Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 12, Number 1 (2016), pp. 741-746

http://www.ripublication.com

Panel Data Regression Model for Case of *Dengue* Hemorrhagic Fever (DHF) in Bogor

Zamahsary Martha[1], Budi Susetyo[2] and Muhammad Nur Aidi[3]

Departement of Statistics Bogor Agricultural University,
Jalan Pajajaran, Kampus IPB Baranangsiang,
Bogor 16151, Indonesia.

Abstract

Dengue Hemorrhagic Fever (DHF) is acute febrile disease caused by *dengue* virus infection and transmitted to humans through the bite of *Aedesaegypti* and *Aedesalbopictus* mosquitoes. DHF causes symptoms of sudden high fever, bleeding and cause shock resulting in death. DHF is one health problems in Indonesia. Bogor has an average altitude of 190 m to 330 m above sea level, the high rainfall and density. This condition makes Bogor has great potential spread of DHF. DHF is found throughout the year in Bogor. Resident in the same area could be suffering from DHF in the next year because of *Aedesaegypti* and *Aedesalbopictus* mosquitoes as *dengue* virus spreaders continue to grow and spread throughout the year. This research aims to determine the best panel data regression model in the case of the factors that influence the number of patients with *dengue*in Bogor from 2009 to 2013. Result, fixed effects model using transformed response variable and the addition of independent variable previous year the number of DHF patients is a model that can describe the influence of independent variable on the number of DHF patients in Bogor from 2009 to 2013. Variables that influence of DHF are density, mobility, the average age of DHF patients and the number of DHF patients previous year with a coefficient of determination (R^2) of 72.76% and MSE of 0.2763.

Keywords: *Dengue* Hemorrhagic Fever, Fixed Effects Model, Panel Data Regression.

Introduction

Dengue Hemorrhagic Fever (DHF) is a disease caused by *dengue* virus and can cause high fever, bleeding, shock and death [5]. DHF is a health problem in Bogor, Indonesia. The number of DHF cases in Bogor as much as 849 and8 people died in 2013. From 68 villages in Bogor, most of them was found DHF cases each year [1].

DHF is found throughout the year in Bogor. Resident in the same area could be suffering from DHF in the next year because *Aedesaegypti* and *Aedesalbopictus* mosquitoes as *dengue* virus spreaders continue to grow and spread throughout the year. Based on these characteristics, the combination of cross section data and time series data called the panel data. There are three methods for estimating panel data regression model, Those are Pooled Least Square (PLS), Fixed Effects Model (FEM) and Random Effects Model (REM). The research aim to determine the best panel data regression model in the case of the factors that influence the number of DHF patients in Bogor for period 2009 to 2013 [3].

Methodology

Data:

The data usedfromthe Health Department of Bogorand Central Bureau Statistics of Bogor. Response variable used is the number of DHF patients(Y). Independent variables were observed among others, density(X_1), mobility(X_2), the average age of DHF patients (X_3) and the number of health centers (X_4). This research conductedin68villagesof Bogorfor period2009 to2013.

Methodology:

The step of this research are:

1. Panel data analysis:
    a. The estimation parameter that includes PLS, FEM and REM models.
    b. Chow test used to select between PLS or FEM model.
    c. Hausman test used to select between FEM or REM model.
2. Assumption model selected test, that includes autocorrelation, multicollinearity, heteroscedasticity and normality.
3. Selection of the best models and interpretation of the model.

Materials and Methods

Pooled Least Square (PLS):

Estimating model parameters in PLS using Ordinary Least Squares(OLS). PLS model is expressed in equation form as follows:

$$\boldsymbol{y_{it} = \beta_0 + \beta X_{it} + u_{it}}$$

$i = 1, 2, \ldots, 68$ and $t = 1, 2, \ldots, 5$

Fixed Effect Model (FEM):
Estimating model parameter in FEM using Least Square Dummy Variable (LSDV), where LSDV is a method used to estimate parameters of linear regression using OLS models involving dummy variable as a variable independent [2]. FEM model is expressed in equation form as follows:

$$\boldsymbol{y_{it} = \beta_{01} + C_2 D_2 + \cdots + C_{68} D_{68} + \beta X_{it} + u_{it}}$$
$$\boldsymbol{\beta_{0i} = \beta_{01} + C_i}, \; i=(2,3,\ldots,N)$$

where D_2=1for location 2 and 0 otherwise, D_{3i}=1 for location 3 and 0 otherwise, and so on. β_{01} is a constant value for location 1 and β_{0i} are difference between the value of location 1 with location *i* [3].

