

Modeling of HIV and AIDS in Indonesia Using Bivariate Negative Binomial Regression

by Fatmawati Fatmawati

Submission date: 07-Feb-2020 08:15PM (UTC+0800)

Submission ID: 1253119638

File name: 5_Paper_IOP_Conf._Ser._Mater._Sci._Eng._546_052079.pdf (1.02M)

Word count: 2405

Character count: 12612

PAPER · OPEN ACCESS

Modeling of HIV and AIDS in Indonesia Using Bivariate Negative Binomial Regression

3
To cite this article: Amin Tohari *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **546** 052079

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Modeling of HIV and AIDS in Indonesia Using Bivariate Negative Binomial Regression

Amin Tohari¹, Nur Chamidah², Fatmawati²

¹PhD Student, Faculty of Sciences and Technology, Airlangga University, Indonesia

²Department of Mathematics, Faculty of Sciences and Technology, Airlangga University, Indonesia

nur-c@fst.unair.ac.id

Abstract. The problem of HIV and AIDS in Indonesia is a frightening health problem with a number of cases that tend to increase each year. The aim of this research is to model the number of HIV and AIDS cases in Indonesia using bivariate negative binomial regression approach. Bivariate negative binomial regression is a regression method for modeling a pair of response variables in the form of count data with negative binomial distribution and correlating to each other. This research uses secondary data from the ministry of health in 2017 about the number of HIV and AIDS cases in Indonesia. From the results of this research, we obtained the deviance value of 38.9197 which was used to describe the goodness of fit test.

Keywords: Bivariate Negative Binomial Regression, HIV, AIDS, Drug Users.

1. Introduction

Regression is used to explain the functional relationship between the response variable and the predictor variable. The discrete response variable in the form of a count data will follow a Poisson distribution that requires variance equal to the mean [1]. The conditions in the field show that over dispersion often occurs, namely a condition where the variance is greater than the mean [2]. When over dispersion occurs; it is better to use negative binomial regression [1, 3-4]. Bivariate count data regression models are used to calculate event that mutually depends on each other [5]. Paired count events that show correlations must be estimated together, and bivariate negative binomial regression models are designed to handle over dispersion cases in bivariate Poisson regression models [6]. Several previous studies that examined negative bivariate binomial regression were conducted by [7-11].

The number of Human Immunodeficiency Virus (HIV) cases and the number of Acquired Immuno Deficiency Syndrome (AIDS) cases in Indonesia are correlated to each other, so they are thought to have a high correlation. Moreover, data on the number of HIV and AIDS cases is a count data, so that the appropriate modeling is using bivariate negative binomial regression. HIV is one of the viruses that lowers the immune system so that people affected by the HIV virus will become subject to various infections and cause AIDS [12]. Since 2006 Indonesia has been categorized as a country with high rates of HIV transmission, the cumulative number of HIV infections reported up to December 2017 as many as 280,623, while the cumulative number of AIDS reported as of December 2017 is 102,667 [13].



Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Published under licence by IOP Publishing Ltd

Based on the above explanation, the number of HIV and AIDS cases will be modeled based on the predictor variables that influence it.

There are many researchers that who have studied about estimation of regression model with more than one response variables. There are many researchers who have studied multiresponse regression model, i.e., kernel and local linear estimator has been used by [14] & [15] for estimating median growth chart children in Surabaya, spline estimator has been studied by [16] for designing standard growth charts children in East Java.

2. Materials and Methods

The data number of HIV and AIDS cases used in this study is secondary data from the Ministry of Health Republic of Indonesia in 2017 [17], whereas drug users data is secondary data from the National Narcotics Agency Republic of Indonesia in 2017 [18].

