Mathematical Journal of Modelling and Forecasting
Vol. 1, No. 2, December 2023, pp. 15-22

Factors Influencing the Poverty Gap Index in West Sumatra Province through Panel Data Regression Analysis

Mutia Syifa[1,*] and Helma[1]

[1] Departement of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, Indonesia
[*]syifamut2308@gmail.com

Abstract. The poverty gap index is one of the measures that can be used to measure a region's poverty level well. This is because the poverty gap index measures the poverty gap by focusing on the expenditure gap of the poor against the poverty line. Based on BPS West Sumatra data for 2017-2022, the IKK, which was initially at 1,00, rose to 1,04 in 2018; in 2019, it fell to 0,94; in 2020, it fell to 0,92, but in 2021 it rose to 1,04 and fell to 0,8 in 2022. The dynamics in the IKK show that the government's efforts to deal with poverty problems have not been optimal. Therefore, a study was conducted to determine the model and factors that affect the IKK in West Sumatra Province using panel data regression analysis. After the research was conducted, the best estimation model obtained was the Fixed Effect Model (FEM). The factors that affected IKK in 2017-2022 are average per capita expenditure and human development index.

Keywords: Poverty gap index, Panel Regression, Fixed Effect Model

1 Introduction

Poverty is a complex and complicated phenomenon and is hot to discuss because poverty reduction is one of the goals of the Indonesian state stated in the preamble of the 1945 Constitution, which is to promote general welfare [1]. Regarding poverty, Indonesia is among the 100 poorest countries in the world. One of the indicators of the success of development in the region can be seen in the poverty rate. One of the measures that can be used to measure the level of poverty in a region can be seen from the poverty gap index.

The poverty gap index measures the average expenditure gap of each poor person relative to the poverty line [2]. Based on data from the Central Statistics Agency (BPS) for 2017-2022, which initially stood at 1.00, it rose to 1.04 in 2018, in 2019 it fell to 0.94, in 2020 it fell to 0.92, but in 2021 it rose to 1.04 and fell to 0.8 in 2022 [3]. Although the poverty gap index in West Sumatra Province decreased in 2022, several regencies/cities in West Sumatra experienced an increase and decrease in the poverty gap index. In 2022, several districts/cities, such as Solok, Sijunjung, Solok City, Sawahlunto, Bukittinggi, and Pariaman, experienced an increase in the poverty gap index. Moreover, several regencies/cities in West Sumatra have a poverty gap index that exceeds the gap index in West Sumatra, namely the Mentawai Islands, Solok Regency, Sijunjung, Lima Puluh Kota, South Solok, West Pasaman, and Bukittinggi City. Changes in the poverty gap index indicate that the government's poverty alleviation efforts have not been practical. Therefore, it is necessary to design new, more strategic steps to control the rise and fall of the poverty gap index in West Sumatra Province. Other factors can also cause the dynamics that occur in the poverty gap index and significantly influence it.

Factors influence the poverty rate in West Sumatra Province. So, an analysis is needed to see the influence of these factors. Based on these problems, a systematic study was conducted to analyze the factors influencing the poverty gap index in the Regency / City of West Sumatra. Using data related to economic growth, per capita expenditure, expected years of schooling, and human development index. It uses statistical analysis tools to analyze data covering several subjects and a specific period. One statistical analysis that can be used is panel data regression analysis. Panel data is a combination of two kinds of information, namely information between units (cross section) on differences between subjects and information between time (time series) that reflects changes in the subject of time [4].

2 Theoretical Basic

2.1 Panel Data Regression Model

Panel data is essentially repeated cross-section data and is able to detect and measure effects that cannot be observed using pure time series data or pure cross-section data [5]. The general model of panel data regression is stated as follows [6]:

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + u_{it} \tag{1}$$

I	=	Cross section/individual unit (i=1,2,3,...,p)
T	=	Time series unit (t= 1,2,3,...,q)
β_0	=	*Intercept*
β	=	*Slope* / Independent variable parameter
Y_{it}	=	The dependent variable of the i individual of period t
X_{it}	=	Independent variable of the i individual of period t
u_{it}	=	Error of the i-th individual i period t

2.2 Estimation of Panel Data Regression

According to [6], panel data regression estimation can be done using three approaches: the common effect model, fixed effect model, and random effect model.

2.2.1 Common Effect Model

The assumption of fixed coefficients across time and individuals, where this model assumes the combined data, shows that each variable's intercept and slope values are the same for each cross-section and time series unit.

