ANKARA YILDIRIM BEYAZIT UNIVERSITY

AYBU Business Journal, 3(1), 86-103
https://dergipark.org.tr/tr/pub/abj
https://aybu.edu.tr/isletme/tr/sayfa/5218/AYBU-BUSINESS-JOURNAL

AYBU Business Journal

Submitted: 22 November 2022
Accepted: 9 January 2023
Published Online: 1 March 2023

RESEARCH ARTICLE

AN APPLICATION ON FINANCE DATA FOR CRITICAL LIMITS OF ASSUMPTIONS IN COUNT DATA

Ali İhsan ÇETİN[a*]

[a]Business School, Ankara Yıldırım Beyazıt University, Ankara, Turkey

Abstract

Regression analysis is used to analyze many cases in real life. The type of data obtained varies according to the type of cases and the variable to be studied. For example, in the most widely used linear regression analysis, the dependent variable must be continuous. Otherwise, the desired results will have a high standard error and will be inconsistent. Alternative regression techniques have been developed according to the types of dependent variable. Two of them are Poisson and Negative Binomial Regression, which are frequently used in case of discrete dependent variables. However, the fact that the dependent variable is discrete does not mean that correct results will be obtained by applying the aforementioned models. Because besides the type of dependent variable, the parameters of the relevant models have been developed and various sub-models have emerged according to its distribution and spread. In this study, a data set containing real data such as HDI, GDP and credit score, which has an crucially important place in the field of finance, was used and the results were compared and interpreted using AIC, RMSE and MAE metrics by applying Poisson, Negative Binomial Regression and their zero-truncated models according to the characteristics of the data set. The empirical results can be interpreted as the negative binomial regression model gives better results when the dependent variable has insufficient distribution, but Poisson regression produces more meaningful results when the assumptions are at the limit. In addition, it was examined whether the number of zeros in the data set is sufficient to go to the Zero Truncated models. As a result, it has been revealed that the Negative Binomial distribution cannot always be used in cases where analysis will be made with Poisson regression, even though there is over- or under-distribution according to the assumptions.

Keyword

Count Data, Poisson, Negative Binomial, Zero Truncated, Finance

*Corresponding author.
Contact: Ali İhsan Çetin aliihsancetin@aybu.edu.tr

To cite this article: Çetin, A. İ., (2023). An Application On Finance Data For Critical Limits Of Assumptions In Count Data. *AYBU Business Journal*, 3(1), 86-103.

1. INTRODUCTION

Regression analysis is a statistical analysis method that has been used and studied for a long time, thanks to its consistency and easy interpretability. With the increase in the importance of data, the increase in access to information and data all over the sectors in the world, statistical analysis techniques, which have become more valuable due to the competitive environment, have started to be used with advanced estimators or reinterpreted with high-developed parameters. Others are being replaced by machine learning models that are more suitable for the real world and do not require assumptions.

Classical linear regression analysis focuses on establishing a connection between the target variable and the features and revealing the properties of this connection. Here, the structure of the target variable is vital for choosing the right regression analysis. Because it is inevitable to obtain inconsistent, biased, ineffective and high standard error results as a result of wrong model selection, and there are many assumptions such as normality, homoscedasticity, multicollinearity and linearity that can hinder the analysis in classical linear regression analysis. Looking at real-life data, the most common type of response variable is the continuous variable. However, a lot of work has been done on data sets with discrete response variable. There are also target variables of the count data type, which is a subtype of the discrete type. Count data regression models can produce convincing results when the response variable is the number of times a case occurs. Some alternative models based on poisson and negative binomial distributions, which are more suitable for count data, have been proposed.

Poisson regression (PR) and negative binomial regression (NBR) models are used in many fields such as management and organization, traffic, biology, finance, medicine, actuarial. However, even if the case of interest never happened, some problems occurred in the use of the regression methods in question and the desired consistent results could not be produced. The reason for this situation is that there is overdispersion in the data due to the large number of zeros and the variance grows. In such a case, classical PR and NBR models may not be sufficient for count data modeling. As an alternative, zero-truncated methods have been developed for this purpose. Zero-Truncated Poisson Regression (ZPR) and Zero-Truncated Negative Binomial Regression (ZNBR) have been proposed to deal with this problem in the presence of extreme zeros.

2. LITERATURE REVIEW

One of the most useful and easy methods in the analysis of count data is the Poisson Regression model, which is one of the generalized linear models. While constructing the Poisson regression model, the poisson distribution is used to determine the probabilities of the data. One of the important features of the distribution in question is that the mean of the result obtained is equal to its variance. However, it is observed that the variance exceeds the mean in the applications performed in general. This situation is defined as overdispersion. In these cases, negative binomial regression models work. (Kabacoff, 2015). The most important difference between the classical regression model and the poisson regression model is that in the poisson distribution created for the dependent variable in the poisson regression model, the values consist of non-negative data. However, it fits a discrete distribution.

Graff et al. (2020), created two different poisson regression models to predict the effectiveness of forest fires between one and five days. It has been observed that the created regression models give more accurate results than permanent models in predicting fires.

Gao et al. (2021), in the study titled "Dispersion modeling of outstanding claims with double Poisson regression models", it is aimed to develop a new distribution structure in the double Poisson chain ladder model by ignoring the existing limitations of the extremely dispersed Poisson chain-ladder models, which are frequently used for compensation provisions in insurance. It is concluded that the proposed method is much more flexible than the currently used methods.

