Reparametrization of COM-Poisson Regression Models with Applications in the Analysis of Experimental Data

Replies to Reviewer's Comments

Eduardo E. Ribeiro Jr Walmes M. Zeviani Wagner H. Bonat
Clarice G. B. Demétrio John Hinde

August 20, 2018

1 Replies to Referee 1

Overall reviewer's comment: *Overall, the authors did a good job while presenting theirs results. However, I believe that the manuscript at the current state requires a considerable revision. Mainly, it requires a rigorous discussion with numerical results that when the proposed approach is useful and when it is not.*

Response: We appreciate the referee for his revision. All suggestions have been taken into account. Regarding the referees' main comment, we propose a reparametrization of the COM-Poisson model that is always valid. Only the interpretation of the μ parameter depends on the approximation accuracy. We make it clear in the updated version of the paper. The applicability of the model is discussed at the end of Section 3 through the study of the dispersion (DI), zero-inflation (ZI) and heavy-tail (HI) indexes (Figure 3).

Specific reviewer's comments:

1. *There has been some active research to find a better approximation for normalizing constant $Z(\lambda, \nu)$ from which the expressions for both mean and variance will be derived. Please refer to the Gaunt et al. (2016) for the latest results.*

   Response:

2. *It seems that the paper missing some latest literature on CMP distribution. For example, the distributional properties of CMP were found in Daly and Gaunt (2015).*

   Response:

3. *The likelihood equation (4.1) needs more details. Can we derive score equations? It seems like the likelihood has become even more complicated. Some of the nice properties of the CMP likelihood such as the canonical link are lost.*

   Response: The log-likelihood function cannot be derived in any parametrization. We obtain the Hessian matrix by central finite differences (as said in ...). Regarding the expression of reparametrized log-likelihood (Equation 3.1), although it seems more complicated, this function is better behaved than the function expressed in the original form (see Figure 8). Consider the exponential family of distributions, with density given by

$$f(y) = \exp\{\sigma[\theta y - b(\theta)] - c(y, \phi)\},$$

the COM-Poisson belongs to this family with $\sigma = 1$, $\theta = \log(\lambda)$, $b(\theta) = \log(Z(\lambda, \nu))$ and $c(y, \phi) = -\nu \log(y!)$, only if ν is known. Since we don't have $b'(\theta) = \partial b(\theta)/\partial\theta$ in closed form, the canonical link function has no an important rule. In this case, the link function just maps θ in the parameter space of λ. In the reparametrized version of the COM-Poisson, we adopted log link function to map the parameter space and to have good interpretations (in terms of rate ratios).

4. *For model fitting, the authors used general purpose optimization function in R. Did you observe any non convergence situations during your numerical studies.*

   Response: Yes, we have 37 situations where the algorithm did not converge in the simulation study (see table below). This occurred only for strong overdispersion and small sample size. We report these situations in the updated version of the paper.

