Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/110318

A New Proposed Model for Handling Longitudinal Count Data with Over-Dispersion and Extra Zero Problems

Amany M. Mousa[1], *Ahmed A. El sheikh*[1], *Ahmed M. Gad*[2] *and Fatma El Zahraa S. Salama*[1,*]

[1]Faculty of Graduate Studies for Statistical Research, Cairo University, Cairo, Egypt
[2]Faculty of Economics and Political Science, Cairo University, Cairo, Egypt

Received: 15 Jul. 2021, Revised: 20 Aug. 2021, Accepted: 28 Sep. 2021.
Published online: 1 Sep. 2022.

Abstract: The over-dispersion and the excess zeros are two main problems that are related to the Poisson regression model. To handle these problems, different models like zero inflated negative binomial and zero inflated generalized Poisson were proposed. Lately Conway-Maxwell Poisson (COM-Poisson) had proposed to handle the over-dispersion for count data in cross-sectional case. However, there is no application of the COM-Poisson model in longitudinal case. In this paper, the zero inflated COM-Poisson model was proposed and developed to deal with longitudinal count data. Under two different working correlation structures, exchangeable and autoregressive of order 1, AR(1). The zero-inflated COM-Poisson (ZICP) regression model is considered as modification of the COM-Poisson regression model that allows for an excess of zero counts in the data and over-dispersion problem. this model was compared with the Zero Inflated Poisson regression model and zero inflated negative binomial model. The results show that the COM-Poisson model is very suitable to longitudinal count data, even in presence of dispersion. It gives the smallest AIC values. Also, it is insensitive to the choice of the working structure.

Keywords: COM-Poisson regression model, Count data, Excess of zeros problem, Longitudinal count data, Zero Inflated COM Poisson model, Over-dispersion problem.

1 Introduction

The use of statistical models for the analysis of correlated count data has grown in public health research. Analysis of this data affected by one or more explanatory variables that requires adequate modeling of the correlations underlying the responses repeatedly collected over time. Poisson regression model is often applied for analysis of count data by assuming that the marginal mean and variance of the data are equal but this assumption in most cases is not valid so the Poisson regression might be too obstructive due to this assumption. When the over-dispersion exists in data, the Poisson model might obstruct the estimation efficiency in practice, negative binomial regressions will be the alternative model to handle this problem.

Analysis of longitudinal count data depends on one or more explanatory variables that require adequate modeling of the correlations underlying the responses repeatedly collected over time. Many authors have suggested likelihood-based models for longitudinal count data with over-dispersion. The over-dispersion parameter is modeled using a random mean approach [1],[2],[3]. However, these approaches require the specification of full likelihood function. On the other hand, non-likelihood approaches have been proposed for longitudinal count data such as the generalized estimating equations (GEE). The GEE has been introduced [4] and extended [5],[6][7]. The GEE and its extensions are commonly used under a marginal mean regression framework. The generalized estimating equations require the first two moments and estimation of a few nuisance parameters associated with a working correlation structure. This yields an asymptotically consistent estimator regardless of the working correlation structure. However, it can be inefficient under mis-specified working correlation structure.

To overcome the problem of miss–specified the working correlation structure, quadratic inference functions was developed by extending the inverse of the working correlation structure to a set of predetermined matrices, which avoids additional estimation of the nuisance parameters. The quadratic inference functions results in more efficient estimators than the one ignoring the within-subject correlation, even if the assumed working correlation structure is incorrect. Moreover, quadratic inference functions provide an inference function for model diagnostic tests and goodness of fit tests[8]. This method can handle two different types of working correlation structures: the exchangeable structure and the

*Corresponding author e-mail: fatmafebs90@cu.edu.eg.

autoregressive AR(1) structure. The exchangeable structure means that correlations between different observations on the same subject are identical, no matter how close they are in a time sequence. On the other hand, the autoregressive AR(1) structure means that correlations between observations closer together, in a time sequence, are more correlated than observations further apart.

According to the definition of over-dispersion [9] is the phenomenon that a rise when the empirical variance for the data exceeds the nominal variance under some presumed model. If the over-dispersion is present for count data, the Poisson regression model is not valid for estimation. This problem arises usually through the omission from the regression models of important explanatory variables, existence of outliers, and using inappropriate link function. This violation of the standard model assumptions inflates the residual deviance, produces large residuals, underestimates standard deviations of the estimated parameters, and may lead to biases in the estimates themselves if the omitted variables are correlated with those in the model.

Definition of excess zero in data is exist when often a greater number of zero observations than expected from the Poisson model display in the data, this problem is known as excess zeros [10]. In health research, count outcomes are common and often these counts have a large number of zeros. For such counts, the Poisson regression model is commonly used to explain the relationship between the outcome variable and a set of explanatory variables. However, it is often the case that there are a higher proportion of zero counts than would be predicted by the Poisson distribution, possibly due to a distinct sub- population of subjects whose only response is zero counts.

