Francis *et al.* (2012) amongst several others. However, not much has been extended to the various extensions of the Com-Poisson distribution presented in Imoto (2014), Chakraborty & Ong (2014) and Chakraborty & Imoto (2016). We present in the next sections brief discussions of this class of Com-Poisson distributions.

For the data employed in this study however, the first data set is extracted from the 2013 Nigerian Demographic Health Survey (NDHS). The data comprises a subset of 3980 observations where the response variable is the number of children alive and using as an offset, the total number of children given birth to by women respondents in the six south-west states of Nigeria. The data is under-dispersed. The second data set is the 2003 U.S. Medical Expenditure Panel Survey (MEPS) data set relating to the number of doctor visits (Y=docvis) in 2003 for a number of elderly patients as well as several other covariates relating to patients' characteristics. This data set is over-dispersed.

Our third data set is the example which examines how waste quotas (emps) and the strictness of policy implementation (strict) affect the frequency of waste spill accidents of plants (accident) in Australia. Our focus is on employing the zero-inflated versions of these distributions to this data set that has excess zeros. We present here a brief introductory discussions on the Com-Poisson distributions and its extensions. These are models that are subsequently applied to the three data sets described above.

2. The Com-Poisson Distribution

The Com-Poisson (here referred to as COMP) distribution, first introduced by Conway and Maxwell (1962) and which Shmueli *et al.* (2005) re-introduced has seen a lot of attention in recent times. The distribution is defined for a random variable Y as:

$$f(y;\nu,\lambda) = \frac{\lambda^y}{(y!)^\nu}\frac{1}{Z(\lambda,\nu)}, \quad y = 0,1,2,\cdots, \quad \lambda > 0,\ \nu \geq 0. \tag{1}$$

Where

$$Z(\lambda,\nu) = \sum_{j=0}^{\infty}\frac{\lambda^j}{(j!)^\nu}. \tag{2}$$

is the normalizing term and ν is the *dispersion parameter* such that if $(\nu > 1)$ we have under dispersion, and when $(\nu < 1)$, we have over dispersion. The distribution reduces to the Poisson, Geometric and Bernoulli as $\nu = 1$, $\nu = 0$ and $\nu \to \infty$ distributions respectively. The mean and variance of Y can be obtained respectively from:

$$\begin{aligned} E(Y) &= \frac{1}{Z(\lambda,\nu)}\sum_{j=0}^{\infty}\frac{j\,\lambda^j}{(j!)^\nu}, \text{ and} \\ \text{Var}(Y) &= \frac{1}{Z(\lambda,\nu)}\sum_{j=0}^{\infty}\frac{j^2\,\lambda^j}{(j!)^\nu} - E(Y)^2 \end{aligned} \tag{3}$$

However, the mean and variance of the Com-P distribution do not have closed form expressions, and, consequently, Shmueli *et al.* (2005) provided approximate mean and variance of the distribution as:

$$\begin{aligned} E(Y) &\approx \lambda^{1/\nu} - \frac{\nu-1}{2\nu}, \quad \text{for } \nu \leq 1 \text{ or } \lambda > 10 \\ \text{Var}(Y) &\approx \frac{1}{\nu}\lambda^{1/\nu} \end{aligned} \tag{4}$$

2.1 Class of COM-Poisson Distributions

Chakraborty & Iyamote (2016) introduced the distributions presented in the following sections, their properties and applications to both frequency distributed data,including zero-truncated case.

2.2 Com-Poisson NB-COMNB

The COM-Poisson Negative Binomial distribution Chakraborty & Ong, (2014) has the pdf in (5) with parameters (ν, p, α):

$$f(y;\nu,p,\alpha) = \frac{(\nu)_y\ p^y}{(y!)^\alpha\ {}_1H_{\alpha-1}(\nu,1,p)} = \frac{\Gamma(\nu+y)}{\Gamma(\nu)\ {}_1H_{\alpha-1}(\nu,1,p)}.\frac{p^y}{(y!)^\alpha}; \quad y = 0,1,2,\ldots \tag{5}$$

where

$${}_1H(\nu,1,p) = \sum_{k=0}^{\infty}\frac{(\nu)_k\ p^k}{(k!)^\alpha} = \sum_{k=0}^{\infty}\frac{\Gamma(k+\nu)\,p^k}{\Gamma(\nu)(k!)^\alpha}$$

and the distribution is defined in the parameter space

$$\Theta_{COM-NB} = \{\nu > 0, p > 0, \alpha > 1\} \cup \{\nu > 0, 0 < p < 1, \alpha = 1\}$$

2.3 The Generalized Com-Poisson Distribution-GCOM

Imoto (2014) proposed the generalized Com-Poisson distribution-GCOM with parameters (ν, p, β) and has the pdf:

$$f(y;\nu,p,\beta) = \frac{[\Gamma(\nu+y)]^\beta}{C(\nu,p,\beta)}.\frac{p^y}{y!}; \quad \nu, p > 0; \beta < 1. \tag{6}$$

where

$$C(\nu,p,\beta) = \sum_{k=0}^{\infty}\frac{[\Gamma(\nu+k)]^\beta}{k!}\,p^k$$

With the distribution is defined in the parameter space

$$\Theta_{GCOMP} = \{\nu > 0, p > 0, \beta < 1\} \cup \{\nu > 0, 0 < p < 1, \beta = 1\}$$