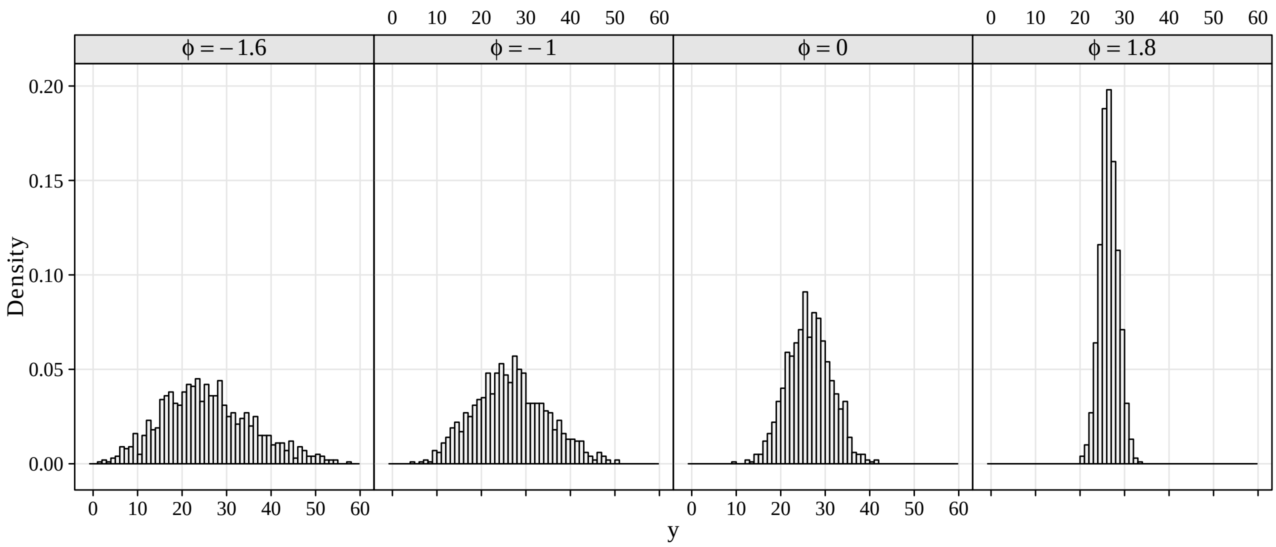
	$\phi = -1.6$	$\phi = -1$	$\phi = 0$	$\phi = 1.8$
$n = 50$	29	0	0	0
$n = 100$	8	0	0	0
$n = 300$	0	0	0	0
$n = 1000$	0	0	0	0

5. *Numerical derivatives with CMP tend to be very slow and prone to errors. It would be good to have a computation times from the method for different sample sizes and dispersion levels.*

   Response:

6. *In Section 5, the simulated counts are ranging from 3 to 27. This is too limited. Please increase the range.*

   Response: The simulation study was performed with averages varying from 3 to 27. Therefore, considering the different scenarios ($\phi = -1.6, \phi = -1, \phi = 0$ and $\phi = 1.8$), we cover a wide range of counts. The Figure below shows the variation of the counts for $\mu = 27$ in the different scenarios.



7. *I believe that the authors simulated data according to the reparametrized distribution. Please generate the data according to the original distribution. Otherwise, the results not useful.*

For details on simulating from original CMP distribution, please refer to Chatla and Shmueli (2018).

Response: We simulated data according to the original parametrization. The steps in the simulation study were i) obtain the $\mu_i = \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})$; ii) Back to the original parametrization $\lambda = (\mu + (\nu - 1)/(2\nu))^\nu$ (Equation 3.1); and iii) simulate data according COM-Poisson distribution using the probability integral transform theorem as implemented in `compoisson` package (Dunn 2012).

8. *For real world data, the over dispersion case is too mild ($\phi = -0.77$). In practice over dispersion can be really high. For example, even lesser than -1.4. It would be nice to have such case, at least in the simulations.*

   Response: In the simulation study we consider the $\phi = -1.6$ and $\phi = -1$ that leads to dispersion indexes greater than four (see Figure 4).

9. *From the presentation point of view, focusing more on applications would be good as there is not much novelty in the methodology.*

   Response: In this paper, we propose a mean-parametrization of the COM-Poisson model. This parametrization introduced $\mu = \lambda^{1/\nu} - (\nu - 1)/(2\nu)$, a simple function of the original parameters λ and ν. We showed that the new parameter space has nice properties like i) the orthogonality between μ and ϕ (and consequently between β_j and ϕ in a regression setting), leading to better both computation and asymptotic (normal-based) inference (Ross 1970); and ii) interpretability of the β_j in terms of the rate ratios since $\mu_i \approx \mathrm{E}(Y_i)$. Moreover, we characterized the COM-Poisson distribution in terms of the dispersion, zero-inflation and heavy-tail indexes that illustrate its applicability in the real count data. These are the methodological novelties of the article besides the exploration of the results. However, in the updated version of the paper, we provided a better discussion of the case studies.