Benz et al. (2021), in the study titled "Estimating the change in soccer's home advantage during the Covid-19 pandemic using bivariate Poisson regression", it was investigated how the matches played without fans affect the home team in an environment where the home team has a great advantage in the matches played with spectators. For this purpose, bivariate Poisson regression models were used by taking data from 17 different leagues. As a result of the research, it is seen that the findings are mixed, the advantage disappears in some leagues, and the advantage increases in others.

In the study titled "Households' Access to Communication and Information Technologies: A Poisson Regression Analysis" prepared by Ercan (2021), it is aimed to investigate the factors that will affect the number of information and communication technology tools in households with the Poisson regression model. In the study, it was concluded that factors such as the city of residence, the difficulty of accessing schools, income status, and the number of students affect the number of technological devices.

Vicuña et al. (2021), in the study titled "Forecasting the 2020 COVID-19 Epidemic: A Multivariate Quasi-Poisson Regression to Model the Evolution of New Cases in Chile", it is aimed to analyze the situation of Covid-19 in Chile, to prevent its spread and to analyze the alternative ways that may be necessary in order to pass the process with the least possible damage. According to the results, the spread of the disease was expected to be higher, but contrary to expectations, it was seen that the rate of spread decreased in the future thanks to the quarantine policies implemented in the country.

İşçi et al. (2021), in the study titled "Comparison of Some Count Models in Case of Excessive Zeros: An Application", it was stated that multi-zero poisson regression and poisson hurdle regression models were used in case the census data had multiple zeros, and negative binomial regression and negative binomial hurdle regression models were used in case of overdispersion. Comparisons of these models were made using a sample data set.

Models except the classical models should be considered as powerful alternatives for modelling count and give better insights to the researchers in applying statistics on working similar data structures Yıldırım et al. (2022).

There are cases where the dependent variable is discrete but not categorical. Such situations are called count data. Count data is among the generalized linear models in practice. There are many models that give precise properties of counting results. However, the Poisson regression is considered the starting point for many analyzes. The Poisson regression model is the most commonly used and simplest method for counting data. In the Poisson regression model, the link function connecting the linear structure of the independent variables to the expected value of the dependent variable is logarithmic. With this model, the probability of counting is determined by the Poisson distribution. The most distinctive feature of the model is that the conditional mean of the result is equal to the conditional variance (Deniz, 2005). However, in practice, sometimes the conditional variance may exceed the conditional mean.

In the Poisson distribution, when the variance is greater than the mean, it is called overdispersion, and when the variance is less than the mean, it is called underdispersion (Cox, 1983). In the case of overdispersion in the dependent variable, two paths are generally followed. The first is by estimating a propagation parameter (α) with it the test statistics and correcting for the residuals. The second is the application of the negative binomial regression model, which is one of the methods that eliminates the effect of overdispersion (Hilbe 2007). In practice, we see that the negative binomial regression model is widely used, while the generalized Poisson regression model and the Poisson quasi-Lindley regression model are also used. In the data set, the goodness-of-fit statistic of deviance is widely used to determine whether there is overdispersion.

Count regression models have been used in many fields from past to present. King (1988) analyzed the party change behavior of members of the House of Representatives in the United States between 1802 and 1876. The number of members of the House of Representatives who changed parties in a year was used as the independent variable.

Michener and Tighe (1992) examined fatal accidents on highways in the United States.

Using the Poisson regression model, Khalat et al. (1997) the differences in fertility levels in Beirut during the war, Burg et al. (1998) examined the rise of male and female academics in the academic labor market.

Şahin (2002) applied Poisson regression for the determinants of strikes in Turkey for the 1964-1998 period and Arısoy and Yaprak (2016) for the 1984-2015 period. Memiş and Önder (2018) compared artificial data with Poisson regression estimation methods.

Data obtained by count generally do not show normal distribution and have a structure starting from 0 and consisting of positive values (Zorn, 1996; Cameron & Travedi, 1998).

Since the dependent variable obtained based on the count does not show a normal distribution, applying linear regression to this type of data causes some statistical problems. The application of linear regression methods to such data may result in biased parameter estimates. The use of methods that do not destroy the original structure of the data eliminates this problem. The dependent variable obtained by count shows the Poisson distribution and Poisson regression is used in its modeling (Ridout et al., 1998).

The most basic feature of the Poisson distribution is that the variance and the mean are equal. In practice, this feature is not always possible (Frome, 1983; Rose et al., 2006). When this equality is not achieved, there are two situations. These; If the variance is greater than the mean, it is called overdispersion, and if the variance is less than the mean, it is called underdispersion (Banik & Kibria, 2008; Jansakul & Hinde, 2009).

There may be various causes of overspread. Some of these are many zero values or unobserved heterogeneity (Rose et al., 2006).

In case of overdispersion, applying Poisson regression leads to biased parameter estimates (Cox, 1983). To avoid this and to obtain more accurate results, negative binomial (Negative Binomial = NB) or generalized Poisson (Generalized Poisson = GP) regression is used (Sileshi, 2008).

While NB takes into account overdispersion, GP is an alternative method to Poisson regression that considers over- and under-dispersion. The existence of zero values in the dependent variable and the intensity of these zero values should be taken into account when modeling the dependent variable. If there is more zero density than expected in the dataset, this is called zero value weight (Zero Inflation=ZI) (Lachenbruch, 2002). In this case, it is recommended to use zero value-weighted regression models for modeling such dependent variables (Khoshgoftaar et al., 2005). In this case, the methods that can be used are zero-weighted Poisson (Zero Inflated Poisson=ZIP), zero-weighted negative binomial (Zero Inflated Negative Binomial=ZINB) and Hurdle models (Cameron and Trivedi, 1998; Hall, 2000; Minami et al., 2007).