2.4 The Extended COM-Poisson (ECOMP) Distribution

The pmf of a random variable Y having the extended COM-Poisson distribution with parameters $(\nu, \alpha, , \beta)$ is given by:

$$f(y; \nu, \alpha, \beta) = \frac{[(\nu)_y]^\beta}{{}_1S^{\beta}_{\alpha-1}(\nu, 1; p)} . \frac{p^y}{(y!)^\alpha} = \frac{[\Gamma(\nu + y)]^\beta}{[\Gamma(\nu)]^\beta \, {}_1S^{\beta}_{\alpha-1}(\nu, 1; p)} . \frac{p^y}{(y!)^\alpha} \tag{7}$$

where

$${}_1S^{\beta}_{\alpha-1}(\nu, 1; p) = \sum_{j=0}^{\infty} \frac{[\Gamma(\nu + j)]^\beta}{[\Gamma(\nu)]^\beta} . \frac{p^j}{(j!)^\alpha}$$

The distribution is defined in the parameter space

$$\Theta_{ECOMP} = \{\nu \geq 0, p > 0, \alpha > \beta\} \cup \{\nu > 0, 0 < p < 1, \alpha = \beta\}$$

Chakraborty & Imoto (2014) have discussed the properties of these distributions in detail, and we would thus not focus on these here.

3. Com-Poisson Type Regression Formulation

We present in this section the application of these distributions in regression situations where we have several covariates. This regression approach is often described as generalized linear Models (GLM regressions. In all cases, we would assume a log link between the parameter, λ, μ or p and the linear predictor $(\mathbf{x}'\boldsymbol{\beta})$. For the COMP (basic) model, is modeled in two ways as:

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_z x_{iz} \tag{8a}$$

$$\log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_z x_{iz} \tag{8b}$$

The formulation in (8a) is designated here as model COMP and models λ and the linear predictor $\mathbf{x}'\boldsymbol{\beta}$. The formulation in (8b) however is based on Guikema and Coffel (2008) alternative parameterization of the COM-Poisson regression model using $\mu = \lambda^{1/\nu}$, the approximate mean of the distribution. This approach leads to the expression in (1) now becoming:

$$f(y; \nu, \mu) = \left(\frac{\mu^y}{y!}\right)^\nu \frac{1}{S(\mu, \nu)}, \tag{9}$$

where, $S(\mu, \nu) = \sum_{j=0}^{\infty} \left(\frac{\mu^j}{j!}\right)^\nu$. This model will be designated here as COM_μ and has its mean and variance defined as: $E(Y) = \mu + 1/2\nu - 1/2$ and $\text{Var}(Y) = \mu/\nu$. Models COMNB, GCOMP and ECOMP are each modeled with the following GLM formulation:

$$\log(p_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_z x_{iz} \tag{10}$$

3.1 Estimation

In all cases, MLE of the above models and those in other sections are 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 & Bates (1995) and the Newton-Raphson optimization algorithm in PROC NLMIXED (**NEWRAP**) are employed.

3.2 Application-Example I

The data for this example comes from the 2013 Nigeria Demographic Health Survey (NDHS). This is a nationwide survey covering the six geographical zones of the country. For our analysis here, we have selected those for zone 6-South West Nigeria comprising of six states: Ekiti, Ondo, Ogun, Osun, Oyo and Lagos. The data concerns the response of women (3980 respondents) relating to current number of living children as the response variables. Other variables chosen for this data are age of respondent (age), age at first child birth (age1), total number of children born (tot), and any previous still birth or miscarriage (1 for yes, 0 for No). Other explanatory variables not used here (preliminary analysis indicate they are not significant) are religion (1,0), Education(1,0), dwelling (urban, rural) and wealth, a categorical variable.

The covariates are:

- age in years
- age1-age at first child birth
- term- any previous still birth or miscarriage (1 for yes, 0 if no)
- lch=number of living children (response variable)
- tch-total number of children given birth to.
- edu-any formal education(1,0)
- rsd-residence (urban=1, rural=0)
- rlg-religion (1-christanity, 0-for others)

We present the first and last five observations for this data set ($n = 3980$).

Obs	age	edu	edu1	rsd	rlg	age1	term	tch	lch	wlt	lof
1	49	2	1	1	1	12	0	9	2	4	2.19722
2	46	1	1	1	1	12	1	7	5	4	1.94591
3	42	1	1	1	1	12	0	4	3	4	1.38629
4	26	3	1	1	1	12	0	2	2	5	0.69315
5	48	2	1	1	1	12	1	6	4	5	1.79176
3976	39	0	0	0	1	37	1	1	1	5	0.00000
3977	43	2	1	1	0	38	0	2	2	5	0.69315
3978	42	1	1	0	1	39	0	1	0	5	0.00000
3979	45	3	1	1	1	39	0	1	1	5	0.00000
3980	49	3	1	1	1	40	0	2	2	5	0.69315

We would employ here the Poisson (P), COMP, COM$_\mu$, COMBNB, GCOMP and ECOMP models to data having covariates. Models COMP and COM$_\mu$ are modeled respectively in (11a) and (11b) respectively as:

$$\log(\lambda_i) = \beta_0 + \beta_1 \text{age} + \beta_2 \text{age1} + \beta_3 \text{term} + \text{lof} \tag{11a}$$

$$\log(\mu_i) = \beta_0 + \beta_1 \text{age} + \beta_2 \text{age1} + \beta_3 \text{term} + \text{lof} \tag{11b}$$

and for the COMBNB, GCOMP and ECOMP models, we have the formulation:

$$\log(p_i) = \beta_0 + \beta_1 \text{age} + \beta_2 \text{age1} + \beta_3 \text{term} + \text{lof} \tag{12}$$

We note here that in all the proposed models above, lof is employed as an *offset*, that is, log(*tch*). Further, our preliminary analysis indicates that none of the other covariates is significant for inclusion in our proposed model.

