Examining the Relationship Between Human Development Index and Socio-Economic Variables: A Panel Data Analysis

Siti Ayu Jalil¹&², Mohamad Nadzrul Kamaruddin²

¹Malaysian Institute of Transport (MITRANS), Universiti Teknologi MARA, Shah Alam, Selangor.
²Faculty Business and Management, Universiti Teknologi MARA, Shah Alam, Selangor.

¹ayu090@salam.uitm.edu.my
²nadzrulkamaruddin@gmail.com

Abstract — Human Development Index (HDI) measures the average achievements from three basic dimensions of human development: a long and healthy life, access to knowledge and a decent standard of living. This study is to investigate the impact of socio-economic variables represented by the three dimensions i.e. mean years of schooling, expected years of schooling, gross domestic product, life expectancy and health expenditure on HDI in fifteen selected developing countries within a 5-year period (2010-2014). The panel data analysis tested the pooled regression model, fixed effects and random effects models. The findings indicated that the Panel Fixed Effects Model (FEM) has proven to be the best model to describe the study. From the FEM model, four predictors have shown significant positive effect on human development index which are, the mean years of schooling, expected years of schooling, life expectancy and GDP per capita whereas, health expenditure is the only variable that shows insignificant relationship. Hence, it can be stated that in these fifteen selected nations despite education and higher GDP are essential to achieve a higher level of HDI, life expectancy is also perceived as a vital indicator to imply a better level of HDI.

Keywords: Human Development Index, Panel data, Fixed effects model, Random effects model

I. Introduction

Human Development Index (HDI) is the most extensively used measure to communicate a country’s development status compared to Gross Domestic Product (GDP). The HDI generally is a measure of development since it captures not only the level of income, but also incorporates measures of health and education (Srinivasan, 1994; Anand and Sen, 2006). Consequently, HDI becomes a statistical tool that measures a country’s overall achievement in terms of its social and economic dimensions referring in particular the level of education attainment, health and standard of living. In other words, it is a comprehensive and holistic measure of human development than per capita income alone. Though rapid economic growth has positive outcome for developing countries whereby the government manages to generate more national income through international trade, which then attracts foreign direct investment into the country and produce more goods and services; yet economic growth is not the main criterion for a developing country to become a developed nation.
Most developed countries have good education facilities and provide better health services to their citizen leading them to achieve a high score of HDI. However, majority of the developing countries focuses less spending on education facilities, consequently limit the poor to have access of their children’s education (UNDP Turkey, 2011) and, they end up participating in the labor force instead to earn income for their family (Naeem, Shaukat, and Ahmed, 2011). Commonly, the developing countries only put emphasis on increasing economic growth while the development of socio economy holds less priority to the country’s objectives. From the perspective of health dimension, developing countries relies heavily on the ability of the developed countries’ researches to identify local, regional and national needs to improve their health outcome. Lack of build in-country health infrastructures for prevention, care and treatment hindered their capabilities to respond to health crises. Therefore, the need to support developing countries health challenges through in-country research capacity development could help to attain long term solutions to critical health needs.

Developing countries can be categorized under upper-middle income, middle income and lower middle income. Basically, from the standard of living perspective, the level of household income and cost of living are essential in ensuring their material well-being. It is believed a higher living standard means that a household would be having a better quality of life. In the case of these countries, inequality may arise when the income distribution is unbalanced which leads to huge gap between high and low-income earners. The increasing gap has led to the rise in the cost of living depressing those poor not only in rural but more significantly those living in urban area. Another main issue in these developing nations is overpopulated causing emerging of slum in urban areas to seek better job with lucrative income. Effort should be place on developing human capital so as to gain higher productivity and generate higher real income that able to cut the cost of living.

Therefore, the objectives of this study is to: i) analyze the relationship of factors i.e. mean years of schooling, expected years of schooling, gross domestic product, life expectancy and health expenditure on the level of HDI, ii) determine the significant factor or factors contributing to the level of HDI scores in fifteen (15) selected developing nations employing a panel data analysis and iii) identify the appropriate model that best explain the analysis.

