

Cek Similarity/1. artikel annals of biology 215-219 (published).pdf

By Amin Tohari

Modelling of HIV and AIDS Cases in Indonesia Using Bi-response Negative Binomial Regression Approach Based on Local Linear Estimator

AMIN TOHARI, NUR CHAMIDAH* AND FATMAWATI

Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Indonesia
*(e-mail : nur-c@fst.unair.ac.id, Phone: +62315936501)

ABSTRACT

As in June 2018, 433 (84.2%) out of 514 districts/cities in 34 provinces in Indonesia reported HIV/AIDS. Therefore, goal of this study was to model the number of HIV and AIDS cases in Indonesia by using bi-response Negative Binomial non-parametric regression based on the local linear estimator. The best model of HIV and AIDS cases with optimal bandwidth of 0.101 and deviance value of 37.73 was obtained. This was smaller than the deviance value of the parametric regression approach i. e. 47.36. It meant that bi-response non-parametric regression approach based on the local linear estimator was better than the parametric regression approach. For example, in Maluku province, if there was an increase in 1% of drug users, the number of HIV cases will increase by 1.51 times than before, and the number of AIDS cases will increase by 1.37 times than before.

Key words : HIV, AIDS, Negative Binomial regression approach, local linear estimator

INTRODUCTION

HIV/AIDS is a contagious disease that occurs in the community that has not found an effective vaccine or drug for the prevention. Globally, in 2017, the number of new HIV infections has declined. New infections (all ages) have gone down from 3.4 million (2.6-4.4 million) in 1996 to 1.8 million (1.4-2.4 million) in 2017. However, in 2020, it should decrease less than 500,000 (UNAIDS, 2018). In 2017, the Asia and Pacific regions were known to be home by around 5.2 million people living with HIV, and Indonesia accounted for nearly three-quarters of them in the region (UNAIDS, 2018). As of June 2018, there were 433 (84.2%) of 514 districts/cities in 34 provinces in Indonesia who reported HIV/AIDS. HIV infections reported as of June 2018 have a cumulative number of 301,959 people (47% of people with HIV/AIDS, namely, ODHA) out of 640,443 and most often in the age group of 25-49 years and 20-24 years. As for the province with the highest number of HIV (human immunodeficiency virus) infections was DKI Jakarta (55,099), followed by East Java (43,399), West Java (31,293), Papua (30,699) and Central Java (24,757) (Pusdatin, 2018). So, it is exciting to discuss what factors growing up ODHA cases in Indonesia. Ardani and Handayani (2017) stated that one of the causes the increase ODHA was the use

of unsterile syringes in drug addicts. Bradley *et al.* (2019) stated that increasing HIV cases in West Virginia countries were highly vulnerable to rapid HIV dissemination through injection drug use. Phanuphak *et al.* (2015) stated that HIV epidemics in Asia remained concentrated in nature, affecting mostly higher risk sub-groups, including men who have sex with men (MSM) and transgender woman (TG), people who inject drugs (PWID), and female sex workers (FSW). Furthermore, Zule *et al.* (2013) and Syvertsen *et al.* (2015) stated that considerable global morbidity and mortality were influenced by drug injection use associated with HIV and AIDS infections. The analysis used to model the number of HIV and AIDS cases with drug users is regression analysis. Regression analysis is used to analyze the functional relationship between response and predictor variables. In this analysis, not all response variables are continuous but also there are discrete response variables. For discrete response variables, Poisson regression model with certain assumptions was used, for example, in the standard Poisson regression model mean and variance of response variable is equal. But, in fact, this equality assumption is often unrealistic because the variance can be less than the mean called as under-dispersion case, and vice versa called as over-dispersion case. The Negative Binomial regression can be the solution to this over-

dispersion (Hilbe, 2011). Data of the number of HIV and AIDS cases follow the Bivariate Poisson distribution. However, a count variable i. e. several HIV and AIDS frequently show over-dispersion conditions. Therefore, in this study, bi-response Negative Binomial model was applied to analyze the number of HIV and AIDS model. Note that there are two approaches in regression analysis. The parametric regression approach will be used if regression function shape is known, and the relationship between response and predictor variables follows certain curves. However, the non-parametric regression approach will be used if the relationship between these variables does not follow a specific pattern. It is only assumed to be continuous and differentiable.

