ADV MATH SCI JOURNAL

Advances in Mathematics: Scientific Journal **9** (2020), no.9, 7341–7351
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
https://doi.org/10.37418/amsj.9.9.81

MODELING DENGUE INCIDENCE IN SOUTHERN MALAYSIA USING FIXED EFFECTS PANEL COUNT MODEL

WAN FAIROS WAN YAACOB[1], NIK NUR FATIN FATIHAH SAPRI, AND YAP BEE WAH

ABSTRACT. Dengue fever is a mosquito-borne viral disease caused by the dengue virus transmitted by female mosquito of Aedes aegypti. It has spread rapidly in almost all regions including Malaysia. This paper analyzed daily panel count data of dengue incidence in southern state of Johor, Malaysia covering 10 districts from the period of 2013 to 2017. The Fixed Effects Poisson and Negative Binomial panel count models were used to model dengue incidence associated with various climatic factors which included rainfall, temperature, humidity and its lags of 7 days, 14 days, 21 days and 28 days. Several model specifications were estimated. Results revealed that there were significant positive relationships between temperature of lag 7 days and 14 days, amount of rainfall of lag 21 days and humidity of lag 14 days with dengue incidence rate. A very high dengue incidence rate was spotted in the district of Johor Bharu. Kulaijaya and Kota Tinggi. This finding can help the health authority to plan for early warning system to detect dengue outbreak.

1. INTRODUCTION

Numerous researchers applied count model and panel count model in modelling the count data such as dengue cases [1-3]. Among widely used models are the generalized linear model of Poisson, Negative Binomial model, generalized additive model of Poisson, regression and auto-regressive model [4-6].

[1]*corresponding author*

2010 *Mathematics Subject Classification.* 62J12.

Key words and phrases. Dengue Incidence, Fixed Effects, Humidity, Negative Binomial, Rainfall.

Though Poison models are preffered in modelling count data type particularly in the dengue cases, it suffers one limitation of equality in the mean and assumption in Poisson model [2,3,7]. Thus, to allow for this overdispersion issues if assumption is violated, researchers used the negative binomial model that is the extension of Poisson model.

Further refinement on the count model has been made using panel count model when dealing with panel count data. Most of the literatures on panel count model include fixed effects and random effects model of Poisson and Negative Binomal which began with the seminal work of Hausman et al. [8]. This model can handle the assumption of independent (or at least uncorrelated) observations in the standard Poisson generalized linear model which may not be valid. There could be strong temporal correlation effects within some areas and there could also be spatial clustering effects in neighboring microregions. Thus, the fixed effects and random effects can allow for such latent effects and correlation structures. Application of these techniques has been applied in modeling dengue disease [9].

Dengue incidence is often influenced by climatic factors such as temperature, amount of rainfall, number of rainy days, relative humidity and wind speed. The relationship between climatic factors and dengue incidence has been studied by many researchers using Poisson, fixed effects and random effects Poison panel count model. The findings were proven to influence dengue incidence [2,6,9]. Hence, the aim of this study is to develop a statistical model to describe the relationship between daily dengue incidence and series of climatic factors using fixed effects negative binomial regression model. Using this model, significant climatic factors were determined to explain the daily dengue incidence in districts of Johor.

2. METHODOLOGY

2.1. **The Data.** iDengue [10] has reported that the dengue outbreak in Johor, Malaysia is alarming. It is one of the states which contribute a high number of dengue incidence. The reported epidemic year for Johor was in the year 2015 with about 15, 743 of dengue incidence over 120, 834 dengue cases reported in Malaysia. This is approximately about 430 cases per 100,000 population. This scenario has proven that dengue outbreak in Johor is worsening. Considering the high incidence rate in Johor, this study focused on dengue incidence in the southern state of Malaysia, which is Johor. The data selected in the study were the daily number of dengue cases in 10 districts of Johor for the period between 1^{st} January 2013 and 31^{st} December 2017. The daily dengue cases were the total count of dengue patients diagnosed based on dengue blood test and confirmed being infected by one of the serotypes of dengue virus (DENV-1, DENV-2, DENV-3 or DENV-4) that causes dengue fever and recorded by the hospital. The data were obtained from Vector Borne and Infectious Diseases Sector (VBIDS), Ministry of Health (MOH) Malaysia.

