IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 6 Issue 8, August 2019
ISSN (Online) 2348 – 7968
www.ijiset.com

Bayesian Panel Data by Means Integrated Nested Laplace

I Gede Nyoman Mindra Jaya[1*], Neneng Sunengsih[2]

[1] Department Statistics, Universitas Padjadjaran, Indonesia

[2] Department Statistics, Universitas Padjadjaran, Indonesia

Abstract
Selection between fixed and random effect model becomes a crucial problem in panel data analysis. Hausman test is a popular tool that usually used to define whether a fixed or random-effect model as the best model for the data. However, in recent years, this method has been criticized. The Hausman may be misleading for some conditions. If the number of time points greater than number of cross-section unit, the Hausman-test tends to wrongly reject the Null-hypothesis of uncorrelated unit effects. Bayesian numerical analysis by means integrated nested Laplace (INLA) is the one alternative that can be used to model panel data. Bayesian approach provides several criteria for model selection between pooled, fixed and random effect model. Those criteria are deviance information criterion (DIC) and marginal predictive likelihood (MPL) and Bayes Factors (BF). Bayesian INLA is applied to model stock price LQ45 on the current ratio (CR) and return on equity (ROE).

Keywords: *Bayesian, Hausman test, Fixed and Random effect, Stock price*

1. Introduction

Bayesian methods are going to more popular in applied and theoretical research [1]. The main reason to use Bayesian statistics is that facilitate the uncertainty in the parameters values. Maximum likelihood or ordinary least square and the other frequentist approach assumes that the values of the parameters are fixed. In fact the parameters values may change over time and conditions. The advantage of Bayesian approach is that accommodate the prior information [2].

Panel data refers to the merging of observations on a cross-section of households, districts, firms, etc. over several time periods [3]. This can be done by collecting a number of households or individuals regularly. Some study used panel data to present more informative results and satisfy the statistical assumptions [4].

Maximum likelihood, ordinary least square and method of moment are the most popular estimators that usually use to estimate the parameter of panel data models and the standard assumptions such as homoscedasticity, non-autocorrelation, and normality must be fulfilled ([5], [6]). Two common models was introduce for panel data structure to take account of the special time structure are fixed and random effects models ([3], [5], [7]). A Hausman (1978) [8] test can be used to choose between the fixed and random effect models, whether fixed effects are needed for controlling of unit heterogeneity or whether more efficient random effects can be use instead ([9], [5]). However, this test have been criticized in simulation studies for both its over rejection of false nulls and underwhelming power ([10], [11]). In other hand, Hausman test will choose one of both models although both models are inadequate description of the data because of in frequentist approach only will decide accepted or rejected the null hypothesis. Additional tests are available that can provide evidence of model adequacy [7].

Bayesian approach might be used as an alternative solution and more flexible for violation of the standard assumptions and gives some model selection criteria might be more useful to select the best model and define the best fit to the data ([12], [13]). The most popular model selection criteria are deviance information criterion (DIC), marginal predictive likelihood (MPL) and pseudo Bayes factor (BF) ([14] [15]).

Markov Chain Monte Carlo (MCMC) is commonly used to estimate the parameter model in Bayesian setting ([16] , [15]). However, MCMC might be high computation cost due to the complexity of the models, particularly the large number of parameters such as panel data. Integrated Nested Laplace Approximation (INLA) was introduced to overcome the computational time in MCMC ([15] [17]). In this study we apply the INLA setting to estimate and evaluate panel model.

The structure of the remainder of this paper is as follows. Section 2 presents the panel data model and summarizes its estimation by INLA. Section 3 applies the method to stock price of LQ45, Indonesia. Section 4 presents the conclusions.

2. Method

2.1. Panel Data

Panel data analysis needs data in panel specific structure. The data are structured as a repetition of observations for each cross-section units (e.g., firm, household, district). Least square estimator may be used to estimate the parameter models and must be fulfilled the standard regression assumption to obtain Best Linear Unbiased Estimate and we called this model as pooled model. However, the units heterogeneity may lead to heteroscedastic problem. In panel data we assume there are unique chrematistics of cross-section units that do not vary over time. This unique characteristics may or may not be correlated with the covariates. Fixed effect and random effect models are the two most popular models in panel data analysis [6]. Fixed effect model is better used when the individual characteristics are correlated with independent variables, and random effect model when the unit characteristics are random. Pooled model may presents unbiased and consistent parameters estimates even when time constant characteristics are present, but random effect model will be more efficient. Using feasible generalized least square which is asymptotically, fixed effect model more efficient than Pooled model when constant characteristics are present. Random effects adjusts for the serial correlation which is induced by unobserved time constant attributes [18]. In frequentist approach, Chow test can be used to choose between pooled model versus fixed effect model, Hausman test to compare fixed and random effect model and Lagrange multiplier test to choose between pooled model versus random effect model.