3.2 Results

The results of applying these model to the NDHS data are presented in Table 1. These results demonstrate that any Comp-Poisson model and its other extensions clearly performs better than the underlying Poisson model. The Poisson model gives a Wald's test statistic. $X^2_W = \left[\sum_{i=1}^{N} \frac{(y_i - \hat{m}_i)^2}{\hat{\sigma}^2_i}\right]$ of 429.9002 on 3976 d.f., giving a dispersion parameter of 0.1081 << 1 clearly indicating under-dispersion in the data. We note that for the Poisson, $\hat{m}_i = \hat{\sigma}^2_i$. The negative-binomial (NB) model will not be suitable in this case as it often leads to non-convergence for under-dispersed data. Under the circumstance therefore, the Com-Poisson (two parameterization approaches) and its extended forms (COMPNB, GCOMP and ECOMP) models discussed earlier are implemented on this data set.

Based on the -2 log-likelihood (-2LL) and the Akaike Information Criteria (AIC), the most parsimonious model would be the Guikema and Coffel (2008) parameterized model COM$_\mu$ with the lowest -2LL and consequently lowest AIC of 7841.6. However, this model produces a very high Wald's goodness-of-fit test statistic of 3568.7497 on 3975 d.f. This GOF value is very much on the high side when compared with those of the other Com-Poisson based models in Table 1.

Table 1. Estimated ML estimates and standard errors in Parentheses for all the models

Parameter	P	COM	COM_μ	COMNB	GCOMP	ECOMP
Intercept	-0.1342	1.5617	0.0748	1.0507	0.5968	1.2408
	(0.0595)	(0.1040)	(0.0191)	(0.3278)	(0.1431)	(0.1127)
age	-0.0033	0.0575	-0.0087	0.0567	0.0538	0.0539
	(0.0012)	(0.0026)	(0.0004)	(0.0026)	(0.0027)	(0.0026)
age1	0.0071	-0.0616	0.0127	-0.0604	-0.0564	-0.0569
	(0.0021)	(0.0039)	(0.0007)	(0.0039)	(0.0040)	(0.0039)
term	-0.0504	-0.2190	-0.0343	-0.2158	-0.2048	-0.2061
	(0.0252)	(0.0400)	(0.0082)	(0.0398)	(0.0394)	(0.0394)
		$\hat{\nu} = 2.7500*$	$\hat{\nu} = 7.9302$	$\hat{\nu} = 1.8535*$	$\hat{\nu} = 0.2733*$	$\hat{\nu} \approx 0.0000$
		(0.0537)	(0.1865)	(0.6457)	(0.0772)	(0.0000)
				$\hat{\alpha} = 3.5294*$	-	$\hat{\alpha} = 2.4777*$
				(0.1351)	-	(0.0700)
					$\hat{\beta} = -1.2517*$	$\hat{\beta} = -0.0453$
					(0.0783)	(0.0247)
-2LL	12014.14	10055	7831.6	10050	10033	10,002
AIC	12022.14	10065	7841.6	10062	10045	10,016
X^2	429.9002	2064.0394	3568.7497	2050.2128	2016.5757	2027.1024
d.f.	3976	3975	3975	3974	3974	3973

The Wald's test statistics computed for the other models are such that $\hat{m}_i$ and $\hat{\sigma}_i^2$ are computed using similar expressions for the means and variances in (3) for the COMP model. Our approach here, for these computations agree with those employed in SAS PROCs GLIMMIX, GENMOD and HPFMM. To further test the validity of our approach in computing these moments from expressions in (3), we have similarly, for the Com-Poisson distribution, computed the approximate means and variances given by expressions in (4) respectively. Results in Table 2 gives the results of this comparison. Only results for the first five and last five observations are presented.

Table 2. Computations of means and variances and Wald test Statistics

Obs#	$\hat{\lambda}$	$\hat{\nu}$	suma	sumb	sumc	mean	var	m1	v1
1	343.458	2.75003	183246242.86	1472526066.24	12390169882.58	8.03578	3.04114	8.04030	3.03942
2	180.572	2.75003	1871424.24	11774976.99	78594404.80	6.29199	2.40801	6.29777	2.40578
3	102.043	2.75003	74431.74	375926.17	2044374.72	5.05062	1.95771	5.05783	1.95490
4	20.322	2.75003	178.54	474.52	1456.24	2.65773	1.09269	2.67147	1.08714
5	173.651	2.75003	1466121.99	9087814.76	59811888.05	6.19854	2.37409	6.20441	2.37183
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3976	3.6985	2.75003	7.130	8.992	15.65	1.26114	0.60383	1.29081	0.58508
3977	10.8989	2.75003	41.514	85.011	210.36	2.04775	0.87379	2.06538	0.86674
3978	4.8375	2.75003	10.229	14.632	27.67	1.43046	0.65879	1.45581	0.64508
3979	5.7489	2.75003	13.226	20.467	40.91	1.54752	0.69817	1.57072	0.68687
3980	13.6086	2.75003	66.876	150.450	401.73	2.24968	0.94607	2.26581	0.93963

In Table 2, the columns labeled **mean** and **var** are based on computations from expressions in (3), while those labeled **m1** and **v1** refer to computations based on approximate results in Shmueli *et al.* (2005) presented in expressions in (4). The means and variances are very similar and this is true for all the 3980 observations in the data. The columns labeled **suma, sumb, sumc** are computed as follows:

$$\text{suma} = Z(\hat{\lambda}, \hat{\nu}) = \sum_{j=0}^{\infty} \frac{\hat{\lambda}^j}{(j!)^{\hat{\nu}}}; \quad \text{sumb} = \sum_{j=0}^{\infty} \frac{j\,\hat{\lambda}^j}{(j!)^{\hat{\nu}}}, \quad \text{sumc} = \sum_{j=0}^{\infty} \frac{j^2\,\hat{\lambda}^j}{(j!)^{\hat{\nu}}}$$

In actuality, the summation converges in the region $200 \leq j \leq 1000$. Thus, the mean is obtained as (sumb/suma) and the variance similarly as (sumc/suma)-(mean*mean). Similarly, the approximate means and variances (m1 & v1) are obtained using expressions in (4). The corresponding Wald test statistic under the latter is 2099.3164 for the Com-Poisson

(λ parameterization based) model. We observe here that the means and variances obtained from either expressions in (3) or (4) are very close, indicating that the procedure employed in the former (which is adopted here across all models) is validated.

