IJCSN - International Journal of Computer Science and Network, Volume 7, Issue 1, February 2018
ISSN (Online) : 2277-5420
www.IJCSN.org

Applied Spatial Durbin Model to Assess Factors Influencing the Number of Foreign Tourists in Indonesia 2011-2015

[1] Dini Arifatin; [2]Anik Djuraidah; [3]Utami Dyah Syafitri

[1] Department of Statistics, Bogor Agricultural University, IPB
Bogor, 16680, Indonesia

[2] Department of Statistics, Bogor Agricultural University, IPB
Bogor, 16680, Indonesia

[3] Department of Statistics, Bogor Agricultural University, IPB
Bogor, 16680, Indonesia

Abstract. Most of the study of the number of foreign tourists in Indonesia base on their origin country, have only focused on panel data analysis. It could be misleading because there may be a spatial influence among the original country. This study analyzed some factors that could affect the number of tourists who came to Indonesia for 2011-2015. The Spatial Durbin Model (SDM) was used in this study. Two types of proximity matrices were used i.e. contiguity matrix and inverse distance matrix. The results showed that the best model was the SDM model with contiguity matrix. The factors which significantly affected the number of tourists were GDP per capita, number of population, and existing the direct flight to Indonesia.

Keywords – *spatial durbin model, spatial panel data, Indonesia, contiguity matrix, foreign tourists, tourism*

1. Introduction

Tourism sector plays an important role in the Indonesian economy for supporting Gross Domestic Product (GDP) and job creation, and also as a source of foreign exchange earnings. By 2015 the tourism sector contributes 10% of national GDP as well as contributes the fourth largest national foreign exchange by business field of US $ 12.6 billion after oil and gas commodities, coal and palm oil [1]. The performance of the tourism sector as a source of foreign exchange earnings is determined by the ability of the government to bring as many foreign tourists (tourists) to Indonesia. The trend of foreign tourists visiting Indonesia by 34 origin countries from 2011-2015 is presented in Figure 1.

The past years study of foreign tourists in Indonesia by country of origin still developing in panel data study. [2] The researcher research using cross-sectional panel data states that the exchange rate of rupiah and geographical distance between Indonesia and the country of origin of foreign tourists significantly influence the number of tourist arrivals to Indonesia. [3] The researcher concluded that the number of visits of foreign tourists in Indonesia by country of origin is influenced by various factors namely GDP per capita, distance of country of origin with Indonesia, relative prices and security dummy variables.

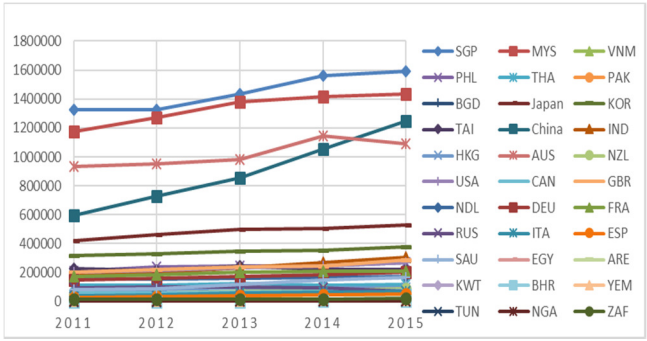

Figure. 1 The number of foreign tourists visits Indonesia by country of origin in 2011-2015

In fact, there was a spatial influence (location) that affects the number of foreign tourists visiting Indonesia. Spatial information is very important because it give an information about the relationship among area. Data

containing spatial elements will not be accurate if only using panel regressions without spatial insertion because it will produce heterogeneous errors due to inter-regional linkages. Cross sectional data between areas and time should be accommodate the spatial variation because of inter-regional linkages.

