EKSAKTA
Journal of Sciences and Data Analysis

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art4

Research Article

# Negative Binomial Panel Regression Modeling on Amount of Crimes In Lampung Province

**Idam Abdurrohim Hasani [1,*], Dian Agustina[2], Winalia Agwil [3]**

[1,2,3] Faculty of Mathematics and Natural Science, University of Bengkulu

[*] *Corresponding author: idamhasani@gmail.com*

Received: 31 July 2023; Accepted: 18 January 2024; Published: 27 January 2024

**Abstract:** Crime in Lampung province is among the 10 highest in Indonesia in 2021. This study aims to obtain a model of the number of crimes and factors influencing it using negative binomial panel regression. The data used is in the form of panel data from the Lampung Province BPS website and publications for 2017-2021. The condition of data on the number of crimes as discrete and overdispersed data makes the negative binomial panel regression method more suitable than Poisson panel regression. Overdispersion is a state where the variance of the data is greater than the mean value of the data. Overdispersion causes the standard error (SE) of the estimated value to decrease, so that variables that should not be significant become significant. The factors thought to be the cause of crime are percentage of poverty ($X_1$), population density ($X_2$), expenditure per capita ($X_3$), unemployment rate ($X_4$), regional gross domestic income ($X_5$), and the average duration of schooling ($X_6$). The results of the analysis obtained for the selected panel data model are the negative binomial random effects (RENB), the influencing factors being $X_1$, $X_3$, $X_4$ and $X_5$. The districts/cities with the largest individual random effects were in the Way Kanan district and the smallest were in Metro City.

**Keywords:** panel data, negative binomial, crime, lampung

## Introduction

In the era of rapid development of science and technology, competition is accelerating. This era impacts how easily a culture enters and leaves a community environment. New values will bring change in society, including negative values such as wanting to fulfill all their desires which cause unrest in the wider community. The number of desires that do not match the abilities (skills) often forces a person to think of quick ways and commit crimes as a solution to get what they want [1].

According to the Statistics Indonesia (BPS), criminal acts are all actions, whether intentional or unintentional, that have occurred or have recently carried out experiments that can cause harm to body, soul, property, honor and others and these actions are punishable by imprisonment and detention. Crime is deviant behavior in society that violates the norms of social life. Based on the data from the Lampung Regional Police Report released by the Lampung Province BPS, the number of crimes in Lampung Province from 2017 to 2021 reached 3,598 cases under the category of crimes against life and body. The number of criminal acts in Lampung province is increasing year by year. This is also in line with data on the risk of the population being exposed to crime in 2021, according to which Lampung province is still above the national average, i.e. per 100,000 inhabitants, around 115 people are at risk of being involved in crime.

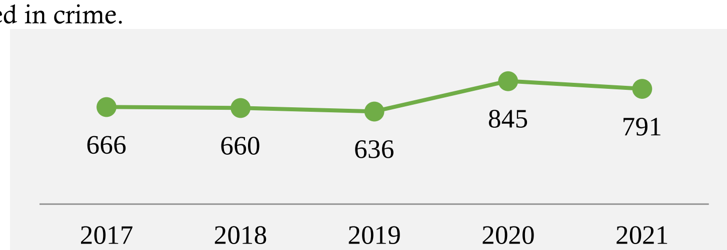


**Figure 1**. *Count of Crime in Lampung Province*

Adri, Karimi and Indrawari [2] stated that the factors behind someone committing a crime are related to social and economic conditions. Poor social and economic conditions tend to increase criminal activity. Apart from that, Harefa [3] stated that the level of public education is also an important factor in someone choosing to commit a crime to get what they want. A low level of education makes a person more vulnerable to committing criminal acts because it is related to a lack of life skills.

Crime will become a social problem that interferes with daily life in society, especially the safety of their life and body. This requires preventive measures to minimize the impact that will result in casualties, both fatalities and property. One form of preventive action includes carrying out a crime-influencing factor analysis, which aims to find out in detail which factors have a significant effect, in order to reduce the number of crimes that will occur in the future. The most widely used methods for discrete data are Poisson panel regression and negative binomial panel regression to handle discrete data with panel data types. However, the Poisson panel regression method becomes less accurate if the equidispersion assumption is not met (overdispersion or underdispersion occurs) [4]. Therefore, according to Cameron and Trivedi [5], negative binomial panel regression is the correct method to deal with discrete data conditions with overdispersion in panel data. This is linear with the state of crime count data which are discrete data and are rare events [6] which experience overdispersion, so when analyzed with the usual panel data method, they will be inaccurate.

Several studies using the negative binomial panel regression model, including those conducted by Sun *et al*. [7] on driving risk assessment using near-miss event based on Poisson panel regression and negative binomial panel regression, which shows that the performance of negative binomial panel regression model is better than Poisson panel regression. In addition, research was also conducted by Adenomon and Akinyemi [8] to analyze TB and HIV cases in West Africa using Poisson panel regression and negative binomial panel regression. The results show that the negative binomial panel regression model with a fixed effect has the highest log likelihood value. Based on the previous explanation, the researchers are interested in modeling the number of crimes in Lampung province using the negative binomial panel regression method. The analysis was performed on factors assumed to have a significant effect. The assumed variables have also been utilized in previous research in the different region, as demonstrated in Study Suliyanto [9], which found that both percentage of poverty, expenditure per-capita, unemployment rate, and overcrowding exerted significant influences on criminality. Furthermore, these variables were employed by Study in Febriani [10] analyzing crime rates in South of Sumatera Province, revealing that regional gross domestic income (PDRB) had a significant impact, and according to Edwart and Azhar [11] found that the average duration of schooling had a signifcant impact because this variable have correlation with economy. The variables mentioned in prior studies will be incorporated into this research, focusing on the research area of Lampung Province. The objective is to determine whether the characteristic factors of criminality in Lampung Province differ from those in other regions that have been previously studied.

## Materials and Methods

### Materials

The data used in this study are secondary data taken from the publication of security statistics and the website of the Indonesian Statistics (BPS) of Lampung Province at the regional level for each district/city. Variables used in this reserach are number of crime as dependent variable (Y), and percentage of poverty ($X_1$), overcrowding ($X_2$), expenditure per capita ($X_3$), unemployment rate ($X_4$), regional gross domestic income ($X_5$), and the average duration of schooling ($X_6$) as independent variables.

### Method

Data analysis used the negative binomial panel regression method. Here are the steps of the analysis performed in this study as follows:

1. Make descriptive statistics on the dependent and independent variables
2. Overdispersion test on the dependent variable
3. Estimation of regression parameters of the negative binomial panel with maximum likelihood estimation (MLE)
4. Detection of Multicollinearity Cases from Independent Variables with VIF Testing Criteria
5. Selection of panel data model with Hausman test, Chow test and Lagrange multiplier (LM) test
6. Simultaneous testing of the regression parameters of the negative binomial panel regression
7. Partial test of the regression parameters of Negative Binomial Panel Regression
8. Calculation of Pseudo-$R^2$

