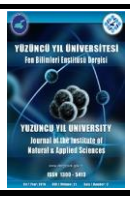
Yuzuncu Yil University
Journal of the Institute of Natural & Applied Sciences
https://dergipark.org.tr/en/pub/yyufbed



Research Article

Zero Truncated Models in Regression Analysis: An Examination of Their Advantages on Small Mean Values†

Rıdvan KARA*[1], Abdullah YEŞİLOVA[2]

[1]Hakkari University, Yüksekova Vocational School, Department of Plant and Animal Production, 30330, Hakkari, Türkiye
[2]Van Yüzüncü Yıl University, Faculty of Agriculture, Department of Biometry and Genetics, 65040, Van, Türkiye
* Corresponding author e-mail: ridvankara@hakkari.edu.tr

Abstract: In this study, two of the most commonly used zero-truncated regression models for modeling positive count data, namely Zero Truncated Poisson and Zero Truncated Negative Binomial, are compared with the classical Poisson and Negative Binomial regression models. The role of the mean of the dependent variable in model selection is examined. Simulations were first conducted using different mean values for the dependent variable, followed by a comparison of model performances using two different real data sets. The real data sets were constructed using crime data published by Turkish Statistical Institude (TSI). AIC, BIC, and residual plots were utilized to determine the most suitable model. The study found that zero-truncated models perform better when the mean of the dependent variable is below 5, compared to classical models.

Keywords: Negative binom, Poisson, Zero truncated negative binom, Zero truncated poisson

Sıfır Değer Kesilmiş Modellerin Regresyon Analizindeki Avantajları: Küçük Ortalama Değerler Üzerine Bir İnceleme

Öz: Bu çalışmada, pozitif sayım verilerinin modellenmesinde en sık kullanılan zero-truncated regresyon modellerinden ikisi Sıfır değer kesilmiş Poisson, Sıfır değer kesilmiş Negatif Binom ile klasik Poisson ve Negatif Binom regresyon modelleri karşılaştırılmış ve bağımlı değişkenin ortalamasının model seçimindeki rolü incelenmiştir. Öncelikle bağımlı değişken için farklı ortalama değerleri kullanılarak simülasyonlar gerçekleştirilmiş, ardından iki farklı gerçek veri seti üzerinden model performansları karşılaştırılmıştır. Gerçek veri setleri, Türkiye İstatistik Kurumu (TÜİK) tarafından yayımlanan suç verileri kullanılarak oluşturulmuştur. En uygun modelin belirlenmesinde AIC, BIC ve residual grafikleri kullanılmıştır. Bağımlı değişkenin ortalamasının 5'ten küçük olduğu durumlarda, sıfır değer kesilmiş modellerin daha üstün bir performans sergilediği tespit edilmiştir.

Anahtar Kelimeler: Negatif binom, Poisson, Sıfır değer kesilmiş negatif binom, Sıfır değer kesilmiş poisson

† This study is derived from a section of the doctoral dissertation titled *"Zero-Truncated Regression Methods and Their Application"* defended by Rıdvan Kara at Van Yüzüncü Yıl University.

Received: 24.11.2024
Accepted: 03.03.2025

How to cite: Kara, R., & Yeşilova, A. (2025). Zero Truncated Models in Regression Analysis: An Examination of Their Advantages on Small Mean Values. *Yuzuncu Yil University Journal of the Institute of Natural and Applied Sciences*, *30*(1), 102-112. https://doi.org/10.53433/yyufbed.1590611

1. Introduction

Count data often deviate from normal distribution and typically start from zero with positive values (Zeileis et al., 2008; Coxe et al., 2009). Such data generally do not satisfy the assumptions of normality and take on positive, greater-than-zero values. This can lead to biased parameter estimates when applying linear regression to count data. Count data typically follow a Poisson distribution, for which Poisson regression is commonly used (Agresti, 1997). However, issues such as overdispersion or underdispersion arise when the equality between the variance and the mean is violated (Cox, 1983). Overdispersion occurs when the variance exceeds the mean, indicating greater variability in the data than expected under a Poisson model, whereas underdispersion refers to the case where the variance is smaller than the mean, suggesting less variability than anticipated. In such cases, alternative models like negative binomial (NB) and generalized Poisson (GP) regressions can be employed (Lawal, 2012; Sáez-Castillo & Conde-Sánchez, 2013; Hilbe, 2014). Zero-truncated regression models also provide an effective approach for positive count data.

Poisson regression is suitable when the variance equals the mean, but it may yield biased estimates and narrow confidence intervals in the presence of overdispersion (Hilbe, 2011). In such situations, NB and GP regressions are preferred (Long, 1997; Jansakul & Hinde, 2002). For count data that do not include zeros, zero-truncated regression models are appropriate. These models estimate probabilities based on positive count data and are utilized in fields such as modeling gene expression in medicine (Thygesen & Zwinderman, 2006). When overdispersion stems from multiple sources, the zero-truncated generalized negative binomial method is preferred (Hilbe, 2011). The comparison of Poisson, negative binomial, and zero-truncated models is often based on AIC and BIC values; the model with the smallest value is considered the best fit (Rose et al., 2006; Sileshi, 2008). Truncated models are widely applied in various fields. For instance, Thygesen & Zwinderman, (2006) and Lee et al. (2003) employed these models in medical studies; Liu et al. (2013) used zero-truncated negative binomial regression with quantile regression to estimate train derailment probabilities; Puza et al. (2008) applied a Bayesian approach to truncated Poisson regression for analyzing illegal immigrant data in the Netherlands; Creel and Loomis (1990) analyzed deer hunting in California; and Van Der Heijden et al. (2003) used zero-truncated regression models to estimate offender populations based on police records in Germany.