2.1.1. Fixed effect model

The fixed effect model can be written as:

$$y_{it} = (\alpha + \mu_i) + \sum_{k=1}^{K} \beta_k x_{itk} + \varepsilon_{it} \qquad (1)$$

where y_{it} denotes the response variable for unit cross-section i and time t, α is an intercept, μ_i individual characteristic which constant over time and sometimes we write $(\alpha + \mu_i) = \alpha_i$. The $k-th$ independent variables for unit cross-section i and time t is denoted by x_{itk} and its slop coefficient is β_k. The random error (ε_{it}) which is assumed independent and identically distribution with zero mean and variance σ^2. For hypothesis testing purpose, ε_{it} is assumed follows normal distribution ([5], [4]). Here, fixed effect term is used due to μ_i is assume as a fixed parameter. Least square dummy variables can be used to estimate the model (1).

2.1.2. Random effect model

The random effect model can be written as:

$$y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{itk} + (\mu_i + \varepsilon_{it}) \qquad (2)$$

In contrast with fixed effect model, in random effect model we assume μ_i is a random component with zero mean and variance σ^2. Generalized least square can be used to estimate the model (2) ([3], [5], [4]).

2.1.3. Model comparison by means frequentist approach

2.1.3.1 Chow test

Chow test is used to choose between pooled and fixed effects model whit the hypothesis $H_0\colon \mu_1 = \cdots = \mu_n = 0$ by performing Chow test with the restricted residual sum squares (RRSS) being that least square on the pooled model and the unrestricted residual sums of square (URSS) being that on the LSDV regression [3].

$$F_0 = \frac{(RRSS - URSS)/(n-1)}{URSS/(nT - n - K)} \sim F_{n-1, n(T-1)-K} \qquad (3)$$

Fixed effect model is selected if the test reject H_0

2.1.3.2. Hausman test

Hausman test is used to choose between fixed and random effects model whit the hypothesis $H_0\colon E(\mu_i | x_{itk}) = 0$. Under the null hypothesis we test [3]:

$$W = (\boldsymbol{\beta}_{RE} - \boldsymbol{\beta}_{FE})' \widehat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\beta}_{RE} - \boldsymbol{\beta}_{FE}) \sim \chi^2(k) \qquad (4)$$

where $\widehat{\boldsymbol{\Sigma}} = \mathrm{Var}(\boldsymbol{\beta}_{RE}) - \mathrm{Var}(\boldsymbol{\beta}_{FE})$ denotes the covariance of an efficient estimator with its difference from an inefficient estimator. Fixed effect model is selected if the test reject H_0

2.1.3.3 Lagrange multiplier test

Lagrange multiplier test is used to choose between pooled and random effects model whit the hypothesis $H_0\colon V(\mu_i) = 0$. Under the null hypothesis we test [3]:

$$LM = \frac{nT}{2(T-1)} \left[\frac{\sum_{i=1}^{n} (\sum_{t=1}^{T} \hat{e}_{it})^2}{\sum_{i=1}^{n} \sum_{t=1}^{T} \hat{e}_{it}^2} - 1 \right]^2 \sim \chi^2(1) \qquad (5)$$

where $\hat{e}_{it}$ is residual from pooled model.
Random effect model is selected if the test reject H_0

2.2. INLA Modeling

2.2.1 Integrated Nested Laplace Approximation: INLA

INLA is a Bayesian numerical method with three stages processes. The first stage defines the observational model $\boldsymbol{\pi}(\boldsymbol{y}|\boldsymbol{\vartheta})$, where $\boldsymbol{y}$ denotes the response variable a vector column. The second stage defines the latent Gaussian field (GMRF) with precision matrix $\boldsymbol{Q}$ and the third stage defines controlling hyperparameter model [19]. For the first stage, we assume that response variable follow Gaussian distribution $y_{it} \sim Gaussian(\alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta}, \sigma^2)$, where $\boldsymbol{y} = [y_{11}, \ldots, y_{nT}]'$.

$$f(\boldsymbol{y}|\boldsymbol{\eta}) = \prod_{i=1}^{n}\prod_{t=1}^{T} f(y_{it}|\boldsymbol{\Phi}) \qquad (6)$$

where $\boldsymbol{\Phi}$ is the vector parameter $\boldsymbol{\Phi} = (\alpha, \boldsymbol{\beta}, \sigma^2)'$

$$f(y_{it}|\eta_{it}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_{it} - \boldsymbol{x}'_{it}\boldsymbol{\beta})^2\right)$$

For fixed effect and random effect model, $y_{it} \sim Gaussian(\alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \mu_i, \sigma^2)$.

