International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

APPLICATION OF SPATIAL REGRESSION ANALYSIS ON HUMAN DEVELOPMENT INDEX SUB-NATIONAL DATA

DEL ROSARIO KATRINA S*, ESTOLANO MARCIAL P[†] AND OLIVO JOHN CLOSTER F

Department of Statistics, Central Luzon State University, Science City of Munoz, Nueva Ecija, Philippines

*Corresponding Author: E Mail: katrinadelrosario2@gmail.com

ABSTRACT

Human Development Index (HDI) is a measure of development that enables international comparison. Through the years, the index was developed until the present measure captures the interest on health, education and the standard of living that provides a more holistic view of the true concept of development. With the continuous strive in improving human lives, it is important to have a closer look on the human development in the sub-national level of a country though complexity increases due to the presence of geographical information. Considering the significance and the difficulty of determining more specific factors of HDI, this study used spatial regression analysis to identify indicators of HDI in the sub-national unit of the Philippines.

As confirmed by Moran's I, spatial dependence exists among the sub-national units in the estimation of HDI and suggested the use of spatial regression. The spatial error regression model was used to determine significant indicators while taking into account the effects of neighboring provinces through their error terms. Results revealed that the key to human development is to be focused in the immunization of children, survival rate in public elementary school children and poverty incidence. It is notable that the explanatory variables capture the dimensions of HDI such as education, health and standard of living.

Keywords: Human development index, spatial regression, sub-national, dimension

INTRODUCTION

The United Nation's Human Development Report in 1990 first introduced the concept of Human Development Index (HDI). In the early years of this index, the concept of development had been mostly focused on the income per capita as a primary measure. Through the years, the index was developed until the present measure captures the interest on health, education and the standard of living that provides a more holistic view of the real concept of development.

Human development as defined by the United Nation's Development Programme (UNDP) is about enlarging human choices. It focuses on the quality of the human lives rather than the wealth of the economy. The longevity of life, education, and quality of living had become the measure to determine the development of human lives.

The HDI represents the summary of human development. Since its introduction in 1990, the measure of the HDI had undergone several criticisms. However, the current measure of this index focuses on three basic dimensions of human development: to live a long and healthy life, the ability to acquire knowledge and the ability to achieve a decent standard of living. The measures of these three dimensions are longevity, knowledge and standard living.

The longevity dimension refers to the long and healthy life that a person lives. It measures the extent to which a newborn infant is expected to live from the prevailing pattern of the age-specific mortality rate. For the knowledge dimension, it is based on the mean years of schooling which refers to the average number of years that adults of age 25 and above had received. The second measure relates to the number of years of education that the school-age children are expected to receive based on an existing pattern of schooling years, and the income per capita defines the last dimension on the standard of living.

The computation of the dimensions of HDI of the provinces is given by:

a. Life expectancy at birth index (I_{life})

$$I_{life} = \frac{(H - H_{\min})}{(H_{max} - H_{\min})} \qquad (1)$$

where, H is the life expectancy at birth (in years) of each province
H_{max} is the maximum year (85 years)
H_{min} is the minimum year (20 years)

The H_{max} and H_{min} values are defined by the United Nations Development Programme (UNDP) and Human Development Network (HDN).

b. Education index ($I_{education}$)

$I_{education}$

$$= \frac{(E_1 \times E_2)^{1/2} - (E_1 \times E_2)^{1/2}{}_{min}}{(E_1 \times E_2)^{1/2}{}_{max} - (E_1 \times E_2)^{1/2}{}_{min}} \quad (2)$$

where E_1 is the mean years of schooling sub-index:

$$E_1 = \frac{(MeanYrs - MeanYrs_{min})}{(MeanYrs_{max} - MeanYrs_{min})} \quad (3)$$

and E_2 is the expected years of schooling sub-index:

$$E_2 = \frac{(ExpYrs - ExpYrs_{min})}{(ExpYrs_{max} - ExpYrs_{min})} \quad (4)$$

The maximum mean and expected years of schooling are based on the actual maximums while the minimum value is set to zero. For the education index, the maximum geometric mean of the sub-indices is equal to the highest geometric mean observed among provinces, and the minimum is equal to zero.

c. Income Index (I_{ncome})

$$I_{income} = \frac{(Y - Y_{min})}{(Y_{max} - Y_{min})} \quad (5)$$

The maximum and minimum values for the income are the highest and lowest values of real income per capita attained by the provinces for a particular reference year. The minimum is set to 90 percent of the actual minimum value observed while the maximum is set at 100 percent of the actual maximum value observed.