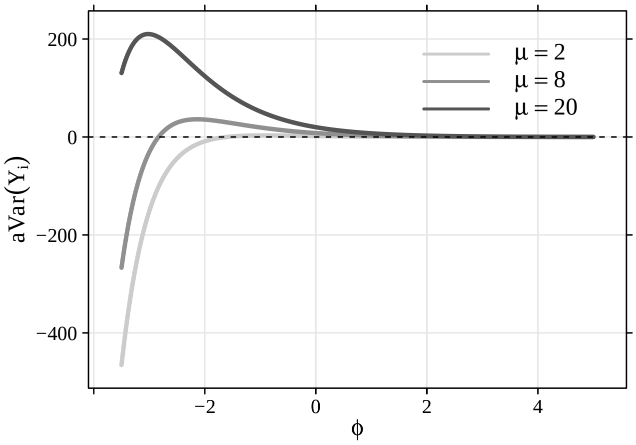
2 Replies to Referee 2

Second paragraph: *The authors suggest the bijective reparameterization, $(\mu_i, \phi) = k(\lambda_i, \nu)$, where $\mu_i = \lambda_i^{1/\nu} - (\nu - 1)/(2\nu)$. (Equation 3.1 plus the sentence after it gives this bijection. Note: The notation $h(\lambda, \nu)$, should probably be changed to $h_\nu(\lambda)$ to better hint that the symbol h^{-1} refers to the inverse with respect to λ, with ν fixed.) The symbol μ_i is presumably used to remind the analyst that this parameter is, by Shmueli et al., approximately equal to $E(Y_i)$; we could write $m(\lambda_i, \nu) = E(Y_i) \approx \mu_i \equiv aE(Y_i)$. Shmueli et al. also showed that $var(Y_i) \approx \lambda_i^{1/\nu} \equiv avar(Y_i)$.*

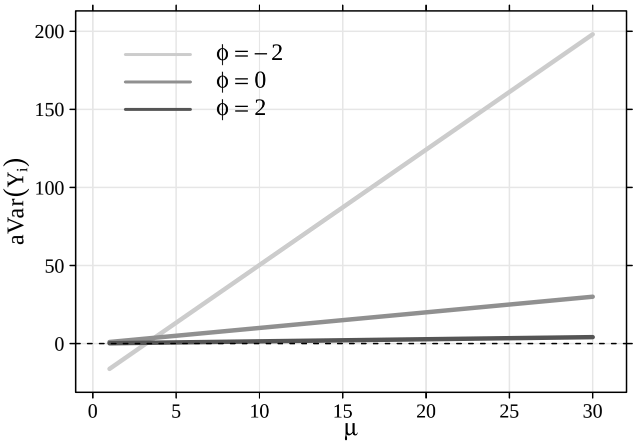
Response: Thank you for this comment. We have changed the notation of the reparameterization function in the updated version.

Third paragraph: *The authors also hint that $var(Y_i) \approx \mu_i \exp(-\phi)$ (see the quadratic error analysis leading to Fig. 1). This is curious given that $avar(Y_i) = \mu \exp(-\phi) + (\exp(\phi) - 1)/(2\exp(2\phi))$ and the latter summand converges to $-\infty$ as $\phi \to -\infty$. This latter summand is 0 or close to 0 for $\phi = 0$ or $\phi > 0$. Taken together, these last two sentences imply that the $\mu_i \exp(-\phi)$ may be a reasonable approximation to the variance when there is equi- or under- dispersion, but it may be unreasonable when there is lots of overdispersion. It seems that some of the results, simulations, and sample analyses reflect this (e.g. Fig 3, Fig 6-8, "deviance function shape under strong overdispersion $\phi = -1.6$ is not as well behaved...", p. 19, etc.). This needs to be addressed.*