There are several spatial data panel models [4], namely Spatial Autoregressive Models (SAR), Spatial Error Models (SEM) and Spatial Durbin Model (SDM). SAR is a spatial regression model that has spatial dependence on the response variable. The SEM is a spatial regression model that has spatial dependence on the error, whereas the SDM model has spatial dependence both on the response and explanatory variables. This study focused on The SDM model because of the auto spatial in the response and the explanatory variables.

This study aims to determine the factors that affected the number of visits of foreign tourists in Indonesia from 34 countries of origin of foreign tourists in 2011-2015 using the SDM model. The weighted matrices were used contiguity matrix and inverse distance matrix. This research was expected to provide benefits for the government as a material information and input for policy formulation in the development of tourism sector in Indonesia.

2. Materials and Methods

2.1 Data

The data was used in this study was took from several sources of publication with the response variable used is the number of foreign tourists visiting in Indonesia by country of origin 2011-2015 from BPS-National Statistics Office of Indonesia. The study considered explanatory variables from the origin country factor, the exchange rate of the rupiah with the origin country of foreign tourists, GDP per capita, population, consumer price index (CPI), total imports, visa-free factor and direct flight factor. The explanatory variables used in this study are presented in Table 1.

The scope of data is 34 countries from seven regions. ASEAN region include Singapore, Malaysia, Vietnam, Philippines and Thailand. The Asian region includes the countries of Pakistan, Bangladesh, Japan, South Korea, Taiwan, China, India, and Hong Kong. The Oceania region includes Australia and New Zealand. The Americas include the United States and Canada. The European Region includes the UK, the Netherlands, Germany, France, Russia, Italy and Spain. The Middle East region includes the countries of Saudi Arabia, Egypt, United Arab Emirates, Kuwait and Bahrain. The African region includes the countries of Yemen, Tunisia, Morocco, Nigeria and South Africa.

Table 1: Description of explanatory variables in this study

Variable	*Unit*	**Abbreviations**
Rupiah exchange rate	IDR	Kurs
GDP per capita	Thousand USD	GDP
Population	Million	Pop
Consumer Price Index	index	Cpi
Import total	Billion USD	Import
Visa free factor	1= Yes; 0= No	BVKS
Direct flight factor	1= Yes; 0= No	Direct

2.2 Statistical Methods

We compared the SDM panel with contiguity weight matrix based on region and inverse distance. To evaluate the suitable SDM method we conducted the step of analysis describe as follows:

1. Explored on response and explanatory variables
2. Perform regression analysis of panel data:
   a. Multicollinearity test
   b. Checking the effect of time and location with Langrange Multiplier (LM) test
   c. Selection of panel data model used with Chow test and Hausman test
   d. Testing error assumption on the model
3. Performed regression analysis of spatial panel data
   a. Spatial contiguity weight matrix with size 34×34. Spatial contiguity weight matrix based on region ($\mathbf{W_1}$) and based on inverse distance ($\mathbf{W_2}$). Based on region divided with seven region Asean, Asia (exclude Asean), America, Oceania, Middle East and Africa will have value 1 if the country in the same region.
   $\mathbf{W_2}$ matrix will have the value based on the row and column of the distance between international airport in two countries.

$$w_{ij} = \begin{cases} \frac{c_{ij}}{\sum_{j=1}^{n} c_{ij}} & , \text{for } i \neq j \\ 0 & , \text{for } i = j \end{cases} \quad (1)$$

   b. Spatial dependency test

Spatial dependency test using Moran Index. The Moran Index is an indicator of spatial dependence to compare the value of a variable at a location with the same variable value at another location. Tests with Moran index are performed on every variable in every year using contiguity matrix.