In statistical modelling, there has been no development of a non-parametric regression model with a local linear approach in the case of bi-response with Negative Binomial distribution. So far, the regression with a parametric approach has been developed by researchers. Some of these studies include Masson (2012), Husain and Bagmar (2015), Hall and Tarko (2019) and Tohari *et al.* (2019). Non-parametric approaches with continuous responses in bi-response and multi-response cases have also been developed by several researchers. They are Chamidah and Saifudin (2013), Chamidah and Lestari (2016), Chamidah and Rifada (2016a, b), Lestari *et al.* (2018 a, b; 2019 a to 2019 f), Ana *et al.* (2019), Chamidah and Lestari (2019), Chamidah *et al.* (2019a, b), Islamiyati *et al.* (2019), Murbarani *et al.* (2019), Puspitawati and Chamidah (2019) and Ramadan *et al.* (2019). Whereas the non-parametric regression of discrete responses is still limited to the Poisson distribution response variable, such studies include Shim and Hwang (2011) used the kernel estimator, Lian *et al.* (2015) used spline estimator, Chee (2018) used kernel estimator, Darnah *et al.* (2019) used local linear estimator and Astuti *et al.* (2013) used local polynomial estimator in generalized Poisson regression. The goal of this study was to estimate the regression function which drew relationship between the number of HIV and AIDS cases and percentage of drug users in Indonesia.

23

MATERIALS AND METHODS

The secondary data used in this study were

data of the number of HIV and AIDS cases and 221g users in Indonesia recorded in 2017 by the Ministry of Health and the National Narcotics Agency of the Republic of Indonesia. To model the data by using bi-response Negative Binomial regression based on local linear estimator following steps were undertaken : (a) Testing the correlation between response variables, (b) Giving paired data (x_i, y_{1i}, y_{2i}) , $i=1,2,\dots,n$ and estimating the parameters of bi-response Negative Binomial regression model (Tohari *et al.*, 2019) by using locally weighted maximum likelihood method and (c) If from step (b) failed then Newton-Raphson method was used.

RESULTS AND DISCUSSION

Coefficient of correlation between response variables was 64.9%. It meant that the correlation between the number of HIV cases and the number of AIDS cases was high enough. Next, optimal bandwidth value based on maximum value of maximum likelihood cross-validation (MLCV) was obtained (Table 1), and plot of MLCV versus bandwidth (h; Fig. 1). Based on Fig. 1, the optimal bandwidth was 0.101 with maximum MLCV of 3310790. Further, the estimated models had different coefficients depending on the location. One of the estimated models was obtained in Maluku province as follows :

$$\hat{\mu}_1 = \exp(6.571 + 0.411(x - 1.16)), 1.059 < x < 1.261 \dots (1)$$

$$\hat{\mu}_2 = \exp(4.610 + 0.313(x - 1.16)), 1.059 < x < 1.261 \dots (2)$$

Equation 1 shows that every addition by 1% of drug users gave the number of HIV cases by 1.51 times over the previous case. Next, Equation 2 shows that every addition by 1% of drug users gave the number of AIDS cases by 1.37 times over the previous case. Estimated models in 1 and 2 were used to

Table 1. Optimal bandwidth and MLCV values

Bandwidth	MLCV
0.001	2.44E-14
0.101	3310790
0.201	2066545
0.301	1373954
0.401	908818
0.501	755331
0.601	712432
0.701	387611
0.801	334703
0.901	301052

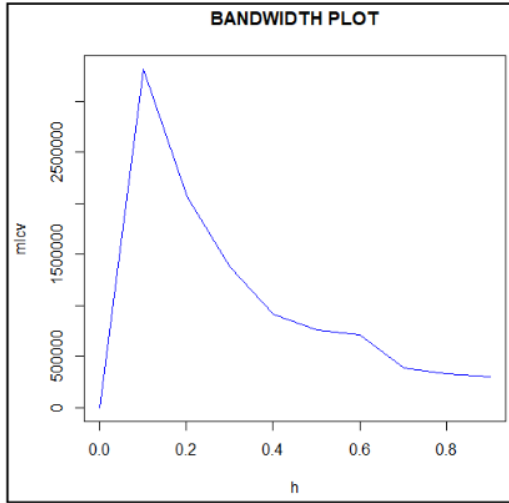


Fig. 1. Plot of MLCV versus bandwidth (h).