Although the COM$_\mu$ has the lowest -2LL and AIC fit statistics, its Wald's GOF is unusually high. This is due to fact that it generally underestimates the expected variances. For Instance, we present in Table 3, the estimated mean and variances under this model using expressions in (3). We see that the procedure gives estimated means that are very close to the estimated mean $\hat{\mu}$ in column 2, however, the variances are all very small compared with those in Table 2.

Table 3. Mean and variances Computations under the COM$_\mu$ Model

Obs	$\hat{\mu}$	$\hat{\nu}$	suma	sumb	sumc	mean	var
1	7.36402	7.93020	1.4496702355E19	1.0033769958E20	7.0795108996E20	6.92142	0.92933
2	5.68133	7.93020	57833332304401	302875077631962	1.6276534003E15	5.23703	0.71735
3	3.47926	7.93020	8553145.31	25929805.92	82365614.60	3.03161	0.43920
4	2.00057	7.93020	500.38	764.24	1313.90	1.52731	0.29314
5	4.78537	7.93020	86996774021.64	377540720887.28	1690983678904.2	4.33971	0.60424
6	4.03625	7.93020	418131530.14	1500441091.15	5597799155.20	3.58844	0.51073
7	4.46736	7.93020	8909552579.89	35826620720.88	149095908016.80	4.02115	0.56476
8	5.04531	7.93020	566997108425.79	2608197501472.0	12359065886056	4.60002	0.63723
9	4.58309	7.93020	20377683836.28	84305118811.50	360579318156.39	4.13713	0.57897
10	6.77855	7.93020	1.8673706994E17	1.1830632949E18	7.6550030331E18	6.33545	0.85556

Because of the inconsistency of the COM$_\mu$ model, we see from Table 1, that of all other models, the GCOMP gives the most parsimonious model with Wald's $X^2 = 2016.5757$ on 3974 d.f. The corresponding -2LL and AIC are 10,033 and 10,045 respectively. The ECOMP model has serious convergence problem and the estimated parameter ν under this model is very small.

For all these models, the effects of age, age at first birth (age1) and previous miscarriage or still birth (term) are all significant. Thus, for ten years increase in age (keeping all other variables constant), the expected number of children alive is $\exp(10 \times 0.0538) = 1.71$,or an average of 1 additional child surviving.

4. GLM with Variable Dispersion Parameters

For all the models applied above to the NDHS data, we have assumed that the dispersion parameters (e.g. ν in the COMP model) are constant and do not depend on the explanatory variables that may or may not necessarily belong to the list of covariates in the main model as for example in model (10). Many often times, the dispersion parameters themselves can be function of these regressions which makes the assumption of constant dispersion parameter untenable. Consequently, in this section, we will model the dispersion parameters ν in the following form:

$$\log \nu = a_0 + a_1 \text{age} + a_2 \text{age1} + a_3 \text{term} \tag{13}$$

where $\{a_0, a_1, a_2, a_3\}$ are additional parameters to be estimated from the data. Further, all or some of these additional parameters may or may not be significant at say, $\alpha = 0.05$. The results of these applications to our models are presented in Table 4.

Table 4. Estimated ML estimates and standard errors in Parentheses for all the models Under variable dispersion Parameters

Parameter	COM	COM_μ	COMNB	GCOMP
Intercept	0.2407	0.0746	1.3346	1.2520
	(0.2641)	(0.0176)	(0.3679)	(0.1667)
age	0.1162	-0.0086	0.0410	0.0457
	(0.0077)	(0.0004)	(0.0065)	(0.0029)
age1	-0.0828	0.0119	-0.0739	-0.0596
	(0.0113)	(0.0007)	(0.0115)	(0.0045)
term	-0.2317	-0.0236	-0.1529	-0.2774
	(0.1213)	(0.0087)	(0.1746)	(0.0484)
Dispersion Parameters				
a_0	0.6364	2.5383	-0.2074	1.0321
	(0.0706)	(0.1496)	(0.5639)	(0.3630)
age	0.0134*	-0.0519*	0.0290*	-0.1781*
	(0.0015)	(0.0031)	(0.0080)	(0.0201)
age1	-0.0022	0.0689*	0.0207	0.1392*
	(0.0025)	(0.0055)	(0.0178)	(0.0243)
term	0.0067	-0.2709*	-0.1225	-0.1415
	0.0288)	(0.0665)	(0.3049)	(0.1985)
			$\hat{\alpha} = 3.3383*$	$\hat{\beta} = -1.4860*$
			(0.1477)	(0.0595)
-2LL	9,995.9	7,391.3	10,036	9,936.1
AIC	10012	7407.3	10,054	9,954.1
X^2	2137.7439	3663.8494	2042.3323	2096.1014
d.f.	3971	3971	3970	3970

We note that we did not implement a variable dispersion parameter for the ECOMP model because results from Table 1 indicate that the dispersion parameter is not significant for this model, its estimate being $\hat{\nu} \approx 0.00$. However, for all the other models, Table 4 gives the results of this application. Clearly, not all covariates are necessarily significant in the dispersion formulation. For instance for the COMP model, only the explanatory variable age is significant while for the COM_μ model, all the explanatory variables are significant. All the significant covariates are asterisk-ed in Table 4.

