Computational Statistics and Data Analysis 99 (2016) 68–80

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A flexible zero-inflated model to address data dispersion

Kimberly F. Sellers [a,b,*], Andrew Raim [b]

[a] Mathematics and Statistics Department, Georgetown University, Washington, DC, USA
[b] Center for Statistical Research & Methodology, U.S. Census Bureau, Washington, DC, USA [1]

HIGHLIGHTS

- Zero-inflated Conway–Maxwell–Poisson models dispersed datasets with excess zeroes.
- Hypothesis test detects statistically significant dispersion in light of excess zeroes.
- Data simulations and examples illustrate flexibility in model fit.

ARTICLE INFO

Article history:
Received 15 December 2014
Received in revised form 7 January 2016
Accepted 13 January 2016
Available online 22 January 2016

Keywords:
Conway–Maxwell–Poisson
Over-dispersion
Under-dispersion
Excess zeroes

ABSTRACT

Excess zeroes are often thought of as a cause of data over-dispersion (i.e. when the variance exceeds the mean); this claim is not entirely accurate. In actuality, excess zeroes reduce the mean of a dataset, thus inflating the dispersion index (i.e. the variance divided by the mean). While this results in an increased chance for data over-dispersion, the implication is not guaranteed. Thus, one should consider a flexible distribution that not only can account for excess zeroes, but can also address potential over- or under-dispersion. A zero-inflated Conway–Maxwell–Poisson (ZICMP) regression allows for modeling the relationship between explanatory and response variables, while capturing the effects due to excess zeroes and dispersion. This work derives the ZICMP model and illustrates its flexibility, extrapolates the corresponding likelihood ratio test for the presence of significant data dispersion, and highlights various statistical properties and model fit through several examples.

Published by Elsevier B.V.

1. Introduction

Numerous datasets contain excess zeroes, thus limiting their ability to be described via a standard distributional model. Accordingly, zero-inflated representations of these distributions have been developed to better describe such a random variable containing excess zeroes. In particular, the zero-inflated Poisson (ZIP) regression (Lambert, 1992; Hall, 2000) is a popular model to describe the relationship between a count response variable and explanatory variables of interest. The ZIP model has been used in a variety of applications, including manufacturing (Lambert, 1992), horticulture (Hall, 2000), zoology (Zipkin et al., 2014), and criminology (Famoye and Singh, 2006). Meanwhile, to address added over-dispersion that may exist in the data (even with accounting for the excess zeroes), the zero-inflated negative binomial (ZINB) model is often selected to address the matter, of which the zero-inflated geometric (ZIG) distribution (as discussed in Pandya et al., 2012, for example) is a special case. The ZIG regression is likewise considered as an alternative model in various applications

* Correspondence to: 306 St. Mary's Hall, Mathematics and Statistics Department, Georgetown University, 3800 Reservoir Road, Washington, DC 20057, USA. Tel.: +1 202 687 8829; fax: +1 202 687 6067.
E-mail address: kfs7@georgetown.edu (K.F. Sellers).

[1] This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

http://dx.doi.org/10.1016/j.csda.2016.01.007
0167-9473/Published by Elsevier B.V.

such as those noted above. These and other zero-inflated models are available for use in the Vector Generalized Linear and Additive Models (VGAM) package (Yee, 2014) available for use in *R* (R Core Team, 2014).

Over-dispersion is a common issue of many datasets. Excess zeroes are often thought of as a cause of data over-dispersion, however excess zeroes do not assure the existence of data over-dispersion. In actuality, excess zeroes reduce the mean of a dataset, thus inflating the dispersion index (variance divided by the mean); however, a severely under-dispersed dataset can still be under-dispersed even with the inclusion of excess zeroes in the data. Through broader examples, Sellers and Shmueli (2013) (in fact) illustrate that datasets with perceived forms of dispersion can actually stem from probability mixtures with different dispersion levels, including cases of over- or under-dispersion. One should therefore consider a flexible distribution that cannot only account for excess zeroes, but also address potential over- or under-dispersion in the distribution mixture. The Conway–Maxwell–Poisson (COM–Poisson) distribution of Conway and Maxwell (1962) is a flexible, two-parameter distribution for count data expressing over- or under-dispersion. Thus, a zero-inflated COM–Poisson (ZICMP) regression model would address the excess zeroes and provide flexibility in modeling data dispersion in a dataset.

This work derives the ZICMP model and illustrates its flexibility and statistical properties. To first further motivate its use, Section 2 provides more details about the COM–Poisson distribution. Section 3 develops the ZICMP regression model. Section 4 illustrates the flexibility of this model to capture data dispersion associated with various simulated zero-inflated structures. Section 5 further demonstrates its adaptability through real and simulated examples. Finally, Section 6 concludes the manuscript with discussion.

2. The Conway–Maxwell–Poisson (COM–Poisson) distribution

The Conway–Maxwell–Poisson (COM–Poisson) distribution of Conway and Maxwell (1962) has the probability mass function

$$P(Y = y) = \frac{\lambda^y}{(y!)^\nu Z(\lambda, \nu)} \quad y = 0, 1, 2, \ldots, \tag{1}$$

for a random variable Y, where $\lambda = E(Y^\nu)$ is the usual rate parameter under the Poisson model, $\nu \geq 0$ is a dispersion parameter, and $Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}$ normalizes the distribution. As the dispersion parameter, $\nu = 1$ denotes equi-dispersion via the Poisson distribution (i.e. $\nu = 1$ implies a Poisson(λ) distribution), while $\nu < 1$ and $\nu > 1$ respectively denote over- and under-dispersion. The COM–Poisson distribution not only captures the Poisson distribution as a special case, but also contains two other classical distributions, namely the geometric distribution with success probability $p_* = 1 - \lambda$ when $\nu = 0$ and $\lambda < 1$, and the Bernoulli distribution with success parameter $\pi = \frac{\lambda}{1+\lambda}$ when $\nu \to \infty$. Thus, this distribution not only captures three classical distributions, but further serves as a flexible bridge distribution for count distributions displaying over- or under-dispersion.

The COM–Poisson distribution has numerous statistical properties, including the ability to be represented as an exponential family (see Shmueli et al., 2005 and Sellers et al., 2011 for details), and moments of the form,

$$\mathrm{E}(Y^{r+1}) = \begin{cases} \lambda[\mathrm{E}(Y+1)]^{1-\nu} & r = 0 \\ \lambda \dfrac{\partial}{\partial \lambda} \mathrm{E}(Y^r) + \mathrm{E}(Y)\mathrm{E}(Y^r) & r > 0. \end{cases} \tag{2}$$

In particular, the expected value and variance are

$$\mathrm{E}(Y) = \lambda \frac{\partial \log Z(\lambda, \nu)}{\partial \lambda} \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu} \quad \text{and} \tag{3}$$

$$\mathrm{Var}(Y) = \frac{\partial \mathrm{E}(Y)}{\partial \log \lambda} \approx \frac{1}{\nu} \lambda^{1/\nu}, \tag{4}$$

where the approximations are especially good for $\nu \leq 1$ or $\lambda > 10^\nu$ (Shmueli et al., 2005). The associated moment generating function of Y is $\mathrm{M}_Y(t) = \frac{Z(\lambda e^t, \nu)}{Z(\lambda, \nu)}$, and its probability generating function is $E(t^Y) = \frac{Z(\lambda t, \nu)}{Z(\lambda, \nu)}$.

3. Zero-inflated COM–Poisson (ZICMP) regression

To formulate our ZICMP model, we consider a random sample, $Y_i,\ i = 1, 2, \ldots, n$, of the form,

$$Y_i \sim \begin{cases} 0 & \text{w.p. } p_i \\ \text{COM–Poisson}(\lambda_i, \nu_i) & \text{w.p. } 1 - p_i, \end{cases}$$

hence

$$P(Y_i = y_i) = \left[p_i + (1 - p_i) \left(\frac{1}{Z(\lambda_i, \nu_i)} \right) \right]^{u_i} \left[\frac{(1 - p_i)\lambda_i^{y_i}}{(y_i!)^{\nu_i} Z(\lambda_i, \nu_i)} \right]^{1-u_i} \tag{5}$$

$$= \left[\frac{p_i \left(Z(\lambda_i, \nu_i) - 1 \right) + 1}{Z(\lambda_i, \nu_i)} \right]^{u_i} \left[\frac{(1 - p_i)\lambda_i^{y_i}}{(y_i!)^{\nu_i} Z(\lambda_i, \nu_i)} \right]^{1-u_i}, \tag{6}$$

where $u_i = 1(0)$ if $y_i = (\neq)0$. The resulting likelihood function then takes the form,