Hypothesis used are:
H_0: $I = 0$ (no spatial autocorrelation)
H_1: $I > 0$ (there is positive spatial autocorrelation)
$I<0$ (there is negative spatial autocorrelation)

The formula of the Moran Index is:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(X_i-\bar{X})(X_j-\bar{X})}{\sum_{i=1}^{n}(X_i-\bar{X})} \quad (2)$$

The Moran index value is 1 to -1. Test statistics are used as follows:

$$Z(I) = \frac{I-E(I)}{S_{E(I)}} \quad (3)$$

$$S_{E(I)} = SQRT\left[\frac{n^2\sum_{ij} w_{ij}^2+3(w_{ij})^2-n\sum_i(\sum_j w_{ij})^2}{(n^2-1)(\sum_{ij} w_{ij})^2}\right] \quad (4)$$

c. Determine the spatial influences using Langrange Multiplier test for spatial panel

4. Testing and estimating regression parameters on the spatial model of the number of foreign tourists visiting Indonesia in 2011-2015 using spatial contiguity weight matrix based on the region ($\mathbf{W_1}$) and inverse distance ($\mathbf{W_2}$).
Spasial Durbin Model Panel Data Model:

$$y_{it} = \rho\sum_{j=1}^{n} w_{ij}y_{ij} + \mathbf{x}_{it}\boldsymbol{\beta} + \sum_{k=1}^{K}\sum_{j=1}^{n} w_{ij}x_{jtk}\boldsymbol{\theta_k} + \mu_i + \varepsilon_{it} \quad (5)$$

Parameter estimation with maximum likelihood function on SDM fixed effect:

$$LogL = -\frac{nT}{2}\log(2\pi\sigma^2) + T\log|I_n-\rho W| - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it}-\rho\sum_{j=1}^{n} w_{ij}y_{ij} - x_{it}\beta - \sum_{k=1}^{K}\sum_{j=1}^{n}\mathbf{w_{ij}x_{jtk}}\boldsymbol{\theta_k} - \mu_i\right)^2 \quad (6)$$

5. Select the best model by comparing Akaike's Information Criterion (AIC), adj-R^2 and Root Mean Squared Error (RMSE) of each model and checking the error assumption.

$$AIC = -2\ln(L) + 2K \quad (7)$$

where:
K: the number of parameters in the model
L: maximum likelihood function of the model

formula of R^2 and adj-R^2 for the spatial panel model are as follows:

$$R^2 = 1 - \frac{\tilde{e}^T\tilde{e}}{(Y-\bar{Y})^T(Y-\bar{Y})} \quad (8)$$

$$\text{adj-}R^2 = 1-(1-R^2)\frac{N-1}{df_{error}} \quad (9)$$

Where $\tilde{e}^T\tilde{e}$ is the sum square of error squares (SSE) model. The RMSE value is obtained through the following formula:

$$RMSE = \left(\frac{\tilde{e}^T\tilde{e}}{df_{error}}\right)^{1/2} \quad (10)$$

Where the df_{error} obtained from N-K-n. While N is the sum of all cases (rows) as number of 34×5 cases and n is the number of regions or locations.

6. Interpreted the model

3. Result and Discussion

3.1 Data exploration

The data described in this study was used 2015, the last year data used in this study. The map of the number of foreign tourist arrivals in Indonesia by 2015 by origin country of foreign tourists is shown in Figure 2. The number of foreign tourists by country mostly come from Africa and Middle East region was the lowest number arrivals. While the number of foreign tourists came from the ASEAN region and Asia was the higher. The number of arrivals described in a map on Figure 2 with categorized its numbers using the quartile.

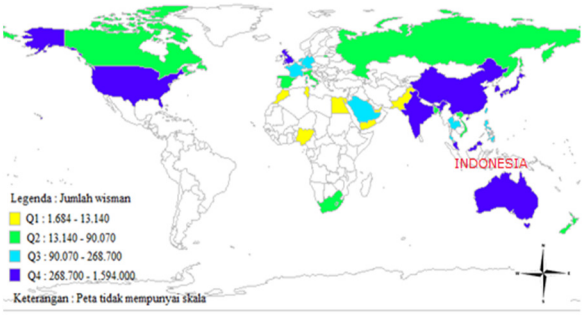

Figure.. 2 Map of the number of visits of foreign tourists in Indonesia by country in 2015

The rupiah exchange rate in 2015 highest was with the Kuwait currency and the lowest was with Vietnam. The GDP variable highest was in Australia and the lowest in Yemen. The population highest was China and the lowest was Singapore. The CPI variable highest was in Nigeria and the lowest was Japan. The import highest was the United States and the lowest was Yemen.