In this paper, a zero inflated COM-Poisson model was proposed and developed to handle the two main problems (the over-dispersion and the excess zeros) in the longitudinal case using two different working correlation (exchangeable and autoregressive ar(1)). Up to now there is no application of this model, in literature, to longitudinal count data. This model enables us to deal with the two main problems of count responses: Second, to suggest using the quadratic inference function under the two different types of working correlation structure; the exchangeable structure and the AR(1) structure. The proposed methods are applied to a real dataset, the epileptic data. The obtained results will be compared with the most common models; zero inflated Poisson regression model and zero negative binomial regression model.

2 COM-Poisson Regression Model for Longitudinal data

Due to the convincing properties of the COM Poisson regression model, it has been widely used to describe under- or over- dispersed count data in cross-sectional set up. However, there exists no application of COM-Poisson model in longitudinal case. The challenge lies in modeling the correlation structure of repeated COM-Poisson counts. The framework was provided [11]for developing and analyzing a COM-Poisson model with AR (1)-type correlation structure under the longitudinal set-up. An extension of the cross-sectional Conway-Maxwell Poisson was presented [12] (COM-Poisson) regression model established [13] a generalized regression model for count data in light of inherent data dispersion – to incorporate random affects for analysis of longitudinal count data. In this paper, we will extend the Com Poisson model for the marginal model. Let Y_{it} be the count response of the i th individual at time t (i= 1, 2, . . . , n ; t= 1, 2, . . . , T). Let X_{it} be the p dimensional vector of covariates corresponding to Y_{it}. Let β is the p dimensional vector of regression parameters. The COM-Poisson regression model can be formulated to use for longitudinal data by replacing y_t with y_{it} and λ_t by λ_{it} in equation (1) where:

$$\lambda_{it}=exp(x'_{it}\beta)$$

and

$$f(y_{it}) = \frac{\lambda_{it}{}^{y_{it}}}{(y_{it}!)^{v} Z(\lambda_{it},v)} \quad , \; y_{it} = 0,1,2, \dots \dots \tag{1}$$

where

$$Z(\lambda_{it},v) \simeq \frac{exp\left(x'_{it}\beta\left(\frac{v-1}{2v}\right)\right)(2\pi)^{\frac{v-1}{2v}}\sqrt{v}}{exp\left(v\, exp\left(\frac{x'_{it}\beta}{v}\right)\right)}.$$

parameter v is the dispersion parameter such that $v=1$, $v<1$ and $v>1$ correspond to equi, over and under-dispersion respectively. The variance and the rth moments can be extracted depend on the above approximation of $Z(\lambda_{it},v)$ so they can take the following formula as follow:

$$E(Y_{it}) \simeq \lambda_{it}^{\frac{1}{v}} - \frac{v-1}{2v} \quad , \qquad var(Y_{it}) \simeq \frac{\lambda_{it}^{\frac{1}{v}}}{v} \tag{2}$$

and

$$E(Y_{it}^{r}) \simeq \lambda_{it}\frac{d}{d\lambda}E(Y_{it}^{r-1}) + E(Y_{it})E(Y_{it}^{r-1}) \quad (3)$$

Since longitudinal count data are correlated, it makes difficult to specify the full likelihood function. The numerical methods are more efficient to find the estimate for this model.

3 The New Proposed Model

In this part, idea to find the estimate of the new model[14]will be extended in this paper. The overall probability of zero counts is the combined probability of zeros from the two processes; these two parts of a zero-inflated model are regarded as a binary model (usually a logit model) and a count model, where the probability of the COM-Poisson can be described as follow to get the new model "Zero Inflated-COM Poisson (ZICP) Model":

$$Y_{it}^{*} \sim \begin{cases} 0 & \text{with probability } \pi_{it} \\ \text{COM} - \text{Poisson} & \text{with probability } (1-\pi_{it}) \end{cases} \quad (4)$$

where Y_{it}^{*} is the outcome of the zero inflated Poisson distribution and π_{it} is the probability of zero (structural) counts from the binary process; in any dataset, whenever $\pi_{it} > 0$. This formulation incorporates more zeros than permitted under the COM-Poisson model. Depending on the probability of probability density function of COM Poisson model, the probability density function of the longitudinal Zero Inflated-COM Poisson Model (ZICP) model can be written [15]:

$$\Pr(Y_{it}^{*} = y_{it}^{*}) = \begin{cases} \pi_{it} + \frac{(1-\pi_{it})}{Z(\lambda_{it},v)} & \text{if } y_{it}^{*} = 0 \\ (1-\pi_{it})\frac{\lambda_{it}^{y_{it}^{*}}}{(y_{it}^{*}!)^{v}Z(\lambda_{it},v)} & \text{if } y_{it}^{*} > 0 \end{cases} \quad (5)$$

3.1 Assumptions of Zero Inflated-COM Poisson Model (ZICP)

- The model assumes that counts are rather generated by two processes. The first process generates zero counts with probability π_{it} while the non-zero counts follow the COM-Poisson distribution with parameter λ and are realized with probability (1- π_{it}).
- Zero counts are generated from two sources based on the probabilities of the two processes. These two parts of a zero-inflated model are regarded as a binary model (usually a logit model) and a count model (in this case COM-Poisson distribution).

