UJM 9 (2) 2020

UNNES Journal of Mathematics

http://journal.unnes.ac.id/sju/index.php/ujm

SEMIPARAMETRIK MULTILEVEL ZERO-INFLATED GENERALIZED POISSON REGRESSION MODELING ON TRAFFIC ACCIDENT DATA

Bani Muhamad Isa✉**, Nur Karomah Dwidayati**

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia
Gedung D7 Lt. 1, Kampus Sekaran Gunungpati, Semarang 50229

Info Artikel

Sejarah Artikel:
Diterima Oktober 2019
Disetujui Desember 2020
Dipublikasikan Desember 2020

Keywords:
Overdispersi, semiparametrik, zero-inflated, multilevel, regresi poisson, zero-inflated generalized poisson

Abstrak

Penelitian ini bertujuan untuk memodelkan data kecelakaan lalu lintas di Kabupaten Temanggung dengan model regresi semiparametrik multilevel *zero-inflated generalized poisson.* Regresi semiparametrik multilevel *zero-inflated generalized poisson* adalah model regresi untuk menganalisis data berdistribusi poisson dengan struktur data bertingkat yang mengalami overdispersi serta terdapat komponen parametrik dan nonparametrik pada variabel bebasnya. Penelitian ini mengunakan variabel banyak kecelakaan sebagai variabel respon, serta variabel banyak pelanggaran traffic light, banyak pelanggaran pengendara tidak punya SIM, banyak keceakaan karena kendaraan tidak fit, banyak kecelakaan karena jalan rusak sebagai variabel bebas. Metode yang digunakan untuk mengestimasi parameter model yaitu dengan metode M*aximum Likelihood Ratio* (MLE) dengan algoritma Ekspektasi Maksimalisasi (EM). Setelah dilakukan estimasi parameter dan uji kesesuaian model deengan Uji Wald, maka didapatkan bentuk model regresi semiparametrik multilevel *zero-inflated generalized poisson* $P(Y_{pqr} = y_{pqr}) = \frac{\exp(-0{,}9756)}{1+\exp(-12{,}4681+22{,}6673T_{1pqr}+u_1(T_{1pqr}-k_1)}$ dengan AIC model count 144.0032 dan AIC model *zero-inflation* −63.0016.

Abstract

This study aims to model the data of traffic accidents in Temanggung Regency with a multilevel zero-inflated generalized poisson semiparametric regression model. Multilevel zero-inflated generalized poisson semiparametric regression is a regression model for analyzing poisson distribution data with stratified data structures that are overdispersed and there are parametric and nonparametric components in the independent variable. This study uses the variable of many accidents as the response variable, as well as the variable of many traffic light violations, many violations of drivers not having a SIM, many accidents because the vehicle is not fit, many accidents due to damaged roads as the independent variable. The method used to estimate the model parameters is the Maximum Likelihood Ratio (MLE) method with the Maximization Expectation (EM) algorithm. After estimating the parameters and the suitability of the test model with the Wald Test, then the model shape is obtained a semiparametric regression multilevel zero inflated generaized poison $P(Y_{pqr} = y_{pqr}) = \frac{exp(-0{,}9756)}{1+exp(-12{,}4681+22{,}6673T_{1pqr}+u_1(T_{1pqr}-k_1)}$ *with AIC count model 144.0032 and AIC zero-inflation model -63.0016.*

How to cite:

Isa, Bani M & Dwidayati, Nur K. 2020. Pemodelan Regresi Semiparametrik Multilevel Zero-Inflated Generalized Poisson. *UNNES Journal of Mathematics*. 9 (2) : 26-33

✉Alamat korespondensi:
E-mail: banimuhamadisa555@gmail.com

p-ISSN 2252-6943
e-ISSN 2460-5859

INTRODUCTION

Regression Analysis is a simple method for analyzing the functional relationship between several variables, namely the response variable (response) or commonly also called the dependent variable (dependent variable) and the independent variable (independent variable). In the regression analysis there are three approaches to estimating the regression curve, namely the parametric regression approach, the non-parametric regression approach, and the semiparametric regression approach. Parametric regression is used if the regression curve follows a certain pattern or forms clear data patterns such as linear, quadratic, and cubic. In nonparametric regression, it is used if there are no known parametric components. While semiparametric regression is used if a regression component contains variables that can be solved by parametric regression but there are variables that must be solved using non-parametric regression analysis

Response variables can be discrete data or count data, while the regression model that can be used to analyze the data count is Poisson regression.