The second stage defines the latent Gaussian field (GMRF) with precision matrix $\boldsymbol{Q}$ (see [19] for detail).

We need to define the hyperprior distribution of the hyperparameter ($\kappa_0 = 1/\sigma^2$). Commonly, the distribution of the inverse of hyperparameter are defined. The inverse is its variance ($\sigma^2 = \frac{1}{\kappa_0}$) and we taking IG(1, 0.00005). Fixed effect model assume μ_i is a fixed and for random effect model $\mu_i \sim N(0, \sigma_\mu^2)$ and $\sigma_\mu^2 \sim IG(1, 0.00005)$

2.2.2 Bayesian model selection

Deviance information criterion (DIC) is the most popular model selection criteria in Bayesian setting. This criteria consider both fit and complexity [20]. It defines as:

$$\mathrm{DIC} = D(\widehat{\boldsymbol{\Phi}}) + 2\mathrm{p_{DIC}}, \qquad (7)$$

where $\widehat{\boldsymbol{\Phi}} = E[\boldsymbol{\Phi}|\mathbf{y}]$ and where $D(\widehat{\boldsymbol{\Phi}})$ is the model's deviance, i.e. $D(\widehat{\boldsymbol{\Phi}}) = -2\log p(\mathbf{y}|\widehat{\boldsymbol{\Phi}})$, and $\mathrm{p_{DIC}}$ denotes the effective number of parameters.

Another method is marginal predictive likelihood (MPL), defined as:

$$\mathrm{MPL} = \sum_{i=1}^{n}\sum_{t=1}^{T} \log(\mathrm{CPO_{it}}). \qquad (8)$$

where conditional predictive ordinate (CPO) is defined as:

$$\mathrm{CPO} - \mathrm{failure_{it}} = \sum_{j=1}^{J} \mathrm{failure}_{it,j}\, p(\boldsymbol{\tau}^{(j)}|\mathbf{y})\Delta_j \qquad (9)$$

with where $\mathrm{failure}_{it,j}\, p(\boldsymbol{\tau}^{(j)}|\mathbf{y})$ indicates the misfit of $p(\boldsymbol{\tau}^{(j)}|\mathbf{y})$ for observation y_{it} at the j^{th} grid, and Δ_j is the corresponding weight. The larger the MPL, the better the prediction.

We also can use the probability integral transform (PIT) is the value of the predicted cumulative distribution function at observation y_{it} [19]:

$$\mathrm{PIT}_{it} = \int p(\hat{y}_{it} \le y_{it}|\boldsymbol{\Phi})p(\boldsymbol{\Phi}|\boldsymbol{y}_{-it})d\boldsymbol{\Phi}. \qquad (10)$$

The PIT histogram indicates the model fit across all panel data. The closer the PIT histogram is to the uniform distribution histogram, the better the fit [19].

Another model selection criterion is the Pesudo Bayes Factor (BF). It is defined as follows. Assume two models, M_1 and M_2. For models M_1 and M_2, the PBF reads [21]:

$$\mathrm{PBF} = \exp(\mathrm{MPL}(M_1) - \mathrm{MPL}(M_2)) \qquad (11)$$

A PBF $<$ 1 indicates that the data favour M_2 over M_1.

Other measures of predictability are the mean absolute error (MAE), the root mean square error (RMSE) and the adjusted pseudo-coefficient of determination ($\widetilde{\mathrm{R}}^2$).

3. Result and Discussion

3.1. Data description

Data used in this study is LQ45 which is obtained from Indonesia stock exchange (https://www.idx.co.id/) on period 2013-2016. Total number of sample is 21 companies incorporated in LQ45. We used stock price as dependent variable (y) and two independent variables are current ratio (x_1) and return on equity (x_2).

Table 1. Descriptive statistics of variables.