Modeling number of HIV and AIDS cases in Indonesia with a percentage of drug users as predictor variable using bivariate negative binomial regression is carried out by the following steps:

- a. Test the correlation for the response variable
- b. Conduct bivariate negative binomial regression parameter estimates

The bivariate negative binomial regression model considers two response variables (y_1 and y_2) which are correlated each other. Suppose that data is given in pairs (x_i, y_{1i}, y_{2i}) , with $i = 1, 2, \dots, n$ and n is the number of observations observed. Response variable y_{1i} and y_{2i} is a discrete type random variable which is assumed to be a bivariate negative binomial distribution and has a regression model as follows [19]:

$$f(y_{1i}, y_{2i} | x_i) = \frac{\Gamma\left(\frac{1}{\alpha} + y_{1i} + y_{2i}\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma(y_{1i} + 1)\Gamma(y_{2i} + 1)} [\mu_1(x_i)]^{y_{1i}} [\mu_2(x_i)]^{y_{2i}} \alpha^{-\frac{1}{\alpha}} \left(\frac{1}{\alpha} + \mu_1(x_i) + \mu_2(x_i)\right)^{-\left(\frac{1}{\alpha} + y_{1i} + y_{2i}\right)} \tag{1}$$

where $\mu_1(x_i) = \exp(x_i^T \beta_1)$, and (2)

$\mu_2(x_i) = \exp(x_i^T \beta_2)$. (3)

The estimation used in the bivariate negative binomial regression model is the maximum likelihood estimation (MLE). To get an estimator β using MLE is with the following steps:

- 1) Taking n random samples $(y_{1i}, y_{2i}, x_i) \quad i = 1, 2, \dots, n$

2) Determine a likelihood function

$$\ell(\beta_1, \beta_2, \alpha) = \prod_{i=1}^n \left\{ \frac{\Gamma\left(\frac{1}{\alpha} + y_{1i} + y_{2i}\right)}{\Gamma\left(\frac{1}{\alpha}\right)\Gamma(y_{1i} + 1)\Gamma(y_{2i} + 1)} [\mu_1(x_i)]^{y_{1i}} [\mu_2(x_i)]^{y_{2i}} \times \alpha^{-\frac{1}{\alpha}} \left(\frac{1}{\alpha} + \mu_1(x_i) + \mu_2(x_i)\right)^{-\left(\frac{1}{\alpha} + y_{1i} + y_{2i}\right)} \right\} \tag{4}$$

- 3) Determine a log-likelihood function

$$L(\beta_1, \beta_2, \alpha) = \sum_{i=1}^n \left\{ \ln \Gamma \left(\frac{1}{\alpha} + y_{1i} + y_{2i} \right) - \ln \Gamma \left(\frac{1}{\alpha} \right) - \ln \Gamma (y_{1i} + 1) - \ln \Gamma (y_{2i} + 1) + y_{1i} \ln (\mu_1(x_i)) + y_{2i} \ln (\mu_2(x_i)) - \frac{1}{\alpha} \ln \alpha - \left(\frac{1}{\alpha} + y_{1i} + y_{2i} \right) \ln \left(\frac{1}{\alpha} + \mu_1(x_i) + \mu_2(x_i) \right) \right\} \quad (5)$$

- 4) Maximize the log-likelihood function by derivate the log-likelihood function to its parameters and equating to zero.

then the first derivative $L(\beta_1, \beta_2, \alpha)$ to β_1 is:

$$\frac{dL(\beta_1, \beta_2, \alpha)}{d\beta_1} = \sum_{i=1}^n \left\{ \left(\frac{y_{1i}}{\mu_1(x_i)} - \frac{(\alpha^{-1} + y_{1i} + y_{2i})}{(\alpha^{-1} + \mu_1(x_i) + \mu_2(x_i))} \right) x_i \mu_1(x_i) \right\} \quad (6)$$

the first derivative $L(\beta_1, \beta_2, \alpha)$ to β_2 is:

$$\frac{dL(\beta_1, \beta_2, \alpha)}{d\beta_2} = \sum_{i=1}^n \left\{ \left(\frac{y_{2i}}{\mu_2(x_i)} - \frac{(\alpha^{-1} + y_{1i} + y_{2i})}{(\alpha^{-1} + \mu_1(x_i) + \mu_2(x_i))} \right) \mu_2(x_i) x_i \right\} \quad (7)$$