Depending on equation (5), the expected value and the variance for the zero inflated COM Poisson model (ZICP) are expressed as:

$$E(Y_{it}^{*}) \simeq (1-\pi_{it})\left[\frac{\lambda_{it}^{\frac{1}{v}}}{v} - \frac{v-1}{2v}\right], \quad Var(Y_{it}^{*}) \simeq \frac{\pi_{it}(E(Y_{it}))^{2}}{(1-\pi_{it})} + (1-\pi_{it}).\frac{\lambda_{it}^{\frac{1}{v}}}{v} \quad (6)$$

3.2 Estimates The New Proposed Model Using Quadratic Inference Functions:

A method of quadratic inference functions (QIF)was introduced [6] that does not involve direct estimation of the correlation parameter and remains optimal even if the working correlation structure is misspecified. The idea is to represent the inverse of the working correlation matrix by the linear combination of basis matrices; a representation that is valid for the most commonly used working correlations. To find the joint estimation of β and v, we develop the joint estimating equations by employing the first two moments of the response variable as:

$$\sum_{i=1}^{n}\frac{\partial\mu_i'}{\partial£}A_i^{-\frac{1}{2}}R(\alpha)_i^{-1}A_i^{-\frac{1}{2}}(f_i - \mu_i) = 0, \quad (7)$$

where $\mu_i = \{E(Y_{i1}), \dots, E(Y_{iT}), E(Y_{i1}^{2}), \dots, E(Y_{iT}^{2})\}'$ are 2t −dimensional vectors and

$E(Y_{it}^{*}) = (1-\pi_{it}).\left[\frac{\lambda_{it}^{\frac{1}{v}}}{v} - \frac{v-1}{2v}\right]$, $E(Y_{it}^{*2}) = (1-\pi_{it})\left[(\frac{\lambda_{it}^{\frac{1}{v}}}{v} - \frac{v-1}{2v})^2 + \frac{\lambda_{it}^{\frac{1}{v}}}{v}\right]$, $£ = (\beta, v)$, $\frac{\partial \mu_i'}{\partial £}$ is(2t) ×(p+1) matrix which can be obtained as in the following lemma , A_i is a $(2t) \times (2t)$ diagonal variance matrix of f_i with

$$var(Y_{it}^{*}) = \frac{\pi_{it}.(E(Y_{it}))^2}{(1-\pi_{it})} + (1-\pi_{it}).\frac{\lambda_{it}^{\frac{1}{v}}}{v}$$

and

$$var(Y_{it}^{*2}) = (1-\pi_{it}).\ \frac{\lambda_{ij}^{\frac{1}{v}}v^2 + 4\lambda_{ij}^{\frac{3}{v}}v^2 + 10\lambda_{ij}^{\frac{2}{v}}v - 4\lambda_{ij}^{\frac{1}{v}}v + 4\lambda_{ij}^{\frac{1}{v}} - 4\lambda_{ij}^{\frac{2}{v}}v^2}{v^3}.$$

Lemma:

Since $E(Y_{it}^{*}) = \theta_{it} = (1-\pi_{it})\left[\frac{\lambda_{it}^{\frac{1}{v}}}{v} - \frac{v-1}{2v}\right]$ and $E(Y_{it}^{*2}) = m_{it} = (1-\pi_{it})\left[(\frac{\lambda_{it}^{\frac{1}{v}}}{v} - \frac{v-1}{2v})^2 + \frac{\lambda_{it}^{\frac{1}{v}}}{v}\right]$, then we have

$$\frac{\partial \theta_{it}}{\partial \beta} = (\,(1-\pi_{it})\,\frac{\lambda_{it}^{\frac{1}{v}} x_{it}^{T}}{v^2} = b, \qquad \frac{\partial \theta_{it}}{\partial v} = (1-\pi_{it}).\frac{(-\lambda_{it}^{\frac{1}{v}} x_{it}^{T}\beta - v\lambda_{it}^{\frac{1}{v}} - 0.5v)}{v^3} = S$$

$$\frac{\partial m_{it}}{\partial \beta} = \left((2\lambda_{it}^{\frac{1}{v}} + 1)/v\right) * b \quad , \qquad \frac{\partial m_{it}}{\partial v} = (2\lambda_{it}^{\frac{1}{v}} + 1)/v) * s + \frac{0.5}{v^2}.(1-\pi_{it}).$$

To obtain the estimators of β and v, specification of the working correlation structure and estimation of α in R(α) are required prior to solving Equation (7). To handle this problem, a class of basis matrices was employed to represent the inverse of $R(\alpha)_i^{-1}$. The QIF is derived by observing that the inverse of the working correlation matrix can be approximated by a linear combination of several basis matrices.

$$R^{-1} \approx \sum_{l=0}^{k} a_l M_l, \tag{8}$$

where M_0 is the identity matrix M_1 ,…, M_k are known basis matrices with 0 or 1 as components and $a_{0,\dots,}a_k$ are unknown coefficients.