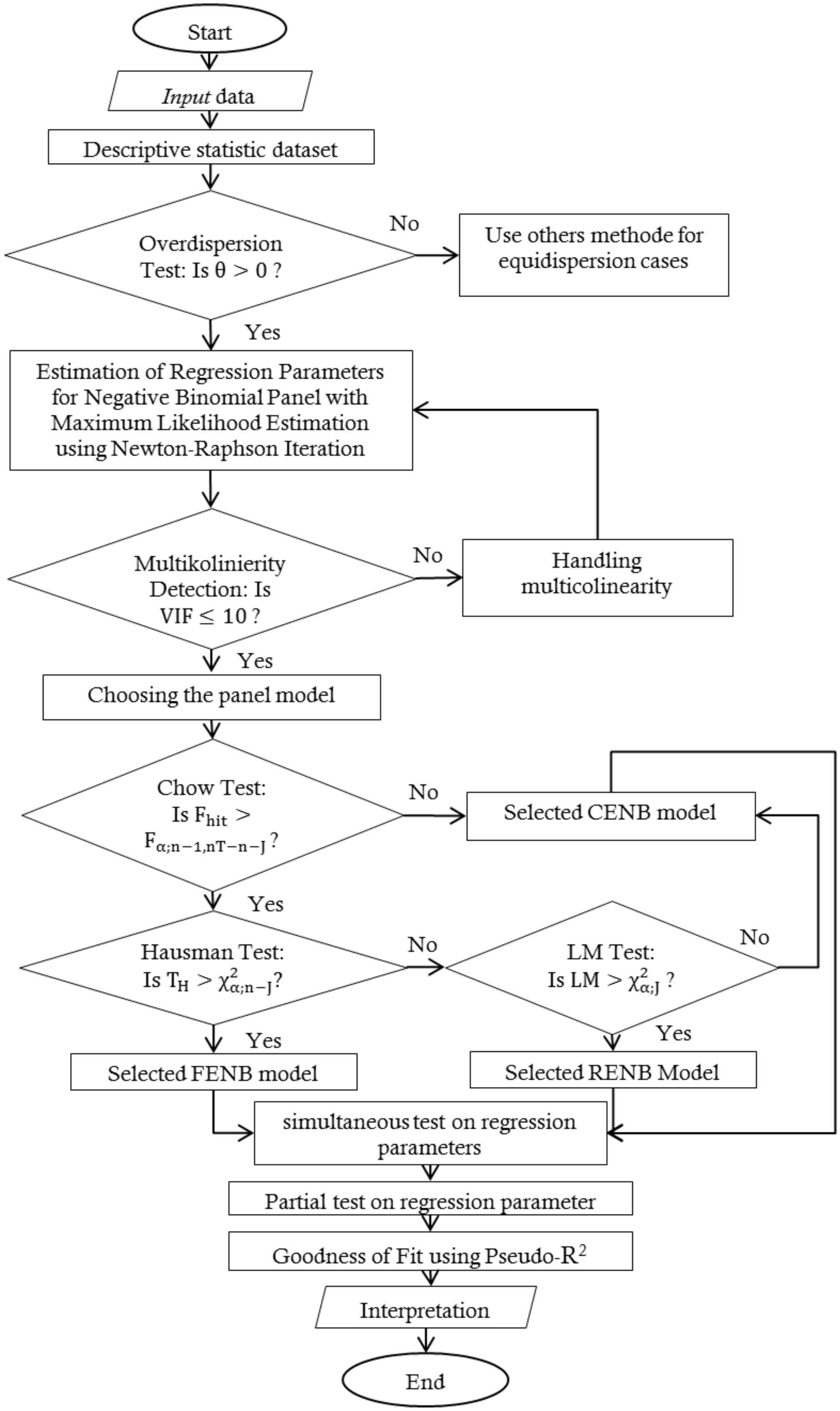


**Figure 2**. Flowchart of Negative Binomial Panel Regression

**Negative Binomial Distribution**

The Negative Binomial distribution is the distribution of the number of Bernoulli trials needed to achieve r successes. The probability mass function of the Negative Binomial distribution is given by:

$$\Pr(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r},\ x=r,\ r+1,\ r+2,\ldots \tag{1}$$

where Pr(X = x) represents the probability of getting r successes in the x-th trial, and p is the probability of success in each trial [12]. In addition to the Negative Binomial distribution being approached as a sequence of Bernoulli trials, Boswell and Patil [13] described twelve other approaches to the Negative Binomial distribution, one of which involves approximating it with the Poisson-Gamma mixture distribution as the Compound Poisson distribution. The probability mass function of Y, which follows a Negative Binomial distribution as a mixture of Poisson and Gamma distributions, is as follows:

$$f(y;\lambda,\theta) = \frac{\Gamma\left(y+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right) y!}\left(\frac{\theta\lambda}{\theta\lambda+1}\right)^{y}\left(\frac{1}{\theta\lambda+1}\right)^{\frac{1}{\theta}} \tag{2}$$

Where $\lambda$ represents the rate parameter of the Poisson distribution, and $\theta$ is the shape parameter of the Gamma distribution [14]. The mean (expected value) and variance of the Negative Binomial distribution formed by the Poisson-Gamma distribution are $E[Y] = \lambda$ and $Var[Y] = \lambda + \theta\lambda^2$ for $\theta > 0$.

**Overdispersion**

Poisson regression is said to exhibit overdispersion when Var[Y] > E[Y]. Colin and Pravin [15] stated that overdispersion testing is conducted using a dispersion test. The hypotheses used in the overdispersion test are as follows $H_0$: $\theta = 0$ (There is no overdispersion in the data distribution)and $H_1$: $\theta > 0$ (There is overdispersion in the data distribution). The test statistic used is: $\chi^2 = \sum_{i=1}^{n} \frac{(y_i-\bar{y})^2}{\bar{y}}$. The testing criterion, with a significance level of $\alpha$, rejects $H_0$ if $\chi^2 > \chi^2_{(\alpha;n-J)}$, where J represents the number of independent variables. The estimated dispersion parameter can be obtained by dividing the value of $\chi^2$ by its degrees of freedom [16].

**Maximum Likelihood Estimator (MLE)**

Maximum Likelihood is a parameter estimation method based on the distributional approach by maximizing the likelihood function to obtain parameter estimators with the highest probability [17]. Let $x_1, x_2, \ldots, x_n$ is random sample have size n from a distribution that have probability function $f(x;\theta)$ which depends on the parameter $\theta \in \Omega$, where $\Omega$ is the parameter space of $\theta$. Thus, the joint likelihood function of $x_1, x_2, \ldots, x_n$, which are independent, can be written as follows:

$$f(x_1, x_2, \ldots, x_n;\theta) = f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta) \tag{3}$$

Joint probability function of equation (3) has $\theta$ parameter so it can be written as likelihood function of $\theta$ notated with $L(\theta) = f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta) = \Pi_{i=1}^{n} f(x_i;\theta)$. Futhermore, find the maximum value of $L(\theta)$. As the step to make easier way, $L(\theta)$ modified into log function notated with $\ell(\theta)$. To get maximum value of $\ell(\theta)$, needed the derivative of $\ell(\theta)$ with respect to $\theta$ and then equating it to 0 and ensuring that its second derivative is less than 0. Thus, the value of $\theta$ from $x_1, x_2, \ldots, x_n$ is an estimate of the population parameter $\theta$, denoted as $\hat{\theta}$. The value of $\theta$ that maximizes $\ell(\theta)$ will become the sought-after maximum likelihood estimate.

**Newton Raphson**

The use of Newton-Raphson is to avoid difficulties in demonstrating the second derivative test because Newton-Raphson is a monotonically increasing function. The algorithm for a single iteration in the Newton-Raphson method is as follows [18]:

1. Determine the first derivative f(x).
2. Set the initial value of $x_0$ and the tolerance value ($\varepsilon$).
3. Substitute the value of $x_n$ into f(x) and f'(x).
4. Calculate the value of $x_{(n+1)}$ using the formula $(x_n - f(x))/f'(x)$.
5. Repeat steps 3 and 4 until the value of f(x) = 0 and $|x_{(n+1)} - x_n| < \varepsilon$ are satisfied.
6. If $|x_{(n+1)} - x_n| < \varepsilon$ is not achieved, the iteration will stop if the number of iterations reaches the maximum iteration specified.

**Panel Negative Binomial Regression**

When the observed data consist of a combination of cross-sectional and time series data, the consideration of panel models into Negative Binomial Regression is necessary [15]. When a discrete data is assumed to follow a Poisson or Negative Binomial distribution, a different model is required [19]:

$$E[Y_{it} \mid X_{itj}, u_i] = u_i . \exp(\beta_1 X_{it1} + \beta_2 X_{it2} + \ldots + \beta_J X_{itJ})$$

where the individual effect $u_i$ becomes multiplicative, no longer additive. In classical panel regression, the classical assumptions must be met, but in count data models, these assumptions can be ignored. However, in more specific cases, another assumption that must be satisfied is whether the distribution of the dependent variable contains overdispersion or not. Some models of Panel Negative Binomial is as follows:

1. Common Effect Negative Binomial (CENB)

   The advantage of using CENB is its ease of use. Anyone with knowledge of cross-sectional techniques can utilize it. However, the weakness of this model is that it disregards individual effects, leading to biased and inconsistent estimates. Therefore, the CENB model is not suitable for situations where individual effects are present. The empirical model for the CENB model is as follows:

$$E\left[Y_{it} \middle| X_{itj}\right] = \exp\left(\beta_0 + \sum_{j=1}^{J} \beta_j X_{itj} + \varepsilon_{it}\right) \tag{4}$$