$$\begin{aligned}
\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{p}; \mathbf{y}) &= \prod_{i=1}^{n} P(Y_i = y_i) \\
\log \mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\nu}, \boldsymbol{p}; \mathbf{y}) &= \sum_{i=1}^{n} \left\{ u_i \log \left(p_i + \frac{1-p_i}{Z(\lambda_i, \nu_i)} \right) \right. \\
&\quad \left. + (1-u_i)[\log(1-p_i) + y_i \log(\lambda_i) - \nu_i \log(y_i!) - \log Z(\lambda_i, \nu_i)] \right\} \quad \text{(from Eq. (5))} \qquad (7) \\
&= \sum_{i=1}^{n} \{ u_i \log (p_i Z(\lambda_i, \nu_i) + (1-p_i)) \\
&\quad + (1-u_i)[\log(1-p_i) + y_i \log(\lambda_i) - \nu_i \log(y_i!)] - \log Z(\lambda_i, \nu_i) \} \quad \text{(from Eq. (6))} \qquad (8) \\
&= \sum_{i=1}^{n} \{ u_i \log (p_i Z(\lambda_i, \nu_i) + (1-p_i)) + (1-u_i)[\log(1-p_i) + y_i \log(\lambda_i) - \nu_i \log(y_i!)] \\
&\quad - \log Z(\lambda_i, \nu_i) \}, \qquad (9)
\end{aligned}$$

where we model the parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)^T$, and $\boldsymbol{p} = (p_1, \ldots, p_n)^T$ via the canonical link generalized linear model ($\log(\boldsymbol{\lambda}) = \boldsymbol{X\beta}$, and $\text{logit}(\boldsymbol{p}) = \boldsymbol{W\zeta}$) and consider a constant dispersion parameter (i.e., $\nu_i \equiv \nu$ for all i). Alternatively, we can further model $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_n)^T$ via the loglinear link, $\log(\boldsymbol{\nu}) = \boldsymbol{G\gamma}$, to ensure non-negativity in the resulting value for ν. For such a formulation, we represent the log-likelihood as

$$\begin{aligned}
\log \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}; \boldsymbol{y}) &= \sum_{i=1}^{n} \{ u_i \log (p_i [Z(\exp(\boldsymbol{X}_i\boldsymbol{\beta}), \exp(\boldsymbol{G}_i\boldsymbol{\gamma})) - 1] + 1) \\
&\quad + (1-u_i)[-\log(1+\exp(\boldsymbol{Z}_i\boldsymbol{\zeta})) + y_i \boldsymbol{X}_i\boldsymbol{\beta} - (\exp \boldsymbol{G}_i\boldsymbol{\gamma}) \log(y_i!)] - \log Z(\exp(\boldsymbol{X}_i\boldsymbol{\beta}), \exp(\boldsymbol{G}_i\boldsymbol{\gamma}))\}.
\end{aligned}$$

For the special case where equi-dispersion holds (i.e. $\nu_i = 1\ \forall i$), ZICMP modeling reduces to ZIP (zero-inflated Poisson) modeling. We see this because, for the special case where equi-dispersion holds, $Z(\lambda_i, \nu_i = 1) = \exp(\lambda_i) = \exp(\exp(\boldsymbol{X}_i\boldsymbol{\beta}))$. Thus, Eq. (7) becomes

$$\begin{aligned}
\log \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}; \mathbf{y}) &= \sum_{i=1}^{n} \left\{ u_i \log \left[\frac{\exp(\mathbf{W}_i\boldsymbol{\zeta}) + \exp(-\exp(\mathbf{X}_i\boldsymbol{\beta}))}{1+\exp(\mathbf{W}_i\boldsymbol{\zeta})} \right] \right. \\
&\quad \left. + (1-u_i)\left[-\log(1+\exp(\mathbf{W}_i\boldsymbol{\zeta})) + y_i\mathbf{X}_i\boldsymbol{\beta} - \log(y_i!) - \exp(\mathbf{X}_i\boldsymbol{\beta})\right] \right\} \\
&= \sum_{i=1}^{n} \{ u_i \log [\exp(\mathbf{W}_i\boldsymbol{\zeta}) + \exp(-\exp(\mathbf{X}_i\boldsymbol{\beta}))] - \log(1+\exp(\mathbf{W}_i\boldsymbol{\zeta})) \\
&\quad + (1-u_i)\left[y_i\mathbf{X}_i\boldsymbol{\beta} - \log(y_i!) - \exp(\mathbf{X}_i\boldsymbol{\beta})\right] \},
\end{aligned}$$

which is the log-likelihood function associated with the ZIP model; see Lambert (1992) and Hall (2000). Analogously, for the special case where $\nu_i = 0$ and $\lambda_i < 1\ \forall i$, the ZICMP model reduces to ZIG (zero-inflated geometric) modeling.

3.1. Parameter estimation

We use the method of maximum likelihood for parameter estimation, i.e. maximizing the log-likelihood (Eq. (9)) with respect to $\boldsymbol{\beta}$, ν, and $\boldsymbol{\zeta}$; recall that we use the loglinear relation, $\log(\boldsymbol{\lambda}) = \boldsymbol{X\beta}$, and the logit model, $\text{logit}(\boldsymbol{p}) = \boldsymbol{W\zeta}$, to capture the probability of a zero-count. While the maximum likelihood estimators (MLEs) do not have a closed form, we can nonetheless determine their values with great accuracy via the bounded nonlinear minimization/optimization procedure in *R*, `nlminb`. We use `nlminb` to determine the MLEs because we wish to consider the relevant parameter space that constrains the constant $\nu \geq 0$; alternative unconstrained optimization functions in *R* (e.g. `nlm` and `optim`) can be used (for example) if we apply a loglinear model to $\boldsymbol{\nu}$. `nlminb` applies a Newton-type algorithm to determine the locale of the minimum of the objective function, therefore we let our objective function equal the negated log-likelihood function. Accordingly, the parameters that minimize the negated log-likelihood function, maximize the log-likelihood; hence, the resulting parameters equal the MLEs. `nlminb` also requires starting values; for ease, we use the ZIP estimates obtained from the `zeroinfl` function in the `pscl` package in *R* and the initial dispersion parameter set equal to 1 (which is consistent with the special case of the ZIP model). The resulting `nlminb` output includes the parameter (`par`) values, which are the MLEs; the objective function, which equals the negative of the log-likelihood value evaluated at the MLEs; and a convergence code and message where we can confirm successful algorithm convergence.

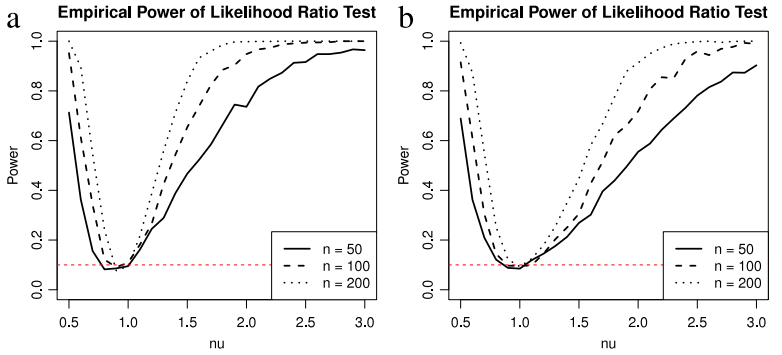


Fig. 1. Power functions for the likelihood ratio test procedure, $H_0 : \nu = 1$ versus $H_1 : \nu \neq 1$, where (a) $\lambda = 2.0$ and $p = 0.01$, and (b) $\lambda = 2.0$ and $p = 0.1$.

As demonstrated by the definition of λ as the νth moment of Y (in Section 2), λ and ν are correlated; more generally, this is true for λ, ν, p, by definition of Y as a zero-inflated COM–Poisson random variable. Given the definitions of λ and p via their respective link functions, we see that the corresponding Fisher Information matrix for $\{\hat{\boldsymbol{\beta}}, \hat{\nu}, \hat{\boldsymbol{\eta}}\}$ is hence a non-orthogonal, block symmetric matrix of the form provided in Appendix. Its components, however, can be computed via the form provided, and the corresponding standard errors associated with the parameter estimates can be obtained by isolating the diagonal components of the inverted resulting Fisher Information matrix; see Appendix.

3.2. Hypothesis testing

Does a statistically significant amount of data dispersion exist so that the ZICMP model is considered more appropriate than a ZIP model? We can address this question via a hypothesis test of the form, $H_0 : \nu = 1$ versus $H_1 : \nu \neq 1$. At this stage, we are not concerned with whether the data are over- or under-dispersed, which explains the consideration of a two-sided test. We can, however, infer knowledge of the dispersion type through the resulting ZICMP dispersion parameter estimate. Because both the `zeroinfl` and `nlminb` functions provide information to easily determine the log-likelihood value evaluated at the MLEs, we can establish a likelihood ratio test of the form, $\Lambda = \frac{L(\hat{\boldsymbol{\beta}}_0, \hat{\nu}_0 = 1, \hat{\boldsymbol{\zeta}}_0)}{L(\hat{\boldsymbol{\beta}}, \hat{\nu}, \hat{\boldsymbol{\zeta}})}$, where

$$-2\log\Lambda = 2\log L(\hat{\boldsymbol{\beta}}, \hat{\nu}, \hat{\boldsymbol{\zeta}}) - 2\log L(\hat{\boldsymbol{\beta}}_0, \hat{\nu}_0 = 1, \hat{\boldsymbol{\zeta}}_0) \sim \chi^2_1, \tag{10}$$

and $\log L(\hat{\boldsymbol{\beta}}, \hat{\nu}, \hat{\boldsymbol{\zeta}})$ and $\log L(\hat{\boldsymbol{\beta}}_0, \hat{\nu}_0 = 1, \hat{\boldsymbol{\zeta}}_0)$ are the respective log-likelihood values associated with the ZICMP and ZIP MLEs, respectively. The ZIP log-likelihood is obtained via the `loglik` component in the `zeroinfl` function, while the ZICMP log-likelihood is the negated value obtained from the `objective` component in the `nlminb` function. Thus, we can compute the test statistic from Eq. (10) and its corresponding p-value to determine statistical significance. A score-type test is likewise possible; under either scenario, the $\nu = 1$ null hypothesis considers the special case of the ZIP model. The likelihood ratio test approach is preferred, however, given the supplied likelihood output provided in `zeroinfl` and `nlminb`, respectively.