In this study, Poisson regression, which is frequently preferred in cases where the dependent variable consists of count data, and negative binomial regression models, which are a generalization of Poisson models, and zero-truncated types of these models were applied. It has been tried to be explained in detail with an example set. Human Development Index and Credit Scoring, which are important macro indicators frequently used in finance, were included in the study as independent variables, and Gross Domestic Product as dependent variables. Model analyzes were made with the open-source Python program. AIC, RMSE, and MAE values for all models were obtained and interpreted.

3. MATERIAL AND METHOD

3.1.Poisson Distribution and Regression Model

The Poisson distribution models the dependent count variable y with the formula (Field, 2009):

$$P((Y = y|\mu) = \frac{e^{-\mu}\,\mu^{y}}{y!} \qquad (3.2)$$

where μ is the mean of the distribution and y is the counting variable expressing the frequency or rate desired to occur (μ > 0, y = 0, 1, 2, …). In log-linear versions of the model, the mean is shown as:

$$\mu_i = exp(\boldsymbol{x_i}'\beta_j) \qquad (3.2)$$

$$= exp(x_{1i}\beta_1)exp(x_{2i}\beta_2) \cdots exp(x_{ki}\beta_k) \qquad (2.3)$$

As can be seen from Eq. (3.3), in Poisson regression, models are created with the assumption that μ parameter is determined by a series of $x_i$variables. Hence, the μ parameter can be represented as an exponential mean function:

$$E(y_i/x_i) = \mu_i = e^{xi\beta j} \qquad (3.4)$$

$$\mu_i = exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k) \qquad (3.5)$$

$$\mu_i = \mu(\boldsymbol{x_i}'\beta) \qquad (3.6)$$

Here, the regression coefficients β_1, β_2, ⋯, β_k are unknown parameters to be estimated using the data set. If the natural logarithm of Eq. (3.2) is taken, a linear shape of the conditional mean according to the $x_i$variables is obtained. Although the dependent variable is a discrete random variable, Poisson regression models are nonlinear models due to their functional form as can be seen in Eq. (3.4).

Regression coefficients are estimated using the maximum likelihood method. The logarithm of the likelihood function is (Field, 2009):

$$ln[L(y, \beta)] = \sum_{i=1}^{n} y_i ln[\mu(\boldsymbol{x}_{\boldsymbol{i}}'\beta)] - \sum_{i=1}^{n} \mu(\boldsymbol{x}_{\boldsymbol{i}}'\beta) - \sum_{i=1}^{n} ln(y_i!) \qquad (3.7)$$

With the solution obtained by taking the differential of Eq. (3.7) with respect to the β parameter, the Poisson maximum likelihood estimators $\hat{\boldsymbol{\beta}}_{\boldsymbol{i}}$ are calculated using the following equation:

$$\sum_{i=1}^{n}(y_i - exp(\boldsymbol{x}_{\boldsymbol{i}}'\beta))x_i = 0 \qquad (3.8)$$

Likelihood equations are created by taking derivatives according to each regression coefficient and equating the result to zero. Doing so leads to the emergence of a set of nonlinear equations that accept no closed form solution. For this reason, iterative algorithms such as Newton-Raphson are used to find regression coefficients that maximize likelihood. It is seen that Fisher iteration method is used frequently in the literature.

One of the consequences of not finding an analytical solution for $\hat{\boldsymbol{\beta}}$ is the difficulty of obtaining exact distribution results for $\hat{\boldsymbol{\beta}}$ estimators. There are several ways to infer for $\hat{\boldsymbol{\beta}}$. First, consider $\hat{\boldsymbol{\beta}}$ as the estimator that maximizes Eq. (3.7) and maximum likelihood theory should be applied. Second, the estimator $\hat{\boldsymbol{\beta}}$ must be considered as defined by Eq. (3.8). These equations have a similar interpretation to Ordinary Least Squares (OLS) estimators. Therefore, the unweighted residuals $(y_i - \mu_i)$ are orthogonal to the estimators. Therefore, as for OLS, it is possible to perform the inference only under assumptions about the mean and possibly variance. This is a generalized linear models approach. Third, since Eq. (3.6) implies the equation $E(y_i - exp(\boldsymbol{x}'\beta))x_i = 0$, an estimator can be defined that is the solution of the moment condition in the sample. This estimator is also the solution of Eq. (3.8). This approach is the moment-based models approach (Cameron and Trivedi, 2013:23).

3.2.Negative Binomial Regression Model

The Poisson regression model is used when the mean of the distribution is equal to its variance. However, this situation is rarely encountered in practice. Negative Binomial regression model is used, which ensures the efficiency of parameter estimations, as a result of overdispersion when the variance of the distribution is greater than the mean of the distribution (Agresti, 2007: 81). Counting variables in applications do not show normal distribution, as they usually have variance greater or less than the mean. In such cases, Negative Binomial regression should be applied instead of Poisson regression model or test statistics and residuals should be corrected with the spread parameter. Negative Binomial regression is a generalization of Poisson's regression, in which the variance is equal to the mean calculated by the Poisson model, loosening the constraining assumption. This model is based on a Poisson-Gamma mixed distribution. For the Negative Binomial model, the variance is given by Eq. (3.9).

$$r_s = 1 - \frac{6\sum_{i=0}^{n} d_i^2}{n(n^2-1)} \quad (3.9)$$

According to this model, the Negative Binomial regression model is expressed as in Eq. (3.10).

$$Var(yi|xi) = \lambda i + \alpha \lambda i^2 \quad (3.10)$$

In this model, $\beta_1, \beta_2, \ldots, \beta_k$ represent unknown parameters.

$$\lambda_i = \exp(\ln(t_i)\beta_{1i}x_{1i} + \beta_{2i}x_{2i}, \ldots, \beta_{ki}x_{ki}) \quad (3.11)$$

3.3. Akaike Information Criterion

The criterion proposed by Akaike and widely used in comparing different models is defined as the Akaike information criterion. Akaike information criterion is expressed as;

$$AIC = 2k - 2ln(\mathrm{L}) \quad (3.12)$$

In this equation, L is the maximum value of the log likelihood function; k represents the number of explanatory variables. Among the existing models, the model with the smallest AIC value calculated by Eq. (3.12) is selected as the appropriate model (Akaike, 1973).

