

Analysis of Factors Affecting the Human Development Index in West Kalimantan Province using Data Panel Data Regression

Uray Hety Humaira, Jaka Nugraha^{*)}

Statistics Department, FMIPA, Islamic University of Indonesia

*Corresponding author: jnugraha@uii.ac.id

Abstrak

Pembangunan di seluruh wilayah Indonesia terus dilakukan, termasuk di Provinsi Kalimantan Barat. Namun demikian sampai dengan tahun 2015, pencapaian pembangunan manusia di tingkat Nasional masih yang cukup rendah dan pencapaian di Kabupaten/Kota cukup bervariasi. Indeks Pembangunan Manusia (IPM) merupakan salah satu ukuran pencapaian pembangunan manusia yang dipengaruhi oleh banyak faktor. Dalam makalah ini dilakukan analisis untuk mengetahui faktor-faktor yang mempengaruhi IPM di Kabupaten dan Kota Provinsi Kalimantan Barat dengan menggunakan Analisis Regresi data panel berdasarkan data dari tahun 2012 sampai tahun 2015. Diperoleh hasil analisis bahwa model regresi terbaik adalah *Fixed Effect Model* dengan dengan nilai R^2 sebesar 99.853%. Variabel yang mempengaruhi IPM adalah variabel Angka Harapan Hidup (AHH), Pendapatan per Kapita Disesuaikan (Pengeluaran), Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Produk Domestik Regional Bruto Atas Harga Konstan (PDRB).

Kata kunci : Regresi Data Panel, Model Efek Tetap, Indeks Pembangunan Manusia.

Abstract

Development in the country is growing including in the West Borneo Province. However in 2015, the achievement of human development at the National level is quite low, while the District and City varied considerably. Human Development Index is one of the parameter for human development that are affected by many factors. In this paper, analysis for identify the factors for human development index in West Kalimantan Province by using Regression Analysis was conducted. Regression was based on time series data from 2012 until 2015. It is found that Fixed Effect Model is the best regression model with the R^2 of 0.99853%. The influencing variables are Life Expectancy (AHH), Adjusted Per Capita (Expenditure), School Average (RLS), School Expectation (HLS), and Gross Regional Domestic Product at Constant Price (GRDP).

Keywords : Panel Data Regression, Fixed Effect Model, Random Effect Model, Human Development Index.

Introduction

HDI is a measure of development of a region. The calculation of *HDI* consists of 3 dimensions, namely health dimension, education dimension, and

expenditure dimension (decent standard of living) (UNDP, 2015).

HDI is used to classify whether a country is a developed country, a developing country or an

Analysis of Factors Affecting the Human Development Index in West Kalimantan Province using Data Panel Data Regression

underdeveloped country, and also to measure the influence of economic policy on quality of life.

By 2015, *HDI* in West Kalimantan Province reaches 65.59% means development level is included in the medium category. With a life expectancy of 69.8 years, the expected length of school is 12.25 years, the average length of school is 6.93 years and per capita expenditure per year is 8279.34 (Thousand Rupiah) (Badan Pusat Statistik, 2015).

Both the *HDI* average in districts / cities, as well as the average *HDI* in the province of West Kalimantan, both experienced a slowing growth. It is seen that the movement of value per year is slower than the national average of *HDI* (Indonesia). The movement from the *HDI* average of District / City and the *HDI* average of West Kalimantan province in each year is 0.68% and 0.81%. Unlike the average national *HDI* value (Indonesia), the annual difference reaches 4% more than the average of District/Municipal *HDI* and *HDI* of West Kalimantan Province.

Meirianti (2016) conducted an analysis using the fixed effect model of *HDI* data in 38 Regencies / Cities in East Java Province in 2010-2014. The results show that the level of poverty,

health spending and economic spending have a positive and significant impact, education spending has a positive but not significant effect on the *HDI*.

This is what causes researchers interested to analyze the factors that affect the deeper *HDI*. By using panel data, we combine time series data and cross-sectional data to provide more data, as well as obtain more complete and diverse information.

Panel Data Regression

A *Hdata setH* containing observations on a single phenomenon observed over multiple time periods is called a “*Htime seriesH*”. A data set containing observations on multiple phenomena observed at a single point in time is called “*Hcross-sectionalH*” (where each individual, sampling unit, data point is observed only once). A data set containing observations on multiple phenomena observed over multiple time periods is called “panel data” (which implies that each individual, sampling unit, data point is observed in more than one time period). Panel data analysis do not require that the time period in which different individuals or sampling units are observed are exactly the same.