Poisson regression model is a standard model used to analyze the count data. The assumption of the Poisson distribution is that the mean (mean) must be equal to the variance (equidispersion). However, such conditions are difficult to fulfill. Often the average value is not the same as the value of the variance. Variance values greater than the mean are called overdispersions. The value of variance that is smaller than the average value is called underdispersion.

Overdispersion of the data can occur because the proportion of zero values is too excessive in the response variable (excess zero). To solve the problem another model is needed. Models that can be used to solve these problems are by utilizing Zero-Inflated regression models, such as Zero-Inflated Poisson, Zero-Inflated Negative Binomial, Zero Inflated Poisson mixed-effect, Zero-Inflated Generalized Poisson, and others.

Multilevel regression is regression with data in the form of hierarchical or stratified data. In some cases, many hierarchical data are found where the data has an excess of zero values. Resulting in an overdispersion of the data

The increase in the number of motorized vehicles over time is directly proportional to the number of traffic accidents. In this study, the authors conducted a study in Temanggung Regency which is one of the densest cities that naturally experience driving safety problems.

This study applies Multilevel Zero-Inflated Generalized Poisson semiparametric regression using traffic accident data in the Temanggung Regency area in 2018. The data uses variables that are assumed to meet parametric and non-parametric components.

This study aims to model the accident data in Temanggung Regency with a multilevel zer-Inflated Generalized Poisson semiparametric regression model, and to know the estimation of many accidents based on the model.

Semiparametric regression is a combination of parametric regression with non-parametric regression. This means that there is a relationship between the response variable and the independent variable that cannot be solved by parametric regression analysis alone but there must be resolved by non-parametric regression. The semiparametric regression model can be written as follows:

$$Y\_i = x\_i\,\beta + f(t\_i) + \varepsilon\_i; \quad i = 1,2,\ldots,n \quad (1)$$

Y_i is the i variable, x_i is a parametric component, f (x_i) is a regression function of unknown regression curve shape and ε_i is a random error with ε_i ~ N (0, σ^2). According to Alain (2009: 224), overdispersion means variance greater than average. The existence of overdispersion can be known through the distance or the difference between the deviation with the degree of freedom. If the difference results in a value greater than one, the model is said to be overdispersed.

Zero-Inflated Generalized Poisson (ZIGP) regression is a method for dealing with overdispersion with the proportion of zero value data being around 65.7% (Famoye and Singh, 2006). ZIGP model as follows:

$$\boldsymbol{P(Y_i = y_i | x_i, z_i)} = \begin{cases} \boldsymbol{\pi_i + (1-\pi_i) f(\mu_i, \omega, y_i), y_i = 0} \\ \boldsymbol{(1-\pi_i) f(\mu, \omega, y_i) \quad , y_i > 0} \end{cases}$$

With f (μ,ω,y_i), y_i = 0,1,2,... is a GP regression model with 0 <π_i <1. In the function μ_i = μ_i (x_i) and π_i = π_i (z_i) fulfill the link function as follows (Famoye and Singh, 2006)

$$\boldsymbol{\log(\mu_i) = \sum_{j=1}^{k} x_{ij}\beta = x^T\beta}$$

and

$$\boldsymbol{logit\ (\pi_i) = \log(\pi_i[1-\pi_i])^{-1} = \sum_{j=1}^{m} Z_{ij}\delta}$$

The MZIGP model considers a hierarchical situation of more than one level. In this study, the hierarchical situation consists of two levels where $y_{ijk}\ (i = 1,2,\ldots,m;\ j = 1,2,\ldots,n_i;\ k =$

$1,2,\ldots,n_{ij}$) is the response variable with ZIGP distribution. The ijk index shows that there are k people in the j-th number combination of the i-th group. The total number of combinations is $\sum_{i=1}^{m} n_i = n$ and the total number of individuals (observations) is $= \sum_{i=1}^{m}\sum_{j=1}^{n_i} n_{ij}$. A combined model for μ and ϕ, ie

$$\boldsymbol{ln(\mu) = x_i^T\beta} \textbf{ and } \boldsymbol{log(\phi) = log\left(\frac{\phi}{1-\phi}\right) = a_i^T\alpha}$$