BVKS variable is the presence or absence of short visa free visa facility. All countries on this study almost got free visa in 2015 except countries of Pakistan, Bangladesh, and Nigeria. Direct variable is the presence or absence of direct flights from the country foreign tourist to Indonesia. All countries in the ASEAN region had direct flights to Indonesia,in 2015, while in Asia there were only several countries had no direct flight to Indonesia, e.g. Pakistan, Bangladesh, and India. Direct flights from the Middle East region were only available in Saudi Arabia and United Arab Emirates, while all countries in Europe, America and Africa did not have direct flights to Indonesia.

3.2 Regression Analysis of Panel Data

The panel data regression model was used to identify the explanatory variables that affect the number of foreign tourists visiting Indonesia. The next step was the multicollinearity test, it used to determine the correlation between the explanatory variables used by calculating the VIF (Variance Inflation Factor) value. VIF values of the seven explanatory variables were less than 5, it was concluded that there was no multicollinearity in all explanatory variables used in this study.

The LM test was performed to determine the effect of time, location or both. Based on the test on this study it was concluded that only the effect of the location has significantly affect the data on the number of visits of foreign tourists. Time effects based on the LM test was not have significant effect.

The Chow test was performed in the selection of pooled or fixed influence models. In the test obtained F_{cal} = 179.09 with p-value = 0.000, so for a while it was concluded that the model in this study was a fixed influence model. The Hausman test was performed to determine between a random model or a fixed model. The test obtained Chi-$Square_{cal}$ = 48.867 with p-value = 0.000, which stated that the appropriate model in this study was a fixed influence model. Based on LM test, Chow test and Hausman test this research used fixed panel model with location effect.

Based on the result of the fixed panel model with the location effect indicated that only GDP and population variables significantly affect the number of foreign tourists visiting Indonesia by country of origin at 5 percent significance level. The value of R^2 = 0.306 and adj-R^2 = 0.091. The result of normality test of residual with Kolmogorov Smirnov test (KS = 0.5353, p-value = 0.000) stated that the residual did not followed the normal distribution. To overcome the nonnormality of residual this study used transformation with natural logarithm function (ln) at response variable.

The result of parameters with transformation of response variable showed that there were two explanatory variables that significantly affect with 5 percent significance level, the rupiah exchange rate and CPI, with the value of R^2 = 0.4942 and adj-R^2 = 0.3374. The result of normality test of residual with KS test (KS =0.363, p-value 0.000) stated that the residual did not followed the normal distribution. The homogeneity test based on the Bartlett test (Bartlett = 151.52, p-value = 0.000) indicated that the residual variation was not homogeneous, so the assumption of homogeneity prohibited. Test of autocorrelation used the plot between the residuals and the countries. The plot showed that it had a pattern by country. From that can be concluded that the assumption of autocorrelation on the residual not fulfilled. To overcome the three unfulfilled assumptions on panel regression was used by inserting spatial elements in the formation of the model.

3.3 Spatial Dependency Test

Spatial dependence on response variable and explanatory variables were tested by the Moran Index in each year. Table 2 showed the results of Moran test on response variable and explanatory variables were significant except population for $\mathbf{W}_2$ and CPI variables for $\mathbf{W}_1$ in 2011-2015.

