Modeling Human Development Index of Bali with Spatial Panel Data Regression

Mita Lasdiyanti, Eka N. Kencana†, and Putu Suciptawati

Abstract—Human development index (HDI) is an index that represents the successfulness of human development in a region. For Bali, one of 34 provinces in Indonesia, the progress of HDI in the period 2010–2017 showed an increasing trend. In the year 2010, the Bali’s HDI is accounted for 70.10, gradually increase to 74.30 in the year 2017. However, in 2017 there are some regions with their HDIs are below of Bali’s HDI, namely Jembrana, Buleleng, Klungkung, Bangli, and Karangasem. The aim of this work is to model the HDI of 9 regencies of Bali so that the main determinant to increase the HDIs especially for the regencies with lower HDIs could be determined. The model consists of one dependent variable (HDI) with three indicators as the independent ones, there are (a) life expectancy, (b) education, and (b) standard of living. By applying spatial panel data analysis, five models were built i.e. CEM, FEM (individual), FEM (time), REM, and spatial autocorrelation FEM to determine the effect of each indicator. The result shows the best model is spatial autocorrelation FEM in which the health index has the biggest influence compare to the others.

Index Terms—Bali, HDI, regression, spatial panel data.

I. INTRODUCTION

One of some prominent issues in national development program that becomes a top priority for Indonesia as well as other countries worldwide is human development. According to the United Nations Development Program, human development is the freedom of choice for humans in improving their quality of life. The choices referred to cover long and healthy life, gain knowledge, and the ability to access the resources needed to obtain a decent standard of living. In measuring the human development successfulness of a country, UNDP introduced an indicator namely the human development index (HDI) [1].

According to Indonesian Statistics Office, the Human Development Index (HDI) of Indonesia has been increasing from 66.5 that was recorded for the 2010 to 70.8 for the year 2017 [2]. Basically, HDI is composed of 3 main indices, namely (a) the health index which is represented by the life expectancy; (b) the education index which is depicted by the average length of school and the length of education expectations; and (c) a decent standard of living index which is figured out through the adjusted per capita expenditure. Refers to these indices, HDI is calculated. According to the HDI’s value, it can be classified into 4 groups, namely: (a) very high HDI if HDI ≥ 80; (b) high HDI if 70 ≤ HDI < 80; (c) medium HDI if 60 ≤ HDI < 70; and (d) low HDI if HDI < 60.

As one of 34 provinces in Indonesia, the economic growth of Bali Province which is based on two leading sectors, tourism and agriculture, exceeds national economic growth. The same trend was observed in the growth of the HDI. In the period 2010-2017 the HDI of Bali Province increased gradually. Regional Statistics Office of Bali Province noted that the average growth of HDI’s Bali in this period was recorded at 0.84 percent per year [3]. In detail, the development of HDI in 9 districts/cities in Bali Province in the period 2010-2017 is shown in Fig. 1.

The importance of the HDI calculation is inseparable from the role of HDI in designing development programs to reduce poverty. The study by Budinirmala et al. [4] in modeling population poverty in Bali using panel data regression showed that each increase in 1 unit of HDI was able to reduce the percentage of poor people as much as 0.250 units. Their research showed HDI has a real effect in reducing the percentage of poor people in Bali Province. Another study in modeling HDI was conducted by Trianggara et al. [5]. Using the spatial panel data method, they examined the factors that influenced the HDI of Central Java Province in the period 2008–2013. Their research justifies the best model in modeling the HDI is the spatial lag fixed effect model with the $R^2$ as much as 99.54 percent. In addition, the variables that significantly influence the HDI are school participation rates and the percentage of poverty.

Noting the importance of HDI, this research is intended to model HDI of Bali. The model is built by involving 3 compiling indexes, namely the health index (HEA), the education index (EDU), and a decent standard of living index (LIV) with a direction to find out how much the influences of these indices in affecting HDI on 9 districts/cities of Bali. The HDI’s data for the period 2010–2017 were used and the model was established by applying spatial panel data analysis techniques.

II. SHORT LITERATURE REVIEW

In this study, the causality between the HDI and the three indices is modeled through two approaches, namely: (a) the Panel Data Regression, the regression equation model on panel data that does not take into account the spatial influence between objects; and (b) the Spatial Panel Data Regression that considers the existence of spatial influences.
A. The Panel Data Regression

In modeling panel data by applying regression techniques, according to Greene [6], can be categorized into three groups namely (a) Common Effect Model (CEM); (b) Fixed Effect Model (FEM); and (c) Random Effect Model (REM). Of these models, CEM is the simplest one.