II. Literature Review

In 1990 a Pakistani economist named Mahbub ul Haq has created HDI, which was then used by the United Nations Development Program (UNDP) to measure a country's development. It measures the achievement of a country via social and economic dimensions based on their people’s health, their level of education attainment and their standard of living. It is seen as the best tool to keep track of the level of development of a country. Since that a number of studies were conducted to provide a more accurate measure of HDI. An existing study by Mazumdar (2003) in his paper “Measuring Human Wellbeing of the Countries: Achievement and Improvement Indices” made an attempt to look at the changing pattern of human well-being of the countries over the period of 1960 to 1994. He has come up with a comparison of the relative position of a country based on the overall achievement index and other composite indices that are used to measure human well-being such as weighted index, physical quality of life index and human development index.

Majority of the studies were carried out not only to investigate factors influencing HDI of a country but also focused on the techniques or methods used for analysis. For instance, Çaglayan-Akay and Van (2017) investigate factors that influence the level of economic development in 130 countries by using Bayesian ordered probit model. In their study, seven selected independent variables such as internet users, rural population, gross domestic product (GDP), life expectancy at birth, health expenditure, and share of expected years of schooling are used to investigate the impact of these variables towards the HDI. Different results were found in both short term and long term periods. For short term period, the variable that had affected the HDI positively are all those selected independent variables. However, in the long term, it is found that only GDP, health expenditure, the share of expected years of schooling and internet users have positive impact towards the HDI while life expectancy at birth and rural population showed a contrary result. Thus, it can be concluded that, HDI is a long term indicator for the country to measure the growth of the society in terms of education, health and standard of living. The result has shown that the expected years of schooling is one of the selected variables that have positive impact for both short and long run periods towards HDI.

Besides, Singariya (2014) supports Çaglayan-Akay and Van (2017) findings since she found countries with high HDI scores will concentrate more on these four determinants (GDP per capita, expected years of schooling, labor participation rate, and life expectancy at birth). According to Uteubayev (2016), majority of Eurasian Economic Union (EEU) members displayed positive outcomes in the HDI dynamics over the past decade based on the achievements in education, no members of EEU has lower index from year 2005. A cross sectional study conducted by Grzech, Patel and Walker (2016) using data of 188 countries found that life expectancy and
education gives the most impact on HDI values implying these elements are the main contributors of HDI improvement. Shah (2016) has examined several socio and macroeconomic variables such as GDP, life expectancy, literacy rate, Gini, fertility rate, CO2 emission and inflation rate on HDI. He found all variables show significant effect on HDI whereby GDP, life expectancy and literacy rate portrayed a positive relationship whilst Gini, fertility rate, CO2 emission and inflation rate portrayed a negative relationship. A more recent research by Arisman (2018) on ASEAN using fixed effects model showed population and per capita income growth do affect the HDI, however rate of inflation and unemployment do not have effect on HDI.

III. Econometric Modelling

This study investigates the relationship of the three dimensions of human development depicted as the socio-economic variables on human development index. The selected socio-economic variables are mean years of schooling (LMYS), expected years of schooling (LEYS) represented as access to knowledge, life expectancy and health expenditure (LLE) signify long and healthy life, whilst gross domestic product (LGDP) per capita (LHEX) to present a decent standard of living. The econometric modeling is presented as follows:

\[
\text{LHDI}_{it} = \beta_1 + \beta_2 \text{LMYS}_{it} + \beta_3 \text{LEYS}_{it} + \beta_4 \text{LLE}_{it} + \beta_5 \text{LGDP}_{it} + \beta_6 \text{LHEX}_{it} + \varepsilon_{it} \tag{1}
\]

where \(a\) is the intercept, \(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\) and \(\beta_6\) are the coefficients of the model, and \(L\) refers to the natural log of each variables whereas \(\mu\) is the error term. On the other hand, \(i\) refers to individual entity and \(t\) is time period. These variables are assumed to be strictly exogenous where each does not depend on the current, composite error term, \(\varepsilon_{it}\). Besides, it also assumed that the error term is normally distributed with \((0, \sigma^2\varepsilon)\) with zero mean and constant variance. It simply means the characteristics of the explanatory variables at any time period the error term will have zero expectation.