3.3 The Basis for The Inverse of The Correlation Matrix

The choice of basis matrices was discussed, if the working correlation structure is exchangeable, then $R^{-1} = \alpha_0 M_0 + \alpha_1 M_1$ where M_1 is 0 on the diagonal and 1 elsewhere. If the working correlation structure is an AR(1) working correlation $R^{-1} = \alpha_0 M_0^{*} + \alpha_1 M_1^{*} + \alpha_2 M_2^{*}$ where $M_0^{*} = I,$ the identity matrix , M_1^{*} has 1 on the sub-diagonal and 0 elsewhere and M_2^{*} has 1 on the two corner components of the diagonal. The advantage of this approach is that it does not require estimation of linear coefficients $a_j's$ which can be viewed as nuisance parameters, since the generalized estimating equation is a linear combination of elements of the estimating functions. Depending on the above linear combination of the elements of the following extended score vector: to find the estimate of β and v we depend on minimizing the function

$$Q_n(\beta, v) = n\ \bar{g}_n(\beta, v) C_n^{-1}(\beta, v)\ \bar{g}_n(\beta, v), \tag{9}$$

Where

$$\bar{g}_n(\beta, v) = \frac{1}{n}\sum_{i=1}^{n} g_i(\beta, v),$$

$$g_i(\beta, v) = \begin{pmatrix} \frac{\partial \mu_i^T}{\partial(\beta, v)} A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}}(f_i - \mu_i) \\ \vdots \\ \frac{\partial \mu_i^T}{\partial(\beta, v)} A_i^{-\frac{1}{2}} M_m A_i^{-\frac{1}{2}}(f_i - \mu_i) \end{pmatrix}, \tag{10}$$

and

$$C_n^{-1}(\beta, v) = \frac{1}{n} \sum_{i=1}^{n} g_i(\beta, v)\, g_i'(\beta, v).$$

The QIF estimator $(\hat{\beta}, \hat{v})$=$argmin\, Q_n(\beta, v)$ and the asymptotic covariance matrix of the estimates $(\beta\ , v)$ can be obtained using equation:

$$\hat{\Sigma} = \left[\left\{\frac{1}{n} \sum_{i=1}^{n} \frac{\partial g_i(\hat{\beta}, \hat{v})}{\partial(\beta, v)}\right\}^T \left\{\frac{1}{n} \sum_{i=1}^{n} g_i(\hat{\beta}\ , \hat{v})\, g_i(\hat{\beta}\ , \hat{v})^T\right\}^{-1} \left\{\frac{1}{n} \sum_{i=1}^{n} \frac{\partial g_i(\hat{\beta}, \hat{v})}{\partial(\beta, v)}\right\}\right]^{-1}$$

4 Applications

The proposed methods are applied to the epileptic data. These data arising from a clinical trial carried. These data consist of 59 epileptics of them 31 patients receive the anti-epileptic drug progabide (treatment group) and the remaining 28 patients receive a placebo (placebo group). The number of epileptic seizures, for each patient, were recorded during the baseline of a period of 8 weeks. Patients were then randomized to one of the two treatments: the anti-epileptic drug progabide or the placebo, in addition to a standard chemotherapy. The number of seizures were then recorded at 4 consecutive two-weeks intervals. Many characteristics of each patient were recorded. These include Y_{it} the number of seizures of patient i at the time point t, "*Treatment*" where 0 for the placebo and 1 for progabide, "*Period*" denotes the measurement period (1,2, …, 8), "*Base*" represents the baseline measurement for each patient, and "*Age*" represents the age of the patient in years.

These data were originally analyzed [16] [17]as a complete data. The data often used as an example for longitudinal data of discrete outcomes, e.g. PROC GENMOD in SAS/STAT User's Guide [18]. Also, these data used as incomplete data. Descriptive statistics of the data including the mean, the standard deviation, the minimum and maximum values and the skewness are presented in Table 1. The results show positive skew in the number of seizures at baseline and during the treatment period. Hence, we need to employ models for count data that take into consideration this feature. The results show that the mean of the number of seizures (8.26) is less than its standard deviation (12.36). This means that over-dispersion is more likely in these data.

Table 1: Descriptive Statistics for the Study Variables.

Variable	Mean	Std. Dev.	Skewness	Min	Max
Y_{ij}	8.26	12.36	4.2	0	102
Treatment	0.53	-	-	0	1
Base	31.22	26.70	2.2	6	151
Age	28.34	6.26	0.3	18	42

Figure1 and figure 2 present the profiles plots of the number of seizures Y_{it} for placebo group and treatment group respectively. From Figure 1 that the placebo group has greater variability between patients. Also, from figure 2 we can see there is an outlier subject in the treatment group (progabide). Also, it can be seen that there are some outliers in the treatment group more than the placebo group, also it can be noted that the change in the group that take placebo are restively stable over the course of the trail while for the progabide group there is a decline in the number of observed seizures. Finding outliers in this data will make the two main problems (over-dispersion and excess zero) exist.