When individual responses from people belonging to different groups are independent, specific correlations to the group and combinations are anticipated. This dependency can be modeled by considering a suitable random effect on linear predictors. As in the ZIGP regression model for calculated data, $log(\mu)$ and $log(\phi/1-\phi)$allow linear dependence on several explanatory variables. Thus, the linear predictor of η_ijk can be expressed as

$$ln(\mu_{ijk}) = \eta_{ijk} = x_{ijk}^T\beta + u_i + v_{ij} \quad (2)$$

and

$$logit(\omega_{ijk}) = \eta_{ijk} = x_{ijk}^T\tau + u_i + v_{ij} \quad (3)$$

or

$$P(Y_{ijk} = y_{ijk}) = \frac{\exp(x_{ijk}^T\beta+\varepsilon_i)}{1+\exp(x_{ijk}^T\tau+\varepsilon_i)} \quad (4)$$

METHODS

The data taken is secondary data that is traffic accident data in Temanggung Regency. Data was collected at the Temanggung District Police Station. The method used in this study is the interview method. One method of data collection is by interviewing, which is getting information by asking questions directly to the source.

In this study the authors interviewed police members related to traffic accident data at the Temanggung District Police Station. The variables observed in this study, according to the nature of the Poisson regression which requires that there be at least a dependent variable with one or more independent variables can be specified as follows.

1. The dependent variable (Y) is the number of traffic accidents in Temanggung Regency
2. The independent variable (X) that will be used in this research is the number of traffic light violations ($X1$), the number of violations motorists do not have a SIM ($X2$), the number of accidents due to vehicles not fit ($T1$), the number of accidents due to damaged roads ($T2$).

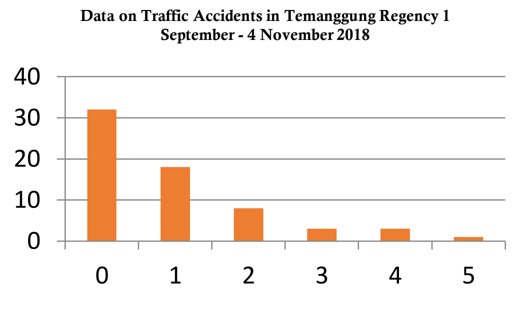
Multilevel data design consists of the number of traffic accidents taken from data over 2 months over 4 days, with monthly data as level 1 (p). Then the data per week as level 2 (q), and data per day as level 3 (r).

In this research, it is explained how the solution to overcome overdispersion by using Multilevel Zero-Inflated Generalized Poisson semiparametric regression analysis method and parameter estimation is done using the maximum likelihood (MLE) method. The steps of the analysis are as follows:

1. Input data
2. Test parametric and non-parametric assumptions
3. Test the assumption of the Poisson distribution
4. Test the assumption of equidispersion
5. Identifying the overdispersion of response data
6. Estimating the parameter values of the Multilevel Zero-Inflated Generalized Poisson semiparametric regression model with MLE
7. Forming a Multilevel Zero-Inflated Generalized Poisson semiparametric regression model
8. Test the significance of model parameters with the Wald test
9. Test the feasibility of the model with AIC

DISCUSSION

Presentation of data on many accidents in Temanggung Regency can be seen in Figure 1.



Based on Figure 1 it can be seen that the data has an excess of zero value (excess zero). The variables in this study use the variable many accidents (Y) as the response variable, and the independent variable is a lot of traffic light violations (X_1), many drivers do not have a SIM (X_2), many accidents because the vehicle is not fit (T_1), many accidents due to road damaged (T_2). Normality test on independent variable data can be seen in Table 1.

Table 1. Test Results for Independent Variables Normality

Variable	Sig. Value
Traffic Light ($X1$)	0,056
No SIM ($X2$)	0,076
Vehicle Not Fit ($T1$)	0,000
Damaged roads ($T2$)	0,000

Based on Table 1, it can be seen that the Traffic Light ($\boldsymbol{X1}$) and No SIM ($\boldsymbol{X2}$) variables have significance values of $\mathbf{0.056} > \mathbf{0.05}$ and $\mathbf{0.076} > \mathbf{0.05}$. Then the variables $\boldsymbol{X1}$ and $\boldsymbol{X2}$ come from population data that are normally distributed so as a parametric component. While the Unfit Vehicle ($\boldsymbol{T1}$) and Damaged Road ($\boldsymbol{T2}$) variables have a significance value of $\mathbf{0,000} < \mathbf{0.05}$ and $\mathbf{0,000} < \mathbf{0.05}$. Then the variables $\boldsymbol{T1}$ and $\boldsymbol{T2}$ come from populations not normally distributed so as a nonparametric component. So the semiparametric regression model used for the data is:

$$\boldsymbol{Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 T_1 + u_1(T_1 - k_1) + \beta_4 T_2 + u_2(T_2 - k_2) + \varepsilon_i}$$

Testing of means and variances can be seen in Table 2.