Variables	Min	Max	Mean	SD
Price	343.00	63,900.00	10,588.69	13,455.09
Current Ratio	0.45	9.72	2.23	1.70
Return on Equity	0.03	1.84	0.23	0.30

Table 1 shows the descriptive statistics of the price (P), current ratio (CR) and returns on equity (ROE). The minimum price of LQ45 stock is IDR 343, and the maximum price is IDR 63,900 with the average IDR 10,588.69. The stock price has a high standard deviation (13,455.09) indicates there is a high differential stock price between companies which grouped in LQ45. It is clearly shown in Figure 1.

IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 6 Issue 8, August 2019
ISSN (Online) 2348 – 7968
www.ijiset.com

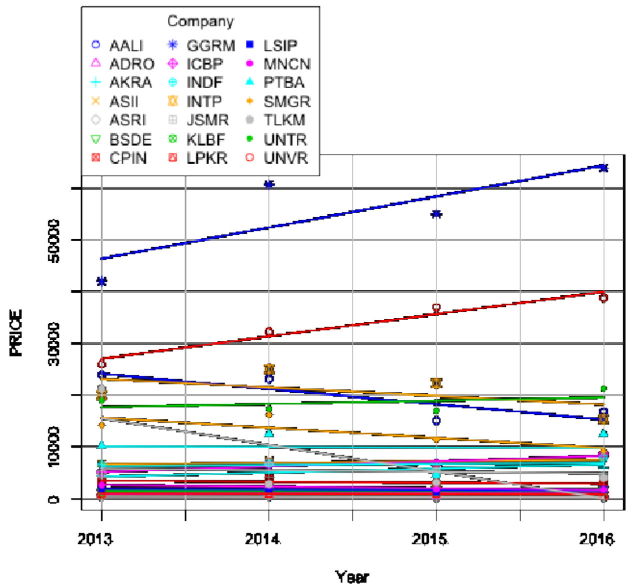

Figure 1. Temporal trend of stock price for each firm (2013-2016)

Some firms have high stock price over 2013 to 2016 and some firms have lower stock price. Every firm seem has linear temporal trend where with some of firms have a positive gradient and the other have negative gradients. The high variability of stock price over firms might and there is no non-linear temporal trend be modeled by means panel data analysis.

There are three candidates model will be evaluates induced: pooled, fixed and random effect model. Two different estimators will be applied, frequentist and Bayesian INLA approaches.

3.3. Data panel modeling

There are three type of panel data model will be constructed: pooled, fixed, and random effects models. We applied frequentist estimator and Bayesian INLA to model panel data included pooled, fixed and random effects models. The results are shown in Table 2.

Table 2. Parameters estimate of panel model by means least square and INLA methods

Parameter	Least square			INLA		
	Pooled	Fixed Effect	Random Effect	Pooled	Fixed Effect	Random Effect
Intercept	10.1315		8.9548	10.1315		8.9467
	(0.3552)		(0.3090)	(0.3491)		(0.2931)
log(Current Ratio)	-0.1541	0.0976	0.0495	-0.1541	0.0994	0.0533
	(0.1962)	(0.1390)	(0.1339)	(0.1928)	(0.1380)	(0.1233)
log(Return on Equity)	0.8210	0.2060	0.2434	0.8210	0.2034	0.2402
	(0.1875)	(0.0885)	(0.0883)	(0.1843)	(0.0879)	(0.0811)

Note: (.) its standard error estimate

Table 3. Model comparison by means frequentist approaches

Test	Statistics	Decision
Pooled vs Fixed H_0: Pooled effect model H_1: Fixed effect model	F = 55.182, df_1 = 20, df_2 = 61, p-value < 2.2e-16	Reject H_0
Fixed vs Random H_0: Random effect model H_1: Fixed effect model	Chisq = 132.91, df = 2, p-value < 2.2e-16	Reject H_0

Table 4. Model comparison by means Bayesian approaches

Model	DIC	MAE	RMSE	R^2	MPL
Pooled	264.5446	0.9079	1.1275	0.2231	-132.1164
Fixed	60.3781	0.1710	0.2580	0.9592	-32.08262
Random	57.8861	0.1744	0.2593	0.9319	-32.53193

ISSN (Online) 2348 – 7968
www.ijiset.com

Table 5. Pseudo Bayes Factor

Test	PBF	Decision
Pooled vs Fixed M_1: Pooled effect model M_2: Fixed effect model	$PBF_{12} = \exp(\text{MPL}(M_1) - \text{MPL}(M_2)) = 0$	M_2: Fixed effect model
Pooled vs Random M_1: Pooled effect model M_3: Random effect model	$PBF_{13} = \exp(\text{MPL}(M_1) - \text{MPL}(M_3)) = 0$	M_2: Random
Fixed vs Random M_2: Random effect model M_3: Fixed effect model	$PBF_{23} = \exp(\text{MPL}(M_2) - \text{MPL}(M_3)) = 1.567$	M_2: Fixed effect model