Response: We thanks the referee for exploring this point. However, we didn't use the approximation of variance ($a\text{Var}(Y_i)$) to repametrize or fit the COM-Poisson model. To compute de variances, that is showed in the Figure 3(a) and (b) we used $Var(Y) = \sum_{y=0}^{500} y^2 p(y) - [\sum_{y=0}^{500} y p(y)]^2$. As highlighted in the Figure 1(b), the approximation is not accurate, mainly for overdispersion ($\phi < 0$). The Figure (a) and (b) below shows the behaviour of $a\text{Var}(Y_i)$ function by fixing μ and ϕ, repectivaly, illustrating the referees' statement. The $a\text{Var}(Y)$ leads to negative values for variance for small μ and $\phi < 0$ that shows that the approximation is not reasonable for this region of the parameter space.



(a)



(b)

The results shown in Figures 7 and 8 reflect that orthogonality property is slightly lost for strong overdispersion and small μ. This is related to accuracy of the $aE(Y_i)$. When $aE(Y_i)$ is accurate, μ represents the expectation of the Y_i and, consequently, μ and ϕ (or β_j and ϕ in a regression setting) are orthogonal. We improved the discussion about this in the current version of the article.

Fifth paragraph: *This last point also hints that a quasi-Poisson model based on the assumption that $var(Y_i) \propto \mu_i$ (see p. 14) may not be reasonable for analyzing overdispersed data if the COMPo model truly holds. Is this the case?*

Response: The quasi-Poisson model is specified by second-moment assumptions (expectation and variance). As highlighted in Figure 3(a) (and in referee's previous comment), the mean-variance relationship for COM-Poisson is linear. Therefore, is suitable to analyse data generated according to the COM-Poisson distribution, using the assumption $Var(Y_i) \propto \mu_i$. The advantage of the COM-Poisson approach is it corresponds to a fully specified probability model allowing to compute the likelihood and its resulting results (deviance, AIC, BIC).

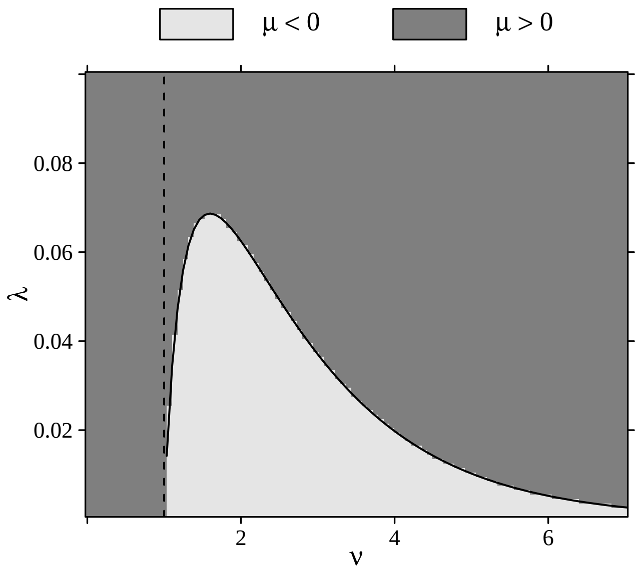
Sixth paragraph: *Consider the example of Section 6.1. If I understand matters correctly, it is surprising that the β estimates under the $[COMPo(\lambda_i, \nu), g(m(\lambda_i, \nu)) = \beta^\top x_i]$ model are so different from those under the $[COMPo(\mu_i, \phi), g(\mu_i) = \beta^\top x_i]$. See the results in Table 2, for example. I would think that if the approximation $m(\lambda_i, \nu) \approx \mu_i$ is reasonable, which your quadratic error results seem to indicate (see Fig 1), then the ML estimates of β would be very similar. Exactly what models are being fitted here? This needs clarification.*