Table 2. Average Test Results and Variance

Mean	Varians
0,9231	1.47836538

Based on Table 2, it is known that the Mean value and the Variance value are different. Then the variable Y does not meet the equidispersion assumption. Then tested whether overdispersion or underdispersion. To find out the existence of overdispersion or underdispersion, there are two ways, first, by comparing the average value with the value of variance.

If the value of the variance is greater than the average, overdispersion occurs. If the value of the variance is smaller than the average, underdispersion occurs. Based on table 2 it is known that the value of variance = 1.47836538> average = 0.9231. Then it can be concluded that there is an overdispersion of the $\boldsymbol{Y}$ response variable data. The second way is by dividing the Pearson chi-square value and residual deviance by the degree of freedom if the value is more than 1 then overdispersion and if the value is less than 1 then underdispersion. By testing using the R 3.3.1 program the results obtained residual deviance value of 36,823 with degrees of freedom (df) 28. Pearson chi square value of 2989.335 then,

a) $\frac{DevianceResidual}{df} = \frac{36.823}{28} = 1{,}315107$

b) $\frac{Pearson\ chi\ sqaure}{df} = \frac{2989.335}{28} = 106.762$

Based on the results of the division of residual deviance values and Pearson chi square values can be known to be more than 1. Then it can be concluded that the Y response variable data has overdispersion. Because of overdispersion in Poisson regression and multilevel, this research used a model to overcome the overdispersion problem, namely the Zero-Inflated Generalized Poisson (ZIGP) regression model.

Multilevel zero-inflated genealized poisson semiparametric regression model is a model used to improve Poisson semiparametric regression models with multilevel structured data that has overdispersed for response variable data due to excess zero. In the traffic accident data in Temanggung Regency as level 1, which is Month, level 2, which is Sunday, and level 3, which is Day. There are 3 months, 10 weeks, 65 days, so the number of data samples is 65 samples.

Based on equations (1) and (4) the shape of the zero-inflated generalized poisson semiparametric regression model equation can be stated as follows:

$$P(Y_{pqr} = y_{pqr}) = \frac{exp(\beta_0 + \beta_1 X_{1ipqr} + \beta_2 X_{2ipqr} + \beta_3 T_{1ipqr} + u_1(T_{1ipqr} - k_1) + \beta_4 T_{2ipqr} + u_2(T_{2ipqr} - k_2))}{1 + \exp(\tau_0 + \tau_1 X_{1pqr} + \tau_2 X_{2pqr} + \tau_3 T_{1pqr} + u_1(T_{1ipqr} - k_1) + \tau_1 T_{2pqr} + u_2(T_{2ipqr} - k_2))}$$

Information:

p : Month ($p = 1, 2, 3$)

q : Week ($q = 1, \dots, 10$)

r : Day ($r = 1, \dots, 65$)

After estimating parameters for the multilevel zero-inflated generalized poisson semiparametric regression model using the R 3.3.1 program by activating the glmmADMB package, the estimation results for the count model in Table 3 are as follows.

Table 3. Estimated Multilevel Zero-Inflated Generalized Poisson semiparametric regression models for the count model (poisson with log link)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.9756	0.3214	-3.04	0.0024 **
β_1	0.0432	0.0412	1.05	0.2948
β_2	0.0121	0.0442	0.27	0.7851
β_3	0.1308	0.5250	0.25	0.8033
β_3	0.3845	0.2997	1.28	0.1994
σ^2 (Month)		1.125e-07		
σ^2 (Week)		0.0001903		
σ^2 (Day)		1.123e-07		

Based on the estimation results shown in Table 3 the count model equation is obtained for μ as follows.

$$\boldsymbol{ln(\mu) = -0,9756 + 0,0432X_{1ipqr} + 0,0121X_{2ipqr} + 0,1308T_{1ipqr} + u_1(T_{1ipqr} - k_1) + 0,3845\,T_{2ipqr} + u_2(T_{2ipqr} - k_2) + v_p + w_{pq}} \quad (5)$$