Table 2 shows the parameters estimate of panel model by means least square and INLA methods. The results are almost similar between frequentist and Bayesian methods. Table the model comparison by means frequentist approach. Using Chow test and Hausman test (chi-square), fixed effect model is the best model. Hausman test is strongly support that the fixed effect is better than random effect. Table 4 and 5 presents the Bayesian model selection criterion. Using Bayesian approach, the fixed and random effect models have similar performance. The DIC, R^2, RMSEA, and MAE are not significantly different. PBF also presents the small values < 3 which indicates both of model have similar performance.

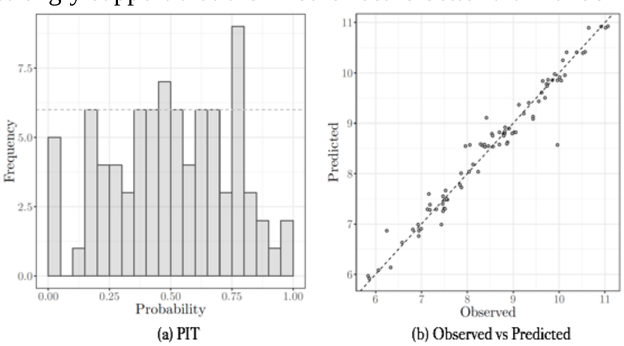

Figure 2. Model evaluation of fixed effect

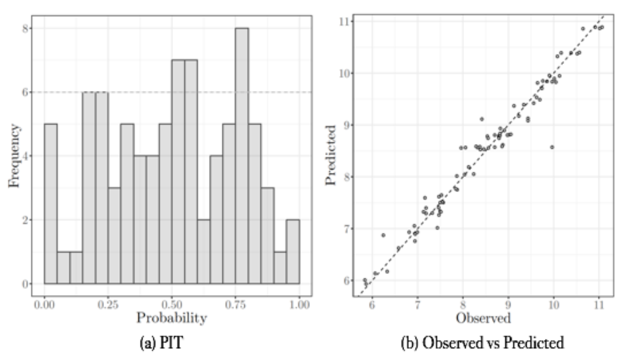

Figure 3. Model evaluation of random effect

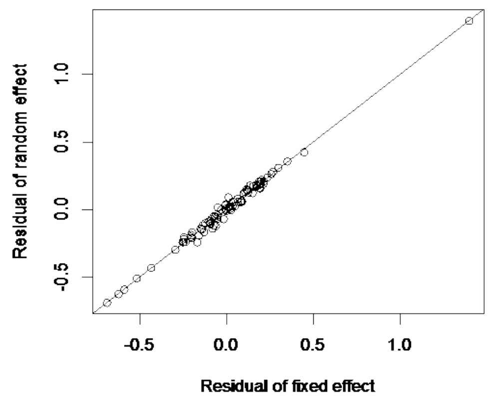

Figure 4. Residual of fixed effect vs random effect

Figure 2 and Figure 3 present PIT and linear plot between observed and predicted values. PIT plots of fixed and random effect models shows uniform pattern which indicates the model fit to the data. This conclusion is also supported by figure observed vs predicted values which shows the perfectly linear pattern. Figure 4 presents the plot between residual of fixed and random effect model which shows the perfectly linear pattern. It indicates the fixed and random effect models are a good model have good predictability and goodness of fit. From the parameter estimates, comparing the regression parameter estimate on its standard error, only the return on equity has a significant effect on the stock price. This result comes from the fixed and random effects model. This result suggests that the evaluation or model selection have to consider the theoretical background and its application, where the statistical tools only for supporting the theoretical and practical aspects.

5. Conclusion

The Hausman may be misleading for some conditions. If the number of time points greater than number of cross-section unit, the Hausman-test tends to wrongly reject the Null-hypothesis of uncorrelated unit effects. Bayesian numerical analysis by means integrated nested Laplace is the one alternative that can be used to model panel data. Bayesian approach provides several criteria for model selection between pooled, fixed and random effect model. Those criteria are deviance information criterion (DIC) and marginal predictive likelihood (MPL) and Bayes Factors (BF). Using Bayesian approach we have similar parameters estimates result with least square approach. However, Bayesian approach found that the fixed effect and random effect models have similar performance in modelling stock price. This result is very different from Hausman statistics which informed that fixed effect model is the best model. However, there is no high different in parameters estimates between fixed and random effect model. Those models concluded that only return on equity has significant effect on the stock price LQ45.