   This model can be estimated using maximum likelihood estimation based on mass probability function of CENB below:

$$f(y_{it};\lambda_i,\theta) = \frac{\Gamma\left(y_i + \frac{1}{\theta}\right)}{\Gamma(y_i+1)\Gamma(\frac{1}{\theta})} \left(\frac{1}{1+\theta.\lambda_i}\right)^{\frac{1}{\theta}} \left(\frac{\theta.\lambda_i}{1+\theta.\lambda_i}\right)^{y_t}$$

   Then, the parameters θ and β will be estimated. The likelihood function for the independent observations $y_1$ , $y_2$, ..., $y_N$ is as follows:

$$L(\lambda_i, \theta) = \Pi_{i=1}^{N} \left( \frac{\Gamma\left(y_i + \frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right) . y_i!} \left(\frac{1}{1+\theta.\lambda_i}\right)^{\frac{1}{\theta}} \left(\frac{\theta.\lambda_i}{1+\theta.\lambda_i}\right)^{y_i} \right) \tag{5}$$

   Known if $\frac{\Gamma\left(y_i+\frac{1}{\theta}\right)}{\Gamma\left(\frac{1}{\theta}\right) . y_i!} = \left(\frac{1}{\theta}\right)^{y_i} \Pi_{r=1}^{y_i-1}(1+\theta r)$, than likelihood function of equation (5) transforms into log likelihood:

$$= \sum_{i=1}^{N} \left\{ \left( \sum_{r=0}^{y_i-1} \ln(1+\theta r) \right) - \ln(y_i!) + y_i \ln(\exp(\mathbf{X_i'\beta}) - \left(\frac{1}{\theta} + y_i\right) . \ln(1+\theta. \exp(\mathbf{X_i'\beta})) \right\} \tag{6}$$

   Then, to maximize the log-likelihood function, it is necessary to take the first derivative of equation (6) with respect to $\boldsymbol{\beta}$ and θ, and then set them equal to zero. The first derivative of equation (6) is:

$$\frac{\partial l(\boldsymbol{\beta},\theta)}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{N} \left( \frac{y_i - \exp(\mathbf{X_i'\beta})\mathbf{X_i'}}{1+\theta. \exp(\mathbf{X_i'\beta})} \right) = 0 \tag{7}$$

$$\frac{\partial l(\boldsymbol{\beta},\theta)}{\partial \theta} = \sum_{i=1}^{N} \left( \frac{y_i}{\theta} + \frac{(\theta^{-2}. \ln(1+\theta. \exp(\mathbf{X_i'\beta})))}{1+\theta. \exp(\mathbf{X_i'\beta})} - \sum_{r=0}^{y_i-1} \left( \frac{\theta^{-2}}{\theta^{-1}+r} \right) \right) = 0 \tag{8}$$

   To find the solutions of equations (7) and (8), the assistance of the Newton-Raphson iterative method is required.

2. Fixed Effect Negative Binomial (FENB)

   The FENB model can be obtained by adding dummy variables for each individual present in the panel. The advantage of the FENB model over CENB is its ability to consistently estimate the individual effects that are correlated with the independent variables [20]. The general empirical model of FENB is as follows:

$$E[Y_{it} \mid X_{itj}, \delta_i] = \delta_i . \exp(\beta_1 X_{it1} + \beta_2 X_{it2} + \ldots + \beta_J X_{itJ}) \tag{9}$$

   Equation (9) in such a way that:

$$\ln\left(E\left[Y_{it} \middle| X_{itj}, \delta_i\right]\right) = \ln(\delta_i) + \left(\beta_1 X_{it1} + \beta_2 X_{it2} + \ldots + \beta_J X_{itJ}\right) \tag{10}$$

   Furthermore, it can be assumed that $c_i = \ln(\delta_i)$ so that equation (10) becomes:

$$\ln\left(E\left[Y_{it} \middle| X_{itj}, c_i\right]\right) = c_i + \left(\beta_1 X_{it1} + \beta_2 X_{it2} + \ldots + \beta_J X_{itJ}\right) \tag{11}$$

Because the fixed effect is estimated with a dummy variable, with a value of 1 for certain groups and 0 for other groups, and if coeffision of dummy ($D_i$) is $\beta_{0i}$, it can change equation (11) into:

$$E\left[Y_{it}\middle|X_{itj}\right]=\exp\left(\beta_{0i}D_i+\beta_1X_{it1}+\beta_2X_{it2}+\ldots+\beta_jX_{itj}\right) \quad (12)$$

The FENB model can be estimated using the Conditional Maximum Likelihood Estimator (CMLE) by maximizing the log-CMLE function value [16]. Meanwhile, the conditional fixed effect likelihood function for the i-th observation is given by the following function [4]:

$$f\left(y_{it}\middle|t;\boldsymbol{\beta},\ \boldsymbol{\beta_0}\right)=\left(\Pi_{t=1}^{T}\left(\frac{\Gamma\left(\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)+y_{it}\right)}{\Gamma\left(\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)\right).\Gamma\left(y_{it}+1\right)}\right)\right)\times\left(\frac{\left(\Gamma\left(\sum_{t=1}^{T}\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)\right)\Gamma\left(\sum_{t=1}^{T}y_{it}+1\right)\right)}{\Gamma\left(\sum_{t=1}^{T}\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)+\sum_{t=1}^{T}y_{it}\right)}\right) \quad (13)$$

And than the conditional log-likelihood of equation (13) is as follows:

$$l_c(\boldsymbol{\beta},\boldsymbol{\beta_0})=\sum_{i=1}^{n}\left\{\begin{matrix}\ln\Gamma\left(\sum_{t=1}^{T}\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)\right)+\ln\Gamma\left(\sum_{t=1}^{T}y_{it}+1\right)-\ln\Gamma\left(\sum_{t=1}^{T}\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)+\sum_{t=1}^{T}y_{it}\right)+\\ \sum_{t=1}^{T}\left[\ln\Gamma\left(\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)+y_{it}\right)-\ln\Gamma\left(\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)\right)-\ln\Gamma\left(y_{it}+1\right)\right]\end{matrix}\right\} \quad (14)$$

Therefore, to maximize the value of $l_c$, equation (14) needs to be differentiated with respect to $\boldsymbol{\beta}$, $\boldsymbol{\beta_0}$, and θ, and then set equal to zero, as stated in [21]:

$$\frac{\partial l(\boldsymbol{\beta},\boldsymbol{\beta_0})}{\partial\boldsymbol{\beta}^{\mathbf{T}}}=\sum_{t=1}^{T}\left\{\begin{matrix}\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right).\left[\Psi\left(\exp\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}+y_{it}\right)-\Psi\left(\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)\right)\right]-\\ \left[\Psi\left(\sum_{t=1}^{T}\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)+\sum_{t=1}^{T}y_{it}\right)-\Psi\left(\sum_{t=1}^{T}\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)\right)\right].\mathbf{X}_{it}\end{matrix}\right\}=0 \quad (15)$$

$$\frac{\partial l(\boldsymbol{\beta},\boldsymbol{\beta_0})}{\partial\boldsymbol{\beta}_{\mathbf{0}}^{\mathbf{T}}}=\sum_{t=1}^{T}\left\{y_{it}.\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)-\ln\left(\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)+\ln\Psi\left(y_{it}\right)-\ln\left(y_i!\right)-\ln\Psi\left(\exp\left(\mathbf{X}_{it}'\boldsymbol{\beta}+\boldsymbol{\beta_0}\mathbf{D_i}\right)\right)\right\}=0 \quad (16)$$

With $\Psi(.)=\frac{\Gamma'(.)}{\Gamma(.)}$ is gamma function. Then, to get the solution of equation (15) and (16) needed Newton Raphson iterasion algoritm.

3. Random Effect Negative Binomial (RENB)

If the individual effects are not correlated with the regression variables, then these effects become entirely part of the error term ε. Therefore, the approach used in the FENB model to account for the absence of different individual effects that would compensate for biases is no longer applicable. The general empirical model of FENB is as follows:

$$E\left[Y_{it}\,|\,X_{itj},\,v_i\right]=v_i.\exp\left(\beta_0+\beta_1X_{it1}+\beta_2X_{it2}+\ldots+\beta_jX_{itj}\right) \quad (17)$$

where it is assumed that $v_i$ is an individual random effect that is uncorrelated with the independent variables. Additionally, two other assumptions are required to construct the maximum likelihood estimator, namely:

a. The dependent variable follows a Negative Binomial distribution [22].
b. The individual effect $v_i$ as a random effect follows a beta distribution with parameters (a, b). Meanwhile, θ = 1+v, where v is the average value of $v_i$, and 1/θ ~ beta (a, b) [19].