Residual analysis is an effective method for determining the best-fitting model, as it directly impacts the accuracy and validity of the model (Draper & Smith, 1998). Residual plots displaying a wide distribution and systematic patterns indicate poor model fit (Montgomery et al., 2012). Another approach to identifying the best model involves examining the AIC and BIC values, with lower values indicating a better model fit (Burnham & Anderson, 2002).

This study investigates the role of the dependent variable's mean in model selection and discusses potential issues arising from the failure to choose an appropriate regression model. By employing AIC, BIC, and residual plots, the study aims to determine the best model theoretically and practically, providing insights into model performance evaluation.

2. Material and Methods

The real data used in this study consists of crime statistics recorded in 81 provinces of Turkey between 2009 and 2013, obtained from the Turkish Statistical Institute (TÜİK). The crimes analyzed include bad treatment, sexual crimes, crimes related with firearms and knifes, homicide, assault, robbery, production and commerce of drugs, use and purchase of drugs, and kidnapping. These statistics, categorized by province in the TÜİK dataset, were grouped into seven geographical regions and incorporated into the regression model as the "region" variable. Additionally, each year was included in the model as the "year" variable. The study utilized a total of 11 variables (considering "year" and "region" as single variables) and 405 data points.

The study employed the R 3.5.2 statistical software and several R packages. The packages used include "msm" and "sandwich" for Poisson regression, "MASS" and "foreign" for negative binomial regressions, and "VGAM," "ggplot2," and "boot" for zero-truncated regression methods. In the simulation study, 1,000 samples, 1 dependent variable, 10 independent variables, and the **"pscl"** package in R were utilized. A sample size of 1,000 was chosen to ensure the statistical stability of the

estimators, maintain computational efficiency, and adopt an approach consistent with the existing literature.

The regression methods applied in this study were Poisson, negative binomial, zero-truncated Poisson (ZTP), and zero-truncated negative binomial (ZTNB) regressions.

2.1. Poisson regression

Poisson regression is one of the generalized linear models in regression analysis. It is used for analyzing count data and contingency tables. When modeling contingency tables, it is often referred to as log-linear regression. In this model, the dependent variable follows a Poisson distribution, and a log link function is used to relate the dependent variable to the independent variables, resulting in a linear equation (Winkelmann, 2008; Yesilova et al., 2010).
When the independent variables are $X_1, X_2, \ldots . X_m$, the log link function is expressed as follows:

$$\log(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m \quad (1)$$

where $X_1, X_2, \ldots, X_m$ is the vector of independent variables and $\beta_0, \beta_1, \beta_2, \ldots, \beta_m$ represents the regression coefficients (Kutner et al., 2004).

Here, μ is an exponential function of the independent variables. The log-likelihood function is defined as (Khoshgoftaar et al., 2005):

$$L(\beta / y_i, x_i) = \sum_{i=1}^{n} [y_i x_i' \beta - \exp(x_i' \beta) - ln y_i!] \quad (2)$$

2.2. Negative binomial regression

Negative binomial (NB) regression is a method used as an alternative to Poisson regression when overdispersion is observed in the dependent variable. It relates the dependent and independent variables through a log link function. The probability function is given as:

$$P_r(Y) = \frac{\Gamma(y_i + \frac{1}{\alpha})(\alpha\mu_i)^{y_i}}{y_i!\,\Gamma(\frac{1}{\alpha})(1+\alpha\mu_i)^{y_i+\frac{1}{\alpha}}} \quad , \quad \alpha > 0 \quad (3)$$

In this equation, μi is the mean of the distribution, α is the NB dispersion parameter, and $\Gamma(.)$ denotes the gamma function (Simo et al., 2007).

The log-likelihood function of the NB regression model can be expressed as (Lawless, 1987):

$$L(\beta, \alpha, y) = \sum_{i=1}^{n} \left[\frac{1}{\alpha}\log(1+\alpha\mu_i) - y_i \log\left(1 + \frac{1}{\alpha\mu_i}\right) + \log\Gamma\left(y_i + \frac{1}{\alpha}\right) - \log\Gamma\left(\frac{1}{\alpha}\right) - \log y_i!\right] \quad (4)$$

2.3. Zero truncated poisson regression

Zero-truncated Poisson (ZTP), also known as positive Poisson, is a distribution where all values are positive and zero (0) is not included, unlike the standard Poisson distribution. The zero-truncated Poisson model is given by (Van Der Heijden et al., 2003):

$$P(y_i | y_i > 0, \lambda) = \frac{P(y_i|\lambda)}{P(y_i > 0, |\lambda)} = \frac{\exp(-\lambda)\lambda^{y_i}}{y_i!\,(1-\exp(-\lambda))} \quad (5)$$

$$P(y_i > 0, |\lambda) = (1 - \exp(-\lambda)), y_i = 1,2, \ldots \quad , \quad i = 1,2, \ldots . N, \quad (6)$$

For the i-th observation, the log-likelihood function is expressed as:

$$I(y_i;\mu_i|y_i>0)=y_i[X^T{}_i\beta]-exp(X^T{}_i\beta)-ln(y_i!)-\ln[1-exp\{-exp(X^T{}_i\beta)\}] \quad (7)$$

Here, y_i represents the observed count outcome, while λ denotes the Poisson rate parameter. The term $X_i^T\beta$ represents the linear predictor in the regression model.