Climatic factors were also used in the analysis which covered temperature, relative humidity and the amount of rainfall with the lag time effect as the explanatory variables. The lag time effects used in the study were lag 7, 14, 21 and 28 days. The lag time effect was taken into consideration in the study as biologically the eggs take a certain period of time to emerge into an adult mosquito, and the influence of climatic surrounding is expected to be visible one or two months later. Temperature refers to the average daily temperature measured in degree Celsius ($^{\circ}C$) recorded by districts in Johor. Meanwhile, relative humidity is defined as atmosphere's water vapor content which exists in gas particles at a certain level of temperature measured in percentage ($\%$). It measures the amount of moisture in the air while rainfall refers to the amount of rainfall recorded everyday by meteorology/weather stations for particular areas. It is measured in millimeter (mm) unit. The sources of climatic variables were collected from Malaysian Meteorological Department (MET) and Climate Data Online (https://power.larc.nasa.gov/data-access-viewer/). The list of variables used in the study is summarized and tabulated in Table 1.

TABLE 1. List of Variables

Variables	Variable Label	Measurement Units
Temperature	Temp	$^{\circ}$C
Rainfall	Rain	mm
Relative Humidity	Humid	$\%$

2.2. **Fixed Effects of Panel Count Model.** Suppose that we have panel count data for i district and each district observe a total of Ti times. Let be the count variable for individual i at time t. Then the expected value of is linked to asset of regressors by:

$$E\left(\eta_{it}\right) = \lambda_{it} = \exp\left(d_i + x'_{it}\beta\right) = \alpha_i\lambda_{it}, \tag{2.1}$$

$i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, where d_i is district specific dummies, $\alpha_i = e^{d_i}$ is the individual specific effect, and x_{it} is a vector of regressors. The Fixed Effects Poisson GLM model can be used to estimate the model parameter via conditional maximum likelihood method. The Fixed Effects Poisson GLM model can be used to estimate the model parameter via conditional maximum likelihood method. However, fitting fixed effects Poisson regression model may also result to overdispersion. Hence, to explicitly allow for overdispersion in panel count model, the fixed effects negative binomial regression model is utilized. In this paper, the fixed effects negative binomial regression model is illustrated using the number of dengue cases which are daily figures for individual district with defined climatic characteristics. When cross sectional heterogeneity exists, the fixed effects model is more appropriate. Based on the fixed effects negative binomial formulation of Hausman et al. [7], the dengue model is expressed as $\log\lambda_i = \alpha_i + \beta x_{it}, i = 1, 2\ldots, n$ and $t = 1, 2, \ldots, T$ where λ_{it} is the expected number of dengue in district i in day t, α_i is the fixed effects associated with i, and β is vectors of parameters to be estimated for the vector of explanatory variables x_{it} Here, the number of dengue, y_{it} for a given time period, t is assumed to follow a negative binomial distribution with parameters $\alpha_i\lambda_{it}$ and ϕ_i, where $\lambda_{it} = \exp\left(x'_{it}\beta\right)$ gives y_{it} as mean $\alpha_i\lambda_{it}/\phi_i$ and variance $(\alpha_i\lambda_{it}/\phi_i) \times (1 + \alpha_i/\phi_i)$. This model allows the variance to be greater than the mean. The parameter α_i is the individual-specific fixed effects and the parameter ϕ_i is the negative binomial overdispersion parameter which can take on any value and varies across individuals. The negative binomial mass function is given by:

$$f\left(y_{it}\right) = \frac{\Gamma\left(\lambda_{it} + y_{it}\right)}{\Gamma\left(\lambda_{it}\right)\Gamma\left(y_{it} + 1\right)} \left(\frac{1}{1 + \frac{\alpha_i}{\varphi_i}}\right)^{\lambda_t} \left(\frac{\alpha_i}{\frac{\varphi_i}{\alpha_i}}\right)^{y_{it}}, \tag{2.2}$$

with conditional joint density that yields the negative hypergeometric distribution. The estimation of vector $\hat{\beta}_{NBFE}$ is done using the maximum likelihood method by maximizing the log likelihood function given by:

$$L_i = \sum_{i=1}^{N} \begin{cases} \log\Gamma\left(\sum_{i=1}^{T} \lambda_{i,t}\right) + \log\Gamma\left(\sum_{i=1}^{T} y_{i,t} + 1\right) \\ -\log\Gamma\left(\sum_{i=1}^{T} \lambda_{,t} + \sum_{i=1}^{T} y_{i,t}\right) + \log\Gamma\left(\sum_{i=1}^{T} \lambda_{i,t} + y_{i,t}\right) \\ -\log\Gamma\left(\sum_{i=1}^{T} \lambda_{i,t}\right) - \log\Gamma\left(\sum_{i=1}^{T} y_{i,t} + 1\right) \end{cases} . \tag{2.3}$$

2.3. **Model Implementation.** This study employed fixed effects Poisson regression model and fixed effects negative binomial regression model. The model was fitted using MASS package in R software version $3.5.2$. The data analysis and parameter estimates for the fixed effects Poisson model were estimated using glm 0 function with family = "Poisson". Meanwhile, for the fixed effects negative binomial model, glm.nb() function was used to perform the analysis. A series of models, from full model intercept to spatially and temporal ranging of covariates with geographical district, were tested beginning with maximal model based on all covariates and followed by different subsets of variable including time lag effect for each of the climatic variable. There were 4 different time lag effects that have been considered in the study such as lag of 7 days, 14 days, 21 days and 28 days. The models were then assessed using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), deviance and $R_D^2(pseudo - R^2)$ for the goodness of fit model and it is explained in the next section of this paper.

2.4. **Goodness of Fit Model.** To select appropriate prediction model fit, each model with and without lag of climatic factors was evaluated using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), deviance and $R_D^2(pseudo - R^2)$. The AIC equation is given as follow:

$$AIC = -2\ln(L) + 2k, \tag{2.4}$$

where L is the log likelihood of the fitted model and k is the number of parameters in the fitted model. Meanwhile, BIC is computed as follows:

$$BIC = -2\ln(L) + k\log(n), \tag{2.5}$$

where n is the sample size. Similar to AIC, the model was selected with the minimum BIC value which indicates that the smaller the value AIC and BIC, the better the model.

The deviance, D is an important measure that should be put into consideration [3]. The deviance is the difference between the log-likelihood of the fitted model and saturated model which can be specified as follows:

$$D = 2[\ln(\hat{L}) - \ln(L)], \tag{2.6}$$

where $\ln(\hat{L})$ is the log-likelihood for saturated model and $\ln(L)$ is the log-likelihood for fitted model. Model which is correctly specified will be close to unity, 1. If the value of D is close to $n - p$ (degrees of freedom), the model is considered as adequate. However, if $D > n - p$, then the data is overdispersed which means negative binomial model can be considered to fix such problem.

Since the study proposes using 'linear' model, the proportion of variation in the response variable can be explained by explanatory variables, R^2 which is a typical measure used as a model adequacy. As suggested in the study of Lowe [9], the R^2 is based on decomposition of the deviance which is termed as 'pseudo-R^2:

$$D' = D + \hat{D}, \tag{2.7}$$

where D' is the deviance in the intercept-only model (null model), D is the deviance of fitted model and $\hat{D}$ is explained deviance. Thus, the pseudo- $\mathrm{R}^2, (R_D^2)$ is specified as follows:

$$R_D^2 = 1 - \frac{D'}{\hat{D}}, \tag{2.8}$$

where pseudo-R^2 measures the reduction in deviance due to inclusion of explanatory variables. The interpretation of R_D^2 is similar to common R^2 at which the values closer to 1 indicate a better fit. The next section of this paper is the description of the data used in the study.

3. Results and Discussions

Johor has 10 districts namely Batu Pahat, Johor Bharu, Kluang, Kota Tinggi, Kulaijaya, Ledang, Mersing, Muar, Pontian and Segamat. Among 10 districts in Johor, Johor Bharu showed a high dengue incidence rate as compared to other districts in 2015 since there were about 11,975 dengue cases (790 dengue cases recorded per 100,000 population) recorded over 15,743 dengue cases reported in Johor. Figure 1 presents the annual trend of dengue incidence rate per 100,000 populations by district in Johor.