These are clearly contrasting results. Ideally therefore,it would make sense to remove all non-significant covariates in the dispersion GLM formulation. Thus for the COMP model for instance, the dispersion model should be $\log(\nu) = a_0 + a_1\text{age}$. Again here the COM_μ model gives the lowest -2LL and AIC but much higher Wald's GOF test value. The GCOMP model provides a much better fit with the dispersion parameter modeled as $\log(\nu) = a_0 + a_1\text{age} + a_2\text{age1}$ (a_3 *omitted*). The results for this reduced model when implemented is presented in figure 1.

The SAS System

The NLMIXED Procedure

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	95% Confidence Limits		Gradient
b0	1.2512	0.1667	3980	7.51	<.0001	0.9245	1.5780	-0.00005
b1	0.04570	0.002889	3980	15.82	<.0001	0.04003	0.05136	0.000093
b2	-0.05974	0.004523	3980	-13.21	<.0001	-0.06861	-0.05088	-0.00009
b3	-0.2596	0.04185	3980	-6.20	<.0001	-0.3417	-0.1776	6.141E-7
a0	1.0470	0.3616	3980	2.90	0.0038	0.3381	1.7559	-0.00015
a1	-0.1788	0.02018	3980	-8.86	<.0001	-0.2184	-0.1393	0.000493
a2	0.1386	0.02428	3980	5.71	<.0001	0.09099	0.1862	0.000448
beta	-1.4860	0.05945	3980	-25.00	<.0001	-1.6026	-1.3695	0.000013

Figure 1. Parameter estimates under GCOMP Model

Under this model, -2LL=9936.6, AIC=9952.6 and the Wald's GOF statistic is 2096.3987 on 3971 d.f. (pvalue=1.000), a very good fit.

4.1 Example II

Our second example data here is the U.S. Medical Expenditure Panel Survey (MEPS) data set relating to the number of doctor visits (Y=docvis) in 2003 for a number of elderly patients as well as several other covariates relating to patients' characteristics. The covariates are:

- private insurance coverage (supplemental to Medicare) (0,1)
- medicaid-eligibility for low income Medicaid coverage (0,1)
- female-gender of patients (1 if female, 0 if male)
- actlim-limitation of activity (0,1)
- totchr-number of chronic conditions
- phylim-physical limitation (0,1)
- educyr-number of years of educational attainment.

We present the first and last five observations for this data set (n = 3677).

Obs	docvis	female	phylim	private	medicaid	educyr	actlim	totchr
1	4	1	0	1	0	15	0	3
2	6	1	1	0	0	8	1	2
3	2	1	1	0	1	11	0	2
4	11	0	0	1	0	13	0	3
5	3	1	0	1	0	14	0	1
---	---	---	---	---	---	---	---	---
3671	5	1	1	1	0	16	0	1
3672	2	0	0	0	0	6	1	2
3673	15	1	1	0	1	12	1	3
3674	8	1	1	1	0	9	1	6
3675	6	1	0	1	0	13	0	2
3676	14	1	1	0	0	3	1	2
3677	10	0	1	0	0	4	1	1

We also created the interaction term fem_{edu} of **female** and **educyr**. The baseline Poisson regression model has Pearson's X^2 = 22930.3628 on 3668 d.f., giving a dispersion parameter of 6.2515, which clearly indicates very strong over-dispersion, considering the size of the data. In Table 5 are the estimated parameters, together with their standard errors for the distributions considered here.

Table 5. Estimated ML estimates and standard errors in Parentheses for all the models

Parameter	P	COM	COM_μ	COMNB	GCOMP	ECOMP
Intercept	0.7374	-0.1913	-2.6490	-0.2627	-0.2632	-0.6088
	(0.0361)	(0.1074)	(0.3791)	(0.0371)	(0.0371)	(0.2280)
female	0.2405	0.0494*	0.6838*	0.0499*	0.0499*	0.0520*
	(0.0427)	(0.0188)	(0.2625)	(0.0187)	(0.0187)	(0.0188)
phylim	0.1898	0.0383*	0.5302*	0.0387*	0.0387*	0.0403*
	(0.0165)	(0.0072)	(0.1034)	(0.0072)	(0.0072)	(0.0073)
private	0.1235	0.0230*	0.3179*	0.0228*	0.0228*	0.0233*
	(0.0144)	(0.0062)	(0.0870)	(0.0061)	(0.0061)	(0.0062)
medicaid	0.0777	0.0135	0.1873	0.0131	0.0130	0.0119
	(0.0190)	(0.0079)	(0.1100)	(0.0078)	(0.0078)	(0.00783)
educyr	0.0440	0.0082*	0.1139*	0.0081*	0.0081*	0.0082*
	(0.0027)	(0.0012)	(0.0174)	(0.0012)	(0.0012)	(0.0012)
actlim	0.0836	0.0138*	0.1912*	0.0135	0.0135	0.0136
	(0.0167)	(0.0070)	(0.0968)	(0.0069)	(0.0069)	(0.0070)
totchr	0.2408	0.0404*	0.5588*	0.0391*	0.0391*	0.0383*
	(0.0047)	(0.0022)	(0.0377)	(0.0023)	(0.0023)	(0.0023)
fem_{edu}	-0.0265	-0.0052*	-0.0717*	-0.0052*	-0.0052*	-0.0054*
	(0.0035)	(0.0015)	(0.0058)	(0.0015)	(0.0015)	(0.0015)
		$\hat{\nu} = 0.0722*$	$\hat{\nu} = 0.0722*$	$\hat{\nu} = 1.1661*$	$\hat{\nu} = 1.1763*$	$\hat{\nu} = 29.0697$
		(0.0058)	(0.0058)	(0.0810)	(0.0837)	(23.7274)
				$\hat{\alpha} = 1.0496*$	-	$\hat{\alpha} = 0.1149*$
				(0.0112)	-	(0.0185)
					$\hat{\beta} = 0.9506*$	$\hat{\beta} = 0.1405*$
					(0.0117)	(0.0397)
-2LL	29974.04	21269	21269	21264	21264	21254
AIC	29992.40	21289	21289	21286	21286	21278
X^2	22930.3628	4541.3103	4541.3111	4565.7195	4566.1066	4653.6250
d.f.	3667	3666	3666	3665	3665	3664