4.1 Over-dispersion Test

Over-dispersion can be checked using a formal test for the epileptic data which include number of explanatory variables (treatment, period, base and age). The null and alternative hypotheses are:

$\mathrm{H_0}$: Poisson regression is suitable

$\mathrm{H_A}$: Com − Poisson model is suitable.

a tests statistic to conduct this test based on the quantity $Q_n(\hat{\beta}, \hat{v})$. The test statistic T_n is defined as

T_n=$n\{\, Q_n(\hat{\beta}|v = 0) - Q_n(\hat{\beta}, \hat{v})\}$.

It is known that the Poisson regression is recommended when the over-dispersion parameter v is zero. The test statistic T_n has asymptoticly chi–square distribution with 1 degrees of freedom, i.e.

$T_n \xrightarrow{d} x^2_{(1)}$ Under H_0.

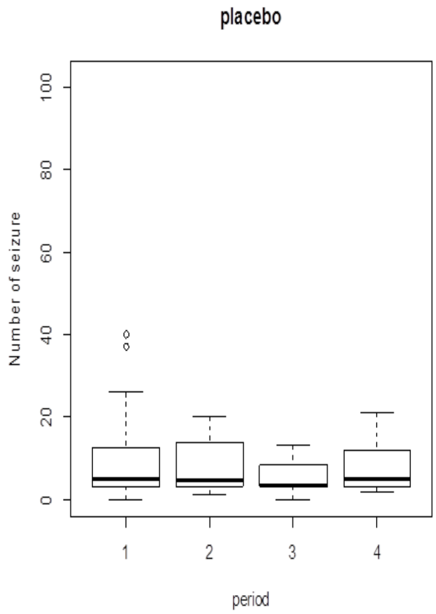

Fig.1: Boxplot of Seizures for Placebo.

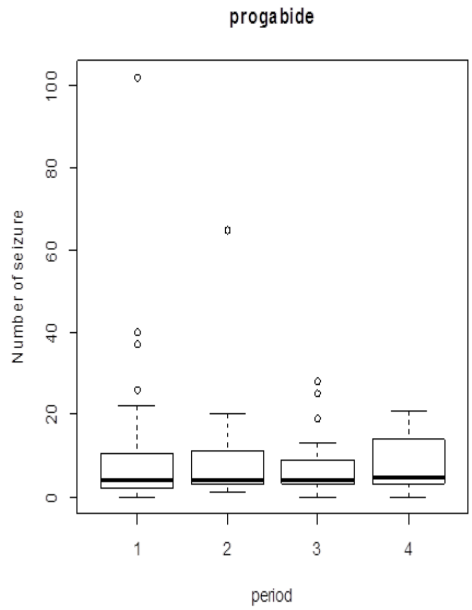

Fig.2: Boxplot of Seizures for Treatment.

The test has been conducted and the calculated test statistics T_n is 9.147. The calculated test is large enough that the tabulated chi-square values at any reasonable level of significance. Hence, we reject the null hypothesis. This means there is significant evidence of over-dispersion. This conclusion suggests that the COM-Poisson model is a suitable than Poisson model to fit the data. This model could be more plausible to these data.

The data are fitted using three models: zero inflated COM-Poisson model, zero inflated negative binomial model and zero inflated Poisson model. The Poisson model is fitted for the sake of comparison with the other two models. The three models are used with two different types of the working correlation structure; the exchangeable structure and the AR(1) structure. The Akaike Information Criteria (AIC)[19] as a goodness of fit statistic. The AIC penalized a model with larger number of parameters and is defined as

$$AIC= -2\ ln(L)+ 2p, \quad (12)$$

where ln(L) is the fitted log-likelihood value and p is the number of predictors (parameters) including the intercept. A relatively small value of AIC is favorable for the fitted model.

The results of the fitted three models under exchangeable structure and AR(1) structure are presented in Table (a.1) and Table (a.2) (see appendix). Table (a.2) show the results of the fitted three models under exchangeable structure. The results in Table (a.2) show that the AIC value of the COM-Poisson model is the smallest. This means that the COM-Poisson model is more appropriate to fit these data. Also, the COM-Poisson model has the smallest stander errors and narrower confidence intervals for all parameters. This means that this model is more efficient than the others. The estimates of the dispersion parameter ν is $\hat{\nu}$=0.4615 and $\hat{\nu}$=0.4938 for the negative binomial and COM-Poisson model, respectively This is evidence of over-dispersion. Both the negative binomial and the COM-Poisson models can account for dispersion beyond that induced by within-subject correlation.
The results in Table (a.2) show higher AIC values comparable to the exchangeable working structure which means that the exchangeable working correlation is more suitable. The AIC value of zero inflated COM-Poisson model still the smallest. This means that the zero inflated COM-Poisson model still the most appropriate model. Over-dispersion still obvious as we have positive dispersion parameter ν. The results of the zero inflated COM-Poisson model under the AR(1) structure are similar to those of the exchangeable structure. Hence, the zero inflated COM-Poisson model is not sensitive to the choice of the working structure.