The log-likelihood for the ZTP model is given by (Hilbe, 2014):

$$I(y_i;\mu_i|y_i>0)=\sum y_i[X^T{}_i\beta]-exp(X^T{}_i\beta)-ln(y_i!)-\ln\left[1-\exp\{-\exp(X^T{}_i\beta)\}\right] \quad (8)$$

2.4. Zero truncated negative binomial regression

Zero-truncated negative binomial (ZTNB) regression is a method used for modeling dependent variables with positive values ($y_i>0$) that also exhibit overdispersion. The log-likelihood for the ZTNB model can be written as:

$$L(\mu;y|y>0)=\sum_{i=1}^{n}\{L_{NB2}-ln[1-\{1+\alpha exp(x_i'\beta)\}^{-\frac{1}{\alpha}}]\} \quad (9)$$

where L_{NB2} is given by (Hilbe, 2014):

$$L(\beta;y,\alpha)=\sum_{i=1}^{n}\{y_i\log\left(\frac{\alpha\exp(x_i'\beta)}{1+\alpha\exp(x_i'\beta)}\right)-\left(\frac{1}{\alpha}\right)\log(1+\alpha\exp(x_i'\beta))+log\Gamma\left(y_i+\frac{1}{\alpha}\right)-log\Gamma(y_i+1)-log\Gamma(\frac{1}{\alpha})\} \quad (10)$$

3. Results

3.1. Simulation study

In the simulation study, one dependent variable and 10 independent variables were generated. The sample size was set to 1,000. The dependent variable had a minimum value of 1 and followed a Poisson distribution with a mean of 100. This mean value was subsequently altered to 50, 10, 5, and 3, and model performances were compared accordingly.

In the second simulation study, the dependent variable was again a count variable with a minimum value of 1 but exhibited overdispersion. The mean values were similarly set to 100, 50, 10, 5, and 3, and model performances under conditions of overdispersion were compared.

Residual plots, along with AIC and BIC criteria, were used to determine the best-performing model. Residual plots for the highest mean of 100 and the lowest mean of 3 were presented to visualize changes in model performance across the two extreme mean values. The goal was to observe the variation in model performance across different mean levels.

The results obtained for models using one dependent variable (consisting of positive values following a Poisson distribution) and 10 independent variables across different mean values are summarized below.

Table 1. AIC and BIC values of Poisson and ZTP regressions for different mean values

MODEL	μ=100		μ =50		μ =10		μ =5		μ =3	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
POISSON	7453	7507.77	6785.1	6839.1	5153.547	5207.532	4342.98	4396.96	3702.74	3756.72
ZTP	7453	7507.77	6785.1	6839.1	5153.455	5207.44	4328.02	4382.00	3589.41	3643.40

For μ=100, the Poisson and ZTP models show similar AIC and BIC values, suggesting similar fit performance. As μ decreases, the ZTP model generally provides a slightly better fit (lower AIC and BIC values) compared to the Poisson model. This indicates that the ZTP model may be more appropriate for datasets where zero values are truncated, as it results in lower AIC and BIC values, signaling better model fit.

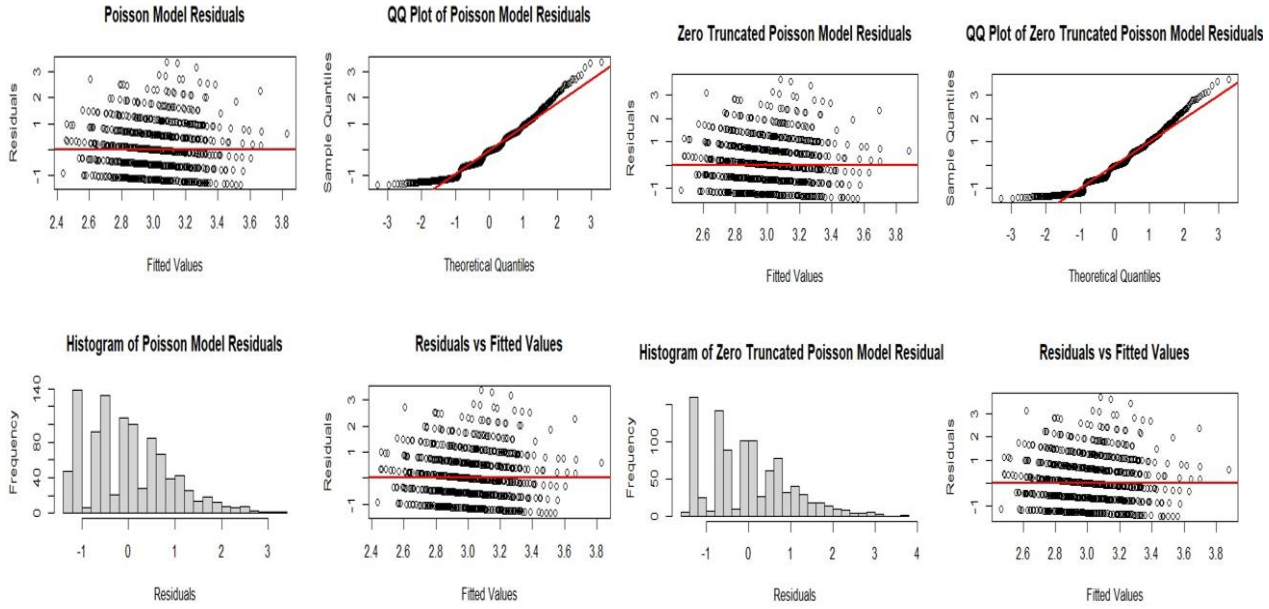


Figure 1. Residual plots for a mean value of 100.