Results in Table 5 demonstrate again that that the Com-Poisson model and its various other extensions clearly perform better than the underlying Poisson model. The most parsimonious models are the COMP and COM_μ models. Although the extended Com-Poisson models have slightly lower AIC and BIC than the COMP models, however, COMP is based on fewer parameters and has the lowest GOF of 4541.3103 on 3666 d.f. We observe here the considerable reductions in the AIC and Wald's test statistic for the Com-Poisson models and its extensions. For all these models, the effect of medicaid is not significant. **actlim**-activity limitation is barely significant in the COMP models but not in the extended models. The interaction term between female and education years is also significant. Thus, for a ten year increase in education (keeping all other variables constant), the expected number of doctors' visits is $\exp(10 \times -0.0052) = \exp(-0.052) = 0.949$,or 5.1% reduction in men visits to doctors. The corresponding value for females would be $\exp\{10 \times (0.0494 - 0.0052)\} = 1.045$. Thus, females expected number of visits will increase by 4.5%. These are based on adopting the COMP model as the most parsimonious.

4.2 Variable Dispersion Parameter Models

For the corresponding variable dispersion models, we present below the summary fit statistics for the models.

	COMP	COM_μ	COMNB	GCOM	ECOMP
-2LL	21135	21219	20983	20977	20906
AIC	21171	21255	21021	21015	20942
X^2	4468.4723	4417.0485	3926.1479	3890.9365	3695.1955
d.f	3659	3659	3657	3657	3657

Although the ECOMP model gives the lowest -2LL and AIC and a most parsimonious value of Wald's X^2, it however most often suffers from convergence and under the circumstance, the GCOMP model may be preferred to the ECOMP model and a typical output under the GCOMP model is presented in figure 2. A test on whether the additional estimated parameters from the variable dispersion model is significant is provided by $-(2LL2 - 2LL1) = (221264 - 20977) = 287$

and is based on (3665-3657)=8 d.f which is highly significant. We note from figure 2 however, that not all the dispersion covariates are significant in the model.

The SAS System

The NLMIXED Procedure

Parameter Estimates								
Parameter	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**	**95% Confidence Limits**		**Gradient**
b0	-0.4694	0.04410	3677	-10.65	<.0001	-0.5559	-0.3830	0.000261
b1	-0.05276	0.03872	3677	-1.36	0.1731	-0.1287	0.02316	0.000077
b2	-0.01071	0.01371	3677	-0.78	0.4349	-0.03760	0.01618	0.000183
b3	-0.02045	0.01347	3677	-1.52	0.1290	-0.04685	0.005952	0.000049
b4	0.02616	0.01465	3677	1.79	0.0742	-0.00256	0.05488	0.000177
b5	0.001228	0.002153	3677	0.57	0.5684	-0.00299	0.005448	0.003778
b6	0.04066	0.01612	3677	2.52	0.0117	0.009068	0.07226	0.000229
b7	-0.01579	0.003902	3677	-4.05	<.0001	-0.02344	-0.00814	0.000490
b8	0.001953	0.003174	3677	0.62	0.5384	-0.00427	0.008177	0.001082
beta	1.0865	0.01236	3677	87.87	<.0001	1.0622	1.1107	0.001207
a0	-0.1452	0.1242	3677	-1.17	0.2424	-0.3887	0.09827	0.000035
a1	0.3972	0.1502	3677	2.64	0.0082	0.1027	0.6917	0.000017
a2	0.1812	0.05344	3677	3.39	0.0007	0.07643	0.2860	0.000011
a3	0.1741	0.05094	3677	3.42	0.0006	0.07428	0.2740	6.984E-6
a4	-0.06853	0.06147	3677	-1.11	0.2650	-0.1890	0.05199	0.000013
a5	0.02583	0.008769	3677	2.95	0.0032	0.008635	0.04302	0.000279
a6	-0.1156	0.06314	3677	-1.83	0.0671	-0.2394	0.008156	0.000023
a7	0.2368	0.01560	3677	15.18	<.0001	0.2063	0.2674	0.000053
a8	-0.02700	0.01227	3677	-2.20	0.0279	-0.05106	-0.00294	0.000139

Figure 2. Parameter estimates under GCOMP Variable Disp. Model

5. Zero-Inflated GLM Models

We present in this section the effect of applying the procedure employed in the last section to data exhibiting excess zeros (like the data in our next example where 64% of the data are zeros. We present here the results of fitting the ZICOMP, ZICOM_μ, ZINBCOMP and ZIGCOMP models are given in Table 6. To accomplish these, we recall that a zero-inflated (ZI) model is a two-part process manifested by the structural zeros part and the process that generates random counts and can be written in the form:

$$\Pr(Y = y|\phi) = \begin{cases} \phi + (1-\phi)\ \Pr(Y=0) & \text{if } y_i = 0 \\ (1-\phi)\ \Pr(Y = y_i)) & \text{if } y_i = 1, 2, \cdots \end{cases} \tag{14}$$

where ϕ is the extra proportion of zeros and Y is the count random variable with specified parameters. ϕ is modeled here in the logit form. Thus, the probability mass function for the ZICOMP, ZICOM_μ, ZINBCOMP and ZIGCOMP models are given respectively in expressions (15) to (19).

