Scientific Journal of King Faisal University: Basic and Applied Sciences

Evaluating the Performance of Mixed Zero-Inflated Poisson Regression Models with Time-dependent and Time-independent Covariates

Gadir Alomair

Quantitative Methods Department, School of Business, King Faisal University, Al Ahsa, Saudi Arabia

LINK	RECEIVED	ACCEPTED	PUBLISHED ONLINE	ASSIGNED TO AN ISSUE
https://doi.org/10.37575/b/sci/230054	11/10/2023	01/02/2024	01/02/2024	01/06/2024
NO. OF WORDS	**NO. OF PAGES**	**YEAR**	**VOLUME**	**ISSUE**
5041	5	2024	25	1

Abstract

One of the issues that researchers may encounter in count data is having many zeros. One of the solutions to model these data is using zero-inflated Poisson (ZIP) regression models. Recently, researchers have started to model longitudinal count data with time-dependent covariates. However, it has not been considered whether a model with time-dependent covariates provides a better fit than a model with time-independent covariates. In this paper, the fit between a mixed ZIP model with time-dependent covariates and a mixed ZIP model with time-independent covariates is compared using simulation. Using the deviance information criterion as a measure of fit, we found that the model with time-dependent covariates exhibits a better fit than the model with time-independent covariates.

KEYWORDS

correlated data, count data, excess zeros, longitudinal, mixed models, model fit

CITATION

Alomair, G. (2024). Evaluating the performance of mixed zero-inflated Poisson regression models with time-dependent and time-independent covariates. *Scientific Journal of King Faisal University: Basic and Applied Sciences*, **25(1)**, 30–34. DOI: 10.37575/b/sci/230054

1. Introduction

One piece of necessary information when analyzing data is knowing the type of data that is used (Miller, 2007). Generalized linear models with either Poisson or negative binomial distributions of the response are usually employed to model count data. However, this does not guarantee that using either Poisson or negative binomial distributions gives the most precise results, as some other characteristics of the count data may yield more sophisticated results using different models (Min and Agresti, 2005). The issue of having many zeros in the count data arises when the observed data show a higher relative frequency of zeros than what is consistent with the Poisson or negative binomial distributions (Cameron and Trivedi, 2013). Zero inflation in the data is common in many disciplines, such as the medical field where counting the number of people who have a rare disease is of interest, counting the number of absences for students in a school, the number of times some bank has been robbed over a certain amount of time, the number of plane crashes, frequencies of crimes in some urban areas, and the number of daily labor deaths (Miller, 2007). Zero-inflated Poisson (ZIP) regression models have been increasingly used in various domains, including the monitoring of social networks. For instance, Motalebi *et al.* (2023) successfully applied this model to observe trends and patterns on social media platforms.

The problem of extra zeros or zero inflation was recognized in the mid-1960s (Lachenbruch, 2002). That was when Weiler (1964) tried to find a way of mixing a continuous distribution with a discrete distribution. One of the definitions of having excess zero counts was identified by Min and Agresti (2005) when they explained it as the lack of fit in a generalized linear model because of the condition of excessive zeros. Analyzing zero-inflated count data as if they come from Poisson or negative binomial distributions has some unwelcome consequences (Miller, 2007).

The consideration of time-dependent covariates in statistical modeling is crucial in many scenarios. Misaii *et al.* (2024), for instance, developed an optimal task-driven maintenance policy considering time-dependent covariates, underscoring the potential advantages of incorporating such covariates in model formulation. The primary aim of this study is to discern whether a mixed ZIP model with time-dependent covariates provides a better fit than a model with time-independent covariates. To achieve this, the two models are compared using simulation studies, and their fits are evaluated using the deviance information criterion (DIC).

The article is organized as follows. Section 2 presents previous related studies concerning the current study, as well as the rationale for conducting the research. Research methodology and simulation settings are presented in Section 3. Results and discussion are presented in Section 4. Concluding remarks are listed in Section 5.