Based on equation (1), this study applies the short-panel balanced data for fifteen countries with a limited five-year period. Using the three competing econometric formulations namely pooled model, random effect model and fixed effect model would provide more information, more variability, more robust analysis and efficiency among the changes in the variables over time. Pooled Regression method can be applied when the groups to be pooled are assumed to be similar or homogenous. In other words, the regression analysis assumed both intercept and slope are the same across units and time thus, the estimation equation is:

\[
\text{LHDI}_{it} = \beta_1 + \beta_2 \text{LMYS}_{it} + \beta_3 \text{LEYS}_{it} + \beta_4 \text{LLE}_{it} + \beta_5 \text{LGDP}_{it} + \beta_6 \text{LHEX}_{it} + \varepsilon_{it} \tag{2}
\]

However, within the context of panel data, two kinds of variation are identified i.e. between cross-section units and within time-series units. Between variance refers to the variability across the unit of observations or cross sectional variations whereas within variance captures how much the overall variance is due to variability within economic units i.e. the time series variations. These variations can be detected with the random and fixed effect models. Basically, both econometric models merely allow heterogeneity or individuality among the countries by allowing them to have its own intercept value, while restricting the slope to be homogenous. To accommodate such heterogeneity, the error term \(\varepsilon_{it}\), is decomposed into two independent components:

\[
\varepsilon_{it} = \lambda_i + \mu_t \tag{3}
\]

\(\lambda_i\) is termed as individual specific effect meaning each nation may have a unique characteristic such as in this case different education system or even years of schooling as well as health system. On the other hand, \(\mu_t\) is a normal error term denoting the remainder disturbance. Hence, the random effects’ equation would be:

\[
\text{LHDI}_{it} = \beta_1 + \beta_2 \text{LMYS}_{it} + \beta_3 \text{LEYS}_{it} + \beta_4 \text{LLE}_{it} + \beta_5 \text{LGDP}_{it} + \beta_6 \text{LHEX}_{it} + \lambda_i + \mu_t \tag{4}
\]
Where \( \lambda_i \) becomes part of the error term and observations is representative of a sample rather than the whole population. To compute the analysis, a crucial assumption to assume in order to ensure consistency is \( \text{E}(x_{it}\lambda_i) = 0 \) for all \( k, t, i \) depicting that \( \lambda_i \) is uncorrelated with \( x_{it} \) i.e. the regressors. It means the individual specific constant terms are randomly distributed across cross-sectional units.

On one hand, the fixed effect model assumes each individual specific effect to have intercepts that may vary across countries. In econometric terms it is when \( \text{Cov}(\lambda_i x_{it}) \neq 0 \) or \( \lambda_i \) is correlated with \( x_{it} \) meaning it relies on the time series variations but is the most flexible in that it allows endogeneity of regressors. Hence, the equation is:

\[
LHDI_{it} = (\beta_1 + \lambda_i) + \beta_2\text{LMYS}_{it} + \beta_3\text{LEYS}_{it} + \beta_4\text{LLE}_{it} + \beta_5\text{LGDP}_{it} + \beta_6\text{LHEX}_{it} + \mu_{it}
\] (5)

(\( \beta_1 + \lambda_i \)) means \( \lambda \) becomes part of constant but still varies by individual units.

The appropriate test to discriminate between the pooled model and the random effect model is to utilize the Breusch-Pagan Lagrangian Multiplier (BP-LM) test. The hypothesis for the test is:

\[
H_0: \sigma^2_\lambda = 0 \quad \text{versus} \quad H_A: \sigma^2_\lambda > 0
\] (6)

Thus, if the level of significance below than 0.05, it means that the proper model that be used for this regression is Random Effect Model and if the result shows that the level of significance is above 0.05, it indicates that the model that be used for this regression is the Pooled Model.