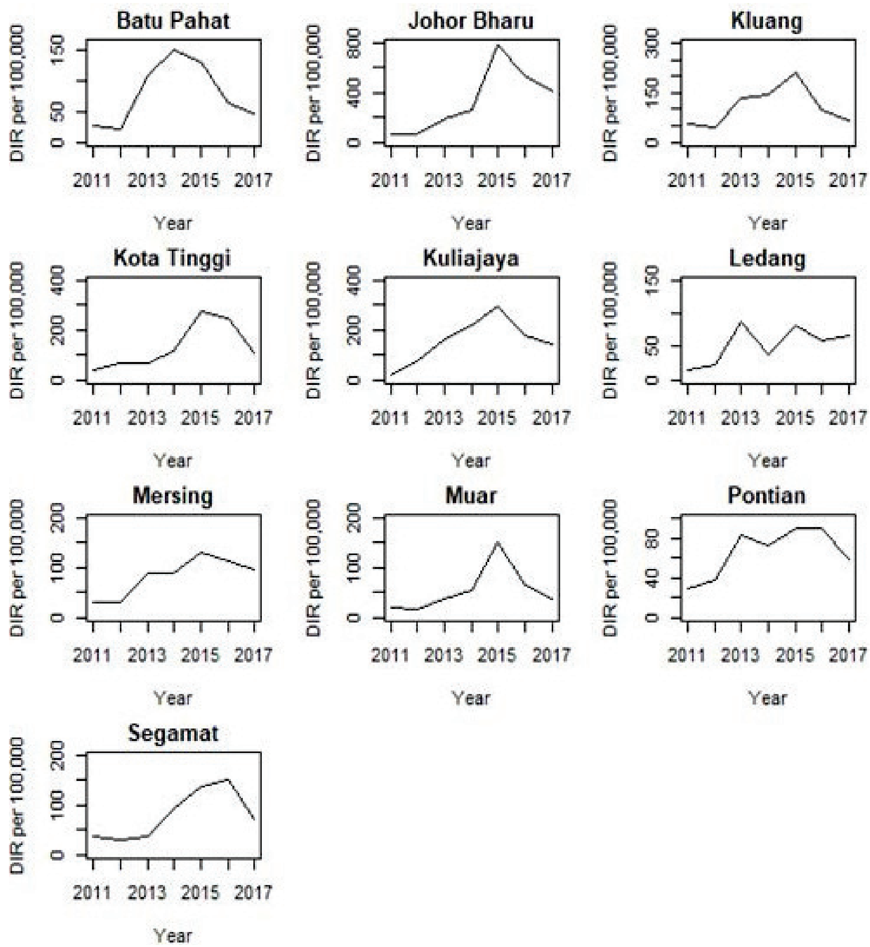

FIGURE 1. Annual Dengue Incidence Rate per 100, 000 Population by District

To further describe the dengue incidence with climatic factors, several models were estimated. The initial model developed begin with the Poisson and followed by Negative Binomial Model of Fixed Effects that involved all possible climatic explanatory variables considered in this study. Then, the modeling process continues by eliminating insignificant variables using backward and stepwise method and the most parsimony model was obtained on the basis of minimizing AIC and BIC value. Table 2 presents the parameter estimates of full model for Fixed Effects Negative Binomial (FENB) and the reduced model for FE Poisson and FENB for the district of Johor.