1) The Common Effect Model: to model the HDI of 9 regencies/cities in Bali \((i = 1, \ldots, 9)\) for the period 2010 (time index \(t = 1\)) to 2017 (time index \(t = 9\)) by the HEA, EDU, and LIV as the predictors; the CEM can be expressed in following equations:

\[
HDI_{it} = \alpha + \beta_1 HEA + \beta_2 EDU + \beta_3 LIV + \mu_{it};
\]

2) The Fixed Effect Model: According to Gujarati [7], the FEM can be differentiated to FEM with fixed individual effect (FEM\(_I\)) and FEM with fixed time effect (FEM\(_T\)). Both types can be expressed as follows:

\[
HDI_{it} = \sum_{j=1}^{9} D_{jt} \alpha_i + \beta_1 HEA + \beta_2 EDU + \beta_3 LIV + \mu_{it};
\]

\[
D_{jt} = \begin{cases} 
1 & \text{if } j = i; \ i = 1, \ldots, 9 \\
0 & \text{otherwise}
\end{cases}
\]

\[
HDI_{it} = \sum_{j=1}^{9} D_{jt} \alpha_t + \beta_1 HEA + \beta_2 EDU + \beta_3 LIV + \mu_{it};
\]

\[
D_{jt} = \begin{cases} 
1 & \text{if } j = t; \ t = 1, \ldots, 9 \\
0 & \text{otherwise}
\end{cases}
\]

3) The Random Effect Model: In FEM, the intercept of regression line \((\alpha_i \text{ or } \alpha_t)\) is always assumed fixed. However, for REM, the intercept is considered a random variable with a mean value of \(\alpha\). So, for an individual regency, the intercept can be expressed as \(\alpha_i = \alpha_I + \epsilon_i; \epsilon_i\) is a random term with zero mean and a variance of \(\sigma^2_\epsilon\). Noting this assumption, the FEM\(_I\) of nine regencies/cities in Bali will share a common intercept value \(\alpha_I\) and an error term \(\epsilon_i\); and their FEM\(_T\) will share an \(\alpha_T\) and an error term \(\epsilon_t\). Refers to these conditions, equation (2) and (3) can be expresses as follows:

\[
HDI_{it} = \alpha_I + \beta_1 HEA + \beta_2 EDU + \beta_3 LIV + \epsilon_i + \mu_{it};
\]

\(i = 1, \ldots, 9\)

\[
HDI_{it} = \alpha_T + \beta_1 HEA + \beta_2 EDU + \beta_3 LIV + \epsilon_t + \mu_{it};
\]

\(t = 1, \ldots, 9\)

B. The Spatial Panel Data Regression

Whenever the spatial differences amongst individual objects are considered in modeling the panel data, the researcher(s) applied the spatial panel data regression. According to Elhorst [8], the standard model for spatial panels data can be expressed in a matrix term as

\[
Y_{it} = X_{it} \beta + \mu_i + \epsilon_{it};
\]

or in stacked form: [9]

\[
Y_t = X_t \beta + \mu + \epsilon_t
\]

Where \(Y_t = (Y_{1t}, \ldots, Y_{Nt})\), \(X_t = (X_{1t}, \ldots, X_{Nt})\), \(\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{Nt})\), and \(\mu = (\mu_1, \ldots, \mu_N)\).

In the spatial panel data models, Elhorst [9] introduced the spatial weight matrix, \(W_{N \times N}\). This matrix represents the spatial arrangement of the spatial units, and its properties can be read in [9]. For the standard model in equation (7), traditionally we can classify the fixed effect spatial model into 2 types i.e. the fixed effect spatial lag model, and the fixed effect spatial error model, with the mathematical formulation as follows:
1) The Fixed Effect Spatial Lag Model:
\[ Y_t = \rho W Y_t + X_t \beta + \mu + \epsilon_t \]

2) The Fixed Effect Spatial Error Model:
\[ Y_t = X_t \beta + \mu + \phi_t \]
\[ \phi_t = \gamma W \phi_t \]

In equation (8), \( \rho \) is named as the spatial autoregressive coefficient while in (9) \( \gamma \) is called as the spatial autocorrelation coefficient.

III. RESEARCH METHOD

To model the HDI’s of Bali by applying spatial panel data regression, following steps are done:
1) build the CEM, FEM, and REM for the data set without taking into account the spatial heterogeneity may exists;
2) choose the best model from the above models;
3) define the spatial matrix \( W \);
4) if there were an evidence for spatial dependence, then build the spatial lag model as well as the spatial error model;
5) choose the best model from both models;
6) interpret the final model.