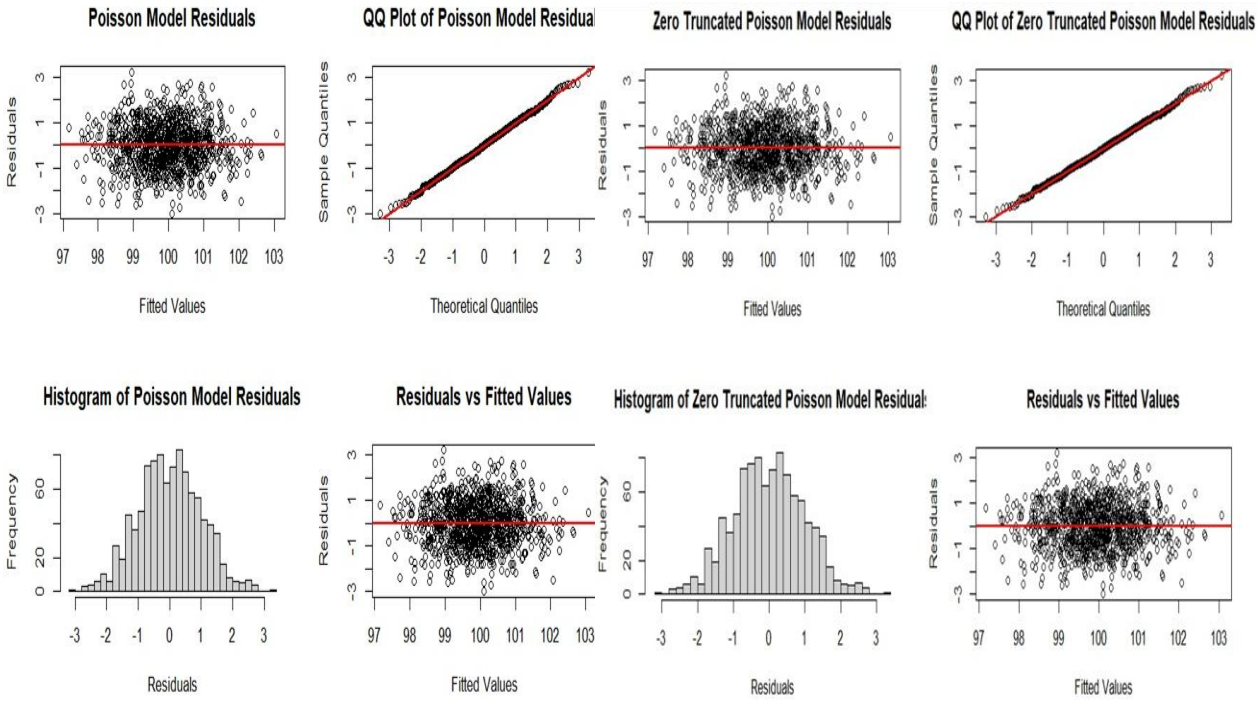


Figure 2. Residual plots for a mean value of 3.

Residual plots exhibit a wide distribution and systematic patterns, which indicate poor model fit. This behavior is observed in the residual plots of the Poisson model. Particularly, when the mean decreases (μ=3), the model demonstrates inadequate fit to the dataset and produces higher prediction errors. In contrast, the Zero-Truncated Poisson (ZTP) model shows better fit to the data, with residuals

distributed more narrowly and homogeneously. The ZTP model is more suitable for count data without zero values, and this suitability becomes more pronounced at lower mean values.

In Table 1, the AIC and BIC values for the models at means of 100, 50, and 10 are identical. As the mean decreases, these values also decrease. However, when the mean is set to 5, the AIC and BIC values begin to diverge, and the differences grow significantly at a mean of 3. For larger mean values, no significant differences are observed in AIC and BIC values across the regression models employed. So, for dependent variables with a mean of 5 or less, the ZTP model is preferable to the Poisson model.

Table 2. AIC and BIC values of NB and ZTNB regressions for different mean values

MODEL	μ=100		μ=50		μ=10		μ=5		μ=3	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
NB	9755.43	9814.32	8446	8505	5785	5844	4781	4840	3995	4054
ZTNB	9755.43	9814.32	8446	8505	5782	5841	4745	4804	3851	3910