3.1 Random Effects Model (REM)
Estimating model parameter in REM using Generalized Least Square (GLS). REM assumptions influence the location are random variables included in the model as a form of residual [4]. REM model is expressed in equation form as follows:

$$\boldsymbol{y_{it} = \beta_{0i} + \beta X_{it} + u_{it}}$$

$$\boldsymbol{\beta_{0i} = \beta_0 + \varepsilon_i}$$

where β_{0i} are value of location *i*, and ε_i is an residual for location *i* [3].

Results and Discussion
Panel Data Analysis:
Panel data regression modeling results can be seen in Table 1.

Table 1: Results of panel data regression modeling

Variable	PLS		FEM		REM	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
intercept	5.4469	0.0456	20.7256	0.0001	11.5252	0.0002
X_1	0.0003	0.0252	-0.0003	0.4187	0.0002	0.3420
X_2	0.0066	0.3777	-0.0181	0.0030	-0.0108	0.0670
X_3	0.2247	0.0389	-0.0267	0.7749	0.0408	0.6461
X_4	4.7682	0.0012	2.0573	0.7111	4.9620	0.0188
R^2		7.69%		61.59%		3.12%

Chow test result showed that the p-value for this test is less than 0.05, it means that the model that chosen is FEM. This shows that villages in Bogor has a different effect on the number of DHF patients.Hausman test result showed that the p-value is close to 0, it means that the model chosen is FEM. Chow and Hausman test showed that the appropriate model to explain the data of density (X_1), mobility (X_2), the average age of

DHF patients (X_3) and the number of health centers (X_4) to sum of DHF patients in Bogor is a FEM model.

Testing Assumptions:
The first assumption testing performed in panel data modeling is multicollinearity test. Multicolinearity test results showed FEM did not have multicollinearity problems with Variance Inflation Factor (VIF)values less than 10. The second test assumption that is autocorrelation test. Durbin-Watson test value that is 2.3545, the value of d_L and d_U at 1.8043 and 1.8399. Durbin-Watson test values are not between d_U and 4-d_U, it means that the model have autocorrelation.The third test is heteroscedasticity test used Breusch Pagan Godfrey(BPG) test. It result explains that residual variance is not homogeneous at significance 5% or it can be concluded that the model is formed having heteroscedasticity problems. The final test is normality test. From the Jarque-Bera test results obtained value of 60.3878, with p-values less than 0.05 or residual is not a normal distribution.

Handling Violations AssumptionsL
In this research,treatment usedisdata transformationto the dependent variable(Y^*), the transformation of data used cubic root transformationandthe additionof newvariablesinthe model, that is the numberof DHF patientsinanearlier timeperiod($Y_{i,(t-1)}$). Results ofparameterestimationinthe modelwiththe handling assumptionsin Table2.

Table 2: Results of analysis using FEM after the handling assumption

Variable	Coefficient	p-value
intercept	2.9891	0.0000
X_1	-0.0003	0.0144
X_2	-0.0006	0.0479
X_3	0.0114	0.0244
X_4	0.3461	0.5145
$Y_{i,(t-1)}$	-0.0089	0.0083
R^2		72.81%
MSE		0.2771

Analysisofthe model using FEM used transformation dependent variable (Y^*_{it}) and additional variable ($Y_{i,(t-1)}$) resultedR^2 of72.81%. Independent variablesthat significantly influencethe number ofDHF patientsisdensity(X_1), mobility(X_2), the average age ofDHF patients(X_3) and the number ofDHF patientsattheearlier time periods($Y_{i,(t-1)}$).

Selection of the best models and interpretation of model:
Further, modeling using FEM with transformation Y* and addition of variable $Y_{i,(t-1)}$with significant variables. The model is obtained as follows:

$\hat{y}^{*}_{\mathrm{it}} = 3.3057 - 0.2893D_2 + \cdots + 1.0915D_{68} - 0.0002X_{1\mathrm{it}} - 0.0006X_{2\mathrm{it}} + 0.0114X_{3\mathrm{it}} - 0.0087Y_{\mathrm{i},(\mathrm{t}-1)}$

Based on the model it showed that the model is already meet all testing assumptions. Multicollinearitytest using VIF showed that the variables of density, mobility, the average age of DHF patients and the number of DHF patient previous period have VIF value less than 10. It means, the modelis hasn't multicollinearity problems. Durbin-Watson test produces a value of 1.9401between d_U and 4-d_U. It means, the model is hasn't autocorrelation. Testing heteroscedasticity with BPG test showed that the model hasn't heteroscedasticity problem. Identify normality on residual with Jarque-Bera test produces p-value of 0.1005shownresidual have normal distribution.