On one part, the Hausman test is conducted to evaluate the consistency of an estimator when compared to an alternative, less efficient, estimator which is already known to be consistent. This test is to determine whether the fixed effect or random effect regression model that appropriately explains this research. The statistical test would be based on the hypothesis:

\[
H_0: \text{Cov}(\lambda_i x_{it}) = 0 \quad \text{(Random)} \quad \text{versus} \quad H_A: \text{Cov}(\lambda_i x_{it}) \neq 0 \quad \text{(Fixed)}
\] (7)

If there is significant result the level of significance should be less than 0.05, it implies that fixed effects model is an appropriate explanation for the analysis. However, if it is insignificant the level of significance is above 0.05, it indicates that the random effect model would be best when analyzing the model.

Sources of Data

Data gathering involves only fifteen selected developing countries that were chosen based on the availability of data collected from the period of 2010 to 2014. The countries are Albania, Algeria, Bahamas, Brazil, Bulgaria, Ecuador, Iran, Malaysia, Mauritius, Mexico, Panama, Romania, Serbia, Thailand and. These data were gathered from the Human Development Data and World Bank 2017. Since each data has different unit of measurement, the data were converted into double linear log in order to minimize and standardize the gap values of each variable so that it is easier to interpret and provide a reliable empirical analysis.

IV. Empirical Results and Discussion

A preliminary analysis was conducted to test the normality distribution of the data and is shown in Table 1. The Jarque-Bera probability values fail to reject the null hypothesis of normal distribution, thus data of these variables meet the normal distribution criterion. However, data of LHDI and LMYS were negatively skewed.

<table>
<thead>
<tr>
<th>Table 1: Descriptive Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLDI</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Median</td>
</tr>
<tr>
<td>Maximum</td>
</tr>
<tr>
<td>Minimum</td>
</tr>
<tr>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Skewness</td>
</tr>
<tr>
<td>Kurtosis</td>
</tr>
<tr>
<td>Jarque-Bera</td>
</tr>
<tr>
<td>Probability</td>
</tr>
</tbody>
</table>
Table 2 shows most of the selected variable correlations score less than 0.60 meaning the variables portray a rather weak relationship. For instance LMYS and LEYS depict a weak negative relationship stating a value of -0.118 whilst LMYS and LLE score +0.0185 having a weak but positive relationship. The highest score stated is -0.5243 between LMYS and LGDP.

<table>
<thead>
<tr>
<th>Economic Variables</th>
<th>LMYS</th>
<th>LEYS</th>
<th>LLE</th>
<th>LGDP</th>
<th>LHEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMYS</td>
<td>1.000000</td>
<td>-0.118022</td>
<td>0.018501</td>
<td>-0.524286</td>
<td>0.374782</td>
</tr>
<tr>
<td>LEYS</td>
<td>-0.118022</td>
<td>1.000000</td>
<td>-0.102311</td>
<td>0.026920</td>
<td>0.052558</td>
</tr>
<tr>
<td>LLE</td>
<td>0.018501</td>
<td>-0.102311</td>
<td>1.000000</td>
<td>-0.181385</td>
<td>0.120080</td>
</tr>
<tr>
<td>LGDP</td>
<td>-0.524286</td>
<td>0.026920</td>
<td>-0.181385</td>
<td>1.000000</td>
<td>-0.069677</td>
</tr>
<tr>
<td>LHEX</td>
<td>0.374782</td>
<td>0.052558</td>
<td>0.120080</td>
<td>-0.069677</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

Next, the key analysis of this study is presented in Table 3 explains the results of three panel regression model which is pooled model, fixed effects model and random effects model. In addition, the table also indicates the significant economic variables that affect the human development index in the developing countries. The purpose of combining all models into a single table is to provide a better picture of the relationship and to identify the significant variables. Nevertheless, the discussion of selecting the specific relevant model to be used in this research will be explained once the Hausman test and Breusch Pagan Lagrange Multiplier test are conducted, respectively.