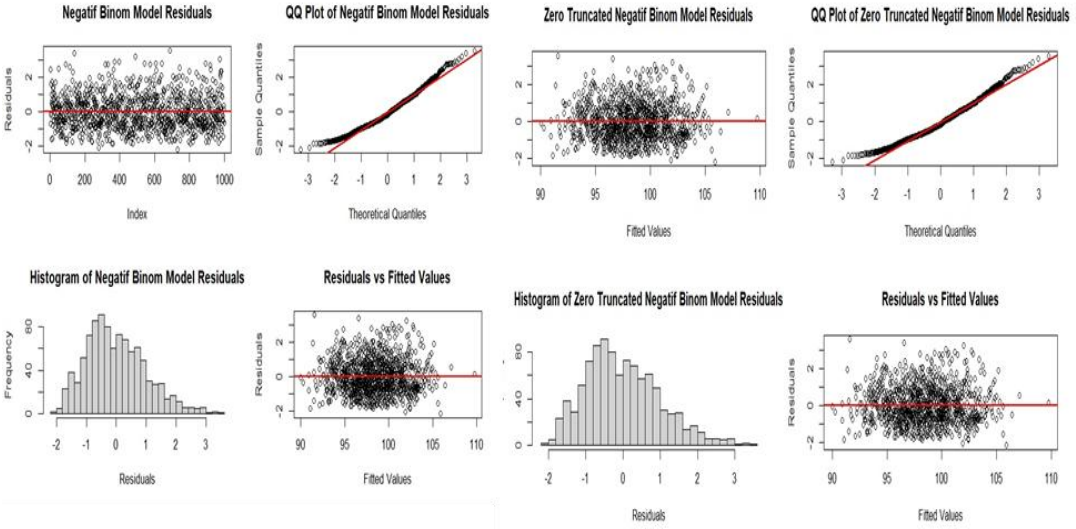


Figure 3. Residual plots for a mean value of 100.

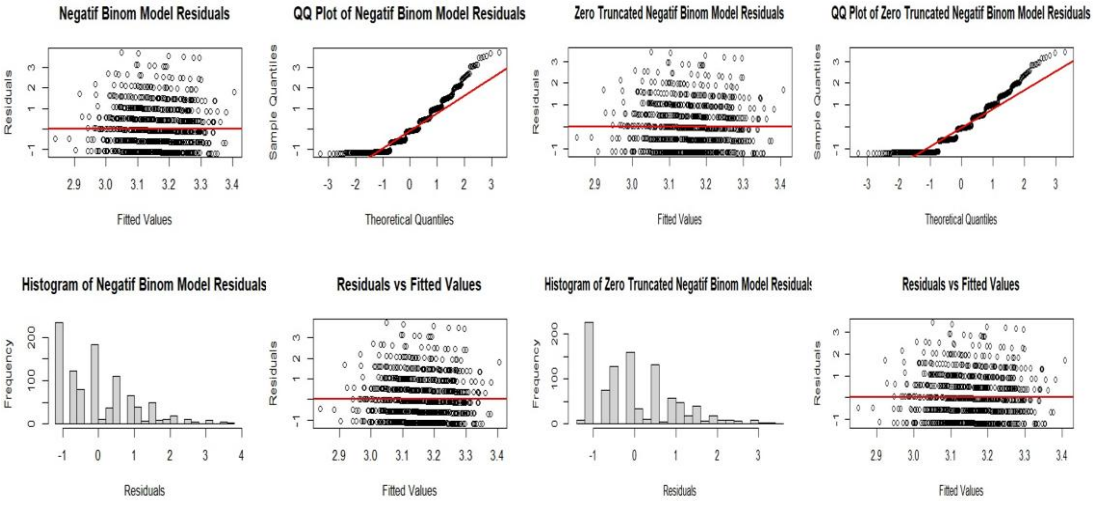


Figure 4. Residual plots for a mean value of 3.

Similar results were obtained in Negative Binomial (NB) and Zero-Truncated Negative Binomial (ZTNB) regressions as in the Poisson application. When the mean is 3, the residual plot of the

ZTNB model shows a narrower and more homogeneous distribution compared to the NB model. This indicates that the ZTNB regression is a more suitable model for the dataset.

Moreover, the AIC and BIC values diverge when the mean is set to 10, and as the mean decreases, it becomes increasingly evident that the ZTNB model is the more appropriate choice.

3.2. Real dataset application

In this study, one of the dependent variables, the crime of bad treatment, was found to consist of positive values, with a minimum value of 1 and a maximum value of 18. Additionally, it exhibited a Poisson distribution with a standard deviation of 1.603 and a mean of 1.54.
Four different regression models were employed in the study, and the results are presented in Table 3.

Table 3. AIC and BIC values of regression models used in the study

Model	AIC value	BIC value
Poisson(PR)	1052.4	1128.43
Negatif binomial(NB)	1054.4	1134.44
Zero truncated Poisson(ZTP)	590.89	666.97
Zero truncated negatif binomial(ZTNB)	1021.81	1101.89