2. Literature Review

According to McCullagh (2019), zero-inflated counts are a special case of overdispersed data where there is more variation in the data than the count distribution, which is Poisson in this study, allows for. There are some negative effects of having high variation in the data, such as inconsistent estimates, biased estimates of the parameters, and inflation of the standard error, which inflates the type I error rate and leads to invalid test statistics (Miller, 2007). Moreover, it was shown that the fit for both the zeros and nonzero counts will be improper if excessive zero is not considered (Perumean-Chaney *et al.*, 2013).

When there are many zeros in the data, there is more variation than the mean, and that is known as over-dispersion (Zorn, 1996). In this case, the assumption of equality between the mean and the variance under the Poisson distribution is violated (McCullagh, 2019). The consequence of having overdispersed data is making incorrect conclusions due to type I error inflation (Agresti, 2006). Therefore, the Poisson distribution needs to be modified when analyzing this type of data to correct the overdispersion in the data (Cameron and Trivedi, 2013).

Several models and associated estimation methods have been proposed for handling zero-inflated data (Kamalja and Wagh, 2018). Those models are usually variations of Poisson or negative binomial distributions. One of the models that can be used to handle excess-

Corresponding Author: Gadir Alomair 00966545547283, galomair@kfu.edu.sa.

zero data is the ZIP model. The ZIP model was proposed by Lambert (1992). In her example of zero-inflated data, 50% of the data were zeros. Lambert assumed that the inflation of zeros in the data can be seen from two different sources. The first source of zeros is produced as a regular count of the Poisson distribution, and the second source results just in "certain zeros." Thus, ZIP is a mixture of two processes: one only generates zeros, and the other generates both zero and nonzero counts. That means two kinds of zeros are modeled: the sampling zeros due to sampling variability under Poisson, and the structural zeros that exceed the frequency of zero counts expected under Poisson (Xia *et al.*, 2018). Therefore, there is a joint distribution for the ZIP model that has two sources of zeros. The ZIP model has two parts: a logistic distributed part, and a Poisson distributed part. As each part of the ZIP model is a generalized linear model, there are three components associated with each distribution (Agresti, 2006). These components are: 1) a random component that states the assumed distribution, which is one of the exponential family distributions, 2) a systematic component that presents the relationship between the parameters and the predictors, and 3) a link function that connects the response mean to the systematic component, which is the logit function for the logistic component and the log function for the Poisson component. The probability density function (PDF) of the ZIP model can be written as:

$$P(y=i)=\begin{cases}\pi+(1-\pi)f_p(0), & if\ i=0\\(1-\pi)f_p(i) & if\ i>0\end{cases} \quad (1)$$

Where f_p is the Poisson PDF and π is the probability of getting zeros (i.e., the proportion of zeros). The first component in the ZIP model has a logistic distribution and results in just zeros. The second component has a Poisson distribution and results in regular counts, where zeros can be one of them (Cameron and Trivedi, 2013).

Lambert (1992) proposed the ZIP model for cross-sectional data where the measurements are taken once. In our paper, we considered longitudinal data instead. Longitudinal data are commonly used nowadays, as they allow the study of changes over time (Lalonde *et al.*, 2013). Moreover, using longitudinal data has been shown to increase the power and the robustness of the model (Zeger and Liang, 1992). Longitudinal data are likely to be dependent, as the measurements of each individual at multiple occasions are repeated. Therefore, the failure to incorporate correlations within the same subject at different time points results in incorrect estimation and biased estimates of the regression coefficients (Ballinger, 2004). Correlated responses are very common in several fields, such as marketing, surveys in healthcare, and behavioral research (Lalonde *et al.*, 2013). Mixed models are often used to account for the within-subject dependence by incorporating a random effect in the model (Min and Agresti, 2005). The idea of extending the ZIP model for cross-sectional data to longitudinal data was proposed by Hall (2000). In his study, Hall considered longitudinal data, but he only incorporated a random effect in the Poisson component of the model. Min and Agresti (2005) suggested adding the random effect to both components of the model (i.e., Poisson and binary components), and hence a mixed ZIP model. Min and Agresti (2005) constructed random intercepts models only. Liu (2007) proposed a correlated random intercepts and slopes model. More studies have then implemented the mixed ZIP model on longitudinal data (Chen and Li, 2016; Burger *et al.*, 2019; Mekonnen *et al.*, 2019; Baghfalaki and Ganjali, 2020; Baghfalaki and Ganjali, 2021).