The interpretation of the results of the estimated model (5) is as follows:

a. The constant value is -0.9756, meaning that if the free variable is zero then the number of accidents is exp (-0.9756) = 0.376966. This is because the number of accidents is influenced by independent variable factors other than the model.
b. Koefisien X_1 bernilai 0,0432, artinya setiap terjadi pelanggaran karena traffic light The coefficient X_1 is 0.0432, meaning that every violation occurs because the traffic light causes an increase in the expected value of the number of accidents, namely exp (0.0432) = 1.044146.
c. The coefficient X_2 is 0.0121, meaning that every violation occurs because the driver does not have a SIM causing an increase in the expected value of the number of accidents that is exp (0.0121) = 1.012173.
d. The coefficient T_1 is worth 0.1308, meaning that every time there is a lack of fit of the driver's vehicle causing an increase in the expected value of the number of accidents that is exp (0.1308) = 1.139739.
e. The coefficient of T_2 is 0.3845, meaning that every time there is damage it causes an increase in the expected value of the number of accidents, namely exp (0.3845) = 1.468879.
f. The v_p variable is a constant value of random effects as the specific value of monthly data for p = 1.2.3 with a variance value of 1.125e-07.
g. The w_{pq} variable is a constant value of random effects as the specific value of data per week for p = 1,2,3 and q = 1,2, ... 10 with a variance value of 0.0001903.

Table 4. Estimated Multilevel Zero-Inflated Generalized Poisson semiparametric regression models for the zero-inflation model (binomial with logit link)

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-12.4681	3.9571	-3.15	0.0016 **
β_1	0.0669	0.4377	0.15	0.8785
β_2	0.0262	0.4691	0.06	0.9555
β_3	22.6673	5.7698	3.93	8.5e-05 ***
β_4	-0.6691	-0.6691	-0.15	0.8792
σ^2 (Bulan)		0.002035		
σ^2 (Minggu)		0.0007097		
σ^2 (Hari)		2073		

Based on the estimation results shown in Table 4 the zero-inflation model equation for $\boldsymbol{\omega}$ is obtained as follows:

$$logit(\omega) = -12,4681 + 0,0669x_{1pqr} + 0,0262x_{2pqr} + 22,6673t_{1pqr} + (-0,6691t_{2pqr}) + y_p + z_{pq} \quad (6)$$

The interpretation of the results of the estimation model (6) is as follows:

a. The constant value is -12.4681, meaning that if the free variable is zero then the number of accidents is exp (-12.4681) = 3.84745E-06. This is because the number of accidents is influenced by independent variable factors other than the model.
b. The coefficient X_1 is 0.0669, meaning that every violation occurs because the traffic light causes an increase in the risk of no accidents ie exp (0.0669) = 1.069188.
c. The coefficient X_2 is 0.0262, meaning that every violation occurs because the driver does not have a SIM causing an increase in the risk of no accidents ie exp (0.0262) = 1.0265462.
d. The coefficient T_1 is worth 22.6673, meaning that every time there is a lack of fit of the driver's vehicle causing an increase in the risk of no accidents ie exp (22,6673) = 6,986,880,408.
e. The coefficient of T_2 is -0.6691, meaning that every time a road is damaged it causes an

increase in the risk of no accidents ie exp yaitu (-0.66691) = 0.5121693.

f. The variable y_p is a constant value of random effects as the specific value of monthly data for p = 1.2.3 with a variance value of 0.002035.

g. The z_{pq} variable is a constant value of random effects as the specific value of data per week for p = 1,2,3 and q = 1,2, ... 10 with a variance value of 0,0007097. Based on equation (4) and Table 3 and Table 4, a zero-inflated generalized poisson semiparametric regression model is obtained as follows:

$$P(Y_{pqr} = y_{pqr}) = \frac{\exp(-0{,}9756 + 0{,}0432X_{1ipqr} + 0{,}0121X_{2ipqr} + 0{,}1308T_{1ipqr} + u_1(T_{1ipqr} - k_1) + 0{,}3845\,T_{2ipqr} + u_2(T_{2ipqr} - k_2))}{1 + \exp(-12{,}4681 + 0{,}0669X_{1pqr} + 0{,}0262X_{2pqr} + 22{,}6673T_{1pqr} + u_1(T_{1ipqr} - k_1) + (-0{,}6691T_{2pqr}) + u_2(T_{2ipqr} - k_2))}$$