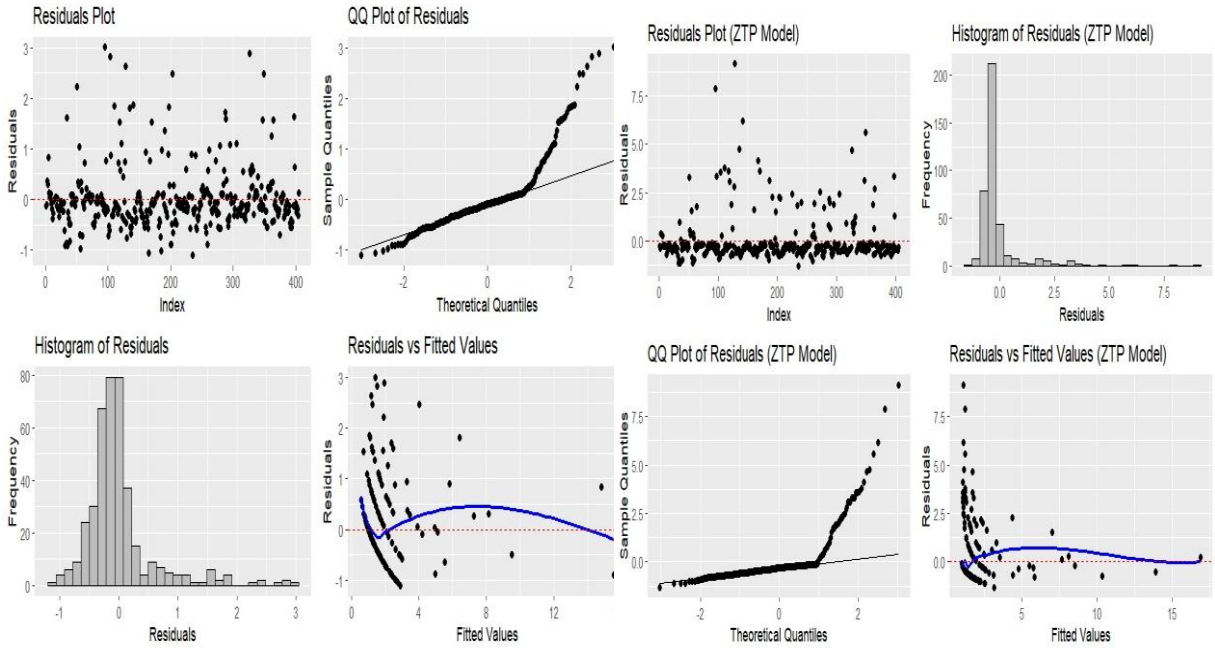


Figure 5. Residual values of Poisson regression and ZTP regression models (Real Data).

The second dependent variable used in the study is "sexual crimes." This variable has a minimum value of 1, a maximum value of 324, a mean of 21.56, and a variance of 1355.98. The variance being significantly larger than the mean indicates overdispersion. The regression models employed and the results obtained are presented in Table 4.

Table 4. Regression methods used and their AIC and BIC values

Model	AIC value	BIC value
Poisson(PR)	3556.7	3632.821
Negatif Binomial(NB)	2622.3	2702.341
Zero Truncated Poisson(ZTP)	3553.009	3629.082
Zero Truncated Negatif Binomial(ZTNB)	2595.403	2675.481

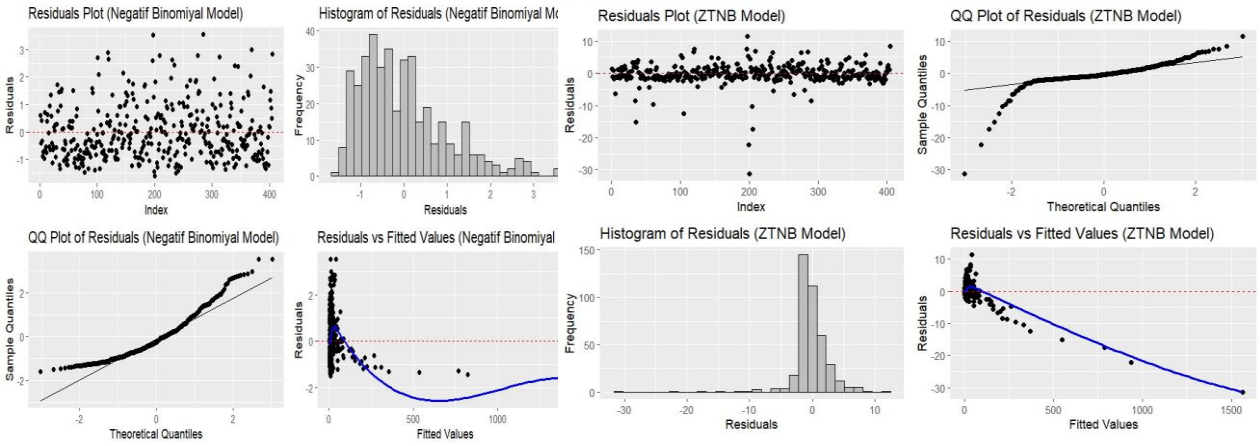


Figure 6. Residual values of negative binomial regression and ZTNB regression models (Real Data).

The residual plot of the negative binomial regression demonstrates a more balanced residual distribution. Systematic errors are less pronounced, and residuals are generally symmetric around zero. In contrast, the residual plots for the ZTNB regression show significant deviations and systematic errors, particularly at the negative extremes.

Table 5 provides the parameter estimates and standard errors for the Poisson and ZTP regression models when the dependent variable is "bad treatment."