TABLE 2. Fixed Effects NB and Poisson GLM

	Full Model (FENB)	FE NB	FE Poisson
Constant	-21.04*** (1.098e+00)	-1.933e+01*** (7.779e-01)	-2.624e+01*** (4.327e-01)
Temp	0.02607 (1.991e-02)		
Temp_lag7	0.08074*** (2.085e-02)	9.624e-02*** (1.890e-02)	2.352e-01*** (1.088e-02)
Temp_lag14	0.06174** (2.092e-02)	5.465e-02*** (1.435e-02)	1.042e-01*** (8.737e-03)
Temp_lag21	0.001446 (2.094e-02)		
Temp_lag28	0.02533 (2.033e-02)		
Humid	0.001261 (3.754e-03)		
Humid_lag7	0.0146*** (3.907e-03)	1.753e-02*** (3.388e-03)	3.695e-02*** (1.870e-03)
Humid_lag14	0.003236 (3.919e-03)		
Humid_lag21	-0.0008583 (3.917e-03)		
Humid_lag28	0.005555 (3.782e-03)		
Rain	-0.0007265 (7.512e-04)		
Rain_lag7	0.0001378 (7.380e-04)		
Rain_lag14	-0.0004199 (7.443e-04)		
Rain_lag21	0.001439* (7.235e-04)	1.437e-03* (7.099e-04)	6.352e-04* (3.791e-04)
Rain_lag28	-0.00007781 (7.472e-04)		
Johor Bharu	1.436*** (3.413e-02)	1.438e+00*** (3.347e-02)	1.484e+00*** (2.217e-02)
Kulaijaya	0.7263*** (3.842e-02)	7.238e-01*** (3.814e-02)	7.517e-01*** (2.869e-02)
Ledang	-0.3227*** (5.763e-02)	-3.414e-01*** (5.686e-02)	-2.276e-01*** (5.056e-02)
Mersing	0.1133* (6.047e-02)	9.818e-02 (6.001e-02)	1.961e-01*** (5.420e-02)
Muar	-3.115e-01*** (4.888e-02)	-3.312e-01*** (4.794e-02)	-1.892e-01*** (3.985e-02)
Pontian	-2.622e-01*** (5.205e-02)	-2.593e-01*** (5.145e-02)	-2.623e-01*** (4.461e-02)
Segamat	1.314e-01** (5.068e-02)	1.008e-01* (4.839e-02)	2.871e-01*** (3.920e-02)
Deviance	17302	17291	33357
R_D^2	0.3137	0.3132	0.4837
AIC	49886.24	49872.11	57642.87
BIC	52956.33	52856.27	60627.02
Likelihood Value	-24553.06	-28439.43	

LR Test: -2(LLPoisson - LLNegative Binomial) = -2 (-28439.43- (-24553.06)) = 7772.74

The maximum likelihood estimates of FE GLM for the coefficient and the estimates of standard error for each coefficient were obtained to calculate the p-value to test the significant of each coefficient estimate. A series of models, from full model intercept to spatially and temporal ranging of covariates with geographical district were tested. Various models were tested beginning with maximal model base on all covariates and follow by different subsets of variable including the time lag effect for each of climatic variable. To select appropriate prediction model fit, each model with and without lag of climatic factors were evaluated using AIC, BIC. The inclusion of factors reflecting districts and time allow the baseline model to vary depending on which district and day of interest. The extra heterogeneity in the data which is not explained by day and districts is handled by the scale parameter k in the Negative Binomial. The overdispersion problem was tested using likelihood ratio test (LRT). Given the null hypothesis of no overdispersion, the results indicate overdispersion exist as the value of likelihood ratio test (7772.74) is greater than the $\chi^2_{1,0.05} = 3.84$. This justify the use of a Negative Binomial rather than Poisson FE GLM. Hence the FE Negative Binomial is preferable with the smallest AIC and BIC. The reference district is Batu Pahat. At 0.05 level of significant, variable temperature of lag 7 days, temperature of lag 14 days, humidity of lag 7 days and rainfall of lag 21 days were found to be statistically significant. This indicate that there is a significant positive relationship between temperature of lag 7 days and 14 days with dengue incidence rate. There is also significant positive relationship between humidity of lag 7 days and amount of rainfall of lag 21 days with dengue incidence rate.

This may be results of warm, humid condition and amount of rainfall that stimulate mosquito development and create mosquito breeding sites. This finding can be used for the development of early warning systems to detect dengue outbreak.

FIGURE 2. Spatial Map of DIR and Fitted DIR

Figure 2 illustrates the relationship between observed and fitted model for 5 years across district in Johor. Overall, the FE NB GLM is able to capture correctly very high DIR in industrial area location such as Johor Bharu and Kulaijaya. However, a very high dengue incidence rate was under predicted in Kota Tinggi for the year 2015 and 2016 when serious epidemics occur in 2015. The high population density, rapid development and uncontrolled urbanization has contributed in increasing the mosquitoes breeding sites in Southern Malaysia.