Fig. 1 illustrates the power associated with the likelihood ratio test of $H_0 : \nu = 1$ versus $H_1 : \nu \neq 1$ in a small simulation study. We draw random samples of size $n \in \{50, 100, 200\}$ from a ZICMP(λ, ν, p) distribution with $\lambda = 2.0, p \in \{0.01, 0.1\}$. For each of these settings, the empirical power of the test procedure,

$$\text{Reject } H_0 \text{ if } 2\log L(\hat{\lambda}, \hat{\nu}, \hat{p}) - 2\log L(\hat{\lambda}_0, \hat{\nu}_0 = 1, \hat{p}_0) \geq \chi^2_1(1-\alpha), \tag{11}$$

is computed, where $\chi^2_1(1-\alpha)$ denotes the $1-\alpha$ quantile of the χ^2_1 distribution. The test (11) is repeated on 1000 samples drawn from each distinct (λ, ν, p), and the empirical power at that point is taken to be the proportion of rejections. Fig. 1(a) plots the power function for $\lambda = 2.0$ and $p = 0.01$ while Fig. 1(b) provides the power function for $\lambda = 2.0$ and $p = 0.1$. For each case, a significance level of $\alpha = 0.1$ is assumed. As expected, the power of the test improves as the sample size n increases. The power also appears to improve for the smaller proportion of systematic zeroes, p. In both settings, the test appears to approximate the correct size.

4. Data simulations

4.1. Model flexibility

We consider various data simulations where we randomly generate 900 values from a count distribution, and add to it 100 zeroes to reflect simulated zero-inflation. Thus, each data representation contains 1000 values with at least 100 zeroes, reflecting some form of a zero-inflated distribution. The underlying simulated distributions considered here include a Poisson, geometric, Bernoulli, and COM–Poisson distribution reflecting over- or under-dispersion, respectively. Table 1

Table 1
True model parameters versus estimated parameters (and associated standard errors provided in parentheses) for various assumed distributions. For model comparisons, the log-likelihood, Akaike's Information Criterion (AIC), goodness of fit (GOF) measures (with conservative associated χ^2 degrees of freedom and p-values) are also provided.

Simulated distribution	ZIP	ZIG	ZINB	ZICMP
ZIG($p_* = 0.3$)	$\hat{\lambda}_*$: 3.2720 (2.4028) $\hat{p}$: 0.3493 (0.5067)	$\hat{p}_*$: 0.2940 (0.3123) $\hat{p}$: 0.1133 (0.7897)	$\hat{\lambda}$: 2.1602 (0.1975) $\hat{p}$: 0.0144 (0.0796) $\hat{\kappa}$: 1.3585 (0.2652)	$\hat{\lambda}$: 0.7060 (1.4072) $\hat{p}$: 0.1133 (1.1301) $\hat{\nu}$: 0.000 (1.1084)
$\log L$	−2215.4710	**−1950.6386**	**−1949.2299**	**−1950.6386**
AIC	4434.9420	**3905.2772**	**3904.4598**	**3907.2772**
GOF	370.6324, 6, <0.0001	**8.8583, 11, 4.6350**	**8.6300, 10, 4.5675**	**8.8583, 10, 4.5456**
ZIP($\lambda = 3$)	$\hat{\lambda}_*$: 3.0544 (1.9840) $\hat{p}$: 0.0859 (0.3631)	$\hat{p}_*$: 0.2637 (0.2637) $\hat{p}$: 0.0000 (0.6975)	$\hat{\lambda}$: 3.0442 (0.0657) $\hat{p}$: 0.0828 (0.0125) $\hat{\kappa}$: 0.0121 (0.0185)	$\hat{\lambda}$: 2.9299 (10.0355) $\hat{p}$: 0.0832 (0.4302) $\hat{\nu}$: 0.9702 (2.4111)
$\log L$	**−1990.3578**	−2187.6236	**−1990.1285**	**−1990.2819**
AIC	**3984.7156**	4379.2472	**3986.2571**	**3986.5639**
GOF	**8.8660, 6, 4.1813**	392.3392, 12, <0.0001	**8.8807, 6, 4.1804**	**8.9864, 6, 4.1743**
"ZIB($\pi = 0.7$)" = Bern($\pi = 0.63$)	$\hat{\lambda}_*$: 0.618 (1.4927) $\hat{p}$: 0.000 (2.0532)	$\hat{p}_*$: 0.6180 (0.6180) $\hat{p}$: 0.000 (2.0583)	$\hat{\lambda}$: 0.618 (0.0249) $\hat{p}$: 0.000 (0.0000) $\hat{\kappa}$: 0.000 (0.0000)	$\hat{\lambda}$: 1.6375 (NA) $\hat{p}$: 0.0046 (NA) $\hat{\nu}$: 33.3248 (NA)
$\log L$	−915.4229	−1075.9896	−915.4229	**−665.0347**
AIC	1834.8458	2155.9793	1836.8459	**1336.0695**
GOF	417.2234, 2, <0.0001	853.5954, 3, <0.0001	417.2426, 1, <0.0001	**0.0001, 1, 4.9998**
ZICMP($\lambda = 8, \nu = 3$)	$\hat{\lambda}_*$: 1.505 (1.6242) $\hat{p}$: 0.000 (0.7073)	$\hat{p}_*$: 0.3992 (0.3992) $\hat{p}$: 0.0000 (1.0516)	$\hat{\lambda}$: 1.5051 (0.0388) $\hat{p}$: 0.0000 (0.0002) $\hat{\kappa}$: 0.0000 (0.0003)	$\hat{\lambda}$: 6.5130 (46.8510) $\hat{p}$: 0.0895 (0.5853) $\hat{\nu}$: 2.7213 (7.7400)
$\log L$	−1410.1430	−1685.0800	−1410.1540	**−1347.6022**
AIC	2824.2861	3374.1599	2826.3079	**2701.2043**
GOF	104.0155, 4, <0.0001	624.6849, 8, <0.0001	104.0495, 3, <0.0001	**6.5697, 1, 4.0104**
ZICMP($\lambda = 2, \nu = 0.25$)	$\hat{\lambda}_*$: 18.0445 (4.4826) $\hat{p}$: 0.1020 (0.3026)	$\hat{p}_*$: 0.0543 (0.0593) $\hat{p}$: 0.0503 (0.3254)	$\hat{\lambda}$: 18.0408 (0.2793) $\hat{p}$: 0.1018 (0.0096) $\hat{\kappa}$: 0.1597 (0.0104)	$\hat{\lambda}$: 1.9869 (2.5119) $\hat{p}$: 0.1006 (0.3033) $\hat{\nu}$: 0.2454 (0.4278)
$\log L$	−4090.7594	−3796.6850	**−3462.9147**	**−3459.1094**
AIC	8185.5189	7597.3699	**6931.8294**	**6924.2187**
GOF	2902.2830, 21, <0.0001	655.8299, 45, <0.0001	**55.3601, 36, 4.0206**	**46.4302, 34, 4.0758**

serves to illustrate the ability of the zero-inflated COM–Poisson distribution to accurately model these classical zero-inflated models, while maintaining enough flexibility to further model other datasets demonstrating over- or under-dispersion. We compare the performance of the ZICMP with other common zero-inflated count models, including zero-inflated Poisson (ZIP), zero-inflated geometric (ZIG), and the zero-inflated negative binomial (ZINB). For ease of illustration, we wish to estimate the constant form of the underlying parameters for each model (i.e. $\hat{\lambda}_{*_i} = \hat{\lambda}_*$ for Poisson, $\hat{p}_{*_i} = \hat{p}_*$ for geometric, $\hat{\mu}_i = \hat{\mu}$ and $\hat{\kappa}_i = \hat{\kappa}$ for negative binomial, and $\hat{\lambda}_i = \hat{\lambda}$ and $\hat{\nu}_i = \hat{\nu}$ for COM–Poisson, and $\hat{p}_i = \hat{p}$ to reflect zero-inflation) in this exercise. By our construction of these simulated data, we expect to obtain a probability estimate, $\hat{p} \approx 0.10$, to reflect the amount of zero-inflation in the simulated data.