while the first derivative $L(\beta_1, \beta_2, \alpha)$ to α is:

$$\frac{dL(\beta_1, \beta_2, \alpha)}{d\alpha} = \sum_{i=1}^n \left\{ \psi \left(\frac{1}{\alpha} + y_{1i} + y_{2i} \right) - \psi \left(\frac{1}{\alpha} \right) + \frac{\ln \left(\frac{1}{\alpha} + \mu_1(x_i) + \mu_2(x_i) \right)}{\alpha^2} - \frac{\frac{1}{\alpha} \left[-\frac{1}{\alpha} \ln \alpha - \left(\frac{1}{\alpha} + y_{1i} + y_{2i} \right) \right]}{\left(\frac{1}{\alpha} + \mu_1(x_i) + \mu_2(x_i) \right)} \right\} \quad (8)$$

Because the results are not close form, so one of the numerical approaches that can be used is the Newton-Raphson method. Through the Newton-Raphson iteration process, a maximum likelihood estimator can be obtained for β , where $\beta^{(m)}$ is a parameter estimate in the m^{th} iteration.

Newton-raphson iteration process algorithm for finding an estimator for β , first determine is the vector g , which is the first derivative of the log likelihood function for its parameters. Next, determine the matrix H , the elements are the second derivative of the parameter.

- c. Test Bivariate Negative Binomial Regression Parameters

According to [11], a measure of goodness of fit for the bivariate negative binomial regression model may be based on the deviance statistic D , the hypothesis used is as follows:

$$H_0 : \beta_{r1} = \beta_{r2} = \dots = \beta_{rk} = 0; r = 1, 2$$

$$H_1 : \text{there is at least one } \beta_{rj} \neq 0; r = 1, 2; j = 1, 2, \dots, k$$

and the deviance statistic is defined as:

$$D = 2(\log L(y; y) - \log L(\mu; y)) \quad (9)$$

The deviance statistic D can be approximated by a chi-square distribution with $n - p$ degrees of freedom.

- d. Interpret the model
e. Make conclusions

3. Result and Discussion

The first step to model bivariate negative binomial regression is to calculate the correlation coefficient between y_1 with y_2 . The results of the calculation of correlation using OSS-R is 0.649 which means that the closeness between the number of HIV cases and the number of AIDS cases is strong, so that bivariate negative binomial regression modeling can be done.

The results of modeling the number of HIV and AIDS cases in Indonesia with the bivariate negative binomial regression approach are as follows:

$$\hat{\mu}_1 = \exp(5.6358 + 0.2648x) \quad (10)$$

$$\hat{\mu}_2 = \exp(3.9606 + 0.2838x) \quad (11)$$

According to the equation (10) it can be interpreted that each addition of 1 percent of drug users will result in an increase in the number of HIV cases in Indonesia amounting to 1.3032 times from the previous case. Likewise in equation (11) it can be interpreted that each addition of 1 percent of drug users will result in an increase in the number of AIDS cases in Indonesia amounting to 1.3282 times from the previous case.

Bivariate negative binomial regression model for the number of HIV and AIDS cases in Indonesia obtained deviance value as a measure of goodness of fit test with predictor variable percentage of drug users that is equal to 38.9197 which is smaller than the $\chi_{(0.05,32)}$ of 46.1943, its means the model is appropriate.

The following will show the plot between the observation and the estimation results in the first response variable and the second response in Figure 1 and Figure 2, where the circle states the observation and the line form states the estimated results of the response variable.