Table 2 Moran Index of the variables

Variables	$\mathbf{W}_1$ Spatial Weight Matrix				
	2011	2012	2013	2014	2015
Tourist	0.52**	0.50**	0.47**	0.46**	0.45**
Kurs.	0.28**	0.26**	0.26**	0.23**	0.21**
GDP	0.36**	0.25**	0.25**	0.36**	0.36**
Pop	0.12*	0.12*	0.12*	0.12*	0.12*
Import	0.22**	0.20**	0.21**	0.21**	0.18**
Cpi	-0.10	-0.08	-0.05	-0.02	-0.02

Table 2 Moran Index of the variables (continuous)

Variables	W_2 Spatial Weight Matrix				
	2011	2012	2013	2014	2015
Tourist	0.21**	0.21**	0.20**	0.19**	0.20**
Kurs.	0.17**	0.16**	0.14**	0.13**	0.12**
GDP	0.08**	0.05**	0.05*	0.08**	0.08**
Pop	-0.01	-0.01	-0.01	-0.01	-0.01
Import	0.07**	0.06**	0.05**	0.05**	0.05**
Cpi	0.02	0.03*	0.05**	0.07**	0.08**

The next test was Langrange Multiplier (LM) test for selecting spatial model. The results of the LM test indicated that there was a spatial lag effect and the effect of spatial error at 5% significance level using either of a regional or distance inverse weighting matrix. Therefore, further predictions of fixed influence models with SDM will be conducted because of the spatial dependence on the response variables and explanatory variables.

3.3 Parameter Estimation

The parameter estimation was using SDM with fixed effect, using spatial weighted matrix based on neighboring (region) and distance inverse matrix. Based on adj- R^2, RMSE, and AIC, the best model was a fixed impact panel SDM model with a spatial weighted matrix W_1. The model formed was as follows:

$$\begin{aligned}\ln(y_{it}) = \; & \mu_i + 0.13\sum_{ij}^{n} \mathrm{w}_{ij}\mathrm{y}_{\mathrm{it}} \\ & -(1.03e-05)Kurs_{it} \\ & +(3.58e-02)GDP_{it} + (1.61e-03)Pop_{it} \\ & +(2.83e-04)Impor_{it} - (1.71e-02)Cpi_{it} \\ & +(8.35e-01)BVKS_{it} + (1.50)Direct_{it} \\ & +(2.70e-05)\sum_{ij}^{n} \mathrm{w}_{ij}\mathrm{Kurs}_{\mathrm{it}} \\ & -(4.84e-02)\sum_{ij}^{n} \mathrm{w}_{ij}GDP_{\mathrm{it}} \\ & +(8.49e-04)\sum_{ij}^{n} \mathrm{w}_{ij}\mathrm{Pop}_{\mathrm{it}} \\ & +(2.69e-04)\sum_{ij}^{n} \mathrm{w}_{ij}Import_{\mathrm{it}} \\ & +(1.83e-02)\sum_{ij}^{n} \mathrm{w}_{ij}\mathrm{Cpi}_{\mathrm{it}} \end{aligned} \quad (8)$$

According to LeSage and Pace (2006), to determine the effect of change of explanatory variables on response variable on spatial regression was measured by direct effect, indirect effect and total effect [5]. The effects on the model are shown in Table 3.

Factors influencing the number of foreign tourists visit to Indonesia which positively and significantly at the 5% significance level were GDP per capita, population, BVKS and direct flight factors. In total, the effect of population, GDP per capita, BVKS and direct flight factor were positive and significant to the number of foreign tourists in Indonesia. While the rupiah exchange rate factor with the country of origin of foreign tourists, and total imports were not significantly affect the number of visits of foreign tourists.