For the Com-Poisson, we have the probability density function:

$$P(\lambda, \beta; y_i) = \begin{cases} \phi + (1-\phi)\dfrac{1}{Z(\lambda_i, \nu)} & \text{if } y_i = 0 \\ (1-\phi)\dfrac{1}{Z(\lambda_i, \nu)}\dfrac{\lambda_i^{y_i}}{(y!)^\nu} & \text{if } y_i = 1, 2, \cdots \end{cases} \tag{15}$$

where

$$Z(\lambda_i, \nu) = \sum_{j=0}^{\infty} \frac{\lambda_i^k}{(k!)^\nu}$$

and the mean and variance of Y_i are respectively given as:

$$E(Y_i) = (1-\phi)\frac{1}{Z(\lambda_i, \nu)} \sum_{j=0}^{\infty} \frac{j\,\lambda^j}{(j!)^\nu} \tag{16a}$$

$$\text{Var}(Y_i) = (1-\phi)\frac{1}{Z(\lambda_i, \nu)} \sum_{j=0}^{\infty} \frac{j^2\,\lambda^j}{(j!)^\nu} - E(Y)^2 \tag{16b}$$

Similarly, for the ZICOM_μ, the probability mass function is given by:

$$P(\lambda, \beta; y_i) = \begin{cases} \phi + (1-\phi)\dfrac{1}{S(\mu_i, \nu)} & \text{if } y_i = 0 \\ (1-\phi)\dfrac{1}{S(\mu_i, \nu)}\left(\dfrac{\mu^y}{y!}\right)^\nu & \text{if } y_i = 1, 2, \cdots \end{cases} \tag{17}$$

where, $S(\mu, \nu) = \sum_{j=0}^{\infty} \left(\frac{\mu^j}{j!}\right)^{\nu}$.

The probability model for the ZINBCOMP is given by:

$$\Pr(Y_i = y) = \begin{cases} \phi + (1-\phi)H^{-1} & \text{if } y = 0 \\ (1-\phi) \quad \dfrac{\Gamma(\nu + y)}{\Gamma(\nu)\,{}_1H_{\alpha-1}(\nu, 1, p)} . \dfrac{p^y}{(y!)^{\alpha}}; & \text{if } y > 1 \end{cases} \tag{18}$$

Similarly, the probability mass function for the ZIGCOMP is given by:

$$\Pr(Y_i = y) = \begin{cases} \phi + (1-\phi)[\Gamma(\nu)]^{\beta}\, C^{-1} & \text{if } y = 0 \\ (1-\phi)\dfrac{[\Gamma(\nu + y)]^{\beta}}{C(\beta, \nu, p)} . \dfrac{p^y}{y!}; & \text{if } y > 1 \end{cases} \tag{19}$$

where H in (18) and C in (19) are defined respectively as:

$${}_1H(\nu, 1, p) = \sum_{k=0}^{\infty} \frac{\Gamma(k+\nu)\, p^k}{\Gamma(\nu)(k!)^{\alpha}} \quad \text{and} \quad C(\beta, \nu, p) = \sum_{k=0}^{\infty} \frac{[\Gamma(\nu + k)]^{\beta}}{k!}\, p^k$$

The pmf model for the ZIECOMP model can also easily be displayed. Consequently, from the above, it is not too difficult to formulate the corresponding log-likelihoods.

5.1 Example Data

We apply these zero-truncated models to the following data example which examines how waste quotas (emps) and the strictness of policy implementation (strict) affect the frequency of waste spill accidents of plants (accident). The data originally came from David Good of Indiana University.

We have reproduced the first and last five observations for these data which has a total of 778 observations. The variables are:

- Accident: number of waste spill accidents recorded in the plant
- Strict: strictness of policy implementation, where,

$$\text{Strict:} \begin{cases} 1 & \text{strict policy} \\ 0 & \text{lenient} \end{cases}$$

- emps: the size of the waste quotas

Table 6. Reported accident numbers with the two covariates

Subj	emps	Strict	accident
1	30	0	11
2	58	1	0
3	34	0	0
4	133	1	0
5	1	0	1
⋮	⋮	⋮	⋮
774	50	1	0
775	112	0	0
776	13	0	3
777	138	1	2
778	36	1	0

A frequency distribution of the data reveals that 498 or 64% of the data are zeros. Clearly, there is overabundance of zeros in this data. We would therefore expect that there are more zeros in these data than would normally be expected from a Poisson model.

We present the results of these analyses in Table 6.