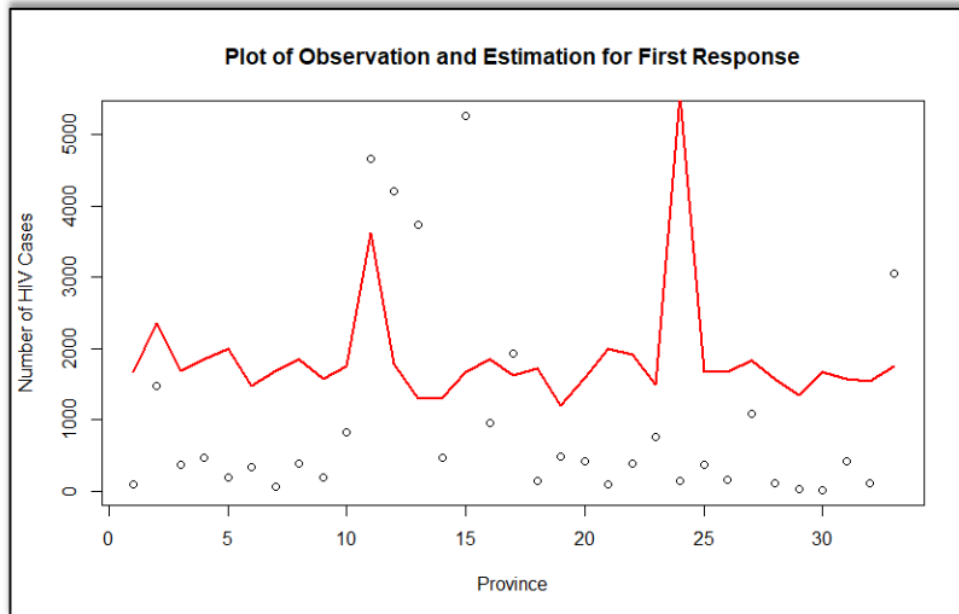


Figure 1. A plot of Observation and Estimation for Number of HIV Cases

Figure 1 above shows a plot between observations and estimation results for the number of HIV cases in Indonesia with 1 predictor variable, namely the percentage of drug users. According to the equation (10), for example in the province of East Java in 2017 the number of HIV cases was 5263 cases, so with an increase of 1 percent of drug users in East Java, it would increase the number of HIV cases in East Java to 6859 cases.