Table 3 The fixed effect HR model with the W_1 spatial weighted matrix

Effect	Variables	Coeff.	Effect Value	p-value
Direct	Kurs.	-1.03e-05	-0.563	0.573
	GDP	3.60e-02	4.474	0.000**
	Pop	1.62e-03	3.763	0.000**
	Import	2.84e-04	1.010	0.312
	Cpi	-1.72e-02	-1.896	0.058*
	BVKS factor	8.39e-01	3.847	0.000**
	Direct factor	1.50e+00	5.534	0.000**
Indirect	Kurs.	-1.47e-06	-0.468	0.640
	GDP	5.14e-03	1.867	0.062*
	Pop	2.31e-04	1.748	0.081*
	Import	4.06e-05	0.840	0.401
	Cpi	-2.45e-03	-1.223	0.221
	BVKS (factor)	1.20e-01	1.628	0.104
	Direct (factor)	2.15e-01	1.749	0.080*
Total	Kurs.	-1.18e-05	-0.558	0.577
	GDP	4.11e-02	4.463	0.000**
	Pop	1.85e-03	3.713	0.000**
	Import	3.25e-04	1.009	0.313
	Cpi	-1.96e-02	-1.877	0.061*
	BVKS (factor)	9.59e-01	3.695	0.000**
	Direct (factor)	1.72e+00	5.018	0.000**

Existence assumption examination was performed on the best model was fixed effect panel HR model with spatial weighted matrix W_1. The assumption of normality used Kolmogorov Smirnov (KS) test obtained value of KS 0.05 with p-value 0.78. Based on the test it can be concluded that the residual was spread normally. The homogeneous homogeneity examination used Bartlett test. Based on Bartlett test, the value of statistic was 37.28 with p-value 0.28. This value indicates that the homogeneous variation met the assumptions of homogeneity. The autocorrelation examination used plot of residuals and the countries. The residuals and the countries plot showed that the pattern was random. This indicates that the assumption of autocorrelation was fulfilled.

4. Conclusion

Data modeling of foreign tourists in Indonesia by country 2011-2015 with SDM fixed effect using a region based spatial weighting matrix was the best model. The factors that influenced the SDM model directly and indirectly

were GDP per capita, population, BVKS factor and direct flight factor.

References

[1] Ministry of tourism affair, "Ministry of Tourism Affair Presentation: Promotion of Tourism Potential for Increasing Indonesia Economy", https://diplomasiekonomi.kemlu.go.id/images/capbuildiplomat/Paparan%20KEMPAR.pdf, 7 February 2017.

[2] A. Putri, "Factors Affecting Tourism in Indonesia Data Panel Analysis from 21 Countries in 2006-2012", thesis, Gadjah Mada University, Jogjakarta, Indonesia, 2014.

[3] S. R. U. Hidayanti, "Analysis of factors affecting the number of tourist visiting Indonesia", thesis, Bogor Agricultural Institute, Bogor, Indonesia, 2016.

[4] J. P. Elhorst, Spatial Panel Data Models, in Fischer M, Getis A (eds) Handbook of Applied Spatial Analysis, Berlin: Springer, 2010.

[5] J. LeSage, and R. K. Pace, Introduction to Spatial Econometrics, Boca Raton: CRC Press Taylor & Francis Group, 2009.

[6] E. D. Fernando, "Identifying and Predicting Turning Points in Australian Inbound Tourism Demand Growth", thesis, School of Economics and Finance, Victoria University Melbourne, Australia, 2010.

[7] B. H. Baltagi, Econometrics Analysis of Panel Data. 3rd edition, West Sussex England: John Wiley and Sons, 2005.

[8] A. South, "rworldmap: A New R package for Mapping Global Data", The R Journal Vol. 3/1, June 2011, pp.35-43.

Dini Arifatin master student in Department of Statistics, Bogor Agricultural University, Indonesia. Her main interest is on spatial analysis.

Dr. A. Djuraidah Currently worked as a lecture in Department of Statistics, Bogor Agricultural University. Her main interests are on spatial analysis and statistical probability theory.

Dr. Utami Dyah Syafitri Currently worked as a lecture in Department of Statistics, Bogor Agricultural University. Her main interest is in Experimental Design, Chemometric and Metabolomics.