Given the three dimension indices, the Human Development Index per province is calculated by taking the geometric mean of the three, i.e.:

$$HDI = (I_{life} \times I_{education} \times I_{income})^{1/3} \quad (6)$$

The value of the Human Development Index ranges from zero to one where higher value refers to a more developed unit.

There had been several studies conducted to determine more bases for the measure of HDI of a country [1]. However, the difficulty in explaining these determinants increases as the study focuses on a sub-national level. The increasing complexity is due to the presence of geographical information affixed in units. This provokes the interest to take into account the effect of geography or location in the study since location had been considered a factor that can usually be applied as a proxy for the interaction between genetic factors, lifestyle, and environment [2]. Today, the advancement in technology provides the researchers an aid to explore the data and study geographic (spatial) pattern and thus gave birth to the interest on spatial data analysis applies explicitly to Human Development Index Sub-National Data.

The existence of spatial autocorrelation between spatial units may cause non-constancy of error variance. The spatial regression that accounts spatial dependence may correct both violations but not in all cases. This study aims to account for the spatial dependence of the sub-national units while satisfying regression conditions to provide a good predictive model for HDI.

As a measure of national development, a country prefers a higher HDI. Accordingly, it is significant to determine more specific factors for the measure of HDI. Focusing on the particular measure to improve the quality of living in sub-national level further promotes national development.

Thus, this study aimed to apply spatial regression analysis onhuman development index in the Philippines subnational data. Specifically, it aimed to perform Exploratory Spatial Data Analysis (ESDA) and spatial regression analysis using Spatial Lag Model (SLM) and Spatial Error Model (SEM).

METHODOLOGY

Source of Data

The data used in this study was from the Philippine Statistics Authority (PSA). The sub-national HDI of 2012, recently released in January 2016, is the response variable and the explanatory variables were obtained from the Philippine Countryside in Figures publication. While the present number of provinces in the country is 81, this study used only 80 units since the province of Shariff Kabunsuan was formerly part of the province of Maguindanao in 2012.

There were sixteen predictor variables initially considered in this study. These were barangay health station ratio, classroom-pupil ratio, Consumer Price Index (CPI), crime rate, crime solution efficiency, employment rate, fully immunized children, health workers, number of immunized pregnant women, Internal Revenue Allotment (IRA), labor force participation rate, percentage IRA to total financial resource, poverty incidence, survival elementary ratio and teacher-pupil ratio.

Software Used

GeoDa was used to create spatial weights and perform spatial regression.It can perform some basic GIS operations and sophisticated spatial statistical models. The shape file format of the Philippine administrative map with provincial boundary used for the analysis was downloaded from the website of PhilGIS.

Method of Analysis

The sequence of the analysis will be as follow:

1. Create a database file to input in GeoDa using MS Excel 2010 (.*xls extension*).

2. Input the shapefile format of the Philippine map with provincial boundaries and the created database on Geoda. Generate spatial weight matrices, and the default threshold distance of 3.633 and nearest four neighbors was used as criteria for the distance–band and k-nearest neighbor weight matrices, respectively.

3. Choose an appropriate spatial weight matrix that will best capture the existence of spatial autocorrelation.

4. If spatial autocorrelation exists, perform Exploratory Spatial Data Analysis to assess the spatial behavior of the response variable.

5. Fit the SLM and SEM incorporating the spatial weight matrices with significant autocorrelation.

6. Perform diagnostic check on the spatial models for the heteroskedasticity and spatial dependence.

7. Compare the models and interpret the human development index using the chosen model.

RESULTS AND DISCUSSIONS

Spatial Weight Matrix

The Rook and Queen Contiguity weight matrices conclude similar neighbors which reflected on the similarity of values. Based on the significance of the Moran's I in each weight matrix, the distance-band was chosen (p<0.01). This spatial weight matrix was then used for the spatial data analysis.