To compare model fits, we consider the log-likelihood, Akaike's Information Criterion (AIC), and the goodness-of-fit (GOF) statistic,

$$\text{GOF}(\theta) = \sum_{\ell=1}^{K} [O_\ell - E_\ell(\theta)]^2 / E_\ell(\theta),$$

based on K categories, $\text{I}_1, \ldots, \text{I}_K$, where category I_ℓ has observed count O_ℓ and expected count

$$E_\ell(\theta) = n \sum_{y=0}^{\infty} I(y \in \text{I}_\ell) g(y \mid \theta)$$

under a given model with density $g(\cdot \mid \theta)$. $\text{GOF}(\theta)$ can be used to test the null hypothesis that the data are a random sample from $g(\cdot \mid \theta)$. When θ is left unspecified in the test and estimated by maximizing the likelihood based on g, $\text{GOF}(\hat{\theta})$ follows a distribution between χ^2_{K-1-q} and χ^2_{K-1} under the null hypothesis (Sutradhar et al., 2008), where q is the number of unknown parameters to be estimated. A smaller p-value indicates evidence against g, and hence a less adequate fit. Table 1 provides the p-values stemming from the χ^2_{K-1-q} distribution as a conservative measure for inference regarding goodness of fit. We merged the possible counts $\{0, 1, 2, \ldots\}$ into K categories $\text{I}_1, \ldots, \text{I}_K$ so that each $E_\ell(\hat{\theta}) \geq 3$. An exception is in the "ZIB" row and "ZICMP' column of Table 1, where 0 and 1 contain virtually all mass of the distribution but at least $K = 5$ categories are needed to carry out the GOF test with $q = 3$ estimated parameters.

For all examples, we see that the zero-inflated COM–Poisson model performs comparably or better than the zero-inflated classical distributions. The added dispersion parameter provides the zero-inflated COM–Poisson model with the added flexibility to better capture data dispersion. In all cases, we are able to best capture the representation from the simulated dataset via the ZICMP model with the maximum likelihood estimates for λ and ν. In particular, for the special case simulations of the ZIG and the ZIP, we see that the results verify the theoretical points noted in Section 2, namely that the $\hat{\nu}$ estimates for these data are respectively $\hat{\nu} = 0.000$ and $\hat{\nu} = 0.9702 \approx 1$ for the ZIG and ZIP data. For the ZIG distribution example, we see that the maximum likelihood estimates obtained via the ZICMP model correspond well with the true values under the simulated ZIG model because we know that the geometric distribution is a special case of the COM–Poisson distribution where $\nu = 0$ and $p_* = 1 - \lambda$; further, we obtain $\hat{p} = 0.1133 \approx 0.10$, reasonably capturing the expected amount of zero-inflation that was simulated. Meanwhile, for the simulated zero-inflated Poisson example, we verify that the ZICMP distribution captures the ZIP distribution for parameters, $\nu = 1$ and λ. In this case, we see that $\hat{\lambda} = 2.9299$, $\hat{\nu} = 0.9702$ produce a mean rate approximately equal to $\hat{\lambda}^{1/\hat{\nu}} = 3.028254 \approx 3$, and $\hat{p} = 0.0832 \approx 0.10$.

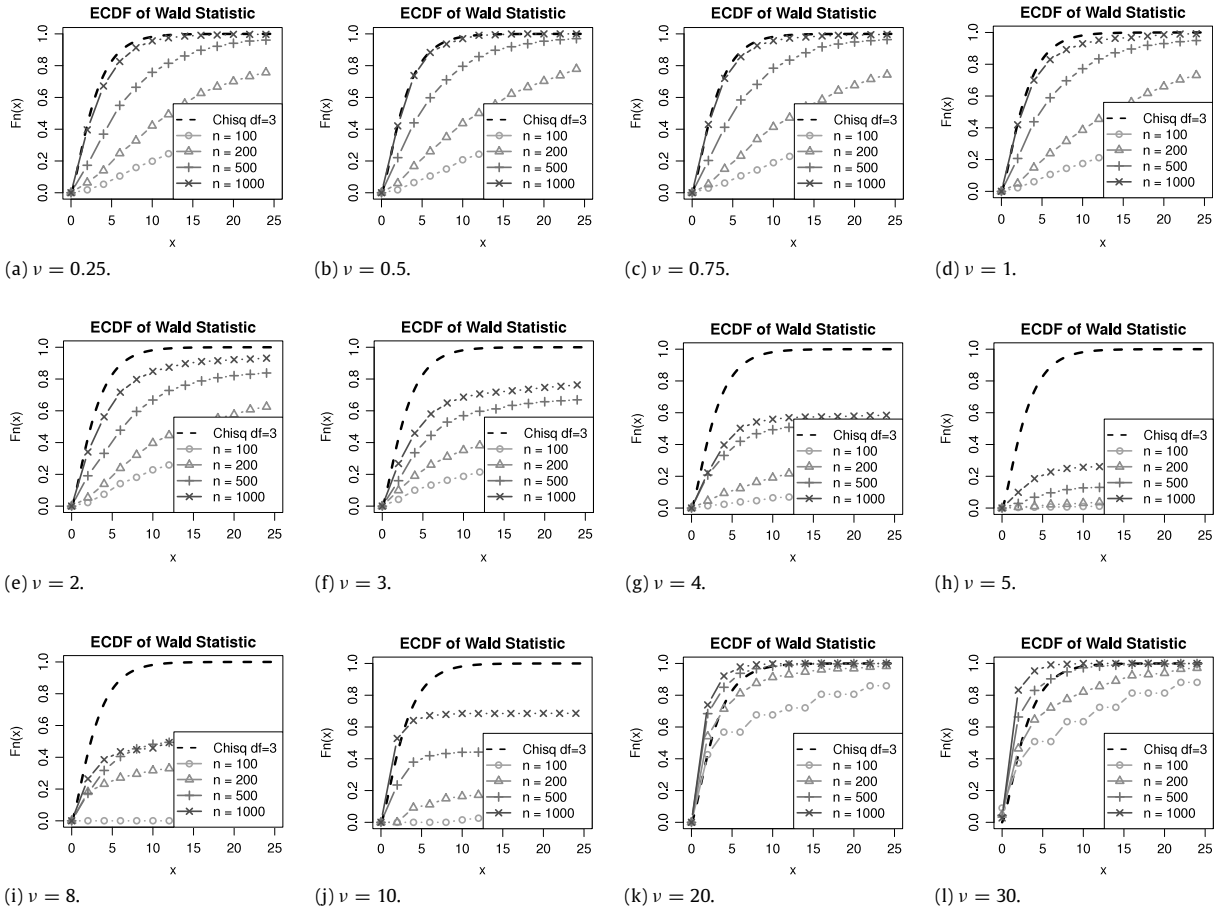
The third special case is most intriguing. By developing a ZICMP model, this implies that we are able to capture three special cases of zero-inflated classical distributions, namely the ZIG, ZIP, and a "zero-inflated Bernoulli" distribution. However, the Bernoulli distribution is itself a distribution that models counts of zero or one, thus what is reflected via a "zero-inflated Bernoulli" distribution is simply a Bernoulli random variable with a modified success probability. Recalling how the data are simulated for this example (ZI-Bernoulli($\pi_* = 0.7$)), we have simulated 900 values from a Bernoulli distribution with $\pi_* = 0.7$ and added 100 zeroes, for a total of 1000 data values in this example dataset. By definition of the success probability associated with this Bernoulli model, we expect to have produced a dataset with approximately $0.7 \times 900 = 630$ ones and $1000 - 630 = 370$ zeroes (i.e. a Bernoulli($\pi = 0.63$)). In actuality, the simulated dataset produced 618 ones and $1000 - 618 = 382$ zeroes, i.e. $\hat{\pi} = 0.618$. Meanwhile, maximum likelihood estimation applied via the ZICMP model produced the estimates, $\hat{\lambda} = 1.6375$, $\hat{\nu} = 33.3248$, and $\hat{p} = 0.0046$. The estimate for ν is consistent with what has been proven to occur in practice across various applications of the COM–Poisson model to various statistical methods. While theory states that the Bernoulli distribution is captured in the special case where $\nu \to \infty$, maximum likelihood estimates in practice produce $\hat{\nu} \geq 30$ (e.g., see examples in Sellers and Shmueli, 2010, Sellers, 2012, and Zhu et al., under review). Meanwhile, $\hat{p} = 0.0046$ implies that approximately 0.46% (i.e. approximately 5 out of 1000 values) of the data contain zeroes. Of the remaining 995 values, $\hat{\lambda} = 1.6375$ and $\hat{\nu} = 33.3248$ imply that COM–Poisson distribution recognizes these data as a Bernoulli model with $\hat{\pi}_* = \frac{1.6375}{1.6375+1} \approx 0.6209$, i.e. $0.6209 \times 995 \approx 618$ ones and $1000 - 618 = 382$ zeroes in the full dataset (precisely the number of zeroes and ones simulated under the Bernoulli model). Thus, these parameter estimates do, in fact, approximate well the simulated data in this example. The Bernoulli model was also considered for model comparison in the case of the "zero-inflated Bernoulli" simulation and produced a resulting estimate of $\hat{\pi} = 0.618$ (with associated standard error, 0.0154), $\log L = -665.0347$, and AIC $= 1332.069$. Thus, the ZICMP and Bernoulli procedures produced equal success probability estimates and log-likelihood values; the difference in AIC values stems from the difference in the number of associated parameters for the two respective distributions. The standard errors associated with the ZICMP model, however, could not be computed because, as $\nu \to \infty$, the ZICMP information matrix approaches singularity. Singularity of the information matrix implies that it cannot be inverted to provide the standard errors associated with the parameter estimates.