And the procces establishment of empirical models such as in FENB, than the structure of the RENB model can be written as follows:

$$E(Y_{it}\,|\,X_{itj})=\exp\left(\beta_0+\sum_{j=1}^{J}\beta_jx_{jit}+v_i\right) \quad (18)$$

Model of RENB as shows at equation (18), will estimated using maximum likelihood with mass probability function is as follows:

$$f(y_{it};\boldsymbol{\beta},\ a,\ b)=\frac{\Gamma(a+b)\Gamma\left(a+\sum_{t=1}^{T}\lambda_{it}\right)\Gamma\left(b+\sum_{t=1}^{T}y_{it}\right)}{\Gamma(a)\Gamma(b)\Gamma\left(a+b+\sum_{t=1}^{T}\lambda_{it}+\sum_{t=1}^{T}y_{it}\right)}\times\Pi_{t=1}^{T}\left[\frac{\Gamma\left(\lambda_{it}+y_{it}\right)}{\Gamma(\lambda_{it}).\Gamma\left(y_{it}+1\right)}\right] \quad (19)$$

With $\lambda_{it}=\exp(\mathbf{X}_{it}'\boldsymbol{\beta})$. Therefore, Maximum likelihood could determined by maximum value that follows log-likelihood function below:

**E-ISSN**: 2720-9326
**P-ISSN**: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art4

$$\ell(\boldsymbol{\beta};\ a,b) = \sum_{i=1}^{n}\left\{\ln\Gamma(a+b) + \ln\Gamma\left(a + \sum_{t=1}^{T}\exp(\mathbf{X}_{it}'\boldsymbol{\beta})\right) + \ln\Gamma\left(b + \sum_{t=1}^{T} y_{it}\right) - \ln\Gamma(a) - \Gamma(b) - \ln\Gamma\left(a + b + \sum_{t=1}^{T}\exp(\mathbf{X}_{it}'\boldsymbol{\beta}) + \sum_{t=1}^{T} y_{it}\right) + \sum_{i=1}^{T}\left[\ln\Gamma(\exp(\mathbf{X}_{it}'\boldsymbol{\beta}) + y_{it}) - \ln\Gamma(\exp(\mathbf{X}_{it}'\boldsymbol{\beta})) - \ln\Gamma(y_{it}+1)\right]\right\} \quad (20)$$

To obtain the estimates of the random effects in the negative binomial model, it is necessary to differentiate equation (20) to $\boldsymbol{\beta}$, a, and b, and then set them equal to zero. The first derivative of equation (20) is:

$$\frac{\partial[\ell(\boldsymbol{\beta},a,b)]}{\partial\boldsymbol{\beta}} = \left\{\sum_{t=1}^{T}\exp(\mathbf{X}_{it}'\boldsymbol{\beta}).\left(\Psi(a + \sum_{t=1}^{T}\exp(\mathbf{X}_{it}'\boldsymbol{\beta})) - \Psi(a + b + \sum_{t=1}^{T}\exp(\mathbf{X}_{it}'\boldsymbol{\beta}) + \sum_{t=1}^{T} y_{it}) + \Psi(\exp(\mathbf{X}_{it}'\boldsymbol{\beta}) + y_{it}) - \Psi(\exp(\mathbf{X}_{it}'\boldsymbol{\beta}))\right).\mathbf{X}_{it}\right\} = 0 \quad (21)$$

$$\frac{\partial[\ell(\boldsymbol{\beta},a,b)]}{\partial a} = \{\sum_{t=1}^{T}\exp(\mathbf{X}_{it}'\boldsymbol{\beta}).\left(\Psi(a + \sum_{t=1}^{T}\exp(\mathbf{X}_{it}'\boldsymbol{\beta})) - \Psi(a + b + \sum_{t=1}^{T}\exp(\mathbf{X}_{it}'\boldsymbol{\beta}) + \sum_{t=1}^{T} y_{it}) + \Psi(\exp(\mathbf{X}_{it}'\boldsymbol{\beta}) + y_{it}) - \Psi(\exp(\mathbf{X}_{it}'\boldsymbol{\beta}))\right) + (\Psi(a) + \Psi(a + \sum_{t=1}^{T}\exp(\mathbf{X}_{it}'\boldsymbol{\beta}))\} = 0 \quad (22)$$

$$\frac{\partial[\ell(\boldsymbol{\beta},a,b)]}{\partial b} = \{\sum_{t=1}^{T}\exp(\mathbf{X}_{it}'\boldsymbol{\beta}).\left(\Psi(a + \sum_{t=1}^{T}\exp(\mathbf{X}_{it}'\boldsymbol{\beta})) - \Psi(a + b + \sum_{t=1}^{T}\exp(\mathbf{X}_{it}'\boldsymbol{\beta}) + \sum_{t=1}^{T} y_{it}) + \Psi(\exp(\mathbf{X}_{it}'\boldsymbol{\beta}) + y_{it}) - \Psi(\exp(\mathbf{X}_{it}'\boldsymbol{\beta}))\right) + (\Psi(b) + \Psi(b + \sum_{t=1}^{T} y_{it})\} = 0 \quad (23)$$

With $\Psi(z) = \frac{d[\log\Gamma(z)]}{dz}$, then to get the solution of equation (15) and (16) needed Newton Raphson iterasion algoritm.

## Multicolinearity

Regression analysis involves multiple independent variables that require the assumption that these variables are not correlated with each other or known as multicollinearity. The assumption of multicollinearity must be maintained because if multicollinearity occurs among the independent variables, it will result in regression estimates having large residuals. Detection of multicollinearity according to Hocking [23] is done by examining the Variance Inflation Factor (VIF). Generally, $VIF_j$ can be expressed as follows:

$$VIF_j = \frac{1}{1-R_j^2} \quad (24)$$

where $R_j^2 = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}$ and j = 1, 2, ..., J, where J is the number of independent variables and $RR_j^2$ is the coefficient of determination for regressing the j-th independent variable on the other independent variables. The VIF value will be equal to one if the independent variable is not linearly related in the regression model. A VIF value greater than 10 indicates the presence of multicollinearity among the independent variables.

## Likelihood Ratio Test

To test the parameters of the Panel Negative Binomial Regression model simultaneously, the Likelihood Ratio Test (LRT) is used. The formulation of the hypothesis for the parameters can be written as follows $H_0: \beta_1 = \beta_2 = \cdots = \beta_J = 0$, dan $H_1: \exists\beta_j \neq 0$ dengan $j = 1,2,\ldots,J$. The test statistic used is the likelihood ratio test, denoted as follows:

$$LRT = -2\frac{\ell_{\widehat{\omega}}}{\ell_{\widehat{\Omega}}} = 2(\ell_{\widehat{\Omega}} - \ell_{\widehat{\omega}}) \quad (25)$$

The rejection region is to reject $H_0$ if LRT > $\chi^2_{(\alpha;1)}$, which indicates that at least one parameter has an effect on the model [24].

## Wald Test

In the analysis process, it is necessary to perform partial parameter testing to determine the significant parameters with the formulation of hypotheses, which are as follows $H_0: \beta_j = 0$, and $H_1: \beta_j \neq 0$. Test statistic of test use $W_j = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$ to compared with $Z_{\frac{\alpha}{2}}$. If $W_j > Z_{\frac{\alpha}{2}}$ concluded that $H_0$ is rejected at significant level α.