Table 7. Parameter Estimates under zero-inflated Com-Poisson Models

Parameters	ZICOMP	ZICOM$_\mu$	ZINBCOMP	ZIGCOMP	ZIECOMP
Intercept	-0.2800	-520.86	-0.1021	-0.0988	-0.4142
	(0.0274)	(829.83)	(0.1185)	(0.1116)	(0.0740)
emps	0.0005	-0.8345	0.0001	0.0001	0.0001
	(0.0005)	(1.6522)	(0.0004)	(0.0004)	(0.0004)
strict	-0.0226	-42.3245	-0.0665	-0.0669	-0.0678
	(0.0415)	(102.50)	(0.0334)	(0.0334)	(0.0332)
ZInflated Parameters					
Intercept	0.0930	0.0935	0.0779	0.0768	0.0501
	(0.1793)	(0.1791)	(0.4011)	(0.4026)	(0.4229)
emps	-0.0220*	-0.0220*	-0.2318*	-0.2335*	-0.2533*
	(0.0071)	(0.0071)	(0.1027)	(0.1032)	(0.1072)
strict	1.5702*	1.5683*	8.3002*	8.3556*	8.9973*
	(0.3295)	(0.3290)	(3.2664)	(3.2819)	(3.3946)
$\hat{\nu}$	≈ 0.000	0.0005	0.3455	0.3300	$\hat{\nu} =\approx 0.0000$
	(0.0000)	(0.0008)	(0.0824)	(0.0961)	(0.0000)
			$\hat{\alpha} = 1.0294$	$\hat{\beta} = 0.9686$	$\hat{\alpha} = -0.0378$
			(0.0503)	(0.0485)	(0.0554)
					$\hat{\beta} = 0.0320$
					(0.0366)
-2LL	2194.0	2194.1	2178.7	2178.6	2173.0
AIC	2208.0	2208.1	2194.7	2194.6	2191.0
X^2	990.2061	988.5184	974.1406	977.4219	2667.7243

Results from Table 6 indicate that the ZICOM$_\mu$ gives parameter estimates whose standard errors are greater than the absolute values of the parameters. This observation is validated by the use of PROC COUNTREG in SAS using the *zeromodel* option. However, the model gives the same test statistics values as the lambda based ZICOMP model. The ZIECOMP model is sometimes intractable and very unattractive because of its convergence issues. Further, parameter estimates are not often conformed with the theoretical justification of the model such as $\hat{\alpha} > \hat{\beta}$. Of all these models, the ZINBCOMP (zero inflated negative binomial Com-Poisson) is the most parsimonious model with Wald's GOF value being 974.1406 on 770 d.f. For this model, both covariates are significant in the zero part of the model. The estimate of the dispersion parameter here is $\hat{\nu} = 0.3455$ which is significant. Of course we could stretch this model further by modeling with a variable dispersion parameter that incorporates the covariates.

6. Conclusions

We have demonstrated here that we can extend the generalized linear modeling approach to the Com-Poisson class of distributions. While the ECOMP model is sometimes difficult to fit in terms of convergence and obtaining initial parameter estimates, we have however, for the examples provided in this paper able to obtain convergence for this distribution. The Generalized Com-Poisson model (GCOMP) and its various forms works well for most data and readily converges. The Com-Poisson re-parameterization by Guikema & Coffelt (2008) produces very large values of Wald's GOF because it sometimes underestimates the true variances. The SAS programs for implementing these models are available from the author.

References

Chakraborty, S., & Imoto, T. (2016). *Extended Conway-maxwell-Poisson distribution and its properties and applications*. *Journal of Statistical Distributions and Applications, 3*(5), 1-199. https://doi.org/10.1186/s40488-016-0044-1

Chakraborty, S., & Ong, S. H. (2015). *A COM-type generalization of the negative binomial distribution*, Accepted in April 2014, (available on line since 07 November 2015) to appear in Communications in Statistics-Theory and Methods.

Conway, R. W., & Maxwell, W. L. (1961). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, *12*, 132-136.

Francis, R. A., Geedipally, S. R., Guikema, S. D., Dhavala, S. S., Lord, D., & Larocca, S. (2012). Characterizing the performance of the Conway-Maxwell-Poisson generalized linear model. *Risk Analysis*, *32*(1), 167-183.

Guikema, S. D., & Coffelt, J. P. (2008). A flexible count data regression model for risk analysis. *Risk Analysis, 28*(1), 213-223.

Imoto, T. (2014). A generalized Conway-Maxwell-Poisson distribution which includes the negative binomial distribution. *Appl. Math. Comput, 247*, 824-834.

Kadane, J., Shmueli, G., Minka, G., Borle, T., & Boatwright, P. (2006). Conjugate analysis of the Conway Maxwell Poisson distribution. *Bayesian Analysis, 1*, 363-374.

2003: *Medical Expenditure Panel Survey (MEPS) data*: US department of Health & Human Services.

Morris, D. S., Sellers, K. F., & Menger, A. (2017). *Fitting a Flexible Model for Longitudinal Count Data Using the NLMIXED Procedure*. SAS Paper (202).

National Population Commission (NPC) [Nigeria] and ICF Macro (2013). Nigeria Demographic and Health Survey 2013. Abuja, Nigeria: National Population Commission and ICF Macro.

Lawal, H. B. (2011). Fitting the Conway-Poisson model to Count data with SAS. *ICASTOR Journal of Mathematical Sciences, 5*, 229-238.

Lord, D., Guikema, S. D., & Geedipally, S. R. (2008). Application of the Conway-Maxwell-Poisson generalized linear model for analyzing motor vehicle crashes. *Accident Analysis and Prevention, 40*(3), 1123-1134.

Pinheiro, J. C., & Bates, D. M.(1995). Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics, 4*, 12-35.

Sellers, K. F., & Shmueli, G. (2010). A flexible regression model for count data. *The Annals of Applied Statistics*, *4*, 943-961.

Sellers, K. F., Borle, S., & Shmueli, G. (2012). The COM-Poisson model for count data: A survey of methods and applications. *Applied Stochastic Models in Business and Industry*, 104-116. https://doi.org/10.1002/asmb.918.

Shmueli, G., Minka, T., Borle, J., & 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.

Copyrights