predict the number of HIV and AIDS cases that occur in certain provinces of Indonesia based on percentage of drug users. As an example, in 2017, the percentage of drug users in Maluku province was 1.16 with 420 HIV cases and 81 AIDS cases. In 2018, according to National Narcotics Agency of the Republic of Indonesia (BNN), the percentage of drug users in Maluku province was 1.9. Thus, in 2018, in Maluku province, the predicted values were 968 for HIV and 127 for AIDS cases. Thus, it can be concluded that HIV cases in Maluku province changed from 420 to 968 cases or the number of HIV cases in 2018 had changed by 1.67 times from those cases in 2017, and AIDS cases in Maluku province changed from 81 to 127 cases or the number of AIDS cases in 2018 has changed by 1.57 times from those cases in 2017. Goodness of fit testing for estimated model gave deviation value of 37.73 which was smaller than 80.23. It meant that the estimated model was appropriate (Figs. 2 and 3).

Figs. 2 and 3 show that the observation data were given in red dot and the estimation results were given in green line for non-parametric regression estimate and in blue line for parametric regression estimates. Comparing the green line plot with blue line plot, it appeared that the green line followed the red dot pattern. This meant that the estimation results by using bi-response Negative Binomial non-parametric regression approach based on local linear estimator was better than that by

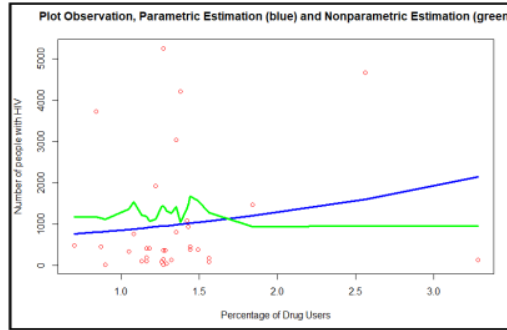


Fig. 2. Plots of estimation results and observations of HIV cases.

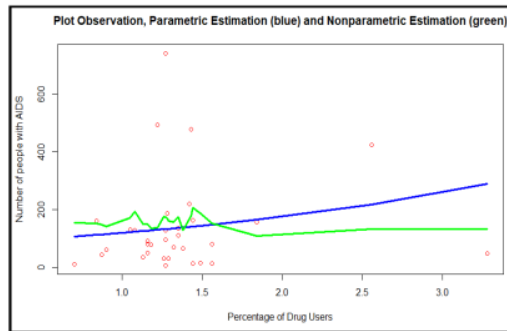


Fig. 3. Plots of estimation results and observations of AIDS cases.

using the bi-response Negative Binomial parametric regression approach.

CONCLUSION

Bi-response Negative Binomial non-parametric regression model approach based on local linear estimator was more appropriate than bi-response negative binomial parametric regression model approach for modelling the number of HIV and AIDS cases in Indonesia, because it had deviance value smaller than that of parametric regression model approach. The obtained model can be used to predict the number of HIV and AIDS cases occurring in certain provinces of Indonesia based on the percentage of drug users. Rising the percentage of drug users will raise the number of HIV and AIDS cases in Indonesia so that the government must continue to suppress drug abuse.

ACKNOWLEDGEMENT

24 The authors thank the Director of the Directorate of Research and Public Service; the

13 Directorate General of Reinforcing of Research and Development and the Ministry of Research, Technology and Higher Education of the Republic of Indonesia for financial support of this research via the Doctoral Dissertation Research (*Penelitian Disertasi Doktor - PDD*) Grant in the fiscal year 2019.