## Selection Panel Model

1. Chow Test
   The Chow test is used to determine the best model between the common effect model and the fixed effects model, with the following hypotheses: $H_0: \beta_{01} = \beta_{02} = \ldots = \beta_{0n} = \beta_{00}$ (Common Effect) and $H_1: \beta_{0i} \neq \beta_{00}$ (Fixed Effect). The test statistic used is [25]:

$$F_{hit}=\frac{\left(\frac{(RSS_1\text{-}RSS_2)}{n\text{-}1}\right)}{\left(\frac{RSS_2}{nT\text{-}n\text{-}J}\right)} \tag{26}$$

   With $RSS = \sum_{i=1}^{n}(y_{it} - \hat{y}_{it})^2$. If $F_{hit} > F_{\alpha;n-1,nT-n-J}$ or p-value < $\alpha$, than $H_0$ is rejected. It's means model selected is fixed effect.
2. Hauman Test
   The Hausman test is used to determine the best model between the random effects model and the fixed effects model, with the following hypotheses: $H_0$: $cor(X_{it}, \varepsilon_i) = 0$ (Random Effects) and $H_1$: $cor(X_{it}, \varepsilon_i) \neq 0$ (Fixed Effects). The test statistic used is [15]:

$$T_H=(\hat{\boldsymbol{\beta}}_{\mathbf{RE}}\text{-}\hat{\boldsymbol{\beta}}_{\mathbf{FE}})^{'}\left[Var[\hat{\boldsymbol{\beta}}_{\mathbf{RE}}\text{-}\hat{\boldsymbol{\beta}}_{\mathbf{FE}}]\right]^{-1}(\hat{\boldsymbol{\beta}}_{\mathbf{RE}}\text{-}\hat{\boldsymbol{\beta}}_{\mathbf{FE}}) \tag{27}$$

   If $T_H > \chi^2_{(\alpha;J)}$, then reject $H_0$ and conclude that the individual effects are correlated with the independent variables.
3. Lagrange Multiplier Test
   The Lagrange Multiplier (LM) test is used to determine the best model between the pooled (common effect) model and the random effects model, with the following hypotheses $H_0 : \sigma_i^2 = 0$ (Common effect) and $H_1 : \sigma_i^2 \neq 0$ (Random effetct). The test statistic used is [26]:

$$LM = \frac{nT}{2(T-1)}.\left[\frac{\sum_{i=1}^{n}(\sum_{t=1}^{T}\varepsilon_{it})^2}{\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}^2} - 1\right]^2 \tag{28}$$

   If the LM value is greater than $\chi^2_{(\alpha;1)}$ or the p-value is less than $\alpha$, then the decision is to reject $H_0$, indicating that the selected model is the random effects model.

## Pseudo R-Sqaure

Goodness of fit proposed by Colin and Pravin [15] measures for exponential family models such as logit, probit, Poisson, negative binomial, and gamma. They suggested pseudo R-Square, also known as McFadden's R square, which can be estimated using:

$$R^2_{McF}=1\text{-}\frac{l_{\hat{\Omega}}}{l_{\hat{\omega}}} \tag{29}$$

Because the value of $R^2_{McF}$ will always increase with the number of independent variables used, an adjusted version is also provided:

$$\bar{R}^2_{McF}=1\text{-}\left(\frac{l_{\hat{\Omega}}\text{-}(J+1)}{l_{\hat{\omega}}}\right) \tag{30}$$

The interpretation of pseudo R-square is that a larger value indicates a better-fitting model, whereas a very small value of the pseudo R-square coefficient indicates a less satisfactory model.

# Result and Discussion

## Descriptive Statistics

According to the Statistics Indonesia (BPS), criminal acts are any intentional or unintentional actions that have occurred or are attempted, which may harm others in terms of their body, soul, property, honor, and others, and such actions are subject to imprisonment and confinement penalties. Criminal acts continue to pose a security threat to society, regardless of whether they are classified as minor or serious offenses, endangering both lives and property. As for the graphic of criminal acts in Lampung Province during the period 2017-2021, they are shown in Figure 3.

**E-ISSN**: 2720-9326
**P-ISSN**: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art4

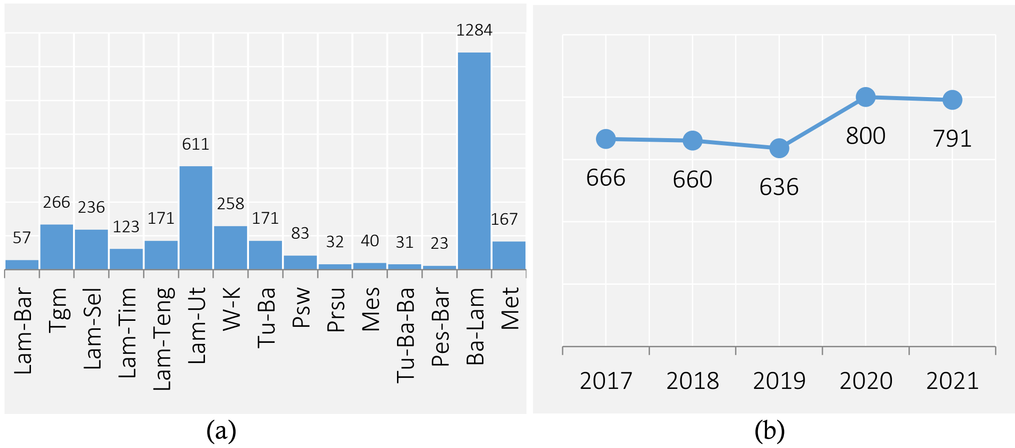


**Figure 3**. Diagram of Amount of Crime in Lampung Province
(a) According to Regency/City (b) According to Year.

Based on Figure 3(a), it can be observed that from the years 2017 to 2021, the regencies/cities in Lampung Province with the highest number of criminal cases were recorded in the city of Bandar Lampung, totaling 1.284 cases. This can be understood considering that Bandar Lampung is the largest district in Lampung Province with high economic activities and significant social inequality among the community. On the other hand, the regency/city with the lowest number of cases was Pesisir Barat Regency, with a total of 23 cases. This phenomenon occurs because of the low population density in Pesisir Barat Regency which results in community activity being quite low and enough for their daily lives [27].The average incidents per regency/city during the 2017-2021 period were recorded at 236,86 cases.

Whereas if the description according to the year of occurrence as Figure 3(b), data on the number of criminal acts in Lampung Province exhibits an increasing trend when examined over time. Although there was a decline in the first three years, starting from 666 cases in 2017, which then decreased to 660 cases in 2018, and further decreased to 636 cases in the subsequent year, there was a considerable increase in the last two years (2020 and 2021). The potential condition contributing to the rise in criminal cases in Lampung Province during these two years was likely influenced by the Covid-19 pandemic situation that occurred in Indonesia during that period. Various policies were implemented to curb economic activities, leading the population to resort to various means to meet their needs, including engaging in criminal activities. Meanwhile, for the numerical summary of the independent variables used, if observed for each year of occurrence in Lampung Province, it is presented in the following Table 1.

**Table 1**. The Numerical Summary of the Independent Variables for All Regencies/Cities per Year During 2017-2022 in Lampung Province.

| Variable | Year | | | | |
|---|---|---|---|---|---|
| | 2017 | 2018 | 2019 | 2020 | 2021 |
| $X_1$ | 13,1 | 12,6 | 12,1 | 11,8 | 12,1 |
| $X_2$ | 666,8 | 675,5 | 684,0 | 803,4 | 812,7 |
| $X_3$ | 9198,8 | 9567,7 | 9816,0 | 9718,5 | 9762,0 |
| $X_4$ | 3,873 | 3,92 | 3,96 | 4,404 | 4,315 |
| $X_5$ | 372817030 | 388812726 | 404876607 | 379221968 | 385370825 |
| $X_6$ | 7,8 | 7,9 | 8,0 | 8,1 | 8,2 |

Base on Table 1, known that the average of all districts/cities for each year during 2017-2021. For percentage of poverty in range 11,8% till highest is 13,1%. Overcrowding per district/city has an increasing trend, from 666,8 people per km in 2017 to 812,7 people per km on average per district/city. In last year during COVID-19, these numbers increased by 0,3% for percentage of poverty, and 9,3 people per km for overcrowding variable. In regards to the per capita expenditure variable ($X_3$), the highest average occurred in the year 2019, with a figure of Rp. 9.816 for each regency/city. Meanwhile, the lowest per capita expenditure figure occurred in the year 2017, amounting to Rp. 9.198,8. The annual average for per capita expenditure from the average of each regency/city

**E-ISSN**: 2720-9326
**P-ISSN**: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art4

in the Lampung Province is Rp. 9.612,6. Subsequently, Table 1 reveals that the variable of unemployment rate ($X_4$) averaged across all regencies/cities reached its highest level in the year 2020 at 4,404%. The lowest unemployment rate, recorded at 3,873%, occurred in the year 2017. The overall average percentage of open unemployment is 4,098%. There is a consistent upward trend in the annual unemployment rates. The variable of regional gross domestic product ($X_5$), which depicts economic activities in a certain region, reached its highest level in the year 2019 at Rp. 404.876.607. Meanwhile, the lowest regional gross domestic product (GDP) occurred in the year 2017, amounting to Rp. 372.817.030. The annual average regional gross domestic product (GDP) for each regency/city in the Lampung Province is Rp. 386.219.831. Over the 5-year period, the trend in the data indicates that the average GDP for each regency/city in the Lampung Province remains fluctuating. Information regarding the variable of average years of schooling is obtained. The longest average duration occurred in the year 2021, with a figure of 8,18 years encompassing all educational levels. On the other hand, the year with the lowest average years of schooling was 2017, with a figure of 7,807 years. The average value for the variable of average years of schooling for each year to complete the educational period is 7,99 years.