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + u_{it} \tag{2}$$

2.2.2 Fixed Effect Model

The slope assumption is constant, but the intercept varies. The fixed effect model is based on the existence of intercept differences between individuals, but the slope regression coefficient is fixed between individuals and over time.

$$Y_{it} = \beta_{0i} + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + u_{it} \tag{3}$$

We can use the dummy variable technique to look at the differences in intercepts. Using the Least Square Dummy Variable (LSDV) method, the number of dummy variables is as many as the number of cross-section units. The FEM equation is as follows:

$$Y_{it} = \beta_{0i} + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + u_{it} \tag{4}$$

2.2.3 Random Effect Model

The random effects model approach will estimate panel data where residuals may correlate across time and individuals. Assuming there are differences in intercepts for each individual, the equation for the random effects model is as follows:

$$Y_{it} = \beta_{0i} + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + u_{it} \tag{5}$$

2.3 The Selection of Panel Data Regression Estimation Method

2.3.1 Chow Test

The Chow test is utilized to determine the best regression model for panel data, comparing the common effect model with the model obtained using the fixed effect model approach [7]. The hypotheses used are as follows:

H_0 : *Common effect model*
H_1 : *Fixed effect model*
The test statistic using F_{count} is expressed as follows:

$$F_{count} = \frac{\frac{(SSE_1\text{-}SSE_2)}{(N\text{-}1)}}{\frac{SSE_2}{(NT\text{-}N\text{-}K)}} \quad (6)$$

2.3.2 Lagrange Multiplier Test

The LM test is employed to determine the best regression model for panel data, comparing the model obtained based on the common effect model with the random effect model approach [8]. The hypotheses used are as follows:

H_0 : *Common effect model*
H_1 : *Random effect model*
The test statistic is expressed as follows:

$$LM = \frac{NT}{2(T\text{-}1)}\left[\frac{\sum_{i=1}^{n}\left(\sum_{i=1}^{T} Te_{it}\right)^2}{\sum_{i=1}^{n}\sum_{t=1}^{T} e_{it}}\text{-}1\right]^2 \quad (7)$$

2.3.3 Hausman Test

Hausman test is used to determine the best regression model in panel data. The model obtained through the random effect model approach is compared with the fixed effect model approach [7]. The hypotheses used are as follows

The test statistic is expressed as follows:

$$\begin{aligned} W &= \hat{q}'\left[var(\hat{q})\right]^{-1}\hat{q} \\ &= \left(\hat{\beta}_{MET} - \hat{\beta}_{MEA}\right)'\left[var\left(\hat{\beta}_{MET} - \hat{\beta}_{MEA}\right)\right]^{-1}\left(\hat{\beta}_{MET} - \hat{\beta}_{MEA}\right) \end{aligned} \quad (8)$$

2.4 Testing Assumptions of Panel Data Regression Models

The equation obtained from the estimation of panel data regression becomes a good model if it statistically satisfies classical assumptions, namely the normality test, multicollinearity test, and heteroscedasticity test [9].

2.4.1 Normality test

The normality test is used to demonstrate that the sample data is derived from a population with a normal distribution or that the population data itself has a normal distribution. The normality test for residuals can formally be detected using the method developed by Jarque-Bera. The test statistic used is as follows [10]:

$$JB = \frac{n}{6}\left[S^2 + \frac{(K\text{-}3)^2}{4}\right] \quad (9)$$

2.4.2 Multicollinearity Test

The Multicollinearity test aims to examine whether there is a correlation between independent variables in a regression model. To detect the presence of multicollinearity in a regression model, one can examine the output in the correlation matrix among independent variables. If the values generated are less than 0.90, it can be stated that there is no multicollinearity. Conversely, if there are values greater than 0.90, multicollinearity is present [11].

2.4.3 Heteroscedasticity Test

The Heteroscedasticity test aims to examine whether there is or isn't a variance difference in residuals between one observation and another in a regression model [12]. The hypotheses used are as follows:

H_0 : No heteroscedasticity issue.
H_1 : There is a heteroscedasticity problem.