The interpretation of the results of the estimation model (7) is as follows:

a. The constant for the count model is -0.9756, and -12.4681 for the zero-inflation model meaning that if the independent variable is zero then the chance of an accident is worth $\frac{exp\ (-0.9756)}{(1 + exp\ (-12.4681)\ )} =$ 0.37696455. This is because the number of accidents is influenced by independent variable factors other than the model.

b. The coefficient X_1 for the count model has a value of 0.0432, and a value of 0.0669 for the zero-inflation model, meaning that the chance of an accident due to traffic light violations is $\frac{exp\ (0.0432)}{(1 + exp\ (0.0669))} = 0.504616$, assuming the other variables are constant.

c. The coefficient X_2 for the count model has a value of 0.0121, and a value of 0.0262 for the zero-inflation model, meaning that the chance of an accident occurring because the driver does not have a SIM is $\frac{exp\ (0.0121)}{(1 + exp\ (0.0262))} = 0.499457$, assuming the other variables are constant.

d. The coefficient of T_1 for the count model is worth 0.1308, and it is worth 22.6673 for the zero-inflation model, meaning that the chance of an accident occurs due to the motor vehicle less fit factor, namely $\frac{exp\ (0.1308 + u\_1\ (T\_1ipqr-k\_1))}{(1 + exp\ (22{,}6673 + u\_1\ (T\_1ipqr-k\_1)))}$, assuming the other variables are constant.

e. The coefficient of T_2 for the count model is 0.3845, and the value of -0.6691 for the zero-inflation model, meaning that the chance of an accident due to a damaged road factor is $\frac{exp\ (0.3845 + u\_2\ (T\_2ipqr-k\_2))}{(1 + exp\ (-0.6691)\ 〖+ U〗\ \_2\ (T\_2ipqr-k\_2)))}'$ assuming the other variables are constant.

In the equation of the form of a zero-inflaed generalized poisson semiparametric regression model (7) the estimated parameters of the smooth function (u_j) and the optimal knot point (k_j) are not searched for because of limitations in this study. Semiparametric regression is a combination of parametric regression and nonparametric regression. The smooth function of nonparametric regression can be formed with the penalized spline approach. In spline regression there are unknown parameters and optimal knot points. The optimal knot point can be searched by looking at the minimum Generalized Cross Validation (GCV).

Furthermore, the significance of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$, the results obtained in Table 5 and Table 6.

Tabel 5 Analisis Uji Wad Count Part

i	$\hat{\beta}_i$	SE	$W_i = z_i^2$	P value	decision
1	0.0432	0.0412	1.1025	0.2948	H_0 accepted
2	0.0121	0.0442	0.0729	0.7851	H_0 accepted
3	0.1308	0.5250	0.0625	0.8033	H_0 accepted
4	0.3845	0.2997	1.6384	0.1994	H_0 accepted

Table 6 Analysis of the Wald *Zero-Iflation* Part Test

i	$\hat{\tau}_i$	SE	$W_i = z_i^2$	P value	Decision
1	0.0669	0.4377	0.0225	0.8785	H_0 diterima
2	0.0262	0.4691	0.0036	0.9555	H_0 diterima
3	22.6673	5.7698	15.4449	8.5e-05 ***	H_0 ditolak
4	-0.6691	-0.6691	0.0225	0.8792	H_0 diterima

Based on Table 5 and Table 6 it is known that for the count part there are no significant variables that influence. Whereas the zero-inflation part only (τ_3)^ has a significant effect. Then the final model of multi-level zero-inflated generalized poisson semiparametric regression is

$$P(Y_{pqr} = y_{pqr}) = \frac{\exp(-0{,}9756)}{1 + \exp(-12{,}4681 + 22{,}6673T_{1pqr} + u_1(T_{1pqr} - k_1)}$$

CONCLUSION

The model of a zero-inflated generalized poisson multilevel semiparametric regression equation for traffic accident data in Temanggung Regency is as follows.

$$\boldsymbol{P(Y_{pqr} = y_{pqr}) = \frac{\exp(-0,9756)}{1 + \exp(-12,4681 + 22,6673T_{1pqr} + u_1(T_{1pqr} - k_1)}}$$

The variable that has a significant effect on accidents is the variable number of accidents because the vehicle is not fit.