4. FINDINGS AND DISCUSSION

In this study, it is aimed to examine the factors affecting the 2011-2020 volatility of GDP through counting models. In the study, the data set was obtained by taking the GDP average of the years in question and counting the years that were 10% below and above the average. As the attributes affecting this variable, the scores given to the countries by the Fitch Credit Rating Agency in the same years and the human development index (HDI) of the countries were used. While creating the dataset, 10 data of 90 countries between the years 2011-2020 were used for this purpose. AIC, RMSE and MAE metrics of Poisson Regression, Negative Binomial Regression, Zero Truncated Poisson Regression and Zero Truncated Negative Binomial Regression models were compared. Generalized linear models were used in Poisson and Negative Binomial Regression analysis.

Fitch Credit Rating denotes credit scores in letters. Credit scores have numerical equivalents in the literature (Genc and Basar, 2019). In Table 1, their equivalents transformed to numerical data are given.

Table 1. Credit Scoring Transformation Table

Credit Ratings					
TE	S&P	Moody's	Fitch	DBRS	Description
100	AAA	Aaa	AAA	AAA	Prime
95	AA+	Aa1	AA+	AA (high)	High grade
90	AA	Aa2	AA	AA	
85	AA-	Aa3	AA-	AA (low)	

80	A+	A1	A+	A (high)	Upper medium grade
75	A	A2	A	A	
70	A-	A3	A-	A (low)	
65	BBB+	Baa1	BBB+	BBB (high)	Lower medium grade
60	BBB	Baa2	BBB	BBB	
55	BBB-	Baa3	BBB-	BBB (low)	
50	BB+	Ba1	BB+	BB (high)	Non-investment grade
45	BB	Ba2	BB	BB	speculative
40	BB-	Ba3	BB-	BB (low)	
35	B+	B1	B+	B (high)	Highly speculative
30	B	B2	B	B	
25	B-	B3	B-	B (low)	
20	CCC+	Caa1	CCC	CCC (high)	Substantial risks
15	CCC	Caa2		CCC	Extremely speculative
10	CCC-	Caa3		CCC (low)	In default with little
5	CC	Ca		CC	prospect for recovery
5	C	C		C	
0	D	/	DDD	D	In default
		/	DD		
			D		
			RD		
			WD		

The histogram graph of the GDP deviation numbers, which is the dependent variable, is given in the Figure 1.

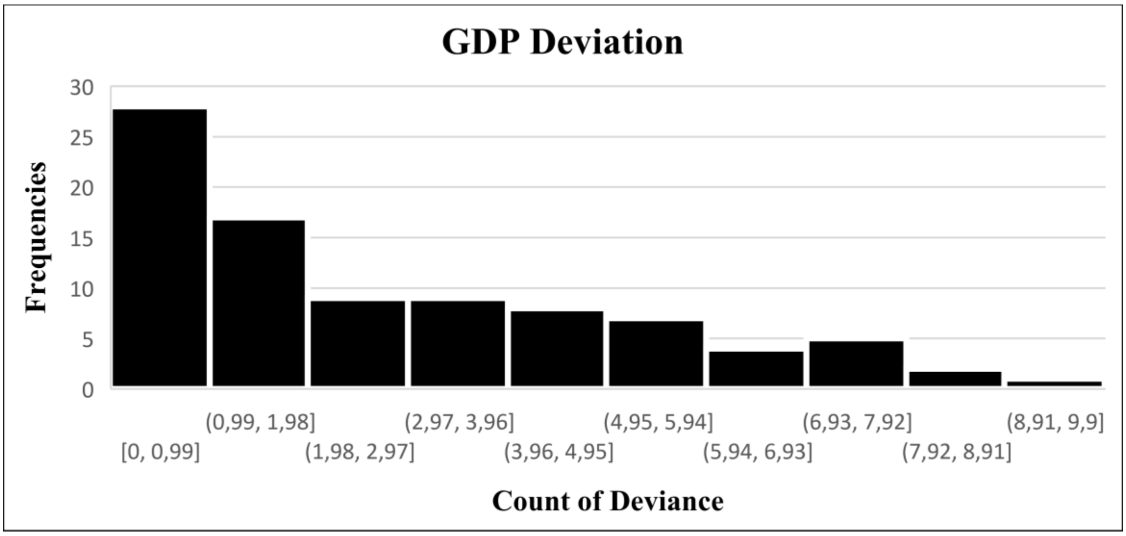

Figure 1. Count of Deviance Frequencies

Frequencies, Relative Frequencies and Poisson probability values for GDP deviation are given in the Table 2.

Table 2. GDP Deviance Frequencies Values

GDP Deviation X=x	**Frequency f_x**	**xf_x**	**Relative Frequency f_x/n**	**ʎ=Weighted Mean**	**Poisson Probability**
0	28	0	0.31	2.37	0.094
1	17	17	0.19	2.37	0.222
2	9	18	0.10	2.37	0.263
3	9	27	0.10	2.37	0.207
4	8	32	0.09	2.37	0.123
5	7	35	0.08	2.37	0.058
6	4	24	0.04	2.37	0.023
7	5	35	0.06	2.37	0.008
8	2	16	0.02	2.37	0.002
9	1	9	0.01	2.37	0.001
Toplam	**90**	**213**	**1.000**		**0.99982**

One of the variables affecting the GDP deviation is the credit score variable. Figure 2 shows the average credit scores for each category of GDP deviations of 90 countries between 2011 and 2020.