<table>
<thead>
<tr>
<th>Model</th>
<th>Pooled Model</th>
<th>Fixed Effects Model (FEM)</th>
<th>Random Effects Model (REM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B&lt;sub&gt;1&lt;/sub&gt;</td>
<td>-1.961985</td>
<td>-3.86504</td>
<td>-3.12018</td>
</tr>
<tr>
<td>β&lt;sub&gt;1&lt;/sub&gt; LMYS</td>
<td>0.166664</td>
<td>(0.0000)***</td>
<td>0.122547</td>
</tr>
<tr>
<td>β&lt;sub&gt;3&lt;/sub&gt; LEYS</td>
<td>0.136170</td>
<td>(0.0000)***</td>
<td>0.195129</td>
</tr>
<tr>
<td>β&lt;sub&gt;4&lt;/sub&gt; LLE</td>
<td>0.172053</td>
<td>(0.0665)*</td>
<td>0.300596</td>
</tr>
<tr>
<td>β&lt;sub&gt;5&lt;/sub&gt; LGDP</td>
<td>0.001102</td>
<td>(0.2887)</td>
<td>0.059240</td>
</tr>
<tr>
<td>β&lt;sub&gt;6&lt;/sub&gt; LHEX</td>
<td>0.028072</td>
<td>(0.0000)***</td>
<td>0.001560</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.859663</td>
<td>0.998909</td>
<td>0.970250</td>
</tr>
<tr>
<td>Prob (F-statistic)</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

Note: *** Significant at the 1% level, ** Significant at the 5% level, * Significant at the 10% level

The pooled regression analysis results in Table 3 indicate that the coefficient of determination $R^2$ for this model is 0.86 whereby approximately 86 per cent of the systematic variation in the dependent variable that been explained by changes in the independent variables. The model is a good fit since there is only 14 per cent of the systematic variation in LHDI for developing countries do not explain by the model. The $p$ value of $F$-statistic is significant at 1 per cent level of significance indicating the overall model is acceptable. The equation for this model is:

$$LHDI_{it} = -1.962 + 0.167 \text{LMYS}_{it} + 0.136 \text{LEYS}_{it} + 0.17 \text{LLE}_{it} + 0.001 \text{LGDP}_{it} + 0.028 \text{LHEX}_{it} + \varepsilon_{it} \ (8)$$

Three explanatory variables namely LMYS, LEYS and LHEX were statistically significant at 1 per cent level of significance implying that these variables have effect on LHDI. LLE is the only variable that is significant at 10 per cent level of significance. However, surprisingly LGDP is statistically insignificant meaning economic growth has no effect on the human development index (LHDI).

The random effects model result displays that the overall study is acceptable whereby the $F$-statistical probability is significant at 1 per cent level of significance. Other than that, the value of $R^2$ (0.970250) reveals that...
about 97 per cent of the systematic variable of LHDI has been explained by the changes in the explanatory variables indicating the model is sufficient and only 3 per cent of the systematic variation in the LHDI does not explain by the error term. All selected variables are statistically significant at 1 per cent level of significance besides showing a positive relationship with LHDI. In other words, the variables portray a direct impact towards the LHDI. The equation of the model is:

$$ L_{HDI_{it}} = -3.120 + 0.157L_{MYS_{it}} + 0.173L_{EYS_{it}} + 0.392L_{E_{it}} + 0.009L_{GDP_{it}} + 0.0167L_{HEX_{it}} + \mu_{it} $$ (9)

The fixed effects model analysis results indicate that the relationship between economic variables and HDI are positive, and have significant impact towards HDI. In addition, the $F$-statistics probability value is statistically significant at 1 per cent level of significance implying the model is acceptable. Although four predictors i.e. LMYS, LEYS, LLE and LGDP are significant, LHEX however has no significant effect on HDI. Hence, the equation can be shown as:

$$ L_{HDI_{it}} = -3.865 + 0.123L_{MYS_{it}} + 0.195L_{EYS_{it}} + 0.301L_{E_{it}} + 0.0592L_{GDP_{it}} + 0.002L_{HEX_{it}} + \mu_{it} $$ (10)

The next section of discussion is to choose which model that explains the best for the analysis. Two statistical tests were conducted i.e. the Breusch-Pagan LM Test to choose between pooled OLS versus random effects model. Whilst the Hausman Test will be conducted to determine either FEM or REM is more appropriate for this study. Applying each technique of testing, Table 4 reveals the results for the two tests.