IV. RESULT AND DISCUSSION

A. Results for Panel Data Models

Without considering the possibility of spatial dependence between units, we made the CEM, FEM, and REM according to equation (1) to equation (5). By assuming the error term in equation (1) \( \mu_{it} \sim iid(0, \sigma_{\mu}^2) \), we get the CEM or the pooled regression estimator by utilizing the OLS method as follows:

| Coefficient Estimator T-Value p-Value |
|-------------------------------------|------------------|------------------|
| Intercept \( \beta_0 \) -1.373 -3.821 0.000 |
| HEA \( \beta_1 \) 33.118 49.461 0.000 |
| EDU \( \beta_2 \) 39.684 143.786 0.000 |
| LIV \( \beta_3 \) 29.611 83.610 0.000 |

| F-Statistic 245.476 p-Value 0.000 Adjusted R² 99.99% |
| Durbin-Watson Statistic 0.644 |
| Residual Sum of Squares 0.212 |

Source: Own calculation (2019)

Table I showed all estimators are significant. In addition, the CEM has a very high of \( R^2 \) value that demonstrates the predictors highly explain the response variable. However, as suggested by [7], the Durbin-Watson (D-W) statistic must be checked. The D-W statistic for the pooled model is equal to 0.644, less than 1.372 as the lower limit of D-W value (70 sample size and 3 regressors). This fact suggests the possibility of autocorrelation and/or spatial autocorrelation in the data matrix [7]. To overcome this problem, we continued to build fixed individual (FEM\(_I\)) and fixed time (FEM\(_T\)) effects. To estimate the coefficients for FEM\(_I\) and FEM\(_T\), we applied the Least Square Dummy Variable (LSDV) method in estimating the parameters. The results are shown on Table II and Table III.

Both tables showed the significance of the fixed effect models. Despite the slightly greater value of the \( R^2 \) for FEM\(_T\), FEM\(_I\) had a very small residual sum squares (RSS) \( \approx 0.087 \) when it was compared to the RSS of FEM\(_T\) as much as 0.206. Guiding by this result, we decided to choose the FEM with individual effect as a representative model of FEM.

To compete the CEM with FEM\(_I\), we applied the Chow test to elaborate the parameter stability of CEM [7]. The hypotheses in this test are:
\[ H_0: \alpha_1 = \cdots = \alpha_9 = \alpha \]
\[ H_1: \exists \alpha_i \neq \alpha_j; i = j = 1, \cdots, 9, i \neq j \]

The hypotheses are tested by using pooltest function in the \texttt{plm} package [10]. The result gave \( F \) value as 10.823 with \( p-value = 0.000 \). Based on the Chow test, the null hypothesis is rejected and we concluded the FEM\(_I\) is a better choice than the CEM to model the data matrix. The last step in searching the best model for the panel data matrix is to build an REM.
Table IV shows the estimates by applying the Generalized Least Squares (GLS) method.

**TABLE IV. The Estimates of REM**

| Coefficient | Estimator | T-Value | p-Value |
|-------------|-----------|---------|---------|
| Intercept   | α         | -1.986  | -2.314  | 0.024   |
| HEA         | β₁        | 34.309  | 21.057  | 0.000   |
| EDU         | β₂        | 39.907  | 112.430 | 0.000   |
| LIV         | β₃        | 29.008  | 39.385  | 0.000   |

F-Statistic 63.279
p-Value 0.000
Adjusted R² 99.96%
Residual Sum of Squares 0.099

Source: Own calculation (2019)

Table IV shows the REM has a slightly greater RSS when compares to the FEM. In addition, the model itself shows all of the estimates have a significant value. To compete this model and the FEM, the Hausman test is applied. Suppose \( \omega_{it} = \epsilon_i + \mu_{it} \), then the hypotheses for the Hausman test can be stated as:

- \( H_0: E(\omega_{it}|X_{it}) = 0; \) (REM is consistent)
- \( H_1: E(\omega_{it}|X_{it}) \neq 0; \) (REM is inconsistent)

The \( \chi^2 \) statistic from the Hausman test as much as 3.890 with \( p-value = 0.274 \). The conclusion is to accept \( H_0 \) or the REM is consistent to model the data. From the panel data model, we conclude the best model is the REM with its estimates has been listed on Table IV. Despite of this conclusion, considering the aim of this work is to model the spatial effect in the data matrix with the result as follows:

**In order to study the spatial effects might exist between regencies/cities in Bali and to evaluate the spatial dependences may exist, then we continue to use FEM in subsequent analysis.**

**B. Results for Spatial Panel Data Models**

The first step in modeling spatially panel data is to build spatial weight matrix \( (W) \). This matrix quantifies the connections between regencies/cities in Bali with its elements could be determined by noting the Bali’s map (Fig. 2).