Spatial Dependence Diagnostic

As the Moran's I statistic tests the spatial autocorrelation in general, the Lagrange Multiplier (LM) tests are more specific. The LM (lag) and LM (error) provide bases in choosing the more

appropriate spatial model. The former refers to the use of the spatial lag model while the latter is on the spatial error model.

Since the Moran's I was highly significant, the Lagrange Multipliers were examined. The analysis suggests the use of SEM with a p-value of 0.04. However, SLM was fitted for further comparison and verification.

Exploratory Spatial Data Analysis (ESDA)

Acknowledging the effect of spatial autocorrelation, a prior examination of spatial characteristics using quantile map is useful for further analysis. Each quantile is represented by a color that allows better visualization of the distribution of quantiles in the study region. Each quantile has a range of values based on the actual observations while the value enclosed in the parentheses refers to the number of units included in each corresponding quantile. **Figure 1** below shows the Philippine map in four quantiles of HDI score.

Clustering of higher HDI was observed in the area of Luzon. Benguet and NCR, with respective first and second HDI ranking, are clustered with sub-national units having relatively high HDI between 0.626 and 0.842. The southern part of the country, on the other hand, has clustering of sub-national units with lower HDI

between 0.217 and 0.473. Included in this cluster are the provinces of Agusan del Sur, Lanao del Sur, Bukidnon, North Cotabato, Maguindanao and Sultan Kudarat.

Global Spatial Autocorrelation

The most common statistics used to determine the existence of spatial autocorrelation is the Moran's I incorporated with a spatial weight matrix. The significant Moran's I of the HDI scores was obtained to be 0.1871 (p<0.01). A positive value for I indicates clustering of high or low HDI scores together.

The four quadrants in **Figure 2** corresponds to the types of spatial autocorrelation that exists among the units. The upper right and the lower left refers to the units with positive spatial autocorrelation with high-high and low-low clustering, respectively. A unit is in high-high quadrant if the unit and its neighbors have high HDIs and in the low-low quadrant, if the unit and its neighbors have low HDIs. On the other hand, the upper-left and lower-right indicate negative spatial autocorrelation where values with low-high or high-low clustering, respectively. A unit with low HDI with neighbors having high HDIs belongs to the low-high quadrant and the unit with high HDI with neighbors having low HDIs belongs to the high-low quadrant. The cluster map below allows better visualization of the distribution of the spatial autocorrelation among the

human development index of the sub-national units of the Philippines. The value enclosed in the parentheses refers to the number of units in each cluster **(Figure 3)**.

The spatial units that fall in the first quadrant with high-high values clustered on the region of Luzon. Twenty-five percent of provinces are in this cluster including Pangasinan, Nueva Ecija, Bulacan, NCR and Bataan among others. Positively spatially dependent units with low-low clustering, however, were observed on the southern part of the country. Most provinces in Mindanao such as Maguindanao, Lanao del Norte, Lanao del Sur, Misamis Occidental Surigao del Norte and other adjacent provinces that cover 29% of the provinces lie in this quadrant.

On the second and fourth quadrants rest the 18% and 11% of the spatial units with a random pattern. Some insignificant autocorrelation comprised of 18% of the units were observed mostly on island provinces. The clustering of the similar level of HDI across the Philippines confirms the significance of spatial dependence.

Spatial Regression

Standard regression procedure aims to provide a good predictive model for a dependent variable using significant explanatory variables. However, cases of spatially dependent data may violate some classical regression assumptions such as

heteroskedasticity and independence. To account spatial information that exists for a given data, the classical regression was modified, and the spatial regression was introduced. This regression procedure accounts for the geographical information in two ways. The first technique is by incorporating the effect of the dependent variable of nearby units or the spatial lag in the model and is called Spatial Lag Model. The second one incorporates the effect of the error terms of neighboring spatial units or the spatial error in the model and is called Spatial Error Model.