**Overdispersion Test**

The early detection of overdispersion is done by examining the mean and variance values of the data. However, to ensure the decision that the data is experiencing overdispersion, testing is employed. Table 2 is the mean and variance values of the dependent variable, along the result of overdispersion test.

**Table 2**. Overdispersion Test Result

| Variable | Mean | Variance | Chi-square | P-value |
|---|---|---|---|---|
| Y | 47,37 | 4246,156 | 66632,8 | $2,2 \times 10^{-16}$ |

Because Chi-square = 6632,8 > $\chi^2_{0,05;9}$ = 16.918 or p-value = $2.2 \times 10^{-16} < \alpha = 0.05$, then the null hypothesis $H_0$ is rejected, it's means that at the 5% level of significance there is enough evidence to state that the data on the amount of criminal cases experience overdispersion. That point supported with the value of mean and variance is very difference

**Binomial Negative Panel Regression Parameter Estimation**

Negative binomial panel regression was performed to obtain the estimation values from panel data with a discrete dependent variable. The modeling was conducted by regressing all research variables. The results of parameter estimation for each negative binomial panel model are as follows Table 3:

**Table 3.** Coefficient of Negative Binomial Panel Regression

| Parameter | CENB | FENB | RENB |
|---|---|---|---|
| Cons | -4,170 | -6,50 | -6,53 |
| $X_1$ | 0,479 | 0,760 | 0,627 |
| $X_2$ | 0,297 | 0,034 | 0,058 |
| $X_3$ | 0,551 | 0,693 | 0,468 |
| $X_4$ | 0,565 | 0,281 | 0,317 |
| $X_5$ | -0,057 | 0,373 | 0,412 |
| $X_6$ | -0,399 | 0,119 | 0,116 |
| $a$ | - | - | 3,725 |
| $b$ | - | - | 28,690 |

Based on Table 3, the estimated values of the regression coefficients for the CENB, FENB, and RENB models are presented. Meanwhile, parameters a and b represent the parameters for the RENB model as the values of the beta distribution parameters.

**Multcolinearity Detection**

The initial detection conducted to assess multicollinearity involved examining the extent of the correlation among independent variables. Meanwhile, as a step to precisely determine the presence of multicollinearity, the multicollinearity testing is conducted using the variance inflation factor (VIF) values. The correlation coefficients among the independent variables can be seen in the following Figure 4:

E-ISSN: 2720-9326
P-ISSN: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art4

**Figure 4**. Correlation Among Independent Variables

Based on Figure 4, most of correlation among independent variables is strong and positive correlation. Only correlation between $X_1$ and other independent variables is weak and negative correlation. However, the magnitude of the correlation values among the research variables is not sufficient evidence to conclude that multicollinearity exists in this data. Than, the multicollinearity testing is conducted using the variance inflation factor (VIF) values by Table 4 below:

**Table 4**. Variance Inflation Factor

| Variable | VIF | Variable | VIF |
|---|---|---|---|
| $X_1$ | 1,42 | $X_4$ | 2,94 |
| $X_2$ | 5,47 | $X_5$ | 1,55 |
| $X_3$ | 2,51 | $X_6$ | 5,69 |

If saw to VIF value at Table 4, they are small than 10, it can be observed that all variables have VIF values less than 10. Therefore, it can be concluded that the independent variables used in this study do not experience multicollinearity.

**Selection of Negative Binomial Panel Model**

The selected model indicates that it satisfies the assumptions present in the data. To obtain the suitable model, the testing was done among the common effect negative binomial, fixed effect negative binomial, and random effect negative binomial using Chow test, Hausman test, and Lagrange Multiplier test.

**Table 5**. Result of Model Selectetion

| Test | Statistic | Statistic Table | p-value | Model Selected |
|---|---|---|---|---|
| Chow test | 74,26 | $F_{(0,05;14;39)}$ = 1,954 | 0,000 | FENB |
| Hausman test | 1,75 | $\chi_{(0,05;6)}$ =12,591 | 0,941 | RENB |
| Lagrange Multiplier test | 137,21 | $\chi_{(0,05;1)}$ = 3,481 | 0,000 | RENB |

Table 5 shows that result of Hausman test and Larange Multiplier test is random effect negative binomial (RENB) model. It's means, model selected to this data panel is RENB that have advantage about specific individual effect and random effect.

**Simultaneous and Partial Tests of Parameters**

Simultaneous and partial tests are carried out repeatedly if there are variables when the partial test results are not significant. In more detail, a simultaneous test was carried out on model 1 and then a partial test was carried out on each variable in model 1. If in the first test there were variables that were not significant, then the variables that were not significant were not used again. Then re-modeling was carried out and simultaneous testing was carried out again without using variables that were not significant in the first test. continued partial testing for the second time. If in the second test there are still variables that are not significant, then the process will be carried out as the previous process. This process will be carried out until we find that there are no variables that are not significant. Result of likelihood ratio test (LRT) for simultaneous test and wild test for partial test in the random effect negative binomial model is can see at the **Table 6** below:

**Table 6**. Result of Likelihood Ratio Test (LRT) and Wald Test

| Likelihood Ratio Test | $\hat{\beta}$ | Model 1 ($M_1$) | | |
|---|---|---|---|---|
| | | Coefficient | $W_j$ | *p-value* |
| **First Test** LRT = 16,21 | $\beta_0$ | -6,5391 | -27,96 | **0,000** |
| | $\beta_1$ | 0,62742 | 4,28 | **0,000** |
| | $\beta_2$ | 0,0586 | 0,59 | 0,558 |
| | $\beta_3$ | 0,46887 | 2,19 | **0,028** |
| | $\beta_4$ | 0,31704 | 2,15 | **0,031** |
| | $\beta_5$ | 0,41329 | 2,98 | **0,003** |
| | $\beta_6$ | 0,11664 | 0,57 | 0,569 |
| | $\hat{\beta}$ | Model 2 ($M_2$) | | |
| | | Coefficient | $W_j$ | *p-value* |
| **Second Test** LRT = 16,43 | $\beta_0$ | -6,5409 | -27,98 | **0,000** |
| | $\beta_1$ | 0,6214 | 4,23 | **0,000** |
| | $\beta_2$ | 0,0543 | 0,54 | 0,592 |
| | $\beta_3$ | 0,5464 | 3,26 | **0,001** |
| | $\beta_4$ | 0,3479 | 2,55 | **0,011** |
| | $\beta_5$ | 0,3932 | 2,93 | **0,003** |
| | $\hat{\beta}$ | Model 3 ($M_3$) | | |
| | | Coefficient | $W_j$ | *p-value* |
| **Third Test** LRT = 17,26 | $\beta_0$ | -6,5352 | -27,89 | **0,000** |
| | $\beta_1$ | 0,60922 | 4,15 | **0,000** |
| | $\beta_3$ | 0,57685 | 3,63 | **0,000** |
| | $\beta_4$ | 0,38896 | 3,40 | **0,001** |
| | $\beta_5$ | 0,3774 | 2,89 | **0,004** |

The testing is performed three times. The first test is carried out on all independent variables (6 independent variables) referred to as Model 1 ($M_1$). Subsequently, the second and third tests are performed on the independent variables that were found to be significant during the previous test. Model 3 is the last model formed. From Table 6, result of LRT from third test is 17,26 > $\chi^2_{0,05;1} = 3{,}841$. Than, null hypothesis is rejected and it's means in this simultaneous testing with a significance level of 5%, there is sufficient evidence that at least one parameter influences the dependent variable. Furthermore, partial tests are needed to determine which variables have an influence. Partial testing in negative binomial panel regression is conducted using the Wald test. At Table 6 shows the result of wald test, which the first test with result $M_1$ that shows $\beta_2$, and $\beta_6$ are not significant. The variables are not significant not included to second test. The result of the second Wald test indicates that the variable $X_2$ is not significant, with a p-value = $0{,}592 > \alpha = 0{,}05$. Therefore, subsequently, the variable $X_2$ is no longer used and the modeling is performed again without $X_2$. Afterward, the third model is also subjected to a Wald test, and it is observed from Table 6 that all variables ($X_1$, $X_3$, $X_4$, $X_5$) are significant. With a p-value < $\alpha$ =0.05, it is decided to reject $H_0$, indicating that in this partial testing with a significance level of 5%, there is evidence that the variables that have an influence on the independent variable are variables $X_1$, $X_3$, $X_4$, and $X_5$.