Hall (2000), Min and Agresti (2005), and Liu (2007) had studied longitudinal data that allow the responses within a subject to be correlated. In their studies, however, they considered the covariates to be the same for all time points. The measurements of these covariates are often taken at the initial time. Thus, any changes that happen to their values are usually ignored, i.e., they are considered time-independent covariates. According to Lalonde *et al.* (2013), for longitudinal studies, as the values of the responses for the same subject change over time, the covariate values can change at different time points as well. Lalonde (2014) considered a time-dependent covariate in his mixed ZIP model by allowing the covariate to change over time just as the response variable.

Zero-inflated models, including variations of Poisson or negative binomial distributions, have been widely applied in different contexts. For example, Hagen *et al.* (2023) used a zero-inflated negative binomial regression model to study parking violations in Berlin, demonstrating the versatility of these models. Chen and Li (2016) constructed a specific case of a two-part zero-inflated Beta regression model with random effects for testing the association between microbial abundance and clinical covariates for longitudinal microbiome data. In their study, they mentioned that covariates can be time-dependent and are not necessarily the same, but no comparison of the performance of time-independent and time-dependent covariates was explored. Mekonnen *et al.* (2019) analyzed determinants of the number of epileptic seizure attacks—count data with a high proportion of zeros with a longitudinal outcome—using zero-inflated, Poisson, and negative binomial family models. In their study, they used some time-independent covariates and one time-dependent covariate as factors that are expected to have some association with epileptic seizure attacks. The comparison, however, was based on the model fit across all models using information criteria: Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC), and no comparison between time-independent and time-dependent covariates was applied. It has been shown that treating covariates as time-dependent in longitudinal studies results in more efficient estimators and stronger statistical inferences (Gibbons and Hedeker, 2006; Zeger and Liang, 1992). In this study, we considered longitudinal data and compared the ZIP model with time-dependent covariates to the ZIP model with time-independent covariates in terms of their fit.

There are several methods to measure the fit of the model to the data. DIC is one of them. It is a Bayesian measure of fit or adequacy and was introduced by Spiegelhalter *et al.* (2002). DIC can be used to compare the fit of different models based on the loss of information, and it is one of the model selection criteria. DIC is based on the posterior distribution of the deviance (Berg *et al.*, 2004). The choice of this particular method to compare the fit of the mixed ZIP model with time-dependent covariates to the fit of a mixed ZIP model with time-independent covariates was because DIC was successfully applied to complex statistical models and is considered a generalization of AIC and BIC (Berg *et al.*, 2004; Neelon *et al.*, 2010). Moreover, it has been shown that it is simpler to calculate than other measures of fit and is widely used (Berg *et al.*, 2004).

Zero-inflation in count data is a problem that can affect the validity of the results if not modeled properly. Moreover, longitudinal studies are commonly used in practice, and they generate correlated counts because of the repeated measures (Pittman *et al.*, 2022). Thus, mixed models are more appropriate for longitudinal data, as they can account for the correlation between observations within the same subject. Several studies have compared the performance of different zero-inflated models with longitudinal data (Min and Agresti, 2005; Chen and Li, 2016; Burger *et al.*, 2019; Mekonnen *et al.*, 2019; Baghfalaki and Ganjali, 2020; Baghfalaki and Ganjali, 2021; Lalonde, 2014; Pittman *et al.*, 2022). Most of these studies have compared the fit of the models based on AIC and BIC. Few studies have used DIC for the fit comparison of the zero-inflated models, while it has been shown that it is an appropriate measure for the fit of more complex statistical models, such as ZIP (Neelon *et al.*, 2010; Pooley and Marion, 2018). Therefore, the purpose of this study is to determine which model better fits the data, investigating whether the model provides new information by considering the possible change of covariates over time.