Meanwhile, the individual parameters of the "zero-inflated Bernoulli" distribution (i.e. ZIB(π, p)) cannot be identified, resulting in singularity of the Fisher information matrix with respect to these parameters (Rothenberg, 1971). In fact, as noted above, we observe that the ZICMP information matrix approaches singularity as well as $\nu \to \infty$. The notion of a "zero-inflated Bernoulli model", however, is non-sensical, so this matter is not a concern. Just as the "zero-inflated Bernoulli" model can be reparametrized as a Bernoulli distribution with a modified success probability, zero-inflation becomes unnecessary for the COM–Poisson model as it approaches a Bernoulli distribution.

The simulated data examples representing over- or under-dispersion, respectively, produce results as expected. Both ZICMP simulations are best estimated via the ZICMP model, producing parameter estimates that are each close to their respective true values. While the under-dispersed example appears to show lacking goodness of fit measures for all models considered, we are reminded that the reported p-values in Table 1 are conservative statistics based on the χ^2_{K-1-q} distribution. Considering the full range of possible degrees of freedom between $K - 1 - q$ and $K - 1$, we find that the corresponding p-values thus range from 0.0104 with 1 degree of freedom (as reported in Table 1) to 0.1605 with 4 degrees of freedom for the ZICMP model; meanwhile, the range in p-values under the other models still results in values less than 0.0001. Hence, we see that the zero-inflated COM–Poisson is the only model choice to offer a reasonable distribution fit for the simulated under-dispersed dataset. Meanwhile, for the simulated over-dispersed example, we see that the zero-inflated negative binomial and zero-inflated COM–Poisson offer comparable measures; this makes sense, given the ability of each of these distributions to address data over-dispersion.

4.2. Large sample MLE properties

We carry out a simulation study to assess the large sample properties of the maximum likelihood estimator (MLE), $\hat{\boldsymbol{\theta}} = (\hat{\lambda}, \hat{\nu}, \hat{p})$, when drawing a random sample from a ZICMP(λ, ν, p) distribution where $\lambda = 2$, $p = 0.1$, and $n \in \{100, 200, 500, 1000\}$ for various values of ν between 0.25 and 30. For each combination of parameters (λ, ν, p) and each n,



(a) $\nu = 0.25$. (b) $\nu = 0.5$. (c) $\nu = 0.75$. (d) $\nu = 1$.

(e) $\nu = 2$. (f) $\nu = 3$. (g) $\nu = 4$. (h) $\nu = 5$.

(i) $\nu = 8$. (j) $\nu = 10$. (k) $\nu = 20$. (l) $\nu = 30$.

Fig. 2. Empirical CDF of Wald statistic for simulated data from ZICMP(λ, ν, p) scenarios with $\lambda = 2$, $p = 0.1$, and various values for ν to assess consistency and asymptotic normality. In each case, the target CDF of χ^2_3 is shown for reference.

$R = 1000$ samples of size n are drawn, and the MLE is computed on each sample yielding $\hat{\boldsymbol{\theta}}^{(r)}$ for $r = 1, \ldots, R$. Wald statistics $W^{(r)} = (\hat{\boldsymbol{\theta}}^{(r)} - \boldsymbol{\theta})^T \mathcal{I}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^{(r)} - \boldsymbol{\theta})$ are then obtained for $r = 1, \ldots, R$. If $\hat{\boldsymbol{\theta}}$ follows the anticipated large sample $\mathrm{N}(\boldsymbol{\theta}, \mathcal{I}^{-1}(\boldsymbol{\theta}))$ distribution, the empirical cumulative distribution function (CDF) of $W^{(1)}, \ldots, W^{(R)}$ should approach the CDF of χ^2_3 as n becomes large. Fig. 2 plots the empirical CDF versus χ^2_3 for different values of ν. We see that the rate of convergence varies depending on the values of ν, approaching convergence faster for $\nu \leq 1$ than for smaller values of $\nu \geq 1$. In fact, the behavior of the $W^{(r)}$ is as expected for $\nu \leq 2$, but changes as ν becomes larger. When $\nu = 5$, the convergence rate slows substantially. As ν increases beyond 5, the rate of convergence to a χ^2_3 distribution appears to increase; in fact, the limiting distribution appears to change as well to be stochastically smaller than χ^2_3. As discussed in Section 4.1, ZICMP(λ, ν, p) approaches a non-identifiable "zero-inflated Bernoulli" distribution with a singular Fisher information matrix. This is likely what is influencing the behavior of our Wald statistic, which is a function of the information matrix.

5. Examples

Loeys et al. (2012) investigated the impact of education level and level of anxious attachment on the number of unwanted pursuit behavior perpetrations in the context of couple separation trajectories. This dataset is clearly over-dispersed as the mean number of unwanted pursuit behavior perpetrations is 2.284 while the associated variance equals 23.302. Further, there are 246 of 387 cases where the number of unwanted pursuit behavior perpetrations equals zero, so the dataset clearly contains an excess number of zeroes. Accordingly, in their work, Loeys et al. (2012) consider various count models, namely Poisson, negative binomial, and their corresponding zero-inflated models to conduct model selection. This section expounds on that work to further consider the COM–Poisson (CMP) and zero-inflated COM–Poisson (ZICMP) models in the model comparison. To allow for reasonable model comparison, a constant dispersion parameter, ν, is estimated. One can, however, consider a model to describe the relationship between various explanatory variables and ν, as described in Section 3. Parameter estimation results are provided in Table 2.

Table 2
Estimated parameters (corresponding standard error in parenthesis; both rounded to three significant digits), log-likelihood, and AIC for various count models considering association between education level and anxious attachment level with the number of unwanted pursuit behavior perpetrations: Poisson (P), negative binomial (NB), COM–Poisson (CMP), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), zero-inflated COM–Poisson (ZICMP), and zero-inflated geometric (ZIG).

	P	NB	CMP	ZIP	ZINB	ZICMP	ZIG
Count component							
(intercept)	0.817*	0.855*	−0.385*	1.921*	1.723*	−0.160*	1.770*
	(0.044)	(0.155)	(0.055)	(0.044)	(0.150)	(0.077)	(0.122)
Education	−0.216*	−0.353	−0.056	−0.350*	−0.490*	−0.068*	−0.476*
	(0.070)	(0.250)	(0.038)	(0.071)	(0.206)	(0.034)	(0.191)
Anxiety	0.422*	0.486*	0.117*	0.133*	0.205	0.023	0.199
	(0.033)	(0.122)	(0.021)	(0.034)	(0.108)	(0.015)	(0.100)
Zero component							
(intercept)				0.673*	0.340	0.418*	0.422*
				(0.142)	(0.210)	(0.167)	(0.159)
Education				−0.232	−0.459	−0.388	−0.416
				(0.222)	(0.297)	(0.268)	(0.271)
Anxiety				−0.483*	−0.520*	−0.524*	−0.503*
				(0.111)	(0.147)	(0.133)	(0.135)
$\hat{\theta}$		0.194			0.821		
		(0.022)			(0.226)		
$\hat{\nu}$			0.000			0.000	
			(0.033)			(0.031)	
# parameters	3	4	4	6	7	7	6
$\log L$	−1388.20	−638.96	−756.92	−802.45	−626.14	−627.17	−626.42
AIC	2782.4	1285.9	1521.84	1616.9	1266.3	1268.3	1264.8