Density coefficient of -0.0002, it means that if the density in village *i* year *t* as many as ten thousand people have an impact on the probability found DHF cases decrease2 persons.Mobility coefficient of -0.0006,it means that if the mobility in villages *i* year *t*have ten thousand people then impact on the probability found DHF cases decrease 6 people. Coefficient of the average age of DHF patients amounted to 0.0114, it means that if in the village *i*year *t*,the average age of DHF patients age 30 years, then probability founded of DHF cases in the village *i* year *t*the average age is 0.3 years.

R^2 value selection model amounted of a 72.76% and the MSE of 0.2736. The results of the village specific effect ($\beta_{0\mathrm{i}}$) shown in Table 3.

Table 3: The value village specific effect of FEM used transformation data dependent variable Y*and the addition of variable $Y_{\mathrm{i},(\mathrm{t}-1)}$on models with significant independent variables

Village	Effects	Village	Effects	Village	Effects
Mulyaharja	3.3057	Tegalgundil	7.3265	Menteng	5.6515
Pamoyanan	3.0165	Tanahbaru	5.4590	Cilendektimur	6.7243
Ranggamekar	4.3028	Cimahpar	3.5021	Cilendekbarat	5.3293
Genteng	1.8374	Ciluar	3.7296	Sindangbarang	4.2433
Kertamaya	0.4232	Cibuluh	5.6255	Margajaya	2.2890
Rancamaya	1.0701	Kedunghalang	5.6510	Balumbangjaya	3.8328
Bojongkerta	1.3281	Ciparigi	6.8351	Situgede	2.2731
Harjasari	3.6106	Paledang	3.7244	Bubulak	3.6098
Muarasari	2.0365	Gudang	7.0244	Semplak	8.6224
Pakuan	2.2637	Babakanpasar	7.7731	Curugmekar	5.5270
Cipaku	4.4267	Tegallega	6.5553	Curug	3.8804
Lawanggintung	5.2773	Babakan	4.7828	Kedungwaringin	7.0608
Batutulis	5.5757	Sempur	5.5329	Kedungjaya	6.4282
Bondongan	7.7148	Pabaton	3.1252	Kebonpedes	8.6740
Empang	7.7469	Cibogor	6.1643	Tanahsareal	4.6378
Cikaret	3.2171	Panaragan	8.9563	Kedungbadak	6.9994

Sindangsari	4.4864	KebonKelapa	8.4462	Sukaresmi	4.9606
Sindangrasa	4.6811	Ciwaringin	4.8126	Sukadamai	5.2181
Tajur	5.2977	Pasirmulya	3.1898	Cibadak	4.3814
Katulampa	4.0909	Pasirkuda	4.0974	Kayumanis	3.7905
Baranangsiang	6.1633	Pasirjaya	3.6331	Mekarwangi	6.0106
Sukasari	7.9735	Gunungbatu	5.4123	Kencana	4.3972
Bantarjati	7.9446	Loji	3.7746		

Conclusion and Remarks

FEM model used the results of the transformation data dependent variables(Y*) and the addition variables one period before ($Y_{i,(t-1)}$) is a model that can describe the influence independent variables to the number of DHF patients in Bogor from period2009 to2013.Based on the results of the model, variables that influence to the number of DHF patients are density, mobility, the average age of DHF patients and the number of DHF patient one period before at 5% significance level. This is best model that can be shown R^2 of 72.76% and MSE of 0.2736.

References

[1] Dinkes Kota Bogor, Dinas Kesehatan Kota Bogor (2014),"Kinerja Dinas Kesehatan Kota Bogor Tahun 2014 Menuju Bogor Kota Sehat", Bogor: PusatInformasidan Humas Dinkes Kota Bogor.

[2] Greene, W.H. (2012),"Econometric Analysis: 7th Ed", New York: Prentice Hall.

[3] Gujarati, D.N. (2009),"Basic Econometrics:5thEd", New York: The McGraw-Hill Companies.

[4] Judge, G.G., Griffith, W.E., Hill, R.C., Lee, T. (1980). "The Theory and Practice of Econometrics", New York: John Wiley & Sons, Inc.

[5] Kemenkes, Kementerian Kesehatan Republik Indonesia(2004),"Tata Laksana Demam Berdarah *Dengue* di Indonesia", Jakarta: PusatInformasidan Humas Kemenkes RI