3. Materials and Methods

3.1. Mixed Zero-inflated Poisson Models:

In the mixed ZIP regression model, the response, Y_{ij}, for subject $i = 1,2,\dots n$ and time $j = 1,2,\dots,T$, is conditionally distributed on the random effect, u_i, using the ZIP distribution with probability π, and mean λ. The random effect component, u_i, is normally distributed with mean zero and variance σ_u^2. The response distribution and the random component can be written as, $Y_{ij}\ |u_i \sim ZIP(\pi,\lambda)$ and $u_i \sim N(0,\sigma_u^2)$, respectively. The conditional PDF for the response is,

$$f_{ZIP}(y;\pi,\lambda) = \begin{cases} \pi + (1-\pi)e^{-\lambda} & y = 0 \\ \frac{(1-\pi)e^{-\lambda}\lambda^y}{y!} & y > 0 \end{cases} \quad (2)$$

The systematic components and link functions for the mixed ZIP regression model are,

$$logit(\pi) = X_l\beta_l + Zu \quad (3)$$

$$ln(\pi) = X_p\beta_p + Zu \quad (4)$$

where X_l and β_l are the design matrix and the parameter vector corresponding to the logistic component, respectively, X_p and β_p are the design matrix and the parameter vector corresponding to the Poisson component, respectively, Z is the random effects design matrix, and u is the random effect parameter vector.

In the current study, two covariates in the mixed ZIP regression models were considered: 1) X_1 is a binary variable, and 2) X_2 is a continuous variable. In both time-dependent and time-independent covariates models, the covariates were treated as fixed effects. However, in the mixed ZIP regression model with time-dependent covariates, it was assumed that the covariates are time-dependent and thus might have different values over time. In the mixed ZIP regression model with time-independent covariates, the covariates were assumed to be time-independent and were measured only at the initial time point. As such, the values were constant over time.

To compare the goodness-of-fit of the two mixed ZIP regression models, the DIC was calculated and compared for each model,

$$DIC = D(\bar{\theta}) + 2p_D \quad (5)$$

where $\bar{\theta}$ is the posterior mean of the parameters of the model, approximated by $\frac{1}{n}\sum_{t=1}^{n}\theta^t$, and $D(\bar{\theta})$ is the deviance evaluated at $\bar{\theta}$, equal to $-2\ log\left(p(y|\bar{\theta})\right)$. A smaller DIC indicates a better fit to the data set.

3.2. Data Generation:

In the current study, no real data were collected. Instead, we generated the data through computer simulation using R, following these steps:

1. A balanced design was considered with 10 time points (T = 10) and three different sample sizes (S = 25, 50, and 100) for each of the 250 replications (M = 250). To start the simulation, we used initial values of $\beta_{l,0} = -.41304, \beta_{l,1} = -.46604, \beta_{l,2} = -.46604, \beta_{p,0} = -.06884, \beta_{p,1} = .03532$, and $\beta_{p,2} = .03532$. The beta values and conditions were used in Lalonde's (2014) study, which have been shown to greatly affect the relative bias, error rate, and power of the models. Some studies also consider the potential impact of measurement errors in the response variable, a complication Zhang and Yi (2023) addressed in their work on ZIP models. Their methodology provides a foundation for dealing with measurement errors.
2. The starting seed was randomly set at 896. The seed increases by one point for each of the 250 replications. We did this to ensure that different data sets were generated for each replication.
3. To simulate the effect of the random component, u_i, on the data set, a variable called *subjectEffect* was created using the *rnorm* function in R with mean zero and variance five (Lalonde, 2014). Then, this variable was added to the mean of the normal distribution, from which the continuous covariate was generated. For the logistic component, the continuous covariate, $X_{l,1}$, was generated using the *rnorm* function with mean 5+2*subjectEffect* and variance three. For the Poisson component, the continuous covariate, $X_{p,1}$, was generated using the *rnorm* function with mean 3-2*subjectEffect* and variance four. The mean and variance values were used in Lalonde (2014).
4. To generate the categorical covariate for the logistic component, $X_{l,2}$, and the Poisson component, $X_{p,2}$, the *rbinom* function was used with a probability of 0.5 and size one.
5. For the ZIP model with time-dependent covariates, the covariate values for each subject were changed from one time point to another. For the ZIP model with time-independent covariates, all the covariate values were fixed for each subject to be equal to the first value obtained by the rnorm and the rbinom function, as it was assumed that the covariates were only measured at the initial time of the study. As such, the covariate values for a subject are the same throughout all 10 time points.
6. To generate the response variable, y, a variable named certain_zero was generated using the rbinom function for each of the 10 time points with size one and probability $1/(1 + exp\left(-(\beta_{l,0} + \beta_{l,1}X_{l,1} + \beta_{l,2}X_{l,2})\right)$. If the value of the variable certain_zero for a subject at a time point is zero, the response, y, was given a zero value. If the value of the variable certain_zero is not zero, the response, y, was given a value calculated using the rpois function with $\lambda = exp\left(-(\beta_{p,0} + \beta_{p,1}X_{l,1} + \beta_{p,2}X_{l,2})\right)$.
7. At the end of each data generation, 10 variables were used in the model fitting: *subject*, y, $X_{l,1}, X_{l,2}, X_{p,1}$, and $X_{p,2}$ for the time-dependent covariates, and $X_{l,1}, X_{l,2}, X_{p,1}$, and $X_{p,2}$ for the time-independent covariates.

3.3. Simulation Procedures and Analysis:

After generating the data, we fitted the mixed ZIP model with time-dependent covariates to the dataset. Then, we extracted the DIC of the model, dependent-DIC. Afterward, we fitted the mixed ZIP model with time-independent covariates to the dataset. Then, we extracted the DIC of the model, independent-DIC. We compared the dependent-DIC to the independent-DIC. If dependent-DIC was less than the independent-DIC, it was recorded as 1; otherwise, we recorded it as 0. We repeated these procedures for M = 250 times. At the end of the simulation procedures, we had 250 dependent-DIC values, 250 independent-DIC values, and calculated the proportions of 0 and 1 for each of the three sample sizes.

If the proportion of 1's over the 250 replication times is more than the proportion of 0's, we conclude that the mixed ZIP model with time-dependent covariates has better fit than the mixed ZIP model with time-independent covariates. Otherwise, it is concluded that the mixed ZIP model with time-independent covariates has a better fit than the mixed ZIP model with time-dependent covariates.

4. Results and Discussion

The number of times that the DIC for the mixed ZIP model with time-dependent covariates is less than the DIC for the model with time-independent covariates is 129 out of 250 replications (52%) when the sample size is 25. The other 121 out of 250 replications yielded that the DIC for the mixed ZIP model with time-independent covariates is less than the DIC for the model with time-dependent. When the sample size is S=50, 68% of the replications have a smaller DIC for the model with time-dependent covariates, and only 32% of the time, the DIC for the time-independent model is smaller. For the largest sample size, which is S=100, the number of times that DIC for time-dependent covariates model is less than DIC for the model with time-independent covariates is 187 out of 250 replications (75%), and only 25% of the time, DIC for the model with time-independent covariates is less than the other (see Table 1).

Table1: Proportions of DIC.