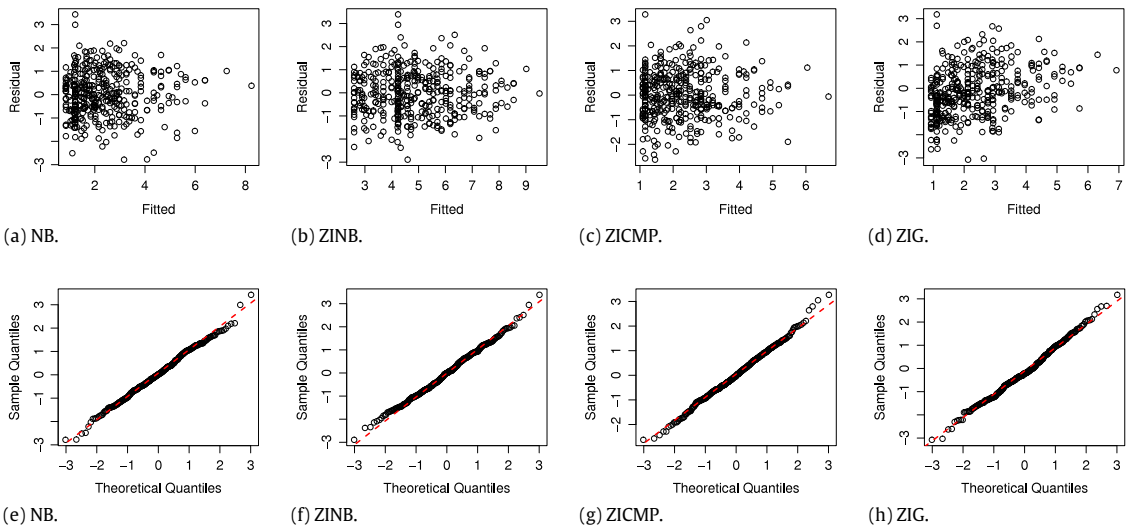
Table 2 identifies those estimated coefficients determined to be statistically significant at the 5% significance level (denoted with an asterisk, *). The difference in models impacts the perceived statistical significance and inference associated with the variables of interest (education and anxiety). While anxiety is always perceived to be a statistically significant variable associating with the number of unwanted pursuit behaviors (UPBs) under a non-zero-inflated model, among zero-inflated models, it is viewed to have a statistically significant association with UPBs only in the case of a ZIP model. Meanwhile, when accounting for zero-inflation, education is inferred to be statistically significantly associated with the number of UPBs under all zero-inflated models (i.e. after accounting for excess zeroes), but only statistically significant in the Poisson model (among non-zero-inflated models). Finally, anxiety is consistently viewed to statistically significantly associate with the probability of no UPBs.

In all cases, the zero-inflated models outperform their standard model counterparts. This makes sense, given the zero-inflated model's ability to account for the excess zeroes. Further, we see that the negative binomial and COM–Poisson models outperform the Poisson model (whether one considers the classical model comparisons or their zero-inflated counterparts). This is not only supported by the decrease in AIC, but can also be seen through the resulting likelihood ratio test comparing ZIP to ZICMP ($-2\log\Lambda = 350.56$ and p-value < 0.0001); the negative binomial and COM–Poisson models are capturing the data over-dispersion present in this dataset (the ZICMP model estimate $\hat{\nu} < 1$), while the Poisson model is restricted to the constraint of perceived data equi-dispersion.

Given the data over-dispersion in this example, we initially find that the zero-inflated negative binomial and zero-inflated COM–Poisson models perform the best, with the ZINB model slightly outperforming the ZICMP model, however the difference in performance appears negligible. In particular, we note that the estimated dispersion parameter for the ZICMP is $\hat{\nu} = 0.000$, i.e. we see that the maximum likelihood estimate for ν implies that the best fit for this dataset reduces to a zero-inflated geometric model because of the significant data over-dispersion present here. In fact, the ZINB and ZICMP models illustrate a powerful relationship—the ZINB is derived by the sum of ZIG (i.e. ZICMP with $\nu = 0$) distributions. Accordingly, it makes sense that the two models performed as illustrated because the ZINB model has slightly added flexibility not provided by the ZICMP($\nu = 0$) model, yet the ZICMP model does a good job trying to compensate for that difference in relation.

Using the ZICMP results as an investigative tool, however, prompts one to consider a zero-inflated geometric (ZIG) regression from VGAM in order to obtain more precise estimates under this special case and reduce the number of parameters used in the model (thus reducing the AIC). In fact, we see that the ZIG model outperforms the ZINB and ZICMP models with regard to AIC because the three models have nearly identical log-likelihood results while the ZIG model uses one less parameter to obtain this result. Further, the resulting inferences associated with the ZINB, ZICMP, and ZIG models are consistent across these optimal models.

Fig. 3 provides residual analysis plots for the randomized quantile residuals stemming from the negative binomial (NB), zero-inflated negative binomial (ZINB), zero-inflated COM–Poisson (ZICMP), and zero-inflated geometric (ZIG) models; Fig. 3(a)–(d) show the respective fitted versus residual plots, and Fig. 3(e)–(h) contain the corresponding QQ plots for each of the four models. We consider these models because they have the smallest AIC in comparison to the other models initially



(a) NB. (b) ZINB. (c) ZICMP. (d) ZIG.

(e) NB. (f) ZINB. (g) ZICMP. (h) ZIG.

Fig. 3. Randomized quantile residual plots for the couples data, stemming from considering the negative binomial (NB), zero-inflated negative binomial (ZINB), zero-inflated COM–Poisson (ZICMP), and zero-inflated geometric (ZIG) models, respectively. (a)–(d) Fitted values versus randomized quantile residuals; (e)–(h) QQ plots of randomized quantile residuals.

considered for model comparison. Further, we consider residual analysis based on the randomized quantile residuals in lieu of the raw residuals because raw residuals associated with our data construct (under any of the considered models) would produce fitted versus residual and QQ plots that contain spurious curves, thus obscuring any meaningful inference (Dunn and Smyth, 1996). In both the fitted versus residual plots and the QQ plots shown in Fig. 3, we see that these considered models describe the data in similar fashion. The fitted versus residual plots appear centered around zero, although the ZIG plot appears to show a slight tendency of association that may deserve further investigation. As well, a few observations may be identified as outliers, or the fitted versus residual plot may contain some heteroskedasticity. These issues, however, can be addressed through model consideration associating the dispersion parameter with the explanatory variables (Sellers and Shmueli, 2009). Meanwhile, the QQ plots associated with each of these model display approximate normality.

To demonstrate the ZICMP's ability to address dispersion, meanwhile, we inflate two datasets each comprised of 100 data values (10–5s, 88–6s, and 2–7s) with 10 or 20 zeroes, respectively. Dataset 1 (which contains 10 zeroes) is clearly under-dispersed with a mean equal to 5.382 and variance approximately equal to 3.027, while Dataset 2 (containing 20 zeroes) appears to be either equi- or over-dispersed (mean equals 4.933; variance equals 5.004). Tables 3 and 4 contain the resulting parameter estimates, log-likelihood and AIC values obtained from modeling these data via Poisson, negative binomial, COM–Poisson distributions, or their zero-inflated analogs. Tables 3 and 4, in particular, demonstrate the impact of applying a (zero-inflated or non-zero-inflated) Poisson or negative binomial model to an under-dispersed dataset; notice that the Poisson and negative binomial models (whether zero-inflated or not, respectively) produce equivalent estimates and log-likelihood values. The negative binomial model is equipped to only address data over-dispersion in that the parameter space only allows for the variance to be greater than or equal to the mean. When trying to model under-dispersed data (i.e. when the variance is less than the mean), the parameter space is restricted from moving into that space which considers a variance less than the mean therefore, at best, it can only maximize the log-likelihood at those parameter estimates that produce equi-dispersion (namely, the Poisson estimates). In contrast, the ZICMP model is able to account for the excess zeroes as well as the data under-dispersion ($\hat{\nu} > 1$); accordingly, it produces the best log-likelihood and AIC measures.

Table 4 is interesting because, based solely on the dispersion index, one perceives this dataset to be approximately equi-dispersed (or even slightly over-dispersed). Through exploratory data analysis and statistical modeling, however, we can actually see and account for the mixture of distributions. In fact, when accounting for the excess zeroes, the dataset is severely under-dispersed ($\hat{\nu} = 38.666$). This dataset thus serves as another example illustrating that datasets with perceived forms of dispersion can actually stem from probability mixtures with different dispersion levels, as discussed in Sellers and Shmueli (2013).

For Datasets 1 and 2, the model comparison between the ZIP and ZICMP models both produce a test statistic value of 281.51 and corresponding p-value <0.0001, demonstrating statistically significant data dispersion exists in the respective datasets. Further, $\hat{\nu} > 1$ in both cases, recognizing the significant data under-dispersion in each dataset.

6. Discussion

This work develops a zero-inflated COM–Poisson regression to model count data containing some form of dispersion (i.e. over- or under-dispersion) and an excess number of zeroes. Such data structures appear frequently in various applications such as psychology, engineering, and business. Excess zeroes are a common cause of data over-dispersion

Table 3
Estimated parameters (and corresponding standard error in parenthesis; both rounded to three significant digits), log-likelihood, and AIC for various count models applied to simulated Dataset 1: Poisson (P), negative binomial (NB), COM–Poisson (CMP), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), and zero-inflated COM–Poisson (ZICMP).