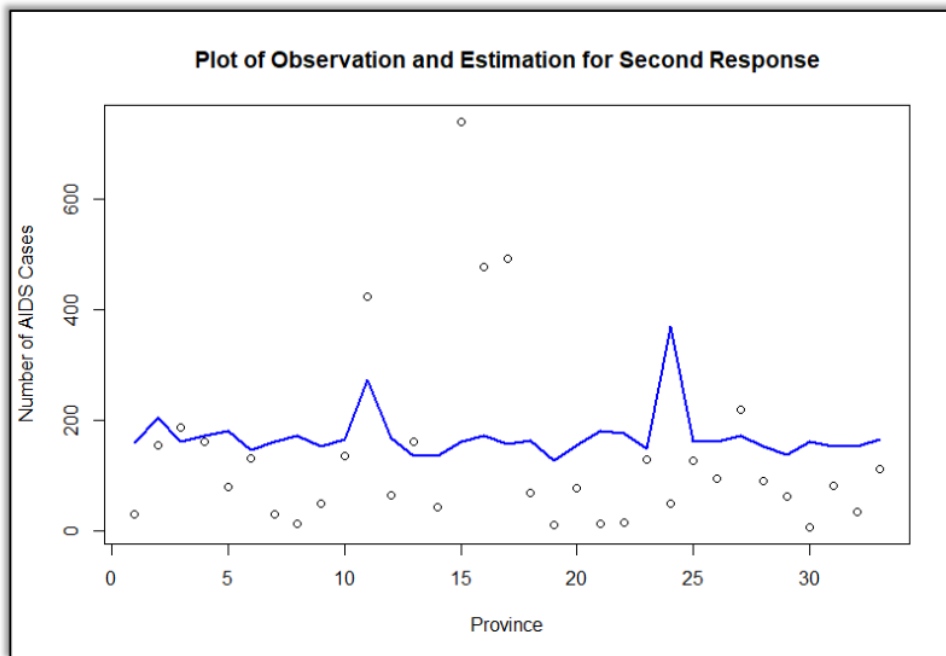


Figure 2. Plot of Observation and Estimation for Number of AIDS Cases

Figure 2 above shows a plot between observation and estimation results for the number of AIDS cases in Indonesia with 1 predictor variable, namely the percentage of drug users. According to the equation (11), for example in the province of East Java in 2017 the number of AIDS cases was 741 cases, so with an increase of 1 percent of drug users in East Java, it would increase the number of AIDS cases in East Java to 985 cases.

4. Conclusion

According to the modeling results of the number of HIV and AIDS cases in Indonesia using bivariate negative binomial regression, the goodness of fit was obtained with a deviance value are smaller than chi-square value, its means the model is appropriate. Suggestions can be given to the Ministry of Health Republic of Indonesia to reduce the percentage of drug users so that the increase in the number of HIV and AIDS cases in Indonesia can be reduced. For further research, bivariate negative binomial regression modeling with a nonparametric approach can be used.

5. Acknowledgement

The author gratefully acknowledges the financial support received from the faculty of sciences and technology Airlangga University (Doctoral Dissertation Research).

References

- [1] Hilbe, J.M. 2011 *Negative Binomial Regression*, 2nd edition. Cambridge University Press
- [2] McCullagh, P. and Nelder, J. A. 1989 *Generalized Linear Models 2nd Edition*. London: Chapman and Hall
- [3] Lawless, J.F. 1987 Negative Binomial and Mixed Poisson Regression. *The Canadian Journal of Statistics*, **15** pp 209-225
- [4] Cameron, A.C. and Trivedi, P. K. 1998 *Regresi Analysis of Count Data*. Cambridge University Press. United Kingdom
- [5] Gurmu, S. and Elder, J. 2007 Generalized bivariate count data regression model. *Economic Bulletin*, **3** (11) pp 1-10
- [6] Chou, N. and Steenhard, D. 2011 Bivariate Count Data Regression Model-A SAS Macro Program. *Proceedings SAS Global Forum 2011*, paper 355-2011
- [7] Edward, C.B. and Gurlan, J. 1961 A Class of Distribution Applicable to Accident. *Journal of the American Statistical Association*, **56** pp 503-517
- [8] Maher, M.J. 1990 A Bivariate Negative Binomial Model to Explain Traffic Accident Migration. *Accident Analysis and Prevention*, **22**(5) pp 487-498
- [9] Iwasaki, M. and Tsubaki, H. 2006 Bivariate negative binomial generalized linear model for environmental count data. *Journal of Applied Statistics*. **33**(9) pp 909-923
- [10] Atella, V. And Deb, P. 2008 Are Primary Care Physicians, Public and Private Sector Specialists Substitutes or Complements? Evidence from a Simultaneous Equations Model for Count Data. *Journal of Health Economic*. **27** pp 770-785
- [11] Famoye, F. 2010 On the Bivariate Negative Binomial Regression Model. *Journal of Applied Statistics*. **37** pp 969 – 981
- [12] Saputro, A.S. 2013 *Pemodelan mixture survival untuk kasus HIV/AIDS*. Universitas Airlangga
- [13] Kemenkes RI. 2018 Laporan Perkembangan HIV-AIDS & Infeksi Menular Seksual (IMS) Triwulan IV Tahun 2017
- [14] Chamidah N and Saifudin T 2013 Estimation of children growth curve based on kernel smoothing in multi-response nonparametric regression *Applied Mathematical Sciences* **7**(37) 1839-1847
- [15] Chamidah N and Rifada M 2016 Local linear estimator in bi-response semiparametric regression model for estimating median growth chart of children *Far East Journal of Mathematical Sciences (FJMS)* **99**(8) 1233-1244
- [16] Chamidah N, Kurniawan A, Zaman B and Muniroh L 2018 Least square spline estimator in multi-response semiparametric regression model for estimating median growth charts of children in East Java, Indonesia *Far East Journal of Mathematical Sciences (FJMS)* **107**(2) 295-307
- [17] Kemenkes RI. 2018 Data dan Informasi Profil Kesehatan Indonesia 2017
- [18] Badan Narkotika Nasional 2017 Executive Summary Survei Penyalahgunaan Narkoba di Indonesia Tahun 2017
- [19] Cheon, S., Song, S.H., and Jung, B.C. 2009 Tests for independence in a bivariate negative binomial model. *Journal of the Korean Statistical Society*, **38** pp 185-90