The \( W_{9x9} \) matrix is built by normalizing the \( i^{th} \) row of the binary matrix \( P_{9x9} \) where \( p_{ij} = 1 \) if the edge of region \( i^{th} \) shares a common edge with region \( j^{th} \). Refers to Fig. 2, the \( P \) and \( W \) are:

\[
P = \begin{pmatrix}
JBR & TBN & EDG & GNR & KLK & BLI & KRG & BLL & DPS \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

\[
W = \begin{pmatrix}
JBR & TBN & EDG & GNR & KLK & BLI & KRG & BLL & DPS \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

In order to study the spatial effects might exist between regencies, we have to test the significance of \( \rho \) coefficient in equation (8) as well the significance of \( \gamma \) in equation (9). The Langrange Multiplier (LM) test can be applied for this task, with the tested hypotheses are [11]:

- \( H_0: \rho = 0; \) (There is no spatial dependence)
- \( H_1: \rho \neq 0; \) (There is a spatial dependence)

These hypotheses were tested by using slxtest function in R splm package [12]. The test result demonstrated the LM Statistic = 28.621 with \( p-value < 0.000 \). We concluded the rejection of the \( H_0 \) and the fixed (individual) spatial autocorrelation (lag) model (FEM₁SAR) is worth to build. In addition, we also tested the \( \gamma \) coefficient in equation (9) to study the existence of error dependencies amongst regencies. The test hypotheses are:

- \( H_0: \gamma = 0; \) (There is no error dependence)
- \( H_1: \gamma \neq 0; \) (There is an error dependence)

From this test we got the LM Statistic = 2.068 with \( p-value = 0.150 \). We must accept the \( H_0 \) hypothesis suggested there is no error dependence amongst observations. Finally, we decided the FEM₁SAR is the best model to reveal the spatial effect in the data matrix with the result as follows:

**TABLE V. The Estimates of FEM₁SAR**

| Coefficient | Estimator | T-Value | p-Value |
|-------------|-----------|---------|---------|
| Intercept   | μ         | -13.044 | -9.853  | 0.000   |
| SAR Coefficient | ρ   | -0.153  | -6.922  | 0.000   |
| HEA         | β₁        | 53.555  | 16.273  | 0.000   |
| EDU         | β₂        | 41.702  | 73.608  | 0.000   |
| LIV         | β₃        | 36.308  | 19.966  | 0.000   |

Adjusted R² 99.99%
Residual Sum of Squares 0.053

Source: Own calculation (2019)

**C. Discussion**

By utilizing the HDIs data from 9 regencies/city in Bali in the period 2010–2017, the spatially regression analysis shows the best model to represents the spatial dependences between neighborhood regions is the fixed (individual) autocorrelation model (FEM₁SAR). Mathematically, this model could be expressed as following equation:

\[
Y_t = \mu + \epsilon_t = \rho W Y_t + X_{it} \beta + \mu + \epsilon_t
\]

\[
Y_t = -0.153 \cdot W Y_t + (\text{HEA}, \text{EDU}, \text{LIV})\begin{pmatrix}
53.555 \\
41.702 \\
36.308
\end{pmatrix} + \mu + \epsilon_t
\]
Referring LeSage [11], the product of $W$ and vector $Y_t$ or $WY_t$ known as a spatial lag and represents an average HDI’s of regions at time $t$ defined as neighbors by the matrix $W$. For example, for the year 2017 the spatial lag for 9 regencies/cities in Bali is calculated as follows:

$$W = \begin{bmatrix}
0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.3 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.3 & 0.3 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.3 & 0.3 & 0.3 & 0.5 & 0 & 0 & 0 & 0 & 0 \\
0.3 & 0.3 & 0.3 & 0.3 & 0.5 & 0 & 0 & 0 & 0 \\
0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.5 & 0 & 0 & 0 \\
0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.5 & 0 & 0 \\
0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.5 & 0 \\
0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.3 & 0.5
\end{bmatrix}$$

The readers are suggested to read [13] in order to understand the characteristic of spatial autocorrelation.

From Table V we get three predictors of HDI have significantly estimator. The greatest coefficient is found for the health index (HEA) and the lowest is found for the standard of living (LIV) with the values as much 53.555 and 36.308, respectively. All of the estimators have positive sign, demonstrate the positive effects. The increase of EDU, HEA, and LIV will cause the increase of HDI value for each regions in Bali.