4. Conclusions

The findings from this study suggest that there exist association between climatic factor and dengue incidence rate in Johor, Malaysia. Climate information allows some of the temporal variability to be captured using FE GLM model. However, the FE GLM unable to fully capture the spatial variability across the districts. Hence the use of random effects may be valuable to allow for unobserved latent structures in the model. For instance, to capture the impact of unknown confounding factors such as introduction new vector control program or misreporting of dengue cases. The findings revealed that the temperature of lag of 7 and 14 days, Humidity of 14 days and amount of rainfall of lag of 21 days found to be significantly associated with the dengue incidence. The spatial map analysis able to reveal the hotspot areas in Southern Malaysia of Johor state such as Johor Bharu, Kulaijaya and Kota Tinggi. Therefore, the findings obtained from this study can provide important information especially to the local authority covering the hotspot district of Johor Bharu, Kulaijaya and Kota Tinggi to plan effective vector control program to safeguard the community from dengue outbreak.

Acknowledgment

We would like to express our appreciation and record our gratitude to the Ministry of Higher Education (MOHE) and Universiti Teknologi MARA for financial support under the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2016/STG06/UITM/02/2). The authors are also grateful to the Editor, the Associate Editor and anonymous referees for their insightful comments and suggestions.

References

[1] Y. Choi, C.S. Tang, L. McIver, M. Hashizume, V. Chan, R.R Abeyasinghe, R Huy: *Effects of Weather Factors on Dengue Fever Incidence and Implications for Interventions in Cambodia*, BMC Public Health., **16**, art.no. 241, (2016).

[2] W.F.W Yaacob, W.W Azaki, M.A Lazim, Y.B Wah: *Modelling Dengue Fever (DF) and Dengue Heemorrhagic Fever (DHF) Outbreak Using Poisson and Negative Binomial Model*, World Academy of Science, Engineering and Technology. **38** (2010), 903–908.

[3] W.F.W. Yaacob, W.W. Azaki, M.A. Lazim, Y.B. Wah: *A Practical Approach in Modelling Count Data*, Proceedings of the Regional Conference on Statistical Sciences 2010, (2010), 176–183.

[4] Y.L. Cheong, K. Burkart, P.J. Leitao, T. Lakes: *Assessing Weather Effects on dengue Disease in Malaysia*, International Journal of Environmental Research and Public Health. **10**(12) (2013), 6319–6334.

[5] L. Lu, H. Lin, L. Tian, W. Yang, J. Sun, Q. Liu : *Time Series Analysis of Dengue Fever and Weather in Guangzhou, China*, BMC Public Health. **9**, art.no. 395, (2009). https://doi.org/10.1186/1471-2458-9-395

[6] R. Tuladhar, A. Singh, A. Varma, D.K. Choudhary: *Climate Factors Influencing Dengue Incidence in an Epidemic Area of Nepal*, BMC Res Notes, **12**, art.no. 131, (2019).

[7] A.C. Cameron, P.K. Trivedi: *Regression Analysis of Count Data (Econometric Society Monographs)*, Cambridge University Press: Cambridge, United Kingdom. (1984).

[8] J.A. Hausman, B.H. Hall, Z. Griliches: *Econometric Models for Count Data with an Application to the Patents - RnD Relationship*, Econometrica. **52**(4) (1984), 909–938.

[9] R. Lowe: *Spatio-temporal Modelling of Climate Sensitive Disease Risk: Towards an Early Warning System for Dengue in Brazil*, Ph.D. thesis, University of Exeter, United Kingdom, (2010).

[10] iDengue.: *Statistik Denggi*, (2019), Available: https://idengue. arsm.gov.my/

Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA Cawangan Kelantan,Kampus Kota Bharu
Lembah Sireh,15050 Kota Bharu, Kelantan, Malaysia
Email address: wnfairos@uitm.edu.my

Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA, .
45400 Shah Alam, Selangor, Malaysia.
Email address: fatinfatihah26@yahoo.com

Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA, .
45400 Shah Alam, Selangor, Malaysia.
Email address: beewah@tmsk.uitm.edu.myz