Modeling of HIV and AIDS in Indonesia Using Bivariate Negative Binomial Regression

ORIGINALITY REPORT

16%

SIMILARITY INDEX

10%

INTERNET SOURCES

15%

PUBLICATIONS

1%

STUDENT PAPERS

PRIMARY SOURCES

1	hal.sorbonne-universite.fr Internet Source	3%
2	www.tandfonline.com Internet Source	2%
3	elar.usfeu.ru Internet Source	2%
4	id.scribd.com Internet Source	1%
5	Byoung Cheol Jung. "Score tests for overdispersion in the bivariate negative binomial models", Journal of Statistical Computation and Simulation, 2008 Publication	1%
6	"Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation", Springer Science and Business Media LLC, 1999 Publication	1%

7	<p>N Chamidah, E Tjahjono, A R Fadilah, B Lestari. "Standard Growth Charts for Weight of Children in East Java Using Local Linear Estimator", Journal of Physics: Conference Series, 2018</p>	1 %
Publication		
8	<p>R. Laroche, B. Dupont, J.E. Touze, H. Taelman, J. Bogaerts, A. Kadio, P. M'pele, A. Latif, P. Aubry, J.P. Durbec, J.F. Sauniere. "Cryptococcal meningitis associated with acquired immunodeficiency syndrome (AIDS) in African patients: treatment with fluconazole", Medical Mycology, 1992</p>	1 %
Publication		
9	<p>Famoye, Felix. "A Multivariate Generalized Poisson Regression Model", Communication in Statistics- Theory and Methods, 2015.</p>	1 %
Publication		
10	<p>journal.stats.id</p>	<1 %
Internet Source		
11	<p>Cheon, S.. "Tests for independence in a bivariate negative binomial model", Journal of the Korean Statistical Society, 200906</p>	<1 %
Publication		
12	<p>Darnah, M I Utoyo, N Chamidah. "Modeling of Maternal Mortality and Infant Mortality Cases in East Kalimantan using Poisson Regression</p>	<1 %

Approach Based on Local Linear Estimator",
IOP Conference Series: Earth and
Environmental Science, 2019

Publication

13

J.-P. Jaffrezou, G. Chen, G. E. Duran, J.-S. Kuhl,
B. I. Sikic. "Mutation Rates and Mechanisms of
Resistance to Etoposide Determined From
Flucattion Analysis", JNCI Journal of the
National Cancer Institute, 1994

<1 %

Publication

14

Ka Zeng, Richard Sherman. "Foreign direct
investment and industry demands for trade
protection", Review of International Political
Economy, 2009

<1 %

Publication

15

M. Ataharul Islam, Rafiqul I Chowdhury.
"Analysis of Repeated Measures Data",
Springer Nature, 2017

<1 %

Publication

16

Wang-Shu. Lu. "Score tests for overdispersion
in poisson regression models", Journal of
Statistical Computation and Simulation, 1997

<1 %

Publication

17

Caliendo, Ciro, Maria Luisa De Guglielmo, and
Maurizio Guida. "A crash-prediction model for
road tunnels", Accident Analysis & Prevention,
2013.

<1 %

Publication

Exclude quotes Off

Exclude matches Off

Exclude bibliography On

Modeling of HIV and AIDS in Indonesia Using Bivariate Negative Binomial Regression

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7