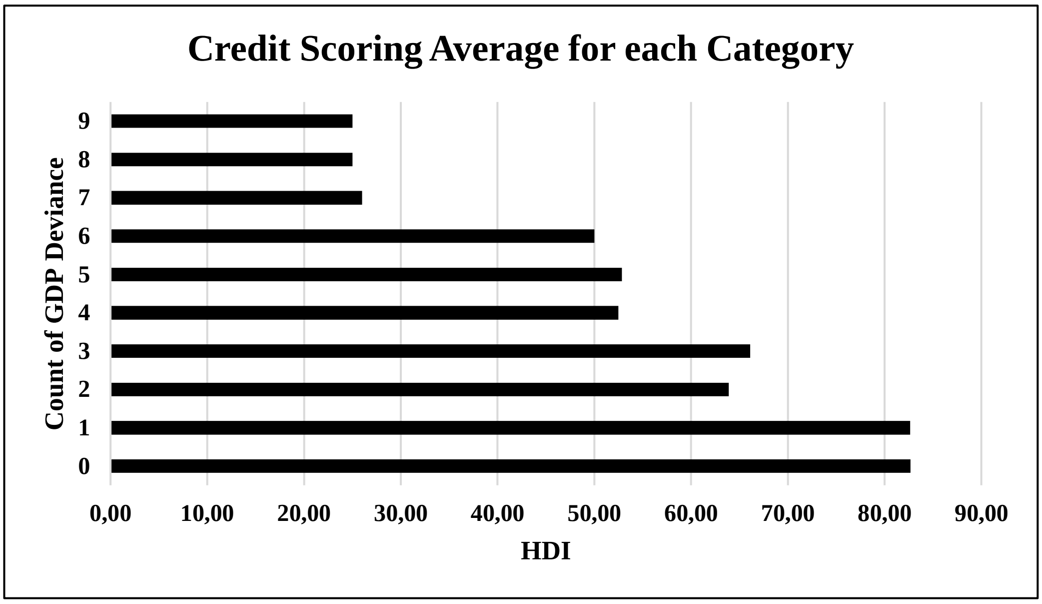

Figure 2. Average Credit Scores for each Category of GDP Deviations

The other variable affecting the GDP deviation is the Human Development Index variable. Figure 3 shows the average HDI for each category of GDP deviations of 90 countries between 2011 and 2020.

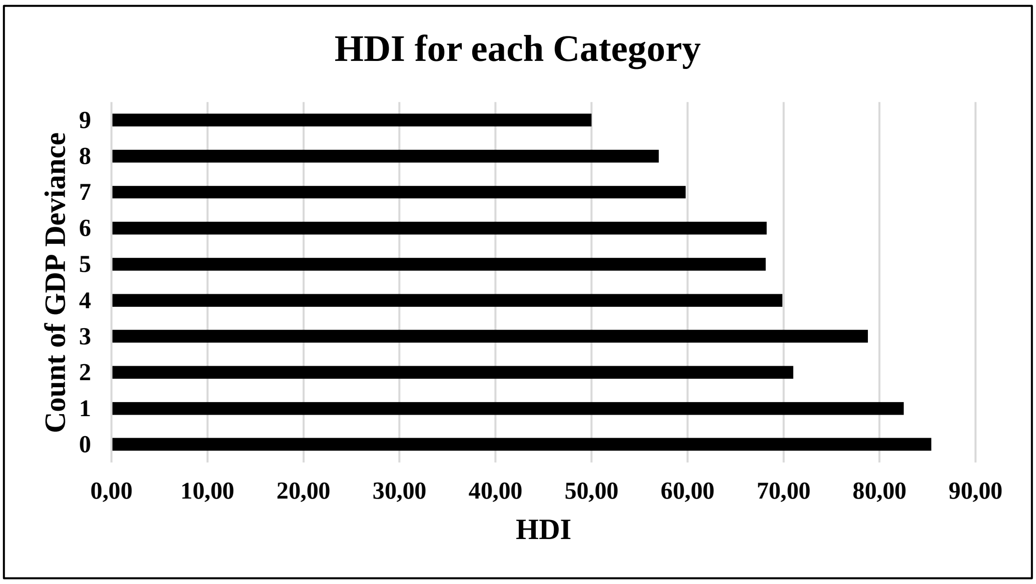

Figure 3. Average HDI for each Category of GDP Deviations

Table 3. Descriptive Statistics

	Count of GDP Deviance	**Human Development Index**	**Credit Score**
count	90.00	90.00	90.00
mean	2.36	77.32	68.44
std	2.44	13.42	23.29
min	0.00	50.00	20.00
25%	0.00	66.00	50.00
50%	1.50	79.50	75.00
75%	4.00	90.00	88.75
max	9.00	95.00	100.00

Although the mean of the dependent variable for the data used in the study is smaller than the variance, since the difference is very small, the difference can be neglected and the analysis can be continued with the assumption of equality. In this case, it is expected that the metrics of Poisson distribution will give better results in the Negative Binomial distribution as explained above. (2.36 <= 2.44). The reason for this situation is the absence of underdispersion or overdispersion in the data. In addition, although the number of zero values is higher than the other values, it is not known whether it requires the use of zero truncated models due to the low difference. For this reason, the analysis was performed with zero-truncated models of the poisson and negative binomial, and the results are given in Table 4.