5 Conclusions

There are some considerations that should be considered when we want to analysis of count data. These considerations should be based on several factors including the features of the available data. One of these features is over-dispersion. The most common model that used to handle over-dispersed data is negative binomial model. The COM-Poisson model emerged in the recent years as a candidate to handle the problem of dispersion also the zero inflated Com Poisson is emerged to handle the problem of extra zeros. It is known that the zero inflated COM-Poisson model is a flexible model for count data in light of data dispersion and extra zeros. In this article, the cross-sectional zero inflated COM-Poisson model was proposed to the case of longitudinal count data context using the method of quadratic inference function. This model was applied, in addition to the zero inflated Poisson model and the zero inflated negative binomial model, to the epileptic data. The results show that the quadratic inference functions can obtain a more efficient estimator than the one that ignores the within-subject correlation, so this method was preferred to use in fitting the COM-Poisson model. The proposed approach yields more efficient results, under the two different working correlation structures, comparable to the two common models for longitudinal count data: the zero inflated Poisson regression model and the zero inflated negative binomial. This is according to AIC values. The results are more similar for the data when we apply the COM-Poisson under the exchangeable working correlation structure and auto regressive AR (1) working correlation structure so the proposed approach even not sensitive to the choice of the working structure.

Acknowledgement

The authors are grateful to the anonymous referee for the careful checking of the details and the constructive comments that improved this paper.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this article.

References

[1] P. J. Toscas, J. M. Faddy, Likelihood-based analysis of longitudinal count data using a generalized Poisson model, *Statistical Modeling: An International Journal.*, **3(2)**, 99–108(2003).

[2] A. R. Fotouhi, Modeling over-dispersion in longitudinal count data in clinical trials with application to epileptic data, *Contemporary Clinical Trials.*, **29 (4)**, 547–554(2008).

[3] G. K. Grunwald, S., Bruce, L.Jiang, M. Strand and N.Rabinovitch, A statistical model for under- or over-dispersed clustered and longitudinal count data, *Biometrical Journal.*, **53 (4)**, 578–594(2011).

[4] K. Y. Liang and S. L. Zeger, Longitudinal data analysis using generalized linear models, *Biometrika.*, **73 (1)**, 13–22(1986).

[5] D. B. Hall and T. A. Severini, Extended generalized estimating equations for clustered data, *Journal of the American Statistical Association.*, **93 (4)**, 1365–1375(1998).

[6] A. B. Qu, G. Lindsay and B. Li, Improving generalised estimating equations using quadratic inference functions, *Biometrika* , **87(4)**, 823–836(2000).

[7] B. Jorgensen and S. J. Knudsen, Parameter orthogonality and bias adjustment for estimating functions, *Scandinavian Journal of Statistics.*, **31(1)**, 93–114(2004).

[8] N. Xiaomeng and R. C. Hyunkeun, Efficient regression modeling for correlated and over-dispersed count data, *Communications in Statistics - Theory and Methods.*, **48**, 6005-6018(2019).

[9] C. B. Dean and E. R. Lundy, *Over-dispersion Statistics*, Wiley, New Jersey (2016).

[10] A. Cameron and P. Trivedi, Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Test, *Journal of Applied Econometrics.*, **1**, 29-53(1986).

[11] V. Jowaheer and N. A. Mamodekhan, Estimating Regression Effects in COM-Poisson Generalized Linear Model, *International Journal of Mathematical, Computational, Physical, Electrical and Computer Engineering.*, **3**.,362-365(2009).

[12] D. S. Morris, K. F. Sellers and A. Menger, Fitting a Flexible Model for Longitudinal Count Data Using the NLMIXED Procedure, SAS Global Forum Proceedings , Cary, NC: SAS Institute.,1–6 (2017).

[13] K. Sellers and G. Shmueli, A Flexible Regression Model for Count Data., *Annals of Applied Statistics.*, **4**, 943-961(2010).

[14] B. Y. Oyindamola, F. A. Rotimi and S. A. Ayoola, Modelling Excess Zeros in Count Data with Application to Antenatal Care Utilisation , *International Journal of Statistics and Probability.*,**7(3)** ,1927-1940(2018).

[15] F. S. Salama, A. M. Gad and A. Mousa, Modeling Longitudinal Count Data with Missing Values: A Comparative Study, *Academic Journal of Applied Mathematical Sciences*, Academic Research Publishing Group., **2**, 19-26(2016).

[16] P. F. Thall and S. C.Vail, Some Covariance Models for Longitudinal Count Data with Over dispersion, *Biometrics.*, **46**, 657-671(1990).