Statistical testing of the panel data regression equation is conducted when all classical assumptions are met, involving significance testing of the regression model (F-test), partial regression coefficient testing (t-test), and determination coefficient (R^2). The F-test for the significance of the regression model is performed to determine whether there is an overall influence of the independent variables on the regression model. This test helps establish a linear relationship between the dependent and independent variables [13]. Subsequently, a t-test is conducted to examine the individual significance of each independent variable on the dependent variable while assuming other variables remain constant [14]. Finally, determining the coefficient of determination is essential to measure a model's effectiveness in explaining the dependent variable. If the coefficient of determination approaches 1, the independent variables are increasingly influential in explaining the dependent variable[15].

3 Method

a. This type of research is applied research. The data used is secondary data obtained from the Central Statistics Agency (BPS) of West Sumatra Province. The data from 2017 to 2022 consists of 19 regencies/cities in West Sumatra Province. The dependent variable (Y) used in this study is the poverty gap index, and the independent variables consisting of economic growth (X_1), average per capita expenditure on food and non-food (X_2), expected years of schooling (X_3), human development index (X_4). The data processing results were obtained using EViews 12.
b. The data analysis steps that will be carried out in this study are as follows:
c. Determine the research variables, where the poverty gap index is the dependent variable in this study. In contrast, the independent variables consist of economic growth, average per capita expenditure on food and non-food, expected years of schooling, and human development index based on 19 districts/cities in 2017-2022.
d. Collect research data obtained from the Central Bureau of Statistics (BPS).
e. Estimate panel data regression models to describe the factors influencing the poverty gap index, using three estimation methods, namely the Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM). Next, select the best model using three specific tests, namely the Chow test, Hausman test, and Lagrange multiplier test.
f. Conduct classical assumption testing, including normality, multicollinearity, and heteroscedasticity tests.
g. Furthermore, testing the significance of the panel data regression model parameters with, among others, the F test, t test, and the coefficient of determination (R^2").
h. Interpret the model and draw conclusions from the analysis results.

4 Results and Discussion

The data used in this study consist of the poverty gap index for 19 regencies/cities in the West Sumatra Province as the dependent variable, with independent variables including economic growth (X_1), average per capita expenditure (X_2), years of schooling expectancy (X_3), and the Human Development Index (X_4). These data were obtained from the official publication of the Central Statistics Agency (Badan Pusat Statistik - BPS) of the West Sumatra Province. The characteristics of each variable and the descriptive statistics of the poverty gap index and its influencing factors, are presented in Table 1. Based on Table 1, the average poverty gap index for regencies/cities in West Sumatra Province from 2017 to 2022 is 0.826579. The highest recorded gap index is 2.84, while the lowest poverty gap index achievement is 0.07.

4.1 The Selection of Panel Data Regression Estimation Method

Specialized tests to select the best panel data regression model for a given problem include the Chow test, Lagrange Multiplier test, and Hausman test.

Tabel 1. Descriptive Statistics of the Poverty Gap Index and Its Influencing Factors in West Sumatra Province for the Years 2017-2022

Variable	N	Descriptive Statistics		
		Mean	Minimum	Maximum
Y	114	0.826579	0.07	2.84
X_1	114	9017819	2312713	47185099
X_2	114	1207496	766704	1864276
X_3	114	13.71719	12.07	16.54
X_4	114	72.27886	59.25	83.29

4.1.1 Chow Test

The data processing results obtained from the Chow test are presented in Table 2.

Tabel 2. The results of the Chow test

Effects Test	Statistic	Df	Prob.
Cross-section F	11.267545	(18.91)	0.0000
Cross-section Chi-square	133.618669	18	0.0000

Based on Table 2, it can be observed that the p-value < α at a significance level of α = 0.05. This indicates that the appropriate model to use is the fixed effect model.

4.1.2 Lagrange Multiplier Test

The data processing results obtained from the Lagrange multiplier test are presented in Table 3.

Table 3. Lagrange Multiplier Test Results

	Test Hypothesis		
	Cross section	Time	Both
Breusch Pagan	82.90970 (0.0000)	3.011212 (0.0827)	85.92092 (0.0000)

Based on Table 3, it can be seen that the cross-section Breusch-Pagan test has a p-value less than α at a significance level of α = 0.05. The cross-section Breusch-Pagan value is 0.0000 < 0.05, so we reject the H_0. Therefore, it can be concluded that the estimation method used in the LM test is the random effect model (REM).

4.1.3 Hausman Test

The data processing results obtained from the Hausman test are presented in Table 4. Based on Table 4, it can be seen that the p-value < α at a significance level of α = 0.05. The p-value is 0.0141 < 0.05, so we reject the H_0. This means that the appropriate model to use is the fixed effect model.