Table 5. Comparison of independent variable estimates for PR and ZTP models when the dependent variable is "bad treatment" crime

VARIABLE	Poisson Parameter Estimates exp (β)	Poisson Standard Errors	ZTP Parameter Estimates exp (β)	ZTP Standard Errors
Sexual crimes	**0.98*** **	0.003	**0.96*** **	0.004
2010	**1.04**	0.13	**1.11**	0.22
2011	**1.01**	0.14	**1.02**	0.26
2012	**0.96**	0.14	**0.74**	0.28
2013	**0.89**	0.16	**0.63**	0.33
Aegean	**0.99**	0.17	**0.96**	0.26
Marmara	**1.30**	0.17	**1.56**	0.24
Black Sea	**0.92**	0.17	**0.48***	0.34
Central Anatolia	**0.96**	0.16	**0.69**	0.28
Eastern Anatolia	**0.81**	0.17	**0.23****	0.48
Sotheast Anatolia	**0.72**	0.2	**0.15*** **	0.54
Homicide	**0.97***	0.001	**1.007*** **	0.001
Crimes related with firearms and knifes	**0.99**	0.001	**1.003**	0.001
Production and commerce of drugs	**1.001**	0.0006	**1.001**	0.0007
Use and purchase of drugs	**1.003****	0.2	**0.99**	0.001
Robbery	**0.99*** **	0.001	**0.99*** **	0.0001
Kidnapping	**1.001**	0.004	**1.004**	0.005
Assault	**1.002*** **	0.0005	**1.004*** **	0.0007

*(p<0.1), **(p<0.01), ***(p<0.001)

In Table 5, significant differences were observed between the parameter estimates of the Poisson and ZTP models. For instance, the "Eastern Anatolia" variable was estimated as 0.81 in the Poisson model and 0.23 in the ZTP model. Similar differences were observed for other variables as well.

The AIC and BIC values presented in Table 3 also reflect this difference, with values of 1052.4 and 1128.43 for the Poisson regression, and 590.89 and 666.97 for the ZTP regression, respectively (Table 3).

Similarly, the "Southeast Anatolia" variable was estimated as 0.72 in the Poisson model, whereas this value dropped to 0.15 in the ZTP model. This indicates that the ZTP model is more sensitive to zero-truncated datasets and better captures the structure of the data.

In contrast, variables such as "Robbery" (0.99 in both models) and "Production and Commerce of Drugs" (1.001 in both models) showed consistent parameter estimates across the models. This suggests that the effect of zero truncation on these variables is limited, and their sensitivity to model selection is relatively low.

Table 6. Comparison of Independent Variable Estimates for NB and ZTNB Models When the Dependent Variable is "Sexual Offenses"

VARIABLE	NB Parameter Estimates exp (β)	NB Standard Errors	ZTNB Parameter Estimates exp (β)	NB Standard Errors
Bad treatment	**1.03**	0.02	**1.03**	0.02
2010	**1.41*** **	0.10	**1.47*** **	0.11
2011	**1.16**	0.10	**1.19**	0.11
2012	**2.07*** **	0.10	**2.18*** **	0.11
2013	**3.09*** **	0.11	**3.35** **	0.12
Aegean	**1.15**	0.12	**1.16**	0.13
Marmara	**1.06**	0.12	**1.06**	0.13
Black sea	**0.65*** **	0.12	**0.64*** **	0.13
Central Anatolia	**0.79**	0.12	**0.78**	0.13
Eastern Anatolia	**0.37*** **	0.12	**0.34*** **	0.14
Southeast Anatolia	**0.36*** **	0.14	**0.33*** **	0.15
Homicide	**1.006*** **	0.001	**1.007*** **	0.001
Crimes related with firearms and knifes	**1.004*** **	0.001	**1.004*** **	0.001
Production and commerce of drugs	**0.99***	0.0005	**0.99***	0.0006
Use and purchase of drugs	**0.99** **	0.001	**0.99** **	0.001
Robbery	**0.99***	0.001	**0.99***	0.001
Kidnapping	**0.99**	0.003	**0.99**	0.003
Assault	**1.001** **	0.0004	**1.001** **	0.0004

*($p<0.1$), **($p<0.01$), ***($p<0.001$)

In Table 6, it was observed that the parameter estimates of the NB regression and ZTNB regression models were very close to each other, with some variables even having identical estimates. The AIC and BIC values for the two methods also showed similar results.

The AIC value for the NB regression was found to be 2622.3, while for the ZTNB regression, it was slightly lower at 2595.403. Similarly, the BIC values were also close, with the NB regression having a BIC value of 2702.341 and the ZTNB regression yielding a BIC value of 2675.481.

These results suggest that both models perform similarly for this dataset, with ZTNB showing a slight advantage in terms of model fit as indicated by its lower AIC and BIC values.

4. Conclusion

In this study, the effects of the mean of the dependent variable on the choice of regression models were examined in models where the dependent variable consisted of positive count data. Both simulation studies and real datasets were utilized to evaluate these effects. Zero-truncated models (ZTP and ZTNB) were compared with classical Poisson and Negative Binomial (NB) models, and model fit was assessed using AIC, BIC values, and residual plots.

The findings indicated that when the mean of the dependent variable was 5 or lower, zero-truncated models provided better fit compared to classical models. Using classical models for datasets with low mean values may result in biased parameter estimates, which could have critical implications in certain fields. For example, in areas such as medicine or pharmacology, where minimal differences in values can lead to significant outcomes, this could directly influence theoretical and application strategies.

The AIC and BIC values obtained in the real data application suggested that the ZTP regression model explained the data more effectively. However, residual plots indicated that the ZTP model could produce larger errors for low-fitness observations. A similar result was observed for the ZTNB regression model. When the mean of the dependent variable was less than 5, zero-truncated models generally performed better in terms of overall fit. If producing large errors for specific observations is acceptable and general fit is prioritized, zero-truncated models can be preferred.