This study considered the spatial regression analysis since a significant spatial autocorrelation exists for the HDI scores of the Philippine provinces. The two spatial regression models were created and then compared further to determine the model that could best predict HDI in sub-national level incorporating distance-band weight matrix.

Spatial Lag Model (SLM)

The parameter estimates in **Table 1** were computed through maximum likelihood estimation. The spatial lag (Lag HDI) refers to the effect of the HDI scores of the neighboring sub-national units, and its spatial coefficient is not significant ($p > 0.05$). The remaining explanatory variables were observed to have a significant effect. The diagnostic check of the model reveals that the assumption on

multicollinearity, normality, and heteroskedasticity had been satisfied.

Spatial Error Model (SEM)

Table 2 shows that the explanatory variables were significant in predicting HDI scores and the effect of the error term of neighboring units is highly significant ($p < 0.01$).

The assumptions of heteroskedasticity and spatial dependence were both satisfied. A comparison of the cluster maps of the predicted HDI of the two models is shown below **(Figure 3, 4)**. Some units were clustered by the spatial models similar to the observed HDI.

To further assess the predictive performance of the spatial models, the Root Mean Square Error (RMSE) of the models were computed. The Spatial Error Model has the lower RMSE of 0.0444 implying that it performs better than the Spatial Lag Model with RMSE of 0.0453.

Generally, the Spatial Error Model with predictors such as poverty incidence, number of fully immunized children, cohort survival rate in public elementary school and the crime rate was the chosen model that may predict the Human Development Index of the sub-national units of the Philippines.
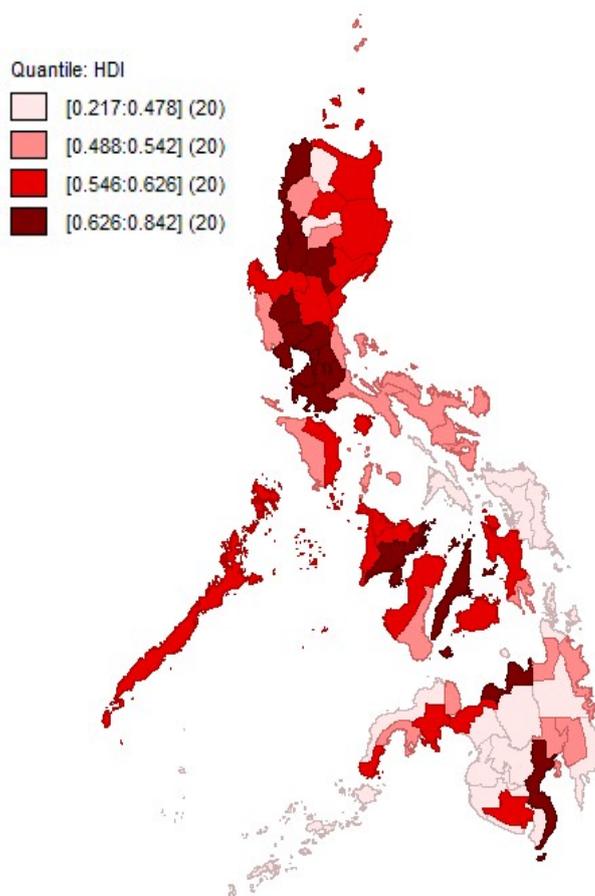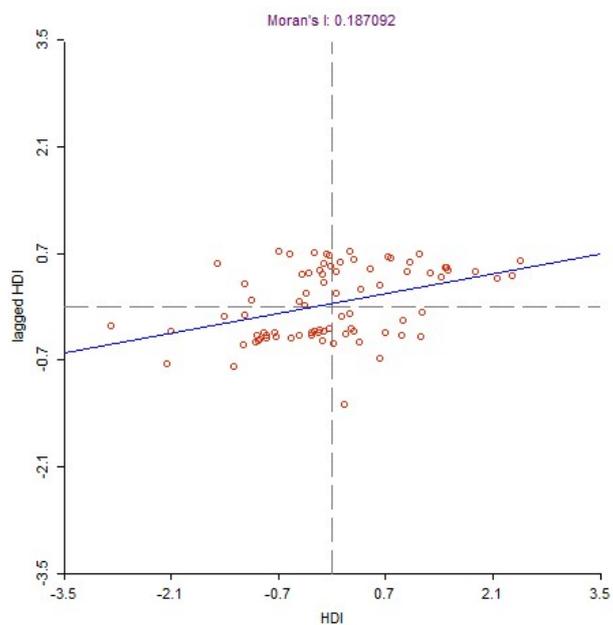
Figure 1. Quantile Map of HDI *(q=4)*