REFERENCES

- Ana, E., Chamidah, N., Andriani, P. and Lestari, B. (2019). Modelling of hypertension risk factors using local linear of additive non-parametric logistic regression. *J. Physics : Conference Series* **1397** : 012067. doi : 10.1088/1742-6596/1397/1/012067.
- Ardani, I. and Handayani, S. (2017). Stigma Terhadap Orang dengan HIV/AIDS (ODHA) sebagai Hambatan Pencarian Pengobatan : Studi Kasus pada Pecandu Narkoba Suntik di Jakarta. *Buletin Penelitian Kesehatan* **45** : 81-88.
- Astuti, E. T., Budiantara, I. N., Sunaryo, S. and Dokhi, M. (2013). Statistical modelling for mortality data using local generalized Poisson regression model. *Int. J. Appl. Mathematics and Statistics* **33** : 92-101.
- Bradley, H., Hogan, V., Brune, C. A., Armstrong, J., Broussard, D., Buchacz, K., Burton, K., Cope, S., Dowsen, E., Garza, G. D., Gerard, A., Granado, M., Gupta, R., Haddy, L., Hoffma, W., Jhonson, S. D., Kirk, N., Lee, C. and Hoots, B. E. (2019). Increased HIV diagnoses in West Virginia counties highly vulnerable to rapid HIV dissemination through injection drug use : a cautionary tale. *Annals of Epidemiology* **34** : 12-17.
- 9 Chamidah, N., Gusti, K. H., Tjahjono, E. and Lestari, B. (2019a). Improving of classification accuracy of cyst and tumor using local polynomial estimator. *TELKOMNIKA* **17** : 1492-1500.
- Chamidah, N. and Lestari, B. (2016). Spline estimator in homoscedastic multi-responses non-parametric regression model in case of unbalanced number of observations. *Far East J. Mathematical Sci. (FJMS)* **100** : 1433-1453.
- 10 Chamidah, N. and Lestari, B. (2019). Estimation of covariance matrix using multi-response local polynomial estimator for designing children growth charts : A theoretically discussion. *J. Physics : Conference Series* **1397** : 012072.
- Chamidah, N. and Rifada, M. (2016a). Local linear estimator in bi-response semi-parametric regression model for estimating median growth chart of children. *Far East J. Mathematical Sci. (FJMS)* **99** : 1233-1244.
- Chamidah, N. and Rifada, M. (2016b). Estimation of median growth curves for children upto two years old based on bi-response local linear estimator. *AIP Conf. Proc.* **1718** : 010001. doi : 10.1063/1.494368.
- Chamidah, N. and Saifudin, T. (2013). Estimation of children growth curve based on kernel smoothing in multi-response non-parametric regression. *Appl. Mathematical Sci.* **7** : 1839-1847.
- Chamidah, N., Zaman, B., Muniroh, L. and Lestari, B. (2019b). Estimation of median growth charts for height of children in East Java Province of Indonesia using penalized spline estimator. *Proc. Global Conf on Engineering and Applied Science (GCEAS), Sapporo, Hokkaido, Japan* **71618** : 68-78.
- Chee, C. S. (2018). A mixture model-based non-parametric approach to estimating a count distribution. *Computational Statistics and Data Analysis*. <http://dx.doi.org/10.1016/j.csda.2016.11.012>.
- Darnah, Utoyo, M. I. and Chamidah, N. (2019). Modelling of maternal mortality and infant mortality cases in east Kalimantan using Poisson regression approach based on local linear estimator. *IOP Conf. Series : Earth and Environ. Sci.* **243** : 012023. <https://doi.org/10.1088/1755-1315/243/1/012023>.
- 4 Hall, T. and Tarko, A. P. (2019). Adequacy of negative binomial models for managing safety on rural local roads. *Accident Analysis and Prevention* **128** : 148-158.
- 20 Hilbe, J. M. (2011). *Negative Binomial Regression*, 2nd edn. Cambridge University Press.
- Hussain, M. M. and Bagmar, M. S. H. (2015). Modelling under-dispersed count data using generalized Poisson regression approach. *Global J. Quantitative Sci.* **2** : 22-29.
- Islamiyati, A., Fatmawati and Chamidah, N. (2019). Penalized spline estimator with multi-smoothing parameters in bi-response multi-predictor non-parametric regression model for longitudinal data. *Songklanakarin J. Sci. and Technol.* rdo.psu.ac.th/sjstweb/Ar-Press/2019June/6.pdf.
- Lestari, B., Anggraeni, D. and Saifudin, T. (2018b). Estimation of covariance matrix based on spline estimator in homoscedastic multiresponse non-parametric regression model in case of unbalanced number of observations. *Far East J. Mathematical Sci. (FJMS)* **108** : 341-355.
- Lestari, B., Chamidah, N. and Saifudin, T. (2019 a). Estimasi Fungsi Regresi Nonparametric Birespon Menggunakan Estimator Smoothing Spline dan Estimator Kernel. *Jurnal Matematika, Statistika & Komputasi* **15** : 20-24.