Table 4. Comparison of Model Metrics

	POISSON	**NEGATIVE BINOMIAL**	**ZIP**	**ZINB**
AIC	**288.49**	**290.83**	304.39	299.53
RMSE	**1.69**	1.76	**1.56**	2.14
MAE	**1.17**	1.21	**1.07**	1.30

As can be seen in the Table 4, Poisson and Negative Binomial Regression according to the AIC criterion are the most suitable models for this data set. On the other hand, when looking at RMSE and MAE, models with low error are Poisson and Zero Truncated Poisson. So, there is no case of under- or over-dispersion in the data because the metrics of the Negative Binomial

Regression model were high. Therefore, if the mean and variance are very close to each other, the analysis can be continued without any dispersion problem and equality. Also, the Zero Truncated models do not give high accuracy predictions based on the results of the AIC measurement metric. This situation can be interpreted as follows; The number of zeros in the data set in the study is not large enough to require the application of Zero Truncated models.

In Figure 4, Figure 5, Figure 6 and Figure 7, there are the measurement results and prediction graphs of the compared models.

```
AIC           : 288.4929360135082
RMSE POISSON  : 1.6901678568964271
MAE  POISSON  : 1.177760300298698
```

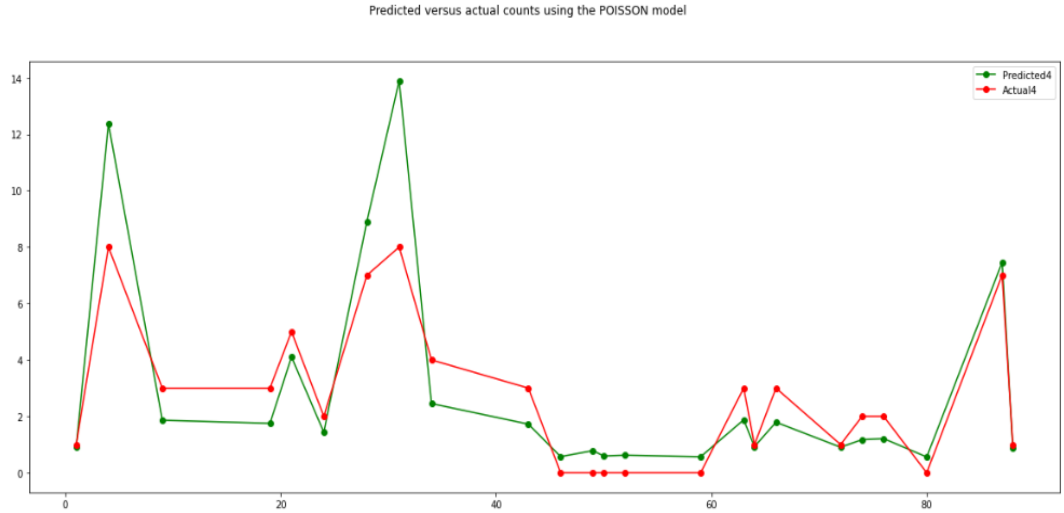

Figure 4. Poisson Model Prediction Graph and Model Metrics

```
AIC           : 290.83450081447063
RMSE NB       : 1.7651409821449673
MAE  NB       : 1.211132389182864
```

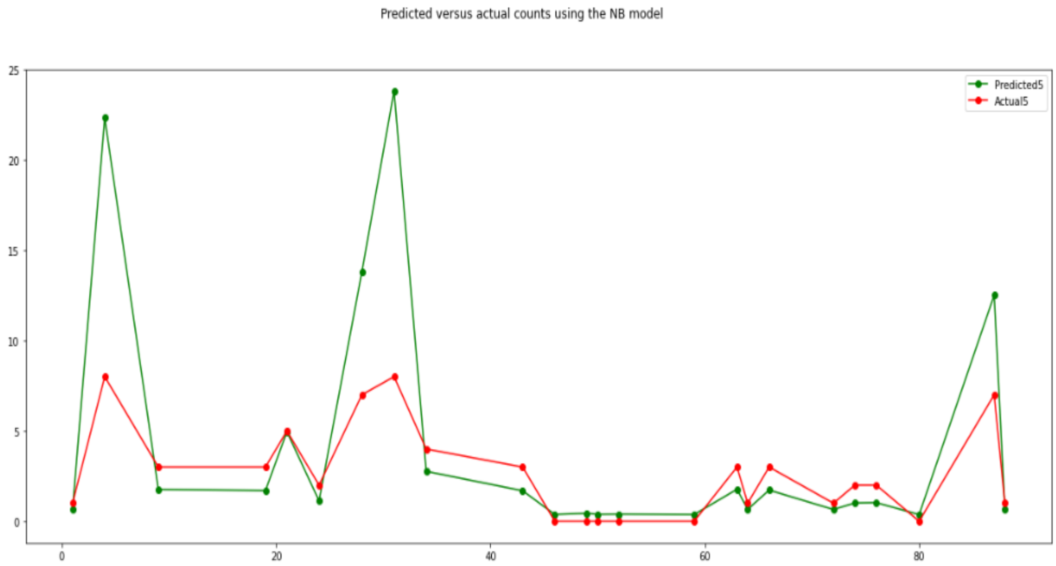

Figure 5. Negative Binomial Model Prediction Graph and Model Metrics

```
AIC         : 304.39648300163407
RMSE ZIP    : 1.5699132696376144
MAE  ZIP    : 1.0746182921318879
```