Table 4. Hausman Test Results

Test Summary	Chi-Sq. Statistic	Chi-Sq.df	Prob.
Cross-section random	12,486851	4	0,0141

After conducting the Chow, Lagrange Multiplier, and Hausman tests, the best model obtained is the fixed effect model estimation method.

4.2 Testing Assumptions of Panel Data Regression Models

After obtaining the best model in this study, the best model is the fixed effect model. The next step is to perform testing on the assumptions of panel data regression.

4.2.1 Normality Test

The results of data processing obtained from the normality test presented in Table 5.

Table 5. Results of Normality Test

Nilai Jarque-Bera	p-value	$\chi^2_{(\alpha.2)}$
0.926961	0.629090	5.99

Based on Table 5, it can be observed that the Jarque-Bera statistic is 0.926961, with a critical value of $\chi^2_{(0,05,2)}$ of 5.99 and a p-value of 0.629090 with significance α of 0.05. This indicates that Jarque-Bera < $\chi^2_{(0,05,2)}$ Therefore, it can be concluded that the normality assumption is satisfied.

4.2.2 Multicollinearity Test

The data processing results obtained from the multicollinearity test are presented in Table 6.

Table 6. Results of Multicollinearity Test

Variable	X_1	X_2	X_3	X_4
X_1	1	0.265702	0.536184	0.326008
X_2	0.265702	1	0.731245	0.875485
X_3	0.536184	0.731245	1	0.864244
X_4	0.326008	0.875485	0.864244	1

Based on Table 6, it can be seen that the correlation coefficients among independent variables are less than 0.9. This means that there is no multicollinearity in the obtained regression model.

4.2.3 Heteroscedasticity Test

The results of data processing obtained from the heteroscedasticity test are presented in Table 7.

Table 7. Results of the Heteroscedasticity Test

Variable	Coefficient	t-Statistic	Prob.
X_1	1.16×10^{-8}	-0.716321	0.4756
X_2	-1.2×10^{-7}	-0.97638	0.3315
X_3	0.081704	0.873222	0.3848
X_4	0.027831	1.059219	0.2923

Based on Table 7, it can be seen that the p-values for X_1, X_2, X_3, X_4 are 0.4756, 0.3315, 0.3848, and 0.2923, respectively, with significance level α f 0.05. This indicates that p-value > α. Therefore, it can be concluded that the decision is to accept H_0, meaning that heteroscedasticity is satisfied.

4.3 Statistical Testing of Panel Data Regression Equations

Statistical testing of panel data regression equations consists of several stages, including:

4.3.1 F-Test

The results obtained from testing the significance of the regression model are presented in Table 8.

Table 8. F-Test Results

Variable	F-Statistic	Probabilitas F	Ket
X_1 X_2 X_3 X_4	29.00020	0.000000	Significant

Based on Table 8, it can be observed that the p-value is 0.0000, which is less than 0.05 p-value < α. This means rejecting H_0, in other words, overall, the independent variables influence the poverty gap index in West Sumatra Province.

4.3.2 t-Test

The results obtained from the partial regression coefficient testing are presented in Table 9.

Table 9. Partial Regression Coefficient Test Results

Variable	Coefficient	t-Statistic	Prob	Keterangan
X_1	$3.23X\ 10^{-8}$	1.023101	0.3090	Not Significant
X_2	$6.75X\ 10^{-7}$	2.684903	0.0086	Significant
X_3	0.160912	0.881662	0.3803	Not Significant
X_4	-0.209718	-4.077205	0.0001	Significant

Based on Table 9, it can be concluded that the variable average per capita expenditure (X_2) and the human development index variable (X_4) are significant variables influencing the poverty gap index in West Sumatra Province. However, for the economic growth variable (X_1) and the years of schooling expectancy variable (X_3) these variables are not significantly influential on the poverty gap index in West Sumatra Province.

4.3.3 Coefficient of Determination

The following are the results for testing the coefficient of determination, as shown in Table 10.

Table 10. Coefficient of Determination Results

Variable	R-square
X_1 X_2 X_3 X_4	0.875172

Based on Table 10 above, the calculation results for the fixed effect model show that the influence of independent variables on the poverty gap index in West Sumatra Province is 0.875172. This can be interpreted as 87.5172% of the poverty gap index is explained by the independent variables, while other variables outside the scope of this research model explain the remaining 12.4828%.