Acknowledgments

This paper is funded by the RFU Unpad contract: 1732 d/UN6.RKT/LT/2018. The authors thank Rector Universitas Padjadjaran and to the anonymous referee whose valuable checking has improved this paper

Reference

[1] R. v. d. Schoot, D. Kaplan, J. Denissen, J. B. Asendorpf, F. J. Neyer and M. A. v. Aken, "A Gentle Introduction to Bayesian Analysis: Applications to Developmental Research," Child Development, vol. 85, no. 3, p. 842–860, 2014.

[2] F. J. Samaniego, A Comparison of the Bayesian and Frequentist Approaches to Estimation, New York: Springer, 2010.

[3] B. H. Baltagi, Econometric Analysis of Panel Data, West Sussex: John Wiley & Sons, 2005.

[4] J. M. Wooldridge, Econometric Analysis of Cross Section and Panel Data, Massachusetts London: The MIT Press, 2001.

[5] W. H. Greene, Econometric Analysis, New Jersey: Prentice Hall, 2006.

[6] U. Morawetz, "Bayesian modelling of panel data with individual effects applied to simulated data," University of Natural Resources and Applied Life Sciences, Vienna, 2006.

[7] K. A. Bollen and J. E. Brand, "A General Panel Model With Random And Fixed Effects: A Structural Equations Approach," Soc Forces, vol. 89, no. 1, pp. 1-34, 2010.

[8] Hausman, "Specification Tests in Econometrics," Econometrica, vol. 46, no. 6, pp. 1251-1271, 1978.

[9] C. Cameron and P. Trivedi, Microeconometrics Methods and Applications, New York: Cambridge University Press, 2005.

[10] T. S. Clark and D. A. Linzer, "Should I Use Fixed or Random Effects?," Political Science Research and Methods, vol. 3, no. 2, pp. 399-408, 2015.

[11] T. Sheytanova, "The Accuracy of the Hausman Test in Panel Data: a Monte Carlo Study," Orebro University, Sweden, 2004.

[12] G. Koop, D. J. Poirier and J. L. Tobias, Bayesian Econometric Methods, New York: Cambridge Press, 2007.

[13] J. P. LeSage, "Spatial econometric panel data model specification: A Bayesian approach," Spatial Statistics, vol. 19, no. 1, pp. 122-145, 2014.

[14] A. Gelman, J. Hwang and A. Vehtari, "Understanding predictive information criteria for Bayesian models," Stat Comput, vol. 24, no. 1, p. 997–1016, 2014.

[15] M. Blangiardo and M. Cameletti, Spatial and Spatio-temporal Bayesian Models with R - INLA, Chichester: John Wiley & Sons, 2015.

[16] I. Jaya, H. Folmer, B. N. Ruchjana, F. Kristiani and A. Yudhie, "Modeling of Infectious Diseases: A Core Research Topic for the Next Hundred Years," in Regional Research Frontiers - Vol. 2 Methodological Advances, Regional Systems Modeling and Open Sciences, USA, Springer International Publishing, 2017, pp. 239-254.

[17] I. Jaya, H. Folmer, B. N. Ruchjana, F. Kristiani and Y. Andriyana, "Modeling of Infectious Disease: A Core Research Topic for The Next Hundred Year," in Regional Research Frontiers, US, Springer, 2017, pp. 681-701.

[18] C. Hsiao, K. Lahiri, L.-F. Lee and M. H. Pesaran, Analysis of panels and limited dependent variable models, Cambridge: Cambridge University Press, 2004.

[19] M. Blangiardo and M. Cameletti, Saptial and Spatio-Temporal Bayesian Model with R-INLA, John Wiley: United Kingdom, 2015.

[20] D. J. Spiegelhalter, N. G. Best, B. P. Carlin and A. v. d. Linde, "Bayesian Measures of Model Complexity and Fit," Journal of the Royal Statistical Society. Series B (Statistical Methodology, vol. 6, no. 4, pp. 583-639, 2002.

[21] A. E. Gelfand, H.-J. Kim, C. Sirmans and S. Banerjee, "Spatial Modeling with Spatially Varying Coefficient Processes," Journal of the American Statistical Association, vol. 98, no. 462, pp. 387-396, 2003.