Response: Note the λ_i does not represente the expectation, nor approximately. The quadratic erros in Figure 1(a) is obtained by $[a\text{E(Y)} - \sum_{y=0}^{500} y p(y)]^2$. The models fitted in the case studies are $[\text{COMPo}(\lambda_i, \nu), \log(\lambda_i) = \beta^\top x_i]$ and $[\text{COMPo}(\mu_i, \phi), \log(\mu_i) = \beta^\top x_i]$. Therefore, the β estimates are only comparable if $\phi = 0$ (Poisson special case) otherwise the β estimates are on different scales. We make it clear which parameters are being placed the linear predictor in the current version of the paper. To understand the relationship of the coefficients, consider $\log(\lambda_i) = \beta^o$ and $\log(\mu_i) = \beta^*$, so β^* expressed in terms of β^o is $\beta^* = \exp(\beta^o)\exp -\phi - (\exp(\phi) - 1)/(2\exp(\phi))$.

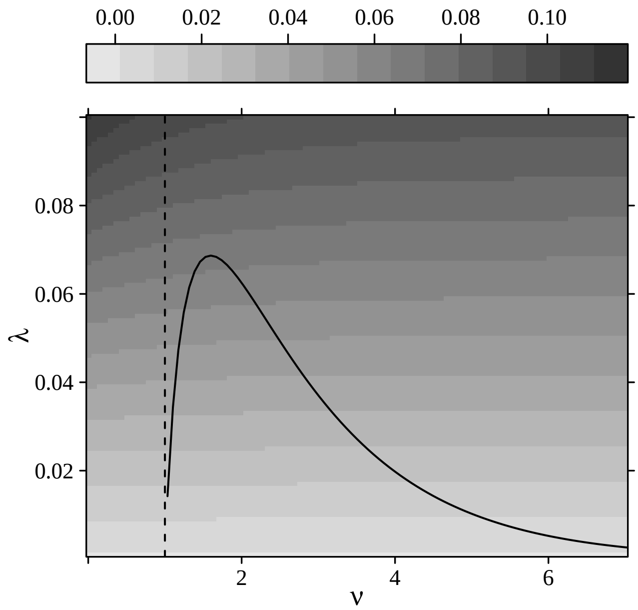
Specific reviewer's comments:

1. *Consider the constraint "$\mu > 0$", after equation (3.2). What constraints does this impose on the λ and ν. Is this reasonable?*

   Response: The constraint $\mu > 0$ implies that $\lambda > [(\nu - 1)/(2\nu)]^\nu$. The Figure below show (a) the constraint parameter space; and (b) the expectated values (obtained by $\sum yp(y)$). The constraint lies on the very small λ and $\nu > 1$ (underdispersion). This parameter region is related to small expected values (smaller than 0.1) and underdispersion. Therefore, although this constraint is undesirable, it does not prejudice the application of the model.

   (a)

   (b)

   We thank the referee for highlighting this point. We have discussed it in the updated version.

2. *The Shmueli approximations to the mean and variance hold under certain conditions. Are those conditions met in practice? For your examples?*

   Response: We used only the mean approximation given by Shmueli to reparametrize the COM-Poisson model. Moreover, as discussed in Figure 1(a), the the errors are close to 0 for the parameter grid evaluated and present no clear relation with regions gives by Shmueli et al. (2005) ($\phi \leq 0$ and $\mu > 10 - (\exp(\phi) - 1)/(2\exp(\phi))$. Except for counts with small averages (close to 0), the reparametrized model fits well. For examples presented in the article, the parameter estimates are summarizes below.