In addition, the residual sum of squares (RSS) of the final model which is equal to 0.053 is lower than RSS of the FEM$_I$ as much as 0.087 (see Table II). This finding proved the FEM$_{LSAR}$ capable to increase the model accuracy or to reduce the error as much as $\frac{0.053-0.087}{0.087} \times 100\% \approx 39.08\%$.

V. Conclusion

This work with its aim is to model the HDIs of 9 regencies/cities in Bali Province of Indonesia as the response and three indices namely the health index, the education index, and the standard of living index as the regressors, spatially, concludes:

1) The three regressors suggest the possibility of autocorrelation and/or spatial autocorrelation that makes the pool model is not appropriate to apply;
2) The fixed individual effect (FEM$_I$) demonstrated the better performance when it compared to the fixed time effect (FEM$_T$). Despite of both models are significant, the RSS of FEM$_I$ is smaller than the RSS of FEM$_T$;
3) The LM test proved there is a spatial dependence amongst neighboring regencies/cities. Considering this dependence, the best model for this work is the spatial autocorrelation with fixed (individual) effect or FEM$_{LSAR}$. 

If we made a comparison between the spatial lag and the real HDIs $\approx (79.99 \cdots 78.32)^T$ and $(70.72 \cdots 83.01)^T$, we find 4 regencies namely JBR, BLI, KRG, and BLL have a positive difference while the others have a negative one. The positive spatial lag indicates the regency takes an 'advantage' from its neighbors in forming its HDI whilst the negative lag indicates the regency 'supports' its neighbors.

According Griffith and Arbia [13], the spatial autocorrelation ($\rho$) has been an interest topic in examining spatial data. They defined $\rho$ as "the tendency for nearby values on a map to be dependent". Both authors wrote a positive $\rho$ has received most of the attention of researchers mostly because rare example of a negative $\rho$. Our study shows the $\rho$ of the final model equals to 0.153 with negative sign. The sign indicates that a tendency of the HDIs of 9 regencies/cities in Bali are diverge based on its macroeconomic predictors such as the educational index (EDU), the health index (HEA), as well as the standard of living index (LIV). The readers are suggested
REFERENCES

[1] UNDP, “Human Development Indices and Indicators: 2018 Statistical Update,” United Nations Development Programme, Report, 2018.
[2] BPS–Statistics Indonesia, Statistical Yearbook of Indonesia 2018, Jakarta: BPS–Statistics Indonesia, 2018, no. ISSN: 0126–2912.
[3] BPS–Provinsi Bali, Indeks Pembangunan Manusia Bali Tahun 2017, Denpasar: BPS–Provinsi Bali, 2018.
[4] K. Budinirmala, P. Suciptawati, K. Jayanegara, and E. N. Kencana, “Memodelkan Kemiskinan Penduduk Provinsi Bali Dengan Regresi Data Panel,” E-Jurnal Matematika, vol. 7, no. 3, pp. 219–225, 2018.
[5] N. Trianggara, R. Rahmawati, and H. Yasin, “Pemodelan Indeks Pembangunan Manusia Menggunakan Spatial Panel Fixed Effect (Studi Kasus: Indeks Pembangunan Manusia Propinsi Jawa Tengah Periode 2008-2013),” Jurnal Gaussian, vol. 5, no. 1, pp. 173–182, 2016.
[6] W. H. Green, Econometric Analysis, 7th ed. New York: Prentice Hall, 2012.
[7] D. N. Gujarati and D. C. Porter, Basics Econometrics, 5th ed. New York: The McGraw-Hill/Irwin, 2009.
[8] J. P. Elhorst, Spatial Panel Data Models, ser. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. New York: Springer, 2010, ch. 2, pp. 377–407.
[9] ——, “Specification and Estimation of Spatial Panel Data Models,” International Regional Science Review, vol. 26, no. 3, pp. 244–268, July 2003.
[10] Y. Croissant and G. Millo, “Panel Data Econometrics in R: The plm Package,” Journal of Statistical Software, vol. 27, no. 2, pp. 1–43, 2008.
[11] J. P. LeSage, “An Introduction to Spatial Econometrics,” Revue d’économie industrielle, vol. 123, pp. 19–44, September 2008.
[12] G. Millo and G. Piras, “splm: Spatial Panel Data Models in R,” Journal of Statistical Software, vol. 47, no. 1, pp. 1–38, April 2012.
[13] D. A. Griffith and G. Arbia, “Detecting negative spatial autocorrelation in georeferenced random variables,” International Journal of Geographical Information Science, vol. 24, no. 3, pp. 417–437, March 2010.