When the mean of the dependent variable exceeded 5, no significant differences were observed between zero-truncated models and classical models. However, slight deviations were noted in the NB models when the mean was slightly above 5, indicating the need for more caution in model selection. This consideration is particularly important for datasets where the dependent variable exhibits overdispersion.

In light of the study's general findings, it is recommended that the mean of the dependent variable be taken into account when modeling positive count-dependent variables. Specifically, zero-truncated models should be preferred for dependent variables with a mean of 5 or less. In such cases, the use of classical methods could lead to biased parameter estimates depending on the research topic, scientific discipline, and dataset characteristics.

Acknowledgements

The authors gratefully acknowledge the financial support provided by the Scientific Research Projects Council of Van Yüzüncü Yıl University under project number FDK-2019-7987.

References

Agresti, A. (1997). *Categorical data analysis*. John Wiley & Sons.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.

Cox, R. (1983). Some remarks on overdispersion. *Biometrika, 70*(2), 269-274.

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment, 91*(2), 121-136. https://doi.org/10.1080/00223890802634175

Creel, M. D., & Loomis, J. B. (1990). Theoretical and empirical advantages of truncated count data estimators for analysis of deer hunting in California. *American Journal of Agricultural Economics, 72*(2), 434-441. https://doi.org/10.2307/1242345

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). Wiley.

Hilbe, J. M. (2011). *Negative binomial regression* (2nd ed.). Cambridge University Press.

Hilbe, J. M. (2014). *Modeling count data*. Cambridge University Press.

Jansakul, N., & Hinde, J. P. (2002). Score tests for zero-inflated Poisson models. *Computational Statistics & Data Analysis, 40*(1), 75-96. https://doi.org/10.1016/S0167-9473(01)00104-9

Khoshgoftaar, T. M., Gao, K., & Szabo, R. M. (2005). Comparing software fault predictions of pure and zero-inflated Poisson regression models. *International Journal of Systems Science, 36*(11), 707-715. https://doi.org/10.1080/00207720500159995

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). *Applied linear regression models* (4th ed.). McGraw-Hill/Irwin.

Lawal, B. H. (2012). Zero-inflated count regression models with applications to some examples. *Quality & Quantity*, *46*, 19-38. https://doi.org/10.1007/s11135-010-9324-x

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, *15*(3), 209-225. https://doi.org/10.2307/3314912

Lee, A. H., Wang, K., Yau, K. K., & Somerford, P. J. (2003). Truncated negative binomial mixed regression modelling of ischaemic stroke hospitalizations. *Statistics in Medicine*, *22*(7), 1129-1139. https://doi.org/10.1002/sim.1419

Liu, X., Saat, M. R., Qin, X., & Barkan, C. P. L. (2013). Analysis of U.S. freight-train derailment severity using zero-truncated negative binomial regression and quantile regression. *Accident Analysis & Prevention*, *59*, 87-93. https://doi.org/10.1016/j.aap.2013.04.039

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Sage Publications. https://doi.org/10.1080/00401706.1998.10485496

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.

Puza, B., Johnson, H., O'Neill, T., & Barry, S. (2008). Bayesian truncated Poisson regression with application to Dutch illegal immigrant data. *Communications in Statistics - Simulation and Computation*, *37*(8), 1565-1577. https://doi.org/10.1080/03610910802117073

Rose, C. E., Martin, S. W., Wannemuehler, K. A., & Plikaytis, B. D. (2006). On the use of zero-inflated and hurdle models for modeling vaccine adverse event count data. *Journal of Biopharmaceutical Statistics*, *16*(4), 463-481. https://doi.org/10.1080/10543400600719384

Sáez-Castillo, A. J., & Conde-Sánchez, A. (2013). A hyper-Poisson regression model for overdispersed and underdispersed count data. *Computational Statistics & Data Analysis*, *61*, 148-157. https://doi.org/10.1016/j.csda.2012.12.009

Sileshi, G. (2008). The excess-zero problem in soil animal count data and choice of appropriate models for statistical inference. *Pedobiologia*, *52*(1), 1-17. https://doi.org/10.1016/j.pedobi.2007.11.003

Simo, T., Esa, L., Anti, M., Jaakko, T., Harri, S., Selina, J., & Timo, A. (2007). Self-reported health problems and sickness absence in different age groups predominantly engaged in physical work. *Occupational and Environmental Medicine*, *64*(11), 739-746. https://doi.org/10.1136/oem.2006.027789

Thygesen, H. H., & Zwinderman, A. H. (2006). Modeling sage data with a truncated gamma-Poisson model. *BMC Bioinformatics*, *7*, 157. https://doi.org/10.1186/1471-2105-7-157

Van Der Heijden, P. G., Cruyff, M., & Van Houwelingen, H. C. (2003). Estimating the size of a criminal population from police records using the truncated Poisson regression model. *Statistica Neerlandica*, *57*(3), 289-304. https://doi.org/10.1111/1467-9574.00232

Winkelmann, R. (2008). *Econometric analysis of count data*. Springer-Verlag Berlin Heidelberg.

Yesilova, A., Kaya, Y., Kaki, B., & Kasap, İ. (2010). Analysis of plant protection studies with excess zeros using zero-inflated and negative binomial hurdle models. *GU Journal of Science*, *23*(2), 131-136.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, *27*(8), 1-25. https://doi.org/10.18637/jss.v027.i08