	P	NB	CMP	ZIP	ZINB	ZICMP
Count component						
(intercept)	1.683	1.683	1.998	1.776	1.776	71.454
	(0.041)	(0.041)	(0.303)	(0.041)	(0.041)	(9.487)
Zero component						
(intercept)				−2.333	−2.333	−2.303
				(0.341)	(0.341)	(0.332)
$\log\hat{\theta}$		11.179			12.836	
		(23.371)			(37.501)	
$\hat{\nu}$			1.177			38.666
			(0.169)			(5.207)
# parameters	1	2	2	2	3	3
$\log L$	−239.55	−239.55	−238.94	−216.40	−216.40	−75.61
AIC	481.10	483.10	481.89	436.73	438.73	157.22

Table 4
Estimated parameters (and corresponding standard error in parenthesis; both rounded to three significant digits), log-likelihood, and AIC for various count models applied to simulated Dataset 2: Poisson (P), negative binomial (NB), COM–Poisson (CMP), zero-inflated Poisson (ZIP), zero-inflated negative binomial (ZINB), and zero-inflated COM–Poisson (ZICMP).

	P	NB	CMP	ZIP	ZINB	ZICMP
Count component						
(intercept)	1.596	1.596	0.948	1.776	1.776	71.455
	(0.041)	(0.042)	(0.173)	(0.041)	(0.041)	(9.487)
Zero component						
(intercept)				−1.626	−1.626	−1.609
				(0.248)	(0.248)	(0.245)
$\log\hat{\theta}$		5.582			12.243	
		(11.483)			(26.507)	
$\hat{\nu}$			0.621			38.666
			(0.099)			(5.207)
# parameters	1	2	2	2	3	3
log L	−291.06	−291.06	−285.24	−236.92	−236.92	−96.17
AIC	584.12	586.11	574.49	477.84	479.84	198.34

(Hilbe, 2008). For any generated dataset, data outcomes that are zeroes add to the sample size but not to the sum total of data observations, thus diminishing the mean of the dataset. Meanwhile, these values still contribute to the variance of the dataset, thus increasing the chance that the variance is greater than the mean. However, it does not necessarily imply the overall dispersion level of a zero-inflated dataset as being overdispersed. Sellers and Shmueli (2013) provide data examples where distribution mixtures can impact the overall level of data dispersion. Because such data can be over- or under-dispersed, the two-parameter COM–Poisson structure allows for more flexibility in describing the relationship between explanatory variables and the response variable—both in the count component and the zero component. In fact, we demonstrate the flexibility of the ZICMP in its ability to capture three special case zero-inflated distributions, namely the ZIP, ZIG, and logistic models. This stems from the distributional structure and statistical properties associated with the COM–Poisson distribution.

While the ZICMP did not outperform the ZINB with respect to AIC in the real example illustration, the difference is small and we further found the ZIG model to slightly outperform both the ZINB and ZICMP models. As well, this example demonstrated our ability to use the ZICMP model as an exploratory tool for model selection. As well, the established results give rise to considering a more flexible zero-inflated COM–Poisson structure that can encompass the ZINB model, namely a zero-inflated sum-of-COM–Poissons (ZIsCMP) model. The sCOM–Poisson distribution is derived from the sum of n independent and identically distributed COM–Poisson random variables, and generalizes the Poisson, binomial, and negative binomial distributions (Sellers, 2015). Accordingly, the development of a zero-inflated sCOM–Poisson model would perform at least as well as the ZINB model and allow for added flexibility in modeling over- or under-dispersion.

This work allows for various extensions, including the inclusion of observation-level dispersion modeling, and longitudinal and/or clustered data settings where data dispersion exists. Choo-Wosoba et al. (2015) extend ZICMP to

consider longitudinal data with clustering; their work can further incorporate the approach of Gumedze and Chatora (2014) to detect outliers in the presence of general data-dispersion.

Acknowledgment

Support for Kimberly Sellers was provided in part by the ASA/NSF/Census Research Program, U. S. Census Bureau Contract #YA1323-14-SE-0122.

Appendix. Fisher information matrix components associated with constant-dispersion ZICMP model coefficients

In order to determine the components of the Fisher information matrix, we decompose the ultimate calculation into its segments involving the distribution parameters, and (from there) the corresponding regression parameters. First, we consider the information matrix associated with the parameters λ, ν, p. Letting $u = I(y = 0)$ and $1 - u = I(y > 0)$, and allowing the shorthand notation Z and $\log Z$ respectively denote the normalizing function $Z(\lambda, \nu)$ and its logarithm, the first derivatives of the zero-inflated COM–Poisson probability mass function are

$$\frac{\partial}{\partial\lambda}\log f(y \mid \theta) = -\frac{1-p}{Zp+(1-p)}\frac{\partial \log Z}{\partial\lambda}u + \frac{y}{\lambda}(1-u) - \frac{\partial \log Z}{\partial\lambda}(1-u),$$
$$\frac{\partial}{\partial\nu}\log f(y \mid \theta) = -\frac{1-p}{Zp+(1-p)}\frac{\partial \log Z}{\partial\nu}u + (1-u)\log\Gamma(y+1) - \frac{\partial \log Z}{\partial\nu}(1-u), \quad \text{and}$$
$$\frac{\partial}{\partial p}\log f(y \mid \theta) = \frac{Z-1}{Zp+(1-p)}u - \frac{1}{1-p}(1-u),$$

while the second derivatives are

$$\frac{\partial^2}{\partial\lambda^2}\log f(y \mid \theta) = -\frac{1-p}{Zp+(1-p)}\frac{\partial^2 \log Z}{\partial\lambda^2}u + \frac{p(1-p)Z}{[Zp+(1-p)]^2}\left(\frac{\partial \log Z}{\partial\lambda}\right)^2 u - \frac{y}{\lambda^2}(1-u) - \frac{\partial^2 \log Z}{\partial\lambda^2}(1-u),$$
$$\frac{\partial^2}{\partial\nu^2}\log f(y \mid \theta) = \frac{p(1-p)Z}{[Zp+(1-p)]^2}\left(\frac{\partial \log Z}{\partial\nu}\right)^2 u - \frac{1-p}{Zp+(1-p)}\frac{\partial^2 \log Z}{\partial\nu^2}u - (1-u)\frac{\partial^2 \log Z}{\partial\nu^2},$$
$$\frac{\partial^2}{\partial p^2}\log f(y \mid \theta) = -\frac{(Z-1)^2}{[Zp+(1-p)]^2}u - \frac{1}{(1-p)^2}(1-u),$$
$$\frac{\partial^2}{\partial\lambda\partial\nu}\log f(y \mid \theta) = -\frac{1-p}{Zp+(1-p)}\frac{\partial^2 \log Z}{\partial\lambda\partial\nu} + \frac{p(1-p)Z}{[Zp+(1-p)]^2}\frac{\partial \log Z}{\partial\lambda}\frac{\partial \log Z}{\partial\nu}u - \frac{\partial^2 \log Z}{\partial\lambda\partial\nu}(1-u),$$
$$\frac{\partial^2}{\partial p\partial\nu}\log f(y \mid \theta) = -\frac{Z}{[Zp+(1-p)]^2}\frac{\partial \log Z}{\partial\nu}u, \quad \text{and}$$
$$\frac{\partial^2}{\partial\lambda\partial p}\log f(y \mid \theta) = \frac{1}{[Zp+(1-p)]^2}\frac{\partial \log Z}{\partial\lambda}u.$$

Thus, the Fisher information matrix for $\{\lambda, \nu, p\}$ has the form

$$I_{\{\lambda,\nu,p\}} = \begin{pmatrix} \mathcal{I}_{\lambda,\lambda} & \mathcal{I}_{\lambda,\nu} & \mathcal{I}_{\lambda,p} \\ \mathcal{I}_{\lambda,\nu} & \mathcal{I}_{\nu,\nu} & \mathcal{I}_{\nu,p} \\ \mathcal{I}_{\lambda,p} & \mathcal{I}_{\nu,p} & \mathcal{I}_{p,p} \end{pmatrix},$$

whose components are

$$\mathcal{I}_{\lambda\lambda} = (1-p)\frac{\partial^2 \log Z}{\partial\lambda^2} - \frac{p(1-p)}{Zp+(1-p)}\left(\frac{\partial \log Z}{\partial\lambda}\right)^2 + \frac{\mu}{\lambda^2},$$
$$\mathcal{I}_{\nu\nu} = (1-p)\frac{\partial^2 \log Z}{\partial\nu^2} - \frac{p(1-p)}{Zp+(1-p)}\left(\frac{\partial \log Z}{\partial\nu}\right)^2,$$
$$\mathcal{I}_{pp} = \frac{1}{Z}\frac{(Z-1)^2}{Zp+(1-p)} + \frac{1}{Z}\frac{Z-1}{1-p},$$
$$\mathcal{I}_{\lambda\nu} = (1-p)\frac{\partial^2 \log Z}{\partial\nu\partial\lambda} - \frac{p(1-p)}{Zp+(1-p)}\frac{\partial \log Z}{\partial\nu}\frac{\partial \log Z}{\partial\lambda},$$
$$\mathcal{I}_{\lambda p} = -\frac{1}{Zp+(1-p)}\frac{\partial \log Z}{\partial\lambda}, \quad \text{and}$$
$$\mathcal{I}_{\nu p} = -\frac{1}{Zp+(1-p)}\frac{\partial \log Z}{\partial\nu}$$

for $\mu = E(y) = (1-p)\lambda \frac{\partial \log Z}{\partial \lambda}$.