		$\hat{\lambda}_i$		
Case study	$\hat{\phi}$	Minimun	Median	Maximum
Cotton experiment	1.58	3.50	8.44	9.48
Nitrofen experiment	0.05	5.94	28.00	32.36
Soybean experiment	−0.78	116.87	174.98	252.29

3. *The symbol μ is used to represent both $E(Y)$ and the parameter $\lambda^{1/\nu} - (\nu - 1)/(2\nu) = aE(Y)$. This causes confusion. Is there a better alternative? As an example, in Simulation study of Section 5, at the bottom of p 14, does the symbol μ represent $E(Y)$ or $aE(Y)$? I assume the former.*

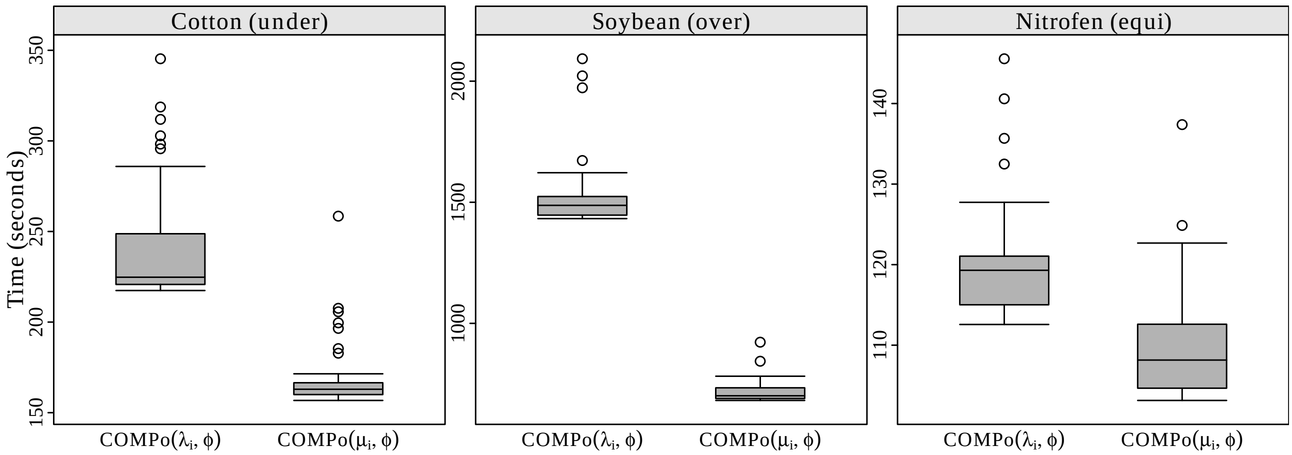
   Response: Thank you for advertising this misunderstanding. At the bottom of page 14 the $\boldsymbol{\mu}$ is $a\mathrm{E}(\boldsymbol{Y})$. We keep symbol μ to the introduced parameter in order to highlight that it is related to the expectation. However, we make clear that $\mu_i = a\mathrm{E}(Y_i)$ to avoid misunderstanding.

4. *The abstract could be improved by making it more concise (shorter). It should be written from the third-person perspective as well.*

   Response:

5. *I assume that the "38% faster" algorithm means that, e.g., instead of 10 seconds it takes just 6.2 seconds. As this is not an order of magnitude, this may not be enough to warrant a change to the "simpler" parameterization given that this latter parameterization models $aE(Y_i)$ rather than $E(Y_i)$.*

   Response: The time to fit the COM-Poisson(μ_i, ϕ) model was compared to the time to fit the COM-Poisson(λ_i, ϕ) model. The computational times for 50 repetitions of fit of the models in the three case studies are presented in Figure below.



The times to adjust under the proposed parameterization are 110% faster than the original, in the overdispersed case. In addition, the nice properties induced by the new parameterization like the orthogonality between μ and ϕ and interpretation of μ are the main advantages to warrant the use of the new parameterization besides computational times.

References

Dunn, J. (2012), *compoisson: Conway-Maxwell-Poisson Distribution*. R package version 0.3. **URL:** *https://CRAN.R-project.org/package=compoisson*

Ross, G. J. S. (1970), 'The efficient use of function minimization in non-linear maximum-likelihood estimation', *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **19**(3), 205–221.

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', *Journal of the Royal Statistical Society. Series C: Applied Statistics* **54**(1), 127–142.