[17] N. E. Breslow and D.G. Clayton, Approximate Inference in Generalized Linear Mixed Models, *Journals of the American Statistical Association.*, **88**, 9-25(1993).
[18] N. C. Cary, *SAS/STAT® 14.1 User's Guide*, SAS Institute Inc, United States of America(2015).
[19] J.M. Hilbe and J.W. Hardin, Modeling Count Data with Generalized Distribution, *stata Journal.*, **14**, 562–579(2014).

Appendix (a)

Table (a.1): Estimates of zero Inflated Poisson Model, Zero Inflated Negative Binomial Model and Zero Inflated COM-Poisson Model under Exchangeable Working structure.

Quadratic inference function under exchangeable working structure						
Zero inflated Poisson model						
	Intercept	Period	Trt	Base	Age	v
Estimate	0.724	-0.0225	-0.142	0.0214	0.016	-
Standard error	0.428	0.011	0.185	0.0023	0.0177	-
CI	(-0.115, 1.562)	(-0.044, -.00094)	(-0.505, 0.2206)	(0.017, 0.0259)	(-0.0186,0.0506)	-
AIC	1541.33					
Zero inflated negative binomial model						
	Intercept	Period	Trt	Base	Age	v
Estimate	0.574	-0.056	-0.041	0.0364	0.024	0.45
Standard error	0.1886	0.047	0.154	0.0045	0.037	0.25
CI	(0.204, 0.944)	(-0.148, 0.036)	(-0..342, 0.261)	(0.027,0.0452)	(-0.0485, 0.0965)	(-0.04, 0.94)
AIC	1323.1					
Zero inflated COM-Poisson model						
	Intercept	Period	Trt	Base	Age	v
Estimate	0.647	-0.162	-0.0652	0.012	0.001	0.49
Standard error	3.358e-02	2.201e-02	1.415e-03	3.51e-04	1.025e-02	0.001
CI	(0.581, 0.713)	(-0.201, -0.1226)	(-0.0679, -.062)	(0.0113, 0.0126)	(-0.019, 0.0211)	(0.488, 0.491)
AIC	1124.52					

Abbreviation: CI is 95% confidence interval

Table (a.2): Estimates of zero Inflated Poisson Model, Zero Inflated Negative Binomial Model and Zero Inflated COM-Poisson Model under under AR(1) Working Structure.

Quadratic inference function under AR(1) working structure						
Zero Inflated Poisson model						
	Intercept	Period	Trt	Base	Age	v
Estimate	0.841	-0.0513	-0.233	0.021	0.019	-
Standard error	0.584	0.0287	0.0184	0.024	0.0115	-

CI	(-0.304, 1.985)	(-0.1075, 0.0049)	(-0.269, -0.196)	(0.026, 0.068)	(-0.003,0.0415)	-
AIC	1635.04					
Zero Inflated negative binomial model						
	Intercept	Period	Trt	Base	Age	v
Estimate	0.5986	-0.0821	-0.287	0.0254	0.023	0.31
Standard error	0.3702	0.0319	0.01793	0.0016	0.011	0.0896
CI	(-0.127, 1.324)	(-0.145, -0.019)	(-0.322, -0.2518)	(0.022,0.0285)	(0.001, 0.044)	(0.134, 0.485)
AIC	1503.2					
Zero inflated COM-Poisson model						
	Intercept	Period	Trt	Base	Age	v
Estimate	0.051	- 0.125	- 0.074	0.011	0.025	0.458
Standard error	0.0215	0.0014	0.0512	0.0047	0.0145	0.0258
CI	(0.0088, 0.0931)	(-0.127, -0.122)	(-0.174, 0.0263)	(0.0017, 0.0202)	(-0.003, 0.053)	(0.407, 0.5085)
AIC	1357.88					

Abbreviation: CI is 95% confidence interval

Appendix (b)

The derivative for the lemma (Zero inflated COM-Poisson regression model)

$$\theta_{it} = (1-\pi_{it})\left[\frac{\lambda_{it}^{\frac{1}{v}}}{v} - \frac{v-1}{2v}\right]$$

since

$$\lambda_{it} = e^{x_{it}^T\beta}$$

then

$$\frac{\partial\theta_{it}}{\partial\beta} = (1-\pi_{it}).\frac{1}{v}\left(\frac{\lambda_{it}^{\frac{1}{v}}x_{it}^T}{v}\right) - zero \ = (1-\pi_{it}).\ \frac{\lambda_{it}^{\frac{1}{v}}x_{it}^T}{v^2}.$$