Subject	DIC for ZIP with time-dependent < DIC for ZIP with time-independent	DIC for ZIP with time-dependent > DIC for ZIP with time-independent
25	52%	48%
	(129/250)	(121/250)
50	68%	32%
	(169/250)	(81/250)
100	75%	25%
	(187/250)	(63/250)

It was found that the mixed ZIP model with time-dependent covariates has a better fit than the mixed ZIP with time-independent covariates. For more than half of the 250 replications at all three sample sizes, ZIP with time-dependent covariates had smaller DIC values than ZIP with time-independent covariates. This means that the ZIP model with time-dependent covariates has a better fit than the model with time-independent covariates. The results showed that by measuring time-dependent covariates over time, we gained new information. This suggests that in a longitudinal study based on count data, if a researcher has a covariate that changes over time, i.e., a time-dependent covariate, the researcher should measure the covariates over time and not just at the initial time. By doing this, models would have a better fit to the data. In addition to that, the effect of the time-dependent covariates on the response can be studied. This, in turn, will give the investigators more information about the data.

The results of a better fit for the time-dependent covariates model were clear, even though the sample sizes considered were not sufficiently large. This emphasizes the importance of accounting for the dependency of covariates whenever it is applicable. The fit appeared to get better for the time-dependent covariates model for larger sample sizes, as expected.

Conducting a diagnostic evaluation of the model's fit can help in identifying issues and guiding directions for its improvement (Tang *et al.*, 2023). Although having a model that fits well does not necessarily mean that the model is superior to others, smaller DIC values indicate better parsimony and a greater chance of being cross validated in future samples (Cameron and Trivedi, 2013). In other words, a better-fitting model closely matches observations, yielding more accurate predictions and less information loss than others (Azen and Walker, 2021). This result was evident with the time-dependent covariates model, as it captured more information about the data than the other model. In addition to the findings, a few points should be noted about the simulated scenarios. First, the simulation considered only two covariates with 10 time points (T = 10). Future studies can consider including a variety of covariates, their interactions, and a range of time points. Moreover, it is recommended to compare the models using different aspects of model comparisons, such as evaluating the models in terms of power and type I error rate. Furthermore, it is of interest in the future to explore how the fit of the models changes for different proportions of zeros in the data. Despite the limitations of this study, the results were as expected, and these findings can be extended in future research.

5. Conclusion

The principal objective was to compare mixed ZIP models with time-dependent and time-independent covariates. This was achieved through the use of simulation studies, followed by a comprehensive evaluation using the DIC.

The findings reveal that the mixed ZIP model with time-dependent covariates provides a superior fit compared to the model with time-independent covariates. This underscores the importance of considering temporal dynamics in covariates when modeling count data, especially in scenarios where a high frequency of zeros is present. Therefore, this research lends further weight to the growing body of evidence that supports the use of time-dependent covariates in ZIP models.

Furthermore, this study contributes to the ongoing discussion on the most effective ways to handle count data with high zero-inflation. Future research could extend our findings by exploring other types of mixed models, applying these methodologies to different types of count data, and testing additional measures of model fit. This research not only contributes to the theoretical understanding of ZIP regression models but also provides practical insights for researchers dealing with zero-inflated count data. This dual contribution positions this study as a valuable resource for both statisticians and practitioners across various fields.

Biography

Gadir Alomair

Quantitative Methods Department, School of Business, King Faisal University, Al Ahsa, Saudi Arabia, 00966545547283, galomair@kfu.edu.sa

Gadir Abdulaziz Alomair, from Saudi Arabia, is an Assistant Professor in the Quantitative Methods department at King Faisal University. Gadir earned her Ph.D. in Applied Statistics from the University of Northern Colorado in the USA, and her master's degree was in Applied Statistics from the same university. She completed her bachelor's degree at King Faisal University. Her areas of expertise include time series modeling and forecasting, longitudinal data analysis, and zero-inflated data analysis.