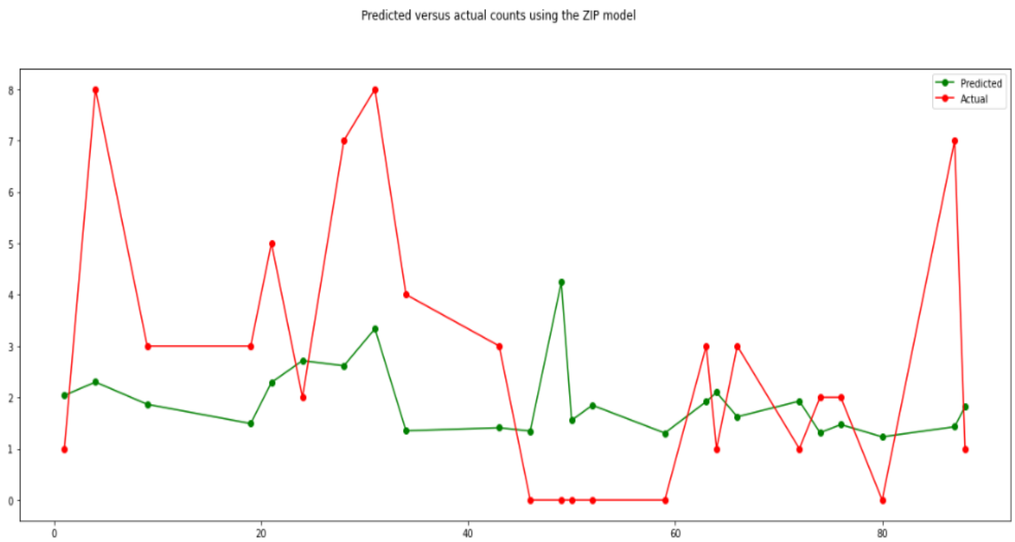

Figure 6. Zero Truncated Poisson Model Prediction Graph and Model Metrics

```
AIC         : 299.5341651402747
RMSE ZINB   : 2.141804500920089
MAE  ZINB   : 1.3049847758937874
```

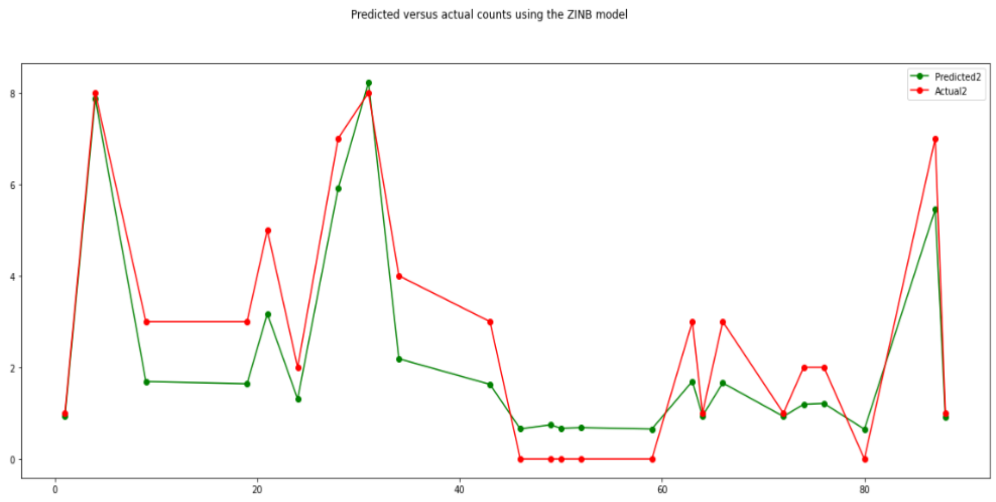

Figure 7. Zero Truncated Negative Binomial Model Prediction Graph and Model Metrics

5. RECOMMENDATION AND FUTURE STUDIES

If the dependent variable is discrete or count data, analyzes using linear regression models will yield inconsistent and highly erroneous results. Therefore, alternative regression models are used for count data. The most well-known among them are Poisson and negative binomial regression models. The most important condition for using the Poisson model is that the conditional variance value is equal to the conditional mean value. This is rare in practice. In many applications, the conditional variance value exceeds the conditional mean value. In such cases it is not correct to use Poisson regression. When such an overdispersion (or underdispersion) occurs, negative binomial regression should be applied or corrected by test statistics and residuals dispersion parameter. In a negative binomial distribution, the variance is assumed to be a square function of the mean. This statement is one of the most important assumptions for the elimination of overdispersion.

The results obtained from the analyzes and the issues that can be focused on in future studies are given below:

• For the data used in the study, the mean of the dependent variable was smaller than the variance (2.36 < 2.44), but since there was little difference, the analyzes applied in both cases were applied in order to talk about any dispersion problem, and as a result, the difference was due to better Poisson Regression measurement results. It can be said that the variance is equal to the mean.

• According to AIC metric results; Poisson Regression gave the lowest value. This means that the multiplicity of zero values in the data is insufficient to apply the Zero Truncated Poisson model. In addition, it has been understood that it is not necessary to apply the Negative Binomial model, which is applied when there is a dispersion problem.

• According to RMSE and MAE metric results; It has been observed that Poisson models make less erroneous predictions than Negative Binomial Models. This once again confirms that there is no dispersion problem.

• In this study conducted with GDP deviation numbers, the variable of credit score was found to be significant in Poisson regression and Negative Binomial regression models, while the variable of HDI was found to be insignificant. However, in the Zero Truncated models, both of the independent variables were found to be insignificant. This can once again be interpreted as the number of zeros being insufficient to go for Zero Truncated models.