Let

$$(1-\pi_{it})\frac{\lambda_{it}^{\frac{1}{v}}x_{it}^T}{v^2} = \mathrm{b}$$

Then

$$\frac{\partial\theta_{it}}{\partial\beta} = \mathrm{b}$$

$$\frac{\partial\theta_{it}}{\partial v} = (1-\pi_{it}).\left(\frac{exp\ (x_{it}^T\beta/v)\left(-\frac{x_{it}^T\beta}{v^2}\right).v - \lambda_{it}^{\frac{1}{v}}}{v^2} - \frac{2v-2(v-1)}{4v^2}\right)$$

$$=(1-\pi_{it}).\left(-\frac{\lambda_{it}^{\frac{1}{v}}x_{it}^{T}\beta}{v^{3}}-\frac{\lambda_{it}^{\frac{1}{v}}}{v^{2}}-\frac{1}{2v^{2}}\right)$$

$$=(1-\pi_{it}).\frac{(-\lambda_{it}^{\frac{1}{v}}x_{it}^{T}\beta-v\lambda_{it}^{\frac{1}{v}}-0.5v)}{v^{3}}=S.$$

Since

$$E(\mathrm{Y}_{\mathrm{it}}^{*2})=m_{it}=(1-\pi_{it})\left[\left(\frac{\lambda_{it}^{\frac{1}{v}}}{v}-\frac{v-1}{2v}\right)^{2}+\frac{\lambda_{it}^{\frac{1}{v}}}{v}\right]$$

$$\frac{\partial m_{it}}{\partial\beta}=2.\ (1-\pi_{it})\left(\frac{\lambda_{it}^{\frac{1}{v}}}{v}-\frac{v-1}{2v}\right)*\frac{\lambda_{it}^{\frac{1}{v}}x_{it}^{T}}{v^{2}}+(1-\pi_{it})\frac{\lambda_{it}^{\frac{1}{v}}x_{it}^{T}}{v^{2}}$$

$$\frac{\partial m_{it}}{\partial\beta}=\left(\frac{2\lambda_{it}^{\frac{1}{v}}-v+1}{v}+1\right)*b$$

$$=\left((2\lambda_{it}^{\frac{1}{v}}+1)/v\right)*b$$

Let

$$\left((2\lambda_{it}^{\frac{1}{v}}+1)/v\right)=a$$

Then

$$=\frac{\partial m_{it}}{\partial\beta}=a*b$$

$$\frac{\partial m_{it}}{\partial v}=(1-\pi_{it}).\left(2\left(\frac{\lambda_{it}^{\frac{1}{v}}}{v}-\frac{v-1}{2v}\right)*\left(\frac{(-\lambda_{it}^{\frac{1}{v}}x_{it}^{T}\beta-v\lambda_{it}^{\frac{1}{v}}-0.5v)}{v^{3}}\right)+(1-\pi_{it}).\frac{(-\lambda_{it}^{\frac{1}{v}}x_{it}^{T}\beta-v\lambda_{it}^{\frac{1}{v}})}{v^{3}}\right.$$

$$=(1-\pi_{it}).\left\{\frac{2\lambda_{it}^{\frac{1}{v}}-v+1}{v}\right\}*\left\{\frac{(-\lambda_{it}^{\frac{1}{v}}x_{it}^{T}\beta-v\lambda_{it}^{\frac{1}{v}}-0.5v)}{v^{3}}\right\}+(1-\pi_{it}).\frac{(-\lambda_{it}^{\frac{1}{v}}x_{it}^{T}\beta-v\lambda_{it}^{\frac{1}{v}})}{v^{3}}$$

$$=\left\{\frac{2\lambda_{it}^{\frac{1}{v}}-v+1}{v}\right\}*(1-\pi_{it})*\left\{\frac{(-\lambda_{it}^{\frac{1}{v}}x_{it}^{T}\beta-v\lambda_{it}^{\frac{1}{v}}-0.5v)}{v^{3}}+(1-\pi_{it}).\frac{(-\lambda_{it}^{\frac{1}{v}}x_{it}^{T}\beta-v\lambda_{it}^{\frac{1}{v}})}{v^{3}}\right.$$

$$=\left\{\frac{2\lambda_{it}^{\frac{1}{v}}-v+1}{v}\right\}*S+(1-\pi_{it}).\frac{(-\lambda_{it}^{\frac{1}{v}}x_{it}^{T}\beta-v\lambda_{it}^{\frac{1}{v}})}{v^{3}}$$

Since

$$(1-\pi_{it})\frac{(-\lambda_{it}^{\frac{1}{v}}x_{it}^{T}\beta-v\lambda_{it}^{\frac{1}{v}})}{v^{3}}=S+(1-\pi_{it})\ \frac{0.5v}{v^{3}}$$

Then

$$\frac{\partial m_{it}}{\partial v}=\left\{\frac{2\lambda_{it}^{\frac{1}{v}}-v+1}{v}\right\}*S+S+(1-\pi_{it})\ \frac{0.5v}{v^{3}}$$

$$=\left\{\frac{2\lambda_{it}^{\frac{1}{v}}-v+1}{v}+1\right\}*S+(1-\pi_{it}).\frac{0.5v}{v^{3}}$$

$=\{2\lambda_{it}^{\frac{1}{v}}+1)/v\}*S+(1-\pi_{it}).\frac{0.5v}{v^3}$

$=a*S+(1-\pi_{it}).\frac{0.5v}{v^3}$.