The panel data is a combination of cross section and time series data, that is, some of the same individuals are observed in a certain period of time. A panel data regression differs from a regular time-series or cross-section regression in that it has a double subscript on its variables, i.e.

$$y_{it} = X_{it}'\beta + e_{it} \quad (1)$$

$$t = 1, 2, 3, \dots, T; i = 1, 2, 3, \dots, N$$

with i denoting households, individuals, firms, countries, etc. and t denoting time. The i subscript, therefore, denotes the cross-section dimension whereas t denotes the time-series dimension. β is $(K+1) \times 1$ and X_{it} is the i th observation on K explanatory variables.

There are three types of panel data regression : Common Effect Model, Fixed Effect Model and Random Effect Model.

a. Common Effect Model (CEM)

The common effect model is estimation which combines time series and cross section data. Parameters β can be estimated using the OLS (Ordinary Least Square) method (Baltagi, 2005). The model in equation (1) is called CEM. e_{it} is an error component which is assumed to have mean 0 and homogeneous variance in time. β are the effect of changes in X that are

assumed to be constant in time and cross-section.

b. Fixed Effect Model

According to Gujarati (2004) the term "fixed effect" is because even though the intercept is different for each subject, each interception of each value does not change over time. According to Rosadi (2011), the FEM model is:

$$Y_{it} = X_{it}'\beta + c_i + d_t + e_{it} \quad (2)$$

with c_i denotes the *unobservable* individual-specific effect, d_t denotes the constants that depend on time t . In this case, the c_i are assumed to be fixed parameters to be estimated and the remainder disturbances stochastic with e_{it} independent and identically distributed IID $(0, \sigma^2)$.

If the model contains the c_i and d_t components, the model is called a two-way FEM, whereas if $c_i = 0$ and $d_t = 0$ the model is called one-way FEM. For one-way FEM, it is often assumed that the component $d_t = 0$, so it has a model:

$$Y_{it} = X_{it}'b + d_t + e_{it} \quad (3)$$

c. Random Effect Model.

When using FEM, we cannot see the influence of various characteristics that are constant in time, or constant between individuals. For this purpose REM can be used, which is generally written as follows:

$$Y_{it} = X_{it}'b + e_{it};$$

$$e_{it} = u_i + v_t + w_{it} \quad (4)$$

The assumptions used for the error component are

$$u_i \sim N(0, \sigma_u^2); \quad v_t \sim N(0, \sigma_v^2)$$

$$w_{it} \sim N(0, \sigma_w^2)$$

Materials and Methods

The data that will be analyzed are *Human Development Index (HDI)*, *Gross Regional Domestic Product at Constant Price (GRDP)*, *Adjusted Per Capita (Expenditure)*, *School Expectation (HLS)*, *School Average (RLS)*, and *Life Expectancy (AHH)* of West Kalimantan Province in 2012-2015. Data is taken from the Badan Pusat Statistik (BPS) report, which is "Kalimantan Barat dalam Angka" in 2012-2015.

The author determine the panel data regression analysis is an appropriate choice to determine whether *Expenditure*, *HLS*, *RLS*, *AHH* and *GRDP* affect the *HDI*.

Result and Discussion

The Average of *Expenditure*, *HLS*, *RLS*, *AHH*, *GRDP* and *HDI* in 14 districts and cities of West Kalimantan Province in 2012 to 2015 are shown in Figure 1.

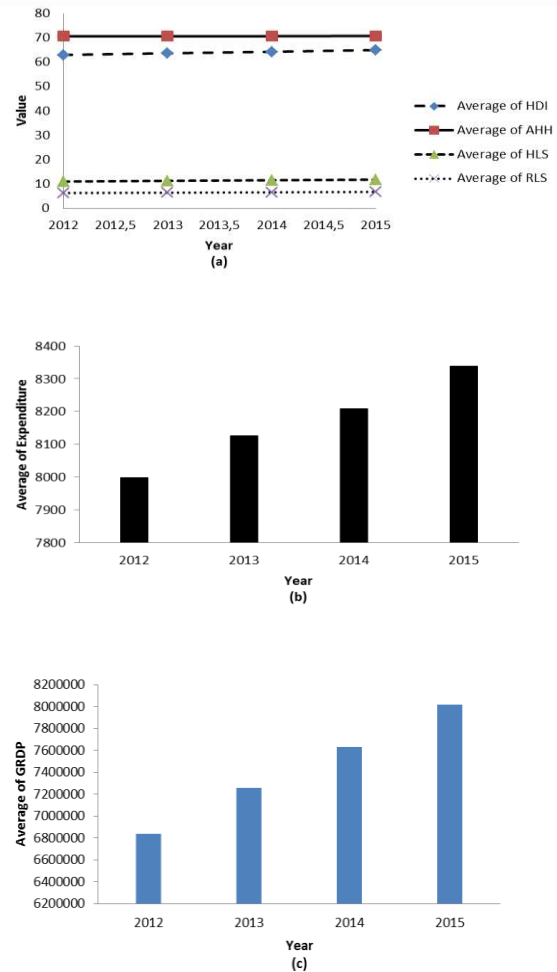


Figure 1 Average of each variable

In Figure 1, it is shown that all observed variables increase every year and the magnitude of the increase varies.