ORCID: 0000-0001-8557-3796

References

Agresti, A. (2006). *An Introduction to Categorical Data Analysis*. 2nd Edition. New York, USA: John Wiley and Sons, Inc. DOI: 10.1002/0470114754

Azen, R. and Walker, C. M. (2021). *Categorical Data Analysis for the Behavioral and Social Sciences*. 2nd Edition. USA: Routledge.

Ballinger, G.A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational research methods*, **7**(2), 127–150. DOI: 10.1177/1094428104263672

Baghfalaki, T. and Ganjali, M. (2020). A transition model for analysis of zero-inflated longitudinal count data using generalized poisson regression model. *REVSTAT-Statistical Journal*, **18**(1), 27–45. DOI: 10.57805/revstat.v18i1.288

Baghfalaki, T. and Ganjali, M. (2021). Approximate Bayesian inference for joint linear and partially linear modeling of longitudinal zero-inflated count and time to event data. *Statistical Methods in Medical Research*, **30**(6), 1484–1501. DOI: 10.1177/09622802211002868

Berg, A., Meyer, R. and Yu, J. (2004). Deviance information criterion for comparing stochastic volatility models. *Journal of Business & Economic Statistics*, **22**(1), 107–120. DOI: 10.1198/073500103288619430

Burger, D.A., Schall, R., Jacobs, R. and Chen, D.G. (2019). A generalized Bayesian nonlinear mixed-effects regression model for zero-inflated longitudinal count data in tuberculosis trials. *Pharmaceutical Statistics*, **18**(4), 420–432. DOI: 10.1002/pst.1933

Cameron, A.C. and Trivedi, P.K. (2013). *Regression Analysis of Count Data*. *England*: Cambridge University press.

Chen, E.Z. and Li, H. (2016). A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. *Bioinformatics*, **32**(17), 2611–2617. DOI: 10.1093/bioinformatics/btw308

Gibbons, R.D. and Hedeker, D. (2006). *Longitudinal Data Analysis* . New York, USA: John Wiley and Sons, Inc.

Hagen, T., Reinfeld, N. and Saki, S. (2023). Modeling of Parking Violations

Using Zero-Inflated Negative Binomial Regression: A Case Study for Berlin. *Transportation Research Record*, **2677**(6), 498–512.

Hagen, T., Reinfeld, N. and Saki, S. (2023). Modeling of Parking Violations Using Zero-Inflated Negative Binomial Regression: A Case Study for Berlin. Transportation Research Record, **2677**(6), 498–512. DOI: 10.1177/03611981221148703

Hall, D.B. (2000). Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study. *Biometrics*, **56**(4), 1030–1039. DOI: 10.1111/j.0006-341x.2000.01030.x

Kamalja, K.K. and Wagh, Y.S. (2018). Estimation in zero-inflated Generalized Poisson distribution. *Journal of Data Science*, **16**(1), 183–206. DOI: 10.6339/jds.201801_16 (1).0010

Lachenbruch, P.A. (2002). Analysis of data with excess zeros. *Statistical Methods in Medical Research*, **11**(4), 297–302. DOI: 10.1191/0962280202sm289ra

Lalonde, T.L., Nguyen, A.Q., Yin, J., Irimata, K. and Wilson, J.R. (2013). Modeling correlated binary outcomes with time-dependent covariates. *Journal of Data Science*, **11**(4), 715–738. DOI: 10.6339/jds.2013.11(4).1195

Lalonde, T.L. (2014). Modeling Longitudinal Count Data with Excess Zeros and Time-Dependent Covariates: Application to Drug Use. In: *American Public Health Association (APHA) 142nd Annual Meeting and Exposition*, New Orleans University, New Orleans, USA. 15-19/11/2014.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**(1), 1–14. DOI: 10.2307/1269547

Liu, H. (2007). Growth curve models for zero-inflated count data: An application to smoking behavior. *Structural Equation Modeling: A Multidisciplinary Journal*, **14**(2), 247–279. DOI: 10.1080/10705510709336746

McCullagh, P. (2019). *Generalized Linear Models*. Routledge. USA: Routledge.