Estimates of many accidents in Temanggung Regency are based on a multi-level zero-inflated generalized poisson semiparametric regression model that is if there is 1 vehicle that is not fit then the chance of a traffic accident is

$$P(Y_{pqr} = y_{pqr}) = \frac{\exp(-0{,}9756)}{1 + \exp(-12{,}4681 + 22{,}6673(1) + u_1(T_{1pqr} - k_1))} = \frac{0.376966}{26882{,}67 + \exp(u_1(T_{1pqr} - k_1)}$$

Based on the results and discussion, it can be suggested to the reader that further research is needed to model the overdispersion case in traffic accident data with the Multilevel Zero-Inflated Generalized Poisson semiparametric regression model and complete the smooth function in nonparametric regression with the spline approach. Multilevel Zero-Inflated Generalized Poisson (MZIGP) semiparametric regression model can be tried using other case studies, especially on data that experiences the occurrence of excess zero.

REFERENCES

Baillo, A, dkk. (2009). Test for Zero-Inflation and overdispersion a New Approach Based on The Stochastic Convex Order. Journal of Computational Statistics & Data Analysis, 1-27.

Dewi, N. C. S., & Budiantara, I. N. (2018). Faktor-Faktor yang Mempengaruhi Angka Kecelakaan Lalu Lintas di Provinsi Jawa Timur Menggunakan Regresi Nonparametrik Spline Truncate. *Jurnal Sains dan Seni ITS Vol. 7 No.2,* 1-8

Dey, D. K., dkk, ed. (2000). *Generalized Linear Models a Bayessian Perspective*. New York: Marcel Dekker

Eshraghian, M.R., dkk. (2016). Multilevel Zero-Inflated Generalized Poisson Regression Modelling for Dispersed Correlated Count Data. *Journal of Statistical Methodology Vol 30*, 1-14

Famoye, F., & Singh, K. P. (2006). Zero Inflated Generalized Poisson Regression Model with an Aplication to Domestic Violence Data. *Journal Teknologi, Vol, 43. No 1, 39-50*

Fathurahman, K. (2010). Pemilihan Model Terbaik Menggunakan Akaike's Information Criterion. *Jurnal Eksponensial, Vol. 1, No. 2*

Fitriani, A. dkk. (2015). Estimasi Model Regresi Semiparametrik Menggunakan Estimator Kernel Uniform. *E-Jurnal Matematika, Vol. 4, No. 4, 176-180*

Fitriani, Ni Luh Ayu., dkk. (2016). Analisis Regresi Multilevel dalam Menentukan Variabel Determinan Nilai Ujian Akhir Nasional Siswa. *Jurnal Matematika Vol 6, 15-21*

Fuad, M. Syaeful., dkk. (2017). Pemodelan Data Kecelakaan Lalu Lintas Menggunakan Metode Regresi Poisson dan Regresi Binomial Negatif. *Prosiding Statistika*

Jansakul, N. & J.P Hinde. (2002). Score Test for Zero-Inflated Poisson Models. *Computational Statistics & Data Analysis, 40 :75-96*

Lee, Andy H., Wang, Kui. (2006). Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. *Statistical Methods in Medical Research, No. 15, 47-61*

Mahmoodi, M., dkk. (2016). Semiparametric Models for Multilevel Overdispersed Count Data With Extra Zero. *Journal of Statistical Method in Medical Research*

Mardiani, L. K. dkk. (2013). Penerapan Regresi Zero Inflated Poisson Untuk Mengatasi

Overdispersi Pada Regresi Poisson. *E-Jurnal Matematika Vol. 2, No. 3, 23-28*

McCullagh, P., and J. A. Nelder. (1989). *Generalized Linear Models, 2nd Ed*. New York: Chapman and Hall

Moustassim, Y., & Ezzahid, E. H. (2012). Poisson Regression and Zero-Inflated Poisson Regression. *Aplication to Private Health Insurance Data Vol 2, 187-204*

Tantular, B., Aunuddin & Wijayanto, H., (2009). Pemilihan Model Regresi Linear Multilevel Terbaik. *Forum Statistika dan Komputasi.Vol. 14 No. 2*

Wallner, Jon A. (2004). A Semiparametric Regression Model for Panel Count Data: When Do Pseudo-likelihood Estimators Become Badly Inefficient. *Proceedings of the Second Seattle Symposium in BioStatistics*