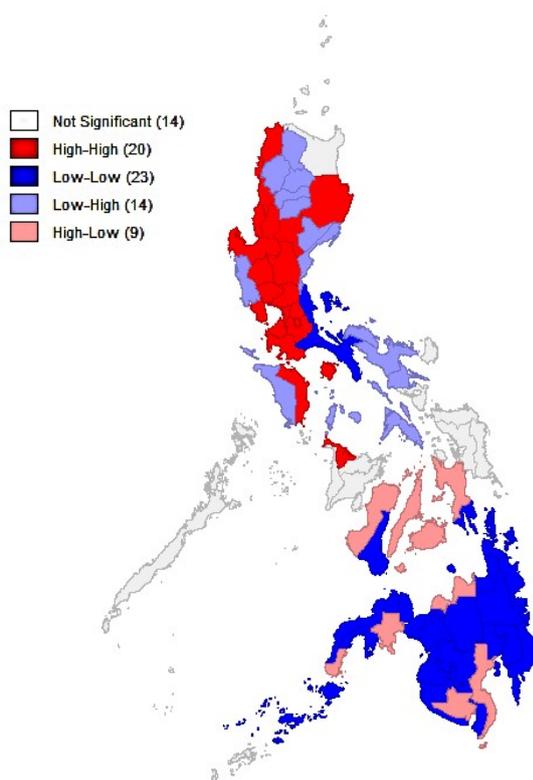**Figure 2: Moran's I Scatterplot of HDI**

Figure 3: Cluster Map of HDI

Table 1: Parameter Estimates of the Spatial Lag Model

Variable	Coefficient	Std. Error	z-value	Prob.
Lag HDI	-0.14821	0.10450	-1.42	0.156
Constant	0.67604	0.06706	10.08	<.001
Crime rate	7.93×10^{-5}	3.07×10^{-5}	2.58	0.010
Fully immunized children	4.18×10^{-7}	1.58×10^{-7}	2.64	0.008
Poverty incidence rate	-0.00552	0.00044	-12.61	<.001
Survival elementary rate	0.00161	0.00034	4.77	<.001