Then proceed with using R Studio application to do regression analysis of panel data, then got output as follows

(1) Modeling

a. Common Effect Model (CEM)

Table 1. F test for CEM

| F-Count | F-Table | P-Value | Decision |
|---------|---------|------------------------------|-------------|
| 4625.44 | 2.40 | < 2.22 x10 ⁻¹⁶ | Significant |

From Table 1, the value of $F_{count} = 4625.44 > F_{table} = 2.40$ and $p\text{-value} \leq 2.22 \times 10^{-16}$ which is less than 0.05 it is decided to reject the initial hypothesis (H_0), which means together independent variables affect the dependent variable.

Table 2. T test for CEM

| Variable | Coefficient | P-Value | Significant Level | Decision |
|-------------|-------------------------|------------------------|-------------------|-----------------|
| Constant | 1.6745 | 3.25×10^{-1} | 0.05 | Not Significant |
| AHH | 4.8628×10^{-1} | 2.20×10^{-16} | 0.05 | Significant |
| Expenditure | 1.1976×10^{-3} | 2.20×10^{-16} | 0.05 | Significant |
| HLS | 1.0871 | 2.20×10^{-16} | 0.05 | Significant |
| RLS | 8.5981×10^{-1} | 2.30×10^{-16} | 0.05 | Significant |
| GRDP | 2.3627×10^{-8} | 7.59×10^{-3} | 0.05 | Significant |

In pooled regression of *AHH*, *Expenditure*, *HLS*, *RLS* and *GRDP* have significant effect on *HDI* variable because the $p\text{-value}$ value is smaller than the significance level value of 5%, so the initial hypothesis H_0 is rejected.

Then obtained value of R^2 (coefficient of determinant) equal to 0.99784. This can be interpreted that in pooled regression analysis of *AHH*, *Expenditure*, *HLS*, *RLS* and *PDRB* influence variable of *HDI* equal to 99.784%. While the rest of 0.216% influenced by other variables that are not analyzed in the regression model.

b. Fixed Effect Model (FEM)

Table 3. F test for FEM

| F-Count | F-Table | P-Value | Decision |
|---------|---------|------------------------|-------------|
| 5016.46 | 2.47 | 2.22×10^{-16} | Significant |

From Table 3, the value of $F_{count} = 5016.46 > F_{table} = 2.47$ and $p\text{-value} < 2.22 \times 10^{-16}$ which is less than 0.05 it is decided to reject the initial hypothesis (H_0). This means that together independent variables affect the dependent variable.

Table 4. T Test for FEM

| Variable | Coefficient | P-Value | Significant Level | Decision |
|-------------|--------------------------|------------------------|-------------------|-------------|
| AHH | 3.7276×10^{-1} | 2.07×10^{-4} | 0.05 | Significant |
| Expenditure | 1.4865×10^{-3} | 2.20×10^{-16} | 0.05 | Significant |
| HLS | 1.1480 | 2.20×10^{-16} | 0.05 | Significant |
| RLS | 1.4236 | 2.30×10^{-16} | 0.05 | Significant |
| GRDP | -9.6664×10^{-8} | 4.08×10^{-7} | 0.05 | Significant |

In the fixed effect model of the *AHH*, *expenditure*, *HLS*, *RLS* and *GRDP* have significant effect on the *HDI* variable because the $p\text{-value}$ is less than the significance level of 5%, so the initial hypothesis H_0 is rejected.

The value of R^2 is 0.99853. This can be interpreted that in the analysis of fixed effect model estimation, only the *AHH*, *expenditure*, *HLS*, *RLS* and *GRDP* variables affecting the *HDI* variable is 99.853%. While the rest of 0.147% influenced by other variables that are not analyzed in the regression model.

c. Random Effect Model

Table 5. F test for REM

| F-Count | F-Table | P-Value | Decision |
|---------|---------|------------------------|-------------|
| 2005.45 | 2.40 | 2.22×10^{-16} | Significant |

From the value $F_{count} = 2005.45 > F_{table} = 2.40$. When viewed from the

value of $p\text{-value} = 2.22 \times 10^{-16}$ which is less than 0.05 it was decided to reject the initial hypothesis (H_0). This means that together independent variables affect the dependent variable.