In working to determine the form of the information matrix associated with our regression model, we first find that

$$\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \log f(y \mid \theta) = \left[\frac{\partial^2}{\partial \lambda^2} \log f(y \mid \theta) \right] \exp(\boldsymbol{x}^T \beta)^2 \boldsymbol{x}\boldsymbol{x}^T + \left[\frac{\partial}{\partial \lambda} \log f(y \mid \theta) \right] \exp(\boldsymbol{x}^T \beta) \boldsymbol{x}\boldsymbol{x}^T,$$

implying that

$$\mathcal{I}_{\beta\beta} = E\left[-\frac{\partial^2}{\partial \lambda^2} \log f(y \mid \theta) \right] \lambda^2 \boldsymbol{x}\boldsymbol{x}^T = \mathcal{I}_{\lambda\lambda} \cdot \lambda^2 \boldsymbol{x}\boldsymbol{x}^T;$$

note that $E\left\{ \left[\frac{\partial}{\partial \lambda} \log f(y \mid \theta) \right] \exp(\boldsymbol{x}^T \beta) \boldsymbol{x}\boldsymbol{x}^T \right\} = 0$. In similar fashion, we find the components

$$\mathcal{I}_{\zeta\zeta} = E\left\{ \left[\frac{\partial^2}{\partial p^2} \log f(y \mid \theta) \right] p^2 (1-p)^2 \boldsymbol{w}\boldsymbol{w}^T + \left[\frac{\partial}{\partial p} \log f(y \mid \theta) \right] g'(\boldsymbol{w}^T \boldsymbol{\zeta}) \boldsymbol{w}\boldsymbol{w}^T \right\} = \mathcal{I}_{pp} \cdot p^2 (1-p)^2 \boldsymbol{w}\boldsymbol{w}^T,$$

$$\mathcal{I}_{\beta\nu} = E\left\{ \left[\frac{\partial^2}{\partial \lambda \partial \nu} \log f(y \mid \theta) \right] \lambda \boldsymbol{x} \right\} = \mathcal{I}_{p\nu} \cdot \lambda \boldsymbol{x},$$

$$\mathcal{I}_{\zeta\nu} = E\left\{ \left[\frac{\partial^2}{\partial p \partial \nu} \log f(y \mid \theta) \right] p(1-p) \boldsymbol{w} \right\} = \mathcal{I}_{p\nu} \cdot p(1-p) \boldsymbol{w}, \quad \text{and}$$

$$\mathcal{I}_{\beta\zeta} = E\left\{ \left[\frac{\partial^2}{\partial \lambda \partial p} \log f(y \mid \theta) \right] \lambda p(1-p) \boldsymbol{x}\boldsymbol{w}^T \right\} = \mathcal{I}_{\lambda p} \cdot \lambda p(1-p) \boldsymbol{x}\boldsymbol{w}^T.$$

Finally, supposing that we have a sample with covariate values $(\boldsymbol{x}_i, \boldsymbol{w}_i)$ for $i \in \{1, \ldots, n\}$, and corresponding $p_i = 1/(1 + e^{-\boldsymbol{w}_i^T \boldsymbol{\zeta}})$ and $\lambda_i = e^{\boldsymbol{x}_i^T \beta}$, the Fisher information matrix has the form

$$\mathbf{I} = \begin{pmatrix} \mathcal{I}_{\beta} & \mathcal{I}_{\beta,\nu} & \mathcal{I}_{\beta,\zeta} \\ \mathcal{I}'_{\beta,\nu} & \mathcal{I}_{\nu} & \mathcal{I}_{\zeta,\nu} \\ \mathcal{I}'_{\beta,\zeta} & \mathcal{I}'_{\zeta,\nu} & \mathcal{I}_{\zeta} \end{pmatrix},$$

where

$$\mathcal{I}_{\beta\beta} = \sum_{i=1}^{n} \mathcal{I}_{\lambda\lambda}(\lambda_i, \nu, p_i) \cdot \lambda_i^2 \boldsymbol{x}_i \boldsymbol{x}_i^T,$$

$$\mathcal{I}_{\zeta\zeta} = \sum_{i=1}^{n} \mathcal{I}_{pp}(\lambda_i, \nu, p_i) \cdot p_i^2 (1-p_i)^2 \boldsymbol{w}_i \boldsymbol{w}_i^T,$$

$$\mathcal{I}_{\beta\nu} = \sum_{i=1}^{n} \mathcal{I}_{p\nu}(\lambda_i, \nu, p_i) \cdot \lambda_i \boldsymbol{x}_i,$$

$$\mathcal{I}_{\zeta\nu} = \sum_{i=1}^{n} \mathcal{I}_{p\nu}(\lambda_i, \nu, p_i) \cdot p_i (1-p_i) \boldsymbol{w}_i,$$

$$\mathcal{I}_{\beta\zeta} = \sum_{i=1}^{n} \mathcal{I}_{\lambda p}(\lambda_i, \nu, p_i) \cdot \lambda_i p_i (1-p_i) \boldsymbol{x}_i \boldsymbol{w}_i^T, \quad \text{and}$$

$$\mathcal{I}_{\nu\nu} = \sum_{i=1}^{n} \mathcal{I}_{\nu\nu}(\lambda_i, \nu, p_i).$$

References

Choo-Wosoba, H., Levy, S.M., Datta, S., 2015. Marginal regression models for clustered count data based on zero-inflated Conway–Maxwell–Poisson distribution with applications. Biometrics. http://dx.doi.org/10.1111/biom.12436.

Conway, R.W., Maxwell, W.L., 1962. A queuing model with state dependent service rates. J. Ind. Eng. 12, 132–136.

Dunn, P.K., Smyth, G.K., 1996. Randomized quantile residuals. J. Comput. Graph. Statist. 5, 236–244.

Famoye, F., Singh, K.P., 2006. Zero-inflated generalized Poisson regression model with an application to domestic violence data. J. Data Sci. 4, 117–130.

Gumedze, F.N., Chatora, T.D., 2014. Detection of outliers in longitudinal count data via overdispersion. Comput. Statist. Data Anal. 79, 192–202.

Hall, D.B., 2000. Zero-inflated Poisson and binomial regression with random effects: A case study. Biometrics 56 (4), 1030–1039.

Hilbe, J.M., 2008. Negative Binomial Regression. Cambridge University Press, New York.

Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics 34 (1), 1–14.

Loeys, T., Moerkerke, B., DeSmet, O., Buysse, A., 2012. The analysis of zero-inflated count data: Beyond zero-inflated Poisson regression. British J. Math. Statist. Psych. 65 (1), 163–180.

Pandya, M., Pandya, H., Pandya, S., 2012. Bayesian inference on mixture of geometric with degenerate distribution: zero inflated geometric distribution. Int. J. Res. Rev. Appl. Sci. 13 (1), 53–65.

R Core Team 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Rothenberg, T.J., 1971. Identification in parametric models. Econometrica 39, 577–591.
Sellers, K.F., 2012. A generalized statistical control chart for over- or under-dispersed data. Qual. Reliab. Eng. Int. 28 (1), 59–65.
Sellers, K.F., 2015. sCOM-Poisson: A flexible count distribution to address dispersion in count data. Working Paper, pp. 1–7.
Sellers, K.F., Shmueli, G., 2009. A regression model for count data with observation-level dispersion. In: Proceedings of the International Workshop on Statistical Modeling, July 20–24, 2009. Cornell University, Ithaca, NY, pp. 337–344.
Sellers, K.F., Shmueli, G., 2010. A flexible regression model for count data. Ann. Appl. Stat. 4 (2), 943–961.
Sellers, K., Shmueli, G., 2013. Data dispersion: Now you see it... now you don't. Comm. Statist. Theory Methods 42, 1–14.
Sellers, K.F., Shmueli, G., Borle, S., 2011. The COM-Poisson model for count data: a survey of methods and applications. Appl. Stoch. Models Bus. Ind. 28, 104–116.
Shmueli, G., Minka, T.P., Kadane, J.B., Borle, S., Boatwright, P., 2005. A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. Appl. Stat. 54 (1), 127–142.
Sutradhar, S.C., Neerchal, N.K., Morel, J.G., 2008. A goodness-of-fit test for overdispersed binomial (or multinomial) models. J. Statist. Plann. Inference 138 (5), 1459–1471.
Yee, Thomas W., 2014. VGAM: Vector Generalized Linear and Additive Models. The University of Auckland, Auckland, New Zealand.
Zhu, L., Sellers, K., Morris, D., Shmuéli, G., 2016. A generalized stochastic process model for count data (under review).
Zipkin, E.F., Leirness, J.B., Kinlan, B.P., O'Connell, A.F., Silverman, E.D., 2014. Fitting statistical distributions to sea duck count data: Implications for survey design and abundance estimation. Stat. Methodol. 17, 67–81.