5 Conclusion

The panel data regression model explaining the factors influencing the poverty gap index in West Sumatra Province for the years 2017-2022, using panel data regression analysis with the best estimation model, is the Fixed Effect Model (FEM). The obtained model is as follows:

$$Y = -0.33541 - 0.330251D_{1i} - 0.281482D_{2i} - 0.493164D_{3i} - 0.30458D_{4i} - 0.319441D_{5i} - 0.037827D_{6i} - 0.368D_{7i} - 0.92714D_{8i} - 0.17567D_{9i} - 0.0478D_{10i} - 0.5567D_{11i} + 0.06078D_{12i} + 0.80491D_{13i} - 0.30621D_{14i} + 0.80246D_{15i} + 1.155213D_{16i} + 1.090398D_{17i} + 0.570195D_{18i} + 0.000000675X_2 - 0.209718\,X_4$$

In the fixed-effect model, the influential independent variables are the average per capita expenditure, which has a positive effect, and the human development index, which harms the poverty gap index. Specifically, an increase of Rp1 in average per capita expenditure can lead to an increase of 0,000000675 in the poverty gap index in West Sumatra Province. Meanwhile, a 1% increase in the human development index can decrease by 0,20. The factors that significantly influence the poverty gap index in West Sumatra Province for the years 2017-2022 are average per capita expenditure (X_2) and the human development index (X_4).

References

[1] Aji, I. W., Adhi, S. W., & Taufiq, A. (2013). Implementasi Pengentasan Kemiskinan Melalui Gerdu Kempling di Kota Semarang. Journal of Politic and Government Studies, 216-225.

[2] Sumbar, B. P. S. (2023). Indikator Kesejahteraan Rakyat Provinsi Sumatera Barat. Padang. BPS Sumbar.

[3] Badan Pusat Statistika 2023. Kemiskinan dan Ketimpangan. Diakses pada 23 desember 2023.https://sumbar.bps.go.id/subject/23/kemiskinan-dan ketimpangan.html#subjekViewTab1.html

[4] Lestari, A., & Setyawan, Y. (2017). Analisis regresi data panel untuk mengetahui faktor yang mempengaruhi belanja daerah di provinsi Jawa Tengah. Jurnal Statistika Industri dan Komputasi, 2(01), 1-11.

[5] Yudiatmaja, F. 2013. Analisis Regresi dengan Menggunakan Aplikasi Komputer Statistika SPSS. Jakarta: Gramedia Pustaka Utama.

[6] D. N. Gujarati, Basic Econometrics. 2009

[7] Baltagi, B. H. (2008). Econometric Analysis of Panel Data. 3rd ed. Chichester. Wiley & Sons Ltd.

[8] Widarjono, Agus.2007, Ekonometrika Teori dan Aplikasi Ekonomi dan Bisnis. Ekonisia,Yogyakarta

[9] Basuki, Agus Tri. (2016). Analisis Data Panel Dalam Penelitian Ekonomi dan Bisnis: Dilengkapi dengan Penggunaan Eviews. Yogyakarta.

[10] Winarno, W. W. (2017). Analisis ekonometrika dan statistika dengan eviews.

[11] I. Ghozali and D. Ratmono. (2017). Analisis Multivariate dan Ekonometrika Teori, Konsep, dan Aplikasi dengan Eviews 10 Edisi 2 Semarang: Badan Penerbit Universitas Diponegoro.

[12] Indri, F. Z., & Putra, G. H. (2022). Pengaruh Ukuran Perusahaan Dan Konsentrasi Pasar Terhadap Kualitas Laporan Keuangan Pada Perusahaan Sektor Industri Barang Konsumsi Yang Terdaftar Di Bursa Efek Indonesia Pada Tahun 2016-2020. Jurnal Ilmu Manajemen, Ekonomi Dan Kewirausahaan, 2(2), 236-252.

[13] Montgomery, D.C., Peck, E. A., & Vining, G.G.(2021). Introduction to linear regression analysis. John Wiley & Sons.

[14] Sriyana, J. (2014). Metode regresi data panel. Yogyakarta: Ekonisia.

[15] Ghozali, Imam. 2011. Aplikasi Analisis Multivariate dengan Program IBM SPSS 19. Edisi Kelima. Program Doktor Ilmu Ekonomi Universitas Diponegoro. CV. NUR LINA.