Table 6. T test for REM

| Variable | Coefficient | P-Value | Significant Level | Decision |
|-------------|--------------------------|------------------------|-------------------|-------------|
| Constant | 6,5667 | 0,06901 | 0.05 | Significant |
| AHH | $3,6528 \times 10^{-1}$ | $4,861 \times 10^{-9}$ | 0.05 | Significant |
| Expenditure | $1,1745 \times 10^{-3}$ | $2,20 \times 10^{-16}$ | 0.05 | Significant |
| HLS | 1,1797 | $2,20 \times 10^{-16}$ | 0.05 | Significant |
| RLS | 1,3643 | $2,30 \times 10^{-16}$ | 0.05 | Significant |
| GRDP | $-4,1899 \times 10^{-8}$ | 0,02473 | 0.05 | Significant |

In the model of random effect of *AHH*, *expenditure*, *HLS*, *RLS* and *GRDP* have significant effect on the *HDI* variable because the p-value value is smaller than the significance level value of 5%, so the initial hypothesis H_0 is rejected (Table 6).

Then obtained value of determinant coefficient (R^2) equal to 0.99504. This can be interpreted that in the analysis of random effect model of *AHH*, *expenditure*, *HLS*, and *GRDP* influence the *HDI* variable of 99.504%. While the rest of 0.496% influenced by other variables that are not analyzed in the regression model.

(2) Model Identification

a. Chow Test

Table 7. The result of Chow Test

| Influence Test | F-count | F-table | p-value | Decision |
|----------------|---------|---------|-------------------------|-------------|
| F test | 118.11 | 2.00 | $< 2.2 \times 10^{-16}$ | Significant |

From the value of $F_{\text{count}} = 118.11 > F_{\text{table}} = 2.00$, and from p-value ($< 2.2 \times 10^{-16}$) less than 0.05 it is decided to reject the initial hypothesis (H_0). This means that the appropriate model is a fixed effect model.

b. Hausmann Test

Table 8. The result of Hausmann Test

| Influence Test | χ^2_{count} | χ^2_{table} | p-value | Decision |
|-----------------|-------------------------|-------------------------|-----------|-------------|
| Chi-Square test | 21.434 | 11.07 | 0.0006705 | Significant |

From Table 8, the value of $\chi^2_{\text{count}} = 21.434 > \chi^2_{\text{table}} = 11.07$, and from p-value = 0.0006705 less than 0.05 it was decided to reject the initial hypothesis (H_0). This means that the appropriate model is a fixed effect model.

c. Lagrange Multiplier Test

Table 9. The result of Lagrange Multiplier Test

| Influence Test | p-value | Decision |
|--------------------------|------------------------|-------------|
| Lagrange Multiplier test | $1,024 \times 10^{-6}$ | Significant |

From Table 9, the value of $p\text{-value} = 1.024 \times 10^{-6}$ is less than 0.05 it was decided to reject the initial hypothesis (H_0). This means that the appropriate model is the Random Effect.

(3) Model Specifications

Table 10. Breusch Pagan test for model specifications FEM

| Hypothesis | χ^2_{count} | χ^2_{table} | P-value | Decision |
|----------------------|------------------|------------------|------------------------|---|
| $H_0 : c_i=0, d_t=0$ | 37.955 | 5.99 | 5.731×10^{-9} | H_0 rejected, there' two ways effect. |
| $H_0 : c_i=0$ | 22.549 | 3.84 | 2.048×10^{-6} | H_0 rejected, the individual effect. |
| $H_0 : d_t=0$ | 15.405 | 3.84 | 8.674×10^{-5} | H_0 rejected, the time effect. |

From Table 10, the value of χ^2_{count} and *P-value* can be seen that the fixed effect model used has a two-way specification, which is influenced by individual effects and time effects. From the results of Chow, Hausmann and Breuch-Pagan test above, it can be concluded that the appropriate model to explain the variables forming *AHH*, *Expenditure*, *HLS*, *RLS* and *GRDP* directly affect *HDI* is a fixed effect model. With the model results are formed namely:

$$\begin{aligned}
 HDI = & 3.7276 \times 10^{-1} AHH + \\
 & 1.4865 \times 10^{-3} Expenditure + 1.1480 \\
 & HLS + 1.4236 RLS - 9.6664 \times 10^{-8} \\
 & GRDP + c_i + d_t
 \end{aligned}$$

and the value of individual and time effects explained in Table 11.