- Lestari, B., Fatmawati and Budiantara, I. N. (2019b). Smoothing spline estimator in multiresponse non-parametric regression for predicting blood pressures and pulse : A theoretical discussion. *Proc. Global Conf. on Engineering and Applied Science (GCEAS), Sapporo, Hokkaido, Japan* **71618** : 8-153.
- Lestari, B., Fatmawati and Budiantara, I. N. (2019c). Spline estimator and its asymptotic properties in multiresponse non-parametric regression model. *Songklanakaraj. Sci. and Technol. (SJST)* **5** : 1177-1199.
- Lestari, B., Fatmawati and Budiantara, I. N. (2019d). Smoothing spline estimator in multiresponse non-parametric regression for predicting blood pressures and heart rate. *Int. J. Academic and Appl. Res.* **3** : 1-8.
- Lestari, B., Fatmawati and Budiantara, I. N. (2019e). Estimation of multiresponse non-parametric regression model using smoothing spline estimator. *Int. J. Acad. and Appl. Res.* **3** : 1-4.
- Lestari, B., Fatmawati, Budiantara, I. N. and Chamidah, N. (2018a). Estimation of regression function in multi-response non-parametric regression model using smoothing spline and kernel estimators. *J. Physics : Conference Series* **1097** : 012091.
- Lestari, B., Fatmawati, Budiantara, I. N. and Chamidah, N. (2019f). Smoothing parameter selection method for multiresponse non-parametric regression model using smoothing spline and kernel estimators approaches. *J. Physics : Conference Series* **1397** : 012064.
- Lian, H., Meng, J. and Zhao, K. (2015). Spline estimator for simultaneous variable selection and constant coefficient identification in high-dimensional generalized varying-coefficient model. *J. Multivariate Analysis* **141** : 81-103.
- Masson, K. (2012). On ridge estimators for the negative binomial regression model. *Economic Modelling* **29** : 178-184.
- Murbarani, N., Swastika, Y., Dwi, A., Aris, **8** and Chamidah, N. (2019). Modelling of the percentage of AIDS sufferers in east Java Province using non-parametric regression approach based on truncated spline estimator. *Indonesian J. Statistics and its Applications* (eISSN : 2599-0802) **3** : 139-147.
- Phanupak, N., Lo, Y. R., Shao, Y., Solomon, S. S., O'Connell, R. J., Tovanabutra, S., Chang, D., Kim, J. H. and Excler, J. L. (2015). HIV epidemic in Asia : Implications for HIV vaccine and other prevention trials. *AIDS Res and Human Retroviruses* **31** : 1-16.
- Pusdatin (2018). Data dan Informasi Profil Kesehatan Indonesia 2017. Kementerian Kesehatan Republik Indonesia.
- Puspitawati, A. and Chamidah, N. (2019). Choroidal neovascularisation classification on fundus retinal image using local linear estimator. *IOP Conf. Series : Materials Science and Engineering* **546** : 052056. doi : 10.1088/1757-899X/546/5/052056.
- Ramadan, W., Chamidah, N., Zaman, B., Muniroh, L. and Lestari, B. (2019). Standard growth chart of weight for height to determine wasting nutritional status in east Java based on semi-parametric least square spline estimator. *IOP Conf. Series : Materials Science and Engineering* **546** : 052063. doi : 10.1088/1757-899X/546/5/052063.
- Shim, J. and Hwang, C. (2011). Kernel Poisson regression machine for stochastic claims reserving. *J. Korean Statistical Society* **40** : 1-9.
- Syvertsen, J. L., Agot, K., Ohaga, S., Strathdee, S. A., Camlin, C. S., Omanga, E., Odonde, P., Rota, G., Akoth, K., Peng, J. and Wagner, K. D. (2015). Evidence of injection drug use in Kisumu, Kenya : Implications for HIV prevention. *Drug and Alcohol Dependence* **151** : 262-266.
- Tohari, A., Chamidah, N. and Fatmawati (2019). Modelling of HIV and AIDS in Indonesia using bivariate negative binomial regression. *IOP Conference Series : Materials Science and Engineering* **546** : 052079. <https://doi.org/10.1088/1757-899X/546/5/052079>.
- UNAIDS (2018). HIV and AIDS Data HUB for Asia-Pacific. World Health Organization.
- Zule, W. A., Cross, H. E., Stover, J. and Pretorius, C. (2013). Are major reductions in new HIV infections possible with people who inject drugs? The case for low dead-space syringes in highly affected countries. *Int. J. Drug Policy* **24** : 1-7.