Based on the results of the partial testing in Table 6**Error! Reference source not found.**, the negative binomial panel regression model with a random effect model is obtained as follows:

$$\hat{\lambda}_{it} = \exp(-6{,}535 + 0{,}609.X_{1it} + 0{,}576.X_{3it} + 0{,}388.X_{4it} + 0{,}377.X_{5it} + v_i)$$

The interpretation of the random effect negative binomial (RENB) model for the average number of criminal incidents ($\lambda$) is as follows:

For every 1% increase in the percentage of the population living in poverty ($X_1$), there will be a 0,609% increase in the number of criminal incidents from the original average number of cases, assuming that other variables are held constant. When a region has a high percentage of the population living in poverty, its community is more

**E-ISSN**: 2720-9326
**P-ISSN**: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art4

likely to engage in criminal activities. These findings are consistent with the study conducted by Putra *et al*. [28], which found a positive correlation between the poverty rate and the average number of criminal incidents. For an increase of 1% in per capita expenditure ($X_3$) will result in a 0,576% increase in the average number of criminal incidents from the original average, assuming that other variables are held constant. This statement is reasonable because when an individual resides in an area with a high per capita expenditure, their needs may increase, and if they are unable to meet the standard of living in their environment, they might resort to engaging in criminal activities to fulfill those needs. These findings align with the research by Ervina [29], which suggests that criminal activities are influenced by increased per capita expenditure. This implies that when a family resides in an environment with a high per capita expenditure, the likelihood of criminal incidents also increases. Whereas every 1% increase in the open unemployment rate ($X_4$), there will be a 0,388% increase in the number of criminal incidents from the original average number of cases, assuming that other variables are held constant. The results of Anata [30] are consistent with these findings, indicating that the open unemployment rate is one of the significant factors influencing the occurrence of criminal activities in society. The last, for every 1% increase in the gross regional domestic income ($X_5$), there will be a 0,377% increase in the number of criminal incidents from the original average number of cases, assuming that other variables are held constant. However, these findings do not align with the concept of benefit and cost [31], which suggests a negative relationship between gross regional domestic income and criminal activities in society. Nevertheless, Purwati and Widyaningsih [32] state that gross regional domestic income only influences certain types of criminal activities, such as theft, robbery, confiscation, and mugging. Therefore, the results of this study, which indicate a positive relationship between gross regional domestic income and the number of criminal incidents, are assumed to be applicable to other types of criminal activities, such as drug-related crimes. This implies that as the gross regional domestic income increases, the number of drug-related criminal incidents will also rise due to the availability of resources to purchase desired goods. To result that the Regional Gross Domestic Product (GDP) and per capita income have a linear influence on the number of criminal cases of drug abuse crimes, suggests that the proposed approach for handling this issue is, according to Lloyd *et al*. [33], providing drug awareness education at the most basic level of education (primary school). This has proven to be effective, as indicated by the results obtained in the United Kingdom (UK). It is worth noting that this proposal is somewhat contradictory to research findings that the variable of average years of schooling, in reality, does not have a significant impact. It should be remembered that the average years of schooling represent the average time someone spends in formal education with general knowledge in natural and social sciences, which does not reflect how much students know and understand the dangers of drug abuse. On the other hand, this effort is a long-term initiative with results expected to be realized around 10-20 years in the future.

The magnitude of the random effect from each district/city in Lampung Province in the negative binomial random effect model can be seen in the following Table 7:

**Table 7**. Random Effect of Each Regency/City

| Regencies/Cities | $v_i$ | Regencies/Cities | $v_i$ |
|---|---|---|---|
| Lampung Barat | 1,03 | Pesawaran | 0,57 |
| Tanggamus | 1,51 | Pringsewu | -3,43 |
| Lampung Selatan | 2,49 | Mesuji | -0,55 |
| Lampung Timur | 0,75 | Tulang Bawang Barat | -2,59 |
| Lampung Tengah | 1,78 | Pesisir Barat | -2,39 |
| Lampung Utara | 2,82 | Bandar Lampung | 2,50 |
| Way Kanan | 12,32 | Metro | 0,45 |
| Tulang Bawang | -0,48 | | |

Table 7 reveals that the largest random effect is observed in Way Kanan Regency, while the smallest random effect is in Metro City. The random effects in the Table 7 show that there are individual random effects that are unobserved and stochastic in nature. This figure provides information about the magnitude of unobserved individual variance in the variables in this panel data. Furthermore, the visual representation of the magnitude of random effects influencing the number of criminal incidents in Lampung Province is displayed in the following Figure 5:

**E-ISSN**: 2720-9326
**P-ISSN**: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art4

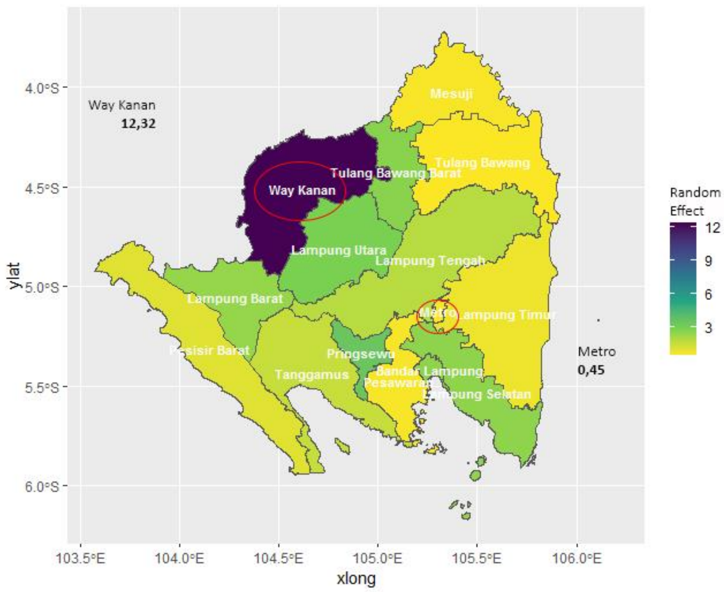


**Figure 5**. Visualization of Random Effect Each Regency/City

When assessing the magnitude of the factors influencing the number of criminal offenses in Lampung Province, the evaluation is based on the magnitude of the resulting pseudo R-square. The pseudo R-square value is obtained from the log-likelihood of the full panel negative binomial regression model ($\ell_{\hat{\Omega}}$) and the log-likelihood of the panel negative binomial regression model without independent variables ($\ell_{\hat{\omega}}$). The calculation results are listed in Table 8:

**Table 8**. Goodness of Fit Result

| $\ell_{\hat{\omega}}$ | $\ell_{\hat{\Omega}}$ | *Pseudo*-$R^2$ ($R^2_{McF}$) |
|---|---|---|
| -33,4715 | -312,29 | 0,8928 |
| | | *Adjusted* ( $\bar{R}^2_{McF}$) |
| | | 0,870 |

Based on Table 8, it can be observed that the Pseudo-R^2 value is 0.8928 or the Adjusted Pseudo-R^2 value is 0.870, indicating that these values are relatively close to 1. These results suggest that the model used is quite effective in explaining the state of the number of criminal offenses in Lampung Province.