From Table 11, the table shows that the district and city of West Kalimantan Province has the largest individual value is Landak District that is equal to 2.4815989. This indicates that the Human Development Index in

Sub-district Landak is higher than the Regency and City in West Kalimantan Province. While the smallest Individual value found in Pontianak with the value of HDI of -5.5247228. Which means, that the value of HDI in the city has the lowest value among districts and cities in West Kalimantan Province.

Table 11. The value of Individual Effects

| Individual | Districts/Cities | Effect value |
|------------|------------------|--------------|
| 1 | Bengkayang | 0.1185357 |
| 2 | Kapuas Hulu | 0.1132201 |
| 3 | Kubu Raya | 0.8627460 |
| 4 | Kayong Utara | -0.1249034 |
| 5 | Ketapang | 0.8302479 |
| 6 | Landak | 0.0796319 |
| 7 | Melawi | 0.0352745 |
| 8 | Mempawah | 0.0991690 |
| 9 | Pontianak | -2.1012999 |
| 10 | Sambas | 0.1494578 |
| 11 | Sanggau | 0.7393289 |
| 12 | Sekadau | 0.0366270 |
| 13 | Singawang | -1.1981008 |
| 14 | Sintang | 0.3600650 |

Table 12. The value of Time Effects

| Period | Year | Effect value |
|--------|------|--------------|
| 1 | 2012 | 10.38715 |
| 2 | 2013 | 10.44532 |
| 3 | 2014 | 10.49128 |
| 4 | 2015 | 10.56705 |

The influence of time is the value that shows the difference in mean of inter-year *HDI* in West Kalimantan Province. This value indicates that each period will have a different constant value depending on the dt value of each year. Which means in the 1st period that is in 2012 has a value of time effect of 10.38715, the second period that is in

2013 has a value of time effect of 10.44532, and so on.

(4) Assumption Test

a. Multicollinearity Test

Table 13. The result of Multicollinearity Test

| Independent Variable | VIF value |
|----------------------|-----------|
| AHH | 1.043987 |
| Expenditure | 2.514837 |
| HLS | 1.925316 |
| PDRB | 1.844396 |

From Table 13, it is known that the overall independent variable has a VIF value < 10 , so there is no multicollinearity problem.

b. Heteroscedasticity Test

Table 14. The result of Heteroscedasticity Test

| Influence Test | P-Value | Significant Level | Decision |
|----------------|---------|-------------------|------------------------|
| Breusch Pagan | 0.1473 | 0.05 | Failed to reject H_0 |

From Table 14, therefore the value of p-value is greater than 0.05 which means accepting the initial hypothesis. This means that there is no problem of heteroscedasticity.

Conclusions

The correct model for data panel is fixed effect model, with two-way effect. Thus, together the *AHH*, *expenditure*, *HLS*, *RLS* and *GRDP* variables, affect *HDI* in 14 districts and cities in West Kalimantan at 99.853%.

the rest of 0.147% *GRDP* is influenced by other factors outside these variables.

Variables *AHH*, *Expenditure*, *HLS*, and *RLS* 14 districts and cities in West Kalimantan Province have a positive and significant relationship with *HDI* 14 districts / cities in West Kalimantan Province. Each increase of 1 (one) rupiah from the value of the variable with the value of constant variable then the value of *HDI* will each increase 3.7276×10^{-3} , 1.4865×10^{-3} , 1.1480, and 1.4236.

References

- Baltagi, B. H., 2005, *Econometrics Analysis of Panel Data* (3rd ed), England, John Wiley & Sons Ltd
- Badan Pusat Statistik, 2012, *Kalimantan Barat Dalam Angka*, Pontianak, BPS
- Badan Pusat Statistik, 2013, *Kalimantan Barat Dalam Angka*, Pontianak, BPS
- Badan Pusat Statistik, 2014, *Kalimantan Barat Dalam Angka*, Pontianak, BPS
- Badan Pusat Statistik, 2015, *Kalimantan Barat Dalam Angka*, Pontianak, BPS.
- Gujarati, 2004, *Basic Econometric*, (4th Ed.), New York, The McGraw-Hill Companies
- Meirianti. 2016. Pengaruh Kemiskinan, Belanja Pemerintah Bidang Pendidikan, Kesehatan dan

Ekonomi Terhadap Tingkat IPM di 38 Kabupaten/Kota Provinsi Jawa Timur, 2010-2014. *Tugas Akhir*. Yogyakarta : Universitas Gajah Mada.

Rosadi, Dedi., 2011, Analisis Ekonometrika dan Runtun Waktu Terapan dengan R, Yogyakarta, Andi Offset

UNDP, 2015, Human Development Report: Human Development Index 2015, New York