• Based on these comments, it can be said that which model will be preferred and which metrics will give more meaningful results in studies where mean and variance are not equal but very close to each other as a subject that can be focused on for future studies.

• In addition, in the literature, there is not a clear statement about how many zeros should be present in the data set, either numerically or as a percentage, in order to be able to apply Zero Truncated models. Future studies can be conducted with a data set suitable for this subject.

REFERENCES

Agresti, A., & Franklin, C. (2007). The art and science of learning from data. Upper Saddle River, New Jersey, 88.

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika, 60(2), 255-265.

Arısoy, İ. ve Yaprak, Ş., (2016). 1984-2015 Türkiye'de Grevlerin Belirleyicileri, Ekonomi Bilimleri Dergisi, 8(2). 130-116.

Banik, S., & Kibria, B. M. G. (2008). On some discrete models and their comparisons: An empirical comparative study. In Proceedings of The 5th Sino-International Symposium on Probability, Statistics, and Quantitative Management KU/FGU/JUFE Taipei, Taiwan, ROC May (Vol. 17, pp. 41-56).

Benz, L.S. & Lopez, M.J. (2021). Estimating the change in soccer's home advantage during the Covid-19 pandemic using bivariate Poisson regression. AStA Advances in Statistical Analysis: A Journal of the German Statistical Society. 1-28.

Burg, B.V.D., Siegers. J., and Ebmer, R.W., (1998). Gender and Promotion in the Academic Labour Market Labour, 12(4), 701-713.

Cameron, A.C. and Trivedi, P.K., (2013). Regression Analysis of Count Data. Cambridge University Press. New York.

Cox, R., (1983). Some Remarks on Overdispersion, Biometrika, 70: 269-274.

Deniz, Ö., (2005). Poisson Regresyon Analizi, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 4(7), 59- 72.

Ercan, U. (2021). Hanehalklarının İletişim ve Bilgi Teknolojilerine Erişimi: Bir Poisson Regresyon Analizi. Akdeniz Iletisim. 35, 402-422.

Field, A. (2009). Discovering Statistics Using SPSS Third Edition (and sex and drugs and rock 'n' roll). SAGE Publications Ltd, London, England.

Frome, E. L. (1983). The analysis of rates using Poisson regression models. Biometrics, 665-674.

Gao, G., Meng, S. & Shi, Y. (2021). Dispersion modelling of outstanding claims with double Poisson regression models. In Insurance Mathematics and Economics November: Part B. 101,572-586.

Genc, E. G., & Basar, O. D. (2019). Comparison of country ratings of credit rating agencies with moora method. Business and Economics Research Journal, 10(2), 391-404.

Graff, C.A., Coffield, S.R., Chen, Y., Foufoula-Georgiou, E., Randerson, J.T. & Smyth, P. (2020). Forecasting daily wildfire activity using poisson regression. IEEE Transactions on Geoscience and Remote Sensing IEEE Trans. Geosci. Remote Sensing Geoscience and Remote Sensing, IEEE Transactions on. 58(7), 4837-4851.

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

Hilbe, J.M., (2007). Negative Binomial Regression. Cambridge, U.K.

İşçi Güneri, Ö., Durmuş, B. & İncekırık, A. (2021). Comparison of some count models in case of excessive zeros: An application. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 20(40), 247-268.

Jansakul, N. and Hinde, J. (2009). Score tests for extra-zero modelsin zero-inflated negative binomial models.Communications inStatistics-Simulation and Computation38,92–108.

Kabacoff, R.I. (2015). R in Action (Second Edt.) Data analysis and graphics with R. Manning. Shelter Island.

Khalat, M., Deep, M. and Courbage, Y., (1997). Fertility Levels and Differentials in Beirut during Wartime: An Indirect Estimation Based on Maternity Registers, Population Studies, 51(1), 85-92.

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), 705-715.

King, G., (1988). Statistical Models for Political Science Event Counts: Bias in Conventional Procedures and Evidence for the Exponential Poisson Regression Model, American Journal of Political Science, 3(3). 838-863.

Lachenbruch, P. A. (2002). Analysis of data with excess zeros. Statistical methods in medical research, 11(4), 297-302.

Memiş, M. and Önder, H., (2018). Poisson Regresyon Tahmin Yöntemlerinin Karşılaştırılması, Black Sea Journal of Engineering and Science, 1(4), 140-146.

Michener, R. and Tighe, C., (1992). Gender and Promotion in the Academic Labour Market. American Economic Review, 82(2). 452-56.

Minami, M., Lennert-Cody, C. E., Gao, W., & Román-Verdesoto, M. (2007). Modeling shark bycatch: the zero-inflated negative binomial regression model with smoothing. Fisheries Research, 84(2), 210-221.

Ridout, M., Demétrio, C. G., & Hinde, J. (1998, December). Models for count data with many zeros. In Proceedings of the XIXth international biometric conference (Vol. 19, pp. 179-192). Cape Town, South Africa: International Biometric Society Invited Papers.

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.

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.

Şahin, H., (2002). Poisson Regresyon Uygulaması: Türkiye'deki Grevlerin Belirleyicileri 1964-1998, Doğuş Üniversitesi Dergisi, 5, 173-180.

Vicuña, M. I., Vásquez C. & Quiroga B. F. (2021). Forecasting the 2020 COVID-19 Epidemic: A multivariate quasi-poisson regression to model the evolution of new cases in Chile. Frontiers in Public Health, 9, 1-7.

Yıldırım, G., Kaçıranlar, S., & Yıldırım, H. (2022). Poisson and negative binomial regression models for zero-inflated data: an experimental study. Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics, 71(2), 601-615.

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