Cek Similarity/1. artikel annals of biology 215-219 (published).pdf

ORIGINALITY REPORT

18%

SIMILARITY INDEX

PRIMARY SOURCES

1	journal.umy.ac.id Internet	70 words — 2%
2	eprints.lancs.ac.uk Internet	43 words — 1%
3	escholarship.org Internet	42 words — 1%
4	Timo Pew, Richard L. Warr, Grant G. Schultz, Matthew Heaton. "Justification for considering zero-inflated models in crash frequency analysis", <i>Transportation Research Interdisciplinary Perspectives</i> , 2020 Crossref	36 words — 1%
5	slidelegend.com Internet	34 words — 1%
6	downloads.hindawi.com Internet	30 words — 1%
7	Xinyi Li, Li Wang, Huixia Judy Wang. "Sparse Learning and Structure Identification for Ultrahigh-Dimensional Image-on-Scalar Regression", <i>Journal of the American Statistical Association</i> , 2020 Crossref	29 words — 1%

8	sinta3.ristekdikti.go.id Internet	28 words — 1%
9	repository.unair.ac.id Internet	24 words — 1%
10	rigeo.org Internet	23 words — 1%
11	Ahmed Ibrahim, E.M. Abu Elgoud, H.F. Aly. "Alginate Modified Graphene Oxide for Rapid and Effective Sorption of Some Heavy Metal Ions from Aqueous Solution", Research Square Platform LLC, 2022 Crossref Posted Content	20 words — 1%
12	Montserrat Fuentes. "Spatial Association between Speciated Fine Particles and Mortality", Biometrics, 9/2006 Crossref	19 words — 1%
13	journal.uinjkt.ac.id Internet	19 words — 1%
14	m.scirp.org Internet	18 words — 1%
15	www.insightsociety.org Internet	18 words — 1%
16	www.isrt.ac.bd Internet	18 words — 1%
17	N Citrasari, K P A Pamungkassari, T W C Putranto, N Hamidah. "Paving block with product 1 material as a substitute of portland cement and landfill mining residue", IOP Conference Series: Earth and Environmental Science, 2019 Crossref	14 words — < 1%

18	coek.info Internet	14 words — < 1%
19	download.garuda.ristekdikti.go.id Internet	13 words — < 1%
20	academic.oup.com Internet	12 words — < 1%
21	Chew-Seng Chee. "A mixture model-based nonparametric approach to estimating a count distribution", Computational Statistics & Data Analysis, 2017 Crossref	10 words — < 1%
22	www.acciontecnicasocial.com Internet	9 words — < 1%
23	www.orientjchem.org Internet	8 words — < 1%
24	Dyah P. Rahmawati, I. N. Budiantara, Dedy D. Prastyo, Made A. D. Octavanny. "Mixed Spline Smoothing and Kernel Estimator in Biresponse Nonparametric Regression", International Journal of Mathematics and Mathematical Sciences, 2021 Crossref	7 words — < 1%
25	philpapers.org Internet	7 words — < 1%

EXCLUDE QUOTES OFF

EXCLUDE MATCHES OFF

EXCLUDE BIBLIOGRAPHY OFF