**Table 4: Statistical Tests among the Models**

<table>
<thead>
<tr>
<th></th>
<th>Poooled OLS versus Random Effect</th>
<th>Random Effect versus Fixed Effect</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breusch-Pagan LM Test</td>
<td>136.3157***</td>
<td>106.0716***</td>
<td>Random</td>
</tr>
<tr>
<td>Hausman Test</td>
<td></td>
<td></td>
<td>Fixed</td>
</tr>
</tbody>
</table>

*Note: ***1% level of significance; **5% level of significance*

Table 4 illustrates the results of the statistical test among the models of which the Lagrangian Multiplier test is significant at 1% level of significance; hence rejecting the null hypothesis revealing a strong evidence to retain country specific effects. Due to this, the random effect model was chosen to be appropriate description then the pooled OLS. To decide between fixed or random effects model, the Hausman test is conducted. Table 4 depicts that the null hypothesis fails to be rejected which means the country specific effects are not correlated with the regressors. Thus, the random effects estimator is inconsistent implying that it is not the appropriate model to explain this analysis. Consequently, the suitable empirical model to be applied for this study is Panel Fixed Effects Model. Accordingly, it also implies that LMYS, LEYS, LLE and LGDP are statistically significant and have impact on LHDI whereas LHEX is insignificant and eventually has no impact on LHDI.

Since the Fixed Effects Model is the best model to describe the analysis, the four predictors that have a direct relationship with Human Development Index can be interpreted in such a way that an increase in any of the variables will lead to an increase in HDI. For instance an increase in the mean years of schooling will lead to an increase in the Human Development Index (Schmidt, Moraes, & Migon, 2015).

V. Conclusion

This research chooses fifteen developing countries which are selected through the World Bank and United Nations Development Program with annual data frequency from 2010 to 2014. Based on the results, the Panel Fixed Effects Model has proven to be the best model to describe the study. From the FEM model, four predictors have shown a significant positive effect on human development index. First, the mean years of schooling has shown a direct significant relationship with HDI verifying that for developing countries to improve their HDI is through providing sufficient facilities for their education system. This was also supported by several empirical findings which is Psacharopoulos & Arriagada (1986) and Arbak (2012).
Second, similarly in every increase of 1 per cent of the expected years of schooling, the index of the LHDI will increase by 0.195 per cent. It can be claimed that for developing countries longer years of schooling can result with better workforce leading to higher productivity as compared to individuals with lesser period of schooling (Psacharopoulos & Arriagada, 1986). Another essential issue to consider is the investment on education or commonly known as investment in human capital that able to improve the skill formation and raises the ability of individuals to produce and work since it is closely connected with the economic development of a nation (Jabbar & Selvaratnam, 2017).

Third, GDP or economic growth has correspondingly shown a positive significant relationship with HDI meaning that the increment of the gross domestic product leads to positive impact on the LHDI (Bintang, Ismail & Indra, 2015). This result has been supported as well by studies conducted by Shome & Tondon (2010) and Abraham & Ahmed (2011). Though GDP may imply a measure of material well-being of a nation, it can still be regarded as improving HDI due to society achieving a better standard of living. The final significant variable is life expectancy that has positive effect on HDI. Klugman (2011) stated life expectancy could be an indicator of a long and healthy life suggesting that people in developing countries see it as an essential predictor to achieve a higher HDI. This finding is also supported by Connolly, Leoz, Gorospe, & Sebastian (2014), Naik & Selvarajan, (2016), and Çaglayan Akay & Van (2017).

However, health expenditure is the only independent variable that shows insignificant direct relationship with HDI. Most developing countries have difficulty to manage proper financial spending on the public’s health expenditure due to insufficient budget allocated for health care (World Bank, 2005). Even if there is sufficient budget allocated for health care, a critical issue that of concerned is the access to health care especially for the poor. In fact education is still the best option to promote health among the people.

In conclusion, socioeconomic policies that are appropriately focused on people’s development will assist the country to achieve a high score of HDI. Though income reflects the basic needs people acquire for food and shelter, from the view of HDI concept, it is merely seen as an instrument that focuses on materialistic development. In the context of this study, it is thus necessary for policymakers from these fifteen developing countries to implement policies as according to their nation’s requirement. HDI can be of useful to discuss and finalize policy choices between nations with approximately identical GDP.

References