## Conclusion

Descriptively, Bandar Lampung City emerged as the area with the highest crime rates during the period 2017-2021 with 1.284 cases, followed by North Lampung with 611 cases. The district with the fewest cases is the West Coast District with only 23 cases during the period 2017-2021. To the best model for data on the number of crimes in Lampung province from 2017 to 2021 using the negative binomial panel regression method is the negative binomial random effects model. The negative binomial random effects model provides a model that takes into account the individual random effects of the constructed model. Then, for the contribution to dealing with the number of crimes in Lampung province, the percentage of poor ($X_1$) has an effect of 0.609%, expenditure per capita ($X_3$) has an effect of 0.576%, the open unemployment rate ($X_4$) has an effect of 0.388% and the gross regional income ($X_5$) has an effect of 0.377% for a 1% increase in each unit of data from variables $X_1$, $X_3$, $X_4$, and $X_5$. Then, among the 15 regencies/cities, the largest random effect was in Way Kanan district with an effect of 12,32 and the smallest effect was Metro City with an effect of 0.45. This means that in Way Kanan District, there is such a large random effect of variables outside of the currently studied variables that it influences the number of crimes that occur in Way Kanan, the same interpretation for Metro City.

## References

[1] S. Rahmalia, Ariusni, dan M. Triani, “Pengaruh Tingkat Pendidikan, Pengangguran, dan Kemiskinan Terhadap Kriminalitas di Indonesia,” *J. Kaji. Ekon. dan Pembang.*, vol. 1, no. 1, hal. 21–36, 2019.

[2] S. Adri, S. Karimi, dan I. Indrawari, “Pengaruh Faktor Sosial Ekonomi terhadap Perilaku Kriminalitas,” *J. Ilm. Adm. Publik*, vol. 5, no. 2, hal. 181–186, 2019.

[3] A. Harefa, “Faktor-faktor Penyebab Terjadinya Tindak Pidana Kekerasan dalam Rumah Tangga,” *J. Panah Keadilan*, vol. 1, no. 1, hal. 18–21, 2021.

[4] A. H. Youssef, M. R. Abonazel, dan E. G. Ahmed, “Estimating the Number of Patents in the World Using Count Panel Data Models,” *Asian J. Probab. Stat.*, hal. 24–33, 2020.

[5] A. C. Cameron dan P. K. Trivedi, “Count Panel Data,” *Oxford Handb. Panel Data*, hal. 233–256, 2013.

[6] D. W. Osgood, “Poisson-Based Regression Analysis of Aggregate Crime Rates,” *J. Quant. Criminol.*, vol. 16, no. 1, hal. 21–43, 2000.

[7] S. Sun, J. Bi, M. Guillen, dan A. M. Pérez-Marín, “Driving risk assessment using near-miss events based on panel poisson regression and panel negative binomial regression,” *Entropy*, vol. 23, no. 7, hal. 1–22, 2021.

[8] M. O. Adenomon dan G. S. Akinyemi, “Statistical Analysis of Tuberculosis and HIV Cases in West Africa Using Panel Poisson and Negative Binomial Regression Models,” *2020 Int. Conf. Math. Comput. Eng. Comput. Sci. ICMCECS 2020*, 2020.

[9] Suliyanto, “Modelling the Amount of Crime in Indonesia Using Geographically Weighted Negative Binomial Regression Approach,” in *Proceeding of the Second International Conference on Pllied Statistics 2016*, 2016, hal. 103–109.

[10] Y. Febriani, “Pengaruh Aspek Sumber Daya Manusia Terhadap Jumlah Kriminalitas di Sumatera Selatan Tahun 2019,” *J. Media Wahana Ekon.*, vol. 18, no. 1, hal. 146, 2021.

[11] A. O. Edwart dan Z. Azhar, “Pengaruh Tingkat Pendidikan, Kepadatan Penduduk Dan Ketimpangan Pendapatan Terhadap Kriminalitas Di Indonesia,” *J. Kaji. Ekon. dan Pembang.*, vol. 1, no. 3, hal. 759, 2019.

[12] F. Liu dan D. Pitt, “Application of bivariate negative binomial regression model in analysing insurance count data,” *Ann. Actuar. Sci.*, vol. 11, no. 2, hal. 390–411, 2017.

[13] Boswell dan Patil, *Statistical Distributions in Scientific Work*, vol. 2. Boston: D. Reidel Publishing Company, 1974.

[14] B. J. Park dan D. Lord, “Adjustment for maximum likelihood estimate of negative binomial dispersion parameter,” *Transp. Res. Rec.*, vol. 2061, no. 1, hal. 9–19, 2008.

[15] C. A. Colin dan T. Pravin, *Regression analysis of count data*, 2 ed. United States: Cambridge University Press, 2013.

[16] J. M. Hilbe, *Negative Binomial Regression*, Second., vol. 4, no. 1. New York: Cambridge University Press, 2011.

[17] S. Angnitha Purba, “Estimasi Parameter Data Berdistribusi Normal Menggunakan Maksimum Likelihood Berdasarkan Newton Raphson,” *J. Sains Dasar*, vol. 9, no. 1, hal. 16–18, 2021.

[18] R. U. Datangeji, A. Warsito, H. I. Sutaji, dan L. A. S. Lapono, “Kajian Distribusi Intensitas Cahaya pada Fenomena Difraksi Celah Tunggal dengan Metode Bagi Dua dan Metode Newton Raphson,” *J. Fis. Fis. Sains dan Apl.*, vol. 4, no. 2, hal. 56–69, 2019, doi: 10.35508/fisa.v4i2.976.

[19] B. H. Baltagi, *The Oxford Handbook of Panel Data*. New York: Oxford University Press, 2015.

[20] J. . Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, 2 ed. London: The MIT Press, 2013.

[21] W. H. Greene, *LIMDEP Econometric Modeling Guide*, 10 ed. New York: Econometric Software Inc, 2012.

[22] J. Hausman, B. H. Hall, dan Z. Griliches, “Econometric Models for Count Data with an Application to the Patents-R & D Relationship,” vol. 52, no. 4, hal. 909–938, 1984.

[23] R. R. Hocking, “Methods and Applications of Linear Models,” in *Technometrics*, vol. 39, no. 3, 1997, hal. 332. doi: 10.2307/1271138.

[24] P. R. Sihombing dan M. Sundari, “Pemodelan Data Diskrit Dengan MenggunakanRegresi Panel Poisson,” 2021.

[25] B. H. Baltagi, *Econometrics*, 5 ed. New York: Springer, 2011.

[26] W. H. Greene, *Econometric Analysis*, 5 ed. New York: Pearson Education Inc, 2002.

[27] Dermawanti, A. Hoyyi, dan A. Rusgiyono, “Faktor-Faktor Yang Mempengaruhi Kriminalitas Di Kabupaten Batang Tahun 2013 Dengan Analisis Jalur,” *J. Gaussian*, vol. 4, no. 2, hal. 247–256, 2015.

[28] A. D. Putra, G. S. Martha, M. Fikram, dan R. J. Yuhan, “Faktor-Faktor yang Memengaruhi Tingkat

**E-ISSN**: 2720-9326
**P-ISSN**: 2716-0459
DOI : 10.20885/EKSAKTA.vol5.iss1.art4

Kriminalitas di Indonesia Tahun 2018," *Indones. J. Appl. Stat.*, vol. 3, no. 2, hal. 123, 2021.

[29] M. E. Ervina, "Hubungan Rata-rata Lama Sekolah, Gini Ratio, dan Pengeluaran Per Kapita dengan Tingkat Kejahatan Tahun 2011-2018," *Euclid*, vol. 7, no. 2, hal. 97, 2020.

[30] F. Anata, "Pengaruh Tingkat Pengangguran Terbuka, PDRB perkapita, Jumlah Penduduk dan Indeks Williamson Terhadap Tingkat Kriminalitas (Studi pada 31 Provinsi Di Indonesia Tahun 2007-2012)," *J. Ilm. Mhs. FEB Univ. Brawijaya*, vol. 1, no. 2, 2013.

[31] D. G. Omotor, "Demographic and Socio-Economic Determinants of Crimes in Nigeria (A Panel Data Analysis)," *J. Appl. Bus. Econ.*, vol. 11, no. 1, hal. 185–195, 2010.

[32] E. Y. Purwanti dan E. Widyaningsih, "Analisis Faktor Ekonomi Yang Mempengaruhi Kriminalitas Di Jawa Timur," *J. Ekon.*, vol. 9, no. 2, 2019.

[33] C. Lloyd, Ruth Joyce, Jane Hurry, Mike, "The Effectiveness of Primary School Drug Education," *Drugs Educ. Prev. Policy*, vol. 7, no. 2, hal. 109–126, 2000.