Table 2: Parameter Estimates of the Spatial Error Model

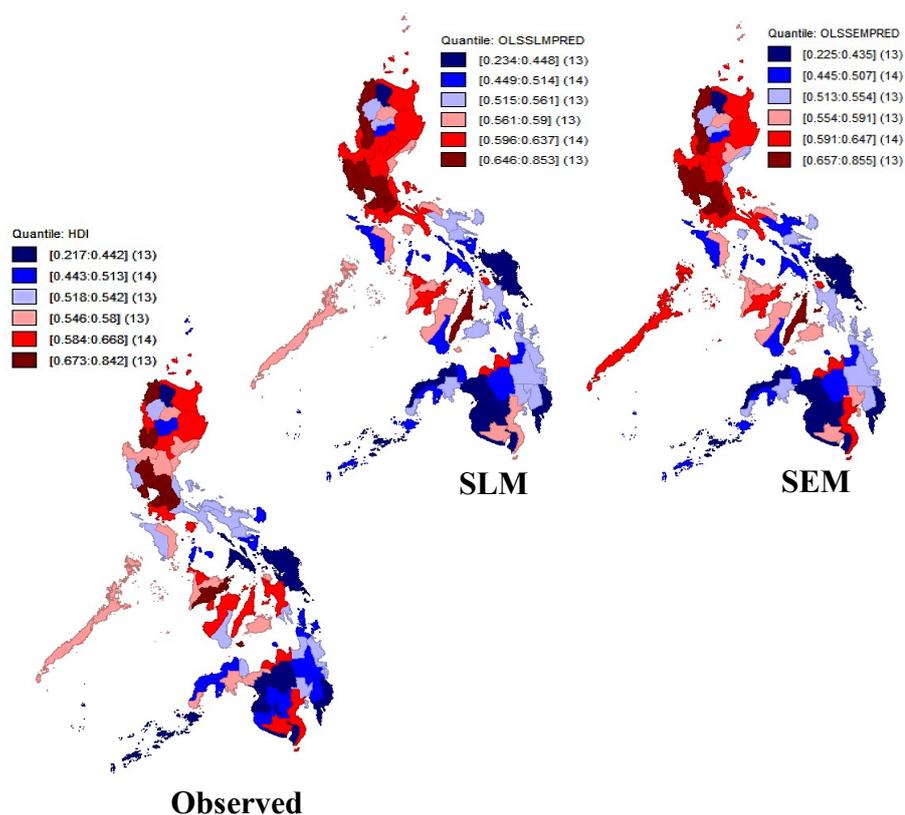| Variable	Coefficient	Std. Error	z-value	Prob.
Lambda	0.54293	0.21775	2.49	0.013
Constant	0.61553	0.03680	16.73	<.001
Crime rate	7.68×10^{-5}	3.01×10^{-5}	2.55	0.011
Fully immunized children	3.87×10^{-7}	1.52×10^{-7}	2.55	0.011
Poverty incidence rate	-0.00581	0.00044	-13.14	<.001
Survival elementary rate	0.00142	0.00034	4.20	<.001

Figure 4: Quantile Map of the Observed and Predicted HDI (*q=6*)

SUMMARY, CONCLUSION, AND RECOMMENDATION

The Human Development Index is a global measure of the development of human lives that enable international comparison. There had been several studies conducted to improve this measure in a different context. In this paper, the HDI of the sub-national units or provinces of the Philippines was regressed using spatial regression to determine more specific indicators of this human development. Spatial regression was performed to identify indicators of HDI while acknowledging the effect of space. There

were sixteen initial explanatory variables reduced using the Stepwise Selection.

The sixteen primary explanatory variables were reduced to four after performing stepwise regression and examining significance. The variables considered for the analysis were the poverty incidence, cohort survival rate in public elementary school, the number of fully immunized children and crime rate.

Diagnostics reveal that spatial dependence exists. Distance-band weight matrix was used to indicate neighborhood in performing spatial regression.

The results of this spatial model reveal that the poverty incidence rate,

number of fully immunized children, cohort survival rate of elementary school children and crime rate statistically significantly affect the score on human development. It is notable that the explanatory variables capture the dimensions of HDI such as education, health, and standard of living. The four significant specific variables represented the broad indicators of these three dimensions. Therefore, to further promote human development, it is vital to account for poverty among the inhabitant.

The education index, on the other hand, has indicators of mean years of schooling and expected years of education. However, the study revealed that the trend on the enrolment rate could be further taken into consideration to determine the factors associated with the number of enrollees. The number of fully immunized children also had a significant positive effect on the HDI. It suggests that health dimension can be improved by giving importance to the children's health as a foundation for a long and healthy life.

It is also of interest to model each of the three dimensions of Human Development Index – long and healthy life, education and standard of living – to determine more specific variables for the dimensions and the HDI as well. Careful must be taken in interpreting the coefficients of the final model as other factors might influence the human development index.

REFERENCES

[1] Rinaldi R, Nurwita E. The spatial dimension of human development index in Indonesia. Center for Economics and Development Studies, Department of Economics, Padjadjaran University Jalan Cimandiri no. 6, Bandung, Indonesia. 2010.

[2] Rezaeian M, Dunn, G, St. Leger S, Appleby L. Geographical epidemiology, spatial analysis, and geographical information systems: a multidisciplinary glossary. J Epidemiol Community Health. 2007; 61(2): 98-102.