Mekonnen, F.H., Lakew, W.D., Tesfaye, Z.D. and Swain, P.K. (2019). Statistical models for longitudinal zero-inflated count data: application to seizure attacks. *African Health Sciences*, **19**(3), 2555–2564. DOI: 10.4314/ahs.v19i3.31

Miller, J.M. (2007). *Comparing Poisson, Hurdle, and ZIP Model Fit Under Varying Degrees of Skew and Zero-inflation*. PhD Thesis, University of Florida, Florida, USA.

Min, Y. and Agresti, A. (2005). Random effect models for repeated measures of zero-inflated count data. *Statistical modelling*, **5**(1), 1–19. DOI: 10.1191/1471082x05st084oa

Misaii, H., Fouladirad, M. and Haghighi, F. (2024). Optimal task-driven time-dependent covariate-based maintenance policy. *Journal of Computational and Applied Mathematics*, **435**(n/a), 115315. DOI: 10.1016/j.cam.2023.115315

Motalebi, N., Owlia, M.S., Amiri, A. and Fallahnezhad, M.S. (2023). Monitoring social networks based on Zero-inflated Poisson regression model. *Communications in Statistics-Theory and Methods*, **52**(7), 2099–2115. DOI: 10.1080/03610926.2021.1945103

Neelon, B.H., O'Malley, A.J. and Normand, S.L.T. (2010). A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use. *Statistical modelling*, **10**(4), 421–439. DOI: 10.1177/1471082x0901000404

Perumean-Chaney, S.E., Morgan, C., McDowall, D. and Aban, I. (2013). Zero-inflated and overdispersed: what's one to do? *Journal of Statistical Computation and Simulation*, **83**(9), 1671–1683. DOI: 10.1080/00949655.2012.668550

Pittman, B., Buta, E., Garrison, K. and Gueorguieva, R. (2023). Models for zero-inflated and overdispersed correlated count data: an application to cigarette use. *Nicotine and Tobacco Research*, **25**(5), 996–1003. DOI: 10.1093/ntr/ntac253

Pooley, C.M. and Marion, G. (2018). Bayesian model evidence as a practical alternative to deviance information criterion. *Royal Society Open Science*, **5**(3), 171519. DOI: 10.1098/rsos.171519

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **64**(4), 583–639. DOI: 10.1111/1467-9868.00353

Tang, W., He, H. and Tu, X.M. (2023). *Applied Categorical and Count Data Analysis*. 2nd Edition. London, England: CRC Press.

Weiler, H. (1964). A significance test for simultaneous quanta1 and quantitative responses. *Technometrics*, **6**(3), 273–285. DOI: 10.2307/1266044

Xia, Y., Sun, J., Chen, D.G., Xia, Y., Sun, J. and Chen, D.G. (2018). Modeling zero-inflated microbiome data. In. Y. Xia, J. Sun, D. Chen, Y. Xia, J. Sun and D. Chen (eds.) *Statistical analysis of microbiome data with R*. Singapore: Springer, ICSA Book Series in Statistics. DOI: 10.1007/978-981-13-1534-3_12

Zeger, S.L. and Liang, K.Y. (1992). An overview of methods for the analysis of longitudinal data. *Statistics in medicine*, **11**(14-15), 1825–1839. DOI: 10.1002/sim.4780111406

Zhang, Q. and Yi, G.Y. (2023). Zero-inflated Poisson models with measurement error in the response. *Biometrics*, **79**(2), 1089–1102. DOI: 10.1111/biom.13657

Zorn, C.J. (1996). Evaluating zero-inflated and hurdle Poisson specifications. *Midwest Political Science Association*, **18**(20), 1–16.