Vygotsky: Jurnal Pendidikan Matematika dan Matematika
7 (2) August 2025, pp. 155 - 168
E-ISSN: 2656-5846 P-ISSN: 2656-2286
Journal Page is available to
https://jurnalpendidikan.unisla.ac.id/index.php/VoJ

Panel Data Regression Modeling of North Sumatra Province's Gross Regional Domestic Product for 2019-2023

Ester Maylani [1*], Rina Filia Sari [1]

[1] Department of Mathematics Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

**Email Correspondence:* ester0703211024@uinsu.ac.id

ARTICLE INFO

Article History

Received	:	11 Jul 2025
Revised	:	11 Aug 2025
Accepted	:	14 Aug 2025
Available Online	:	31 Aug 2025

Keywords:
Data Panel
Gross Regional Domestic Product (GRDP)
Economic Factors

Please cite this article APA style as:
Maylani, E. & Sari, R., F. (2025). Panel Data Regression Modeling of North Sumatra Province's Gross Regional Domestic Product for 2019-2023. *Vygotsky: Jurnal Pendidikan Matematika dan Matematika*, 7(2), pp. 155-168.

ABSTRACT

Regional economic growth is influenced by various factors that need to be analyzed accurately to support the formulation of effective policies. This study aims to analyze the influence of economic factors on the Gross Regional Domestic Product (GRDP) in North Sumatra Province. The main issue raised is the need for an appropriate model to understand the relationship between economic variables and GRDP. This study uses panel data from 33 districts/cities during the period 2019–2023 obtained from official sources. Through Chow, Hausman, and Lagrange Multiplier tests, the Fixed Effect model was selected. The results indicate that population size, number of poor people, and Human Development Index (HDI) significantly influence RDP.

1. Introduction

Development is a continuous process aimed at improving the quality of life of the community, which requires synergy between local government and the community as the main actors (Musolesi, 2025). Economic development is characterized by long-term increases in per capita income and the creation of more jobs (Nandita et al., 2019). as well as reductions in poverty rates and improvements in well-being (Silalahi et al., 2014). In this context, Regional Domestic Product (RDP) was selected as the focus of the panel data regression model because it is the primary indicator for measuring the success of regional economic development (Adityaningrum, 2024). RDP reflects the total gross value added of all economic units in a region, providing a comprehensive overview of regional economic performance (Zebua, 2022). for example, the GRDP of Medan City in 2020 was recorded at 153,669.83 billion rupiah, down from 156,780.58 billion rupiah in the

E-ISSN: *2656-5846*
P-ISSN: *2656-2286*

previous year due to the impact of the COVID-19 pandemic (Hasibuan et al., 2022). This decline has had widespread effects on socio-economic aspects such as increased poverty and unemployment, necessitating in-depth studies on the factors influencing economic growth in Medan City as the economic center of North Sumatra Province (Hasibuan et al., 2022).

Economic growth and poverty analysis using a panel data regression approach is considered appropriate because it can capture the dynamics of data in two dimensions, namely between individuals (cross-section) and over time (time series) (Castro, 2024; Amaliah, 2020). Unlike ordinary regression models, which only analyze data statically and often ignore differences between observation units or the influence of time, panel data regression is able to reveal more complex and realistic relationships between variables. In panel data regression, there are three main models commonly used, namely the Common Effect Model (CEM), Fixed Effect Model (FEM), and Random Effect Model (REM) (Gujarati, 2021). Selecting the most appropriate model is crucial to ensure that the analysis results reflect real-world conditions (Hidaka, 2024). and for this purpose, the Chow, Hausman, and Lagrange Multiplier tests are used (Hutagalung, 2022). This method has been widely applied in regional economic studies to identify the determinants of economic growth (Ratnasari, 2023). Therefore, panel data regression is relevant for analyzing the relationship between GRDP and other economic variables in North Sumatra Province (Hasibuan et al., 2022).

Previous research has provided an important foundation for understanding the relationship between GRDP and poverty levels, such as the study by Zebua (2022)" which showed that GRDP had a negative and significant effect on poverty in North Sumatra during 2011–2020, meaning that an increase in GRDP can reduce poverty levels. This finding aligns with the view that strong economic growth can be a key instrument in poverty alleviation (Elviera & Irawan, 2020). However, economic growth alone is insufficient if it is not accompanied by income distribution and inclusive job creation. Amaliah (2020) also emphasize the need for a more in-depth analysis of other factors that influence GRDP and its impact on poverty. Previous studies have suggested including additional socioeconomic variables such as unemployment and labor force participation in further research. Unfortunately, there are still limitations in studies that specifically develop panel data regression models based on GRDP at the provincial and city/district levels in North Sumatra, particularly in Medan City, which has unique and strategic economic characteristics.

Based on the above, the researcher intends to conduct a study titled "Panel Data Regression Modeling of Regional Domestic Product in North Sumatra Province from 2019 to 2023." This study will focus on applying panel data regression methods to analyze the relationship between GRDP and several socioeconomic indicators. The dependent variable in this study is GRDP at constant prices (ADHK), while the independent variables consist of population size, number of poor people, human development index (HDI), open unemployment rate (OUR), and labor force participation rate (LFPR). The data used are secondary data obtained from the Central Statistics Agency (BPS) and related agencies, covering 33 districts/cities in North Sumatra Province during the period 2019 to 2023. By utilizing panel data, this study is expected to provide a more accurate and dynamic picture of regional economic development. Additionally, the results of the regression analysis will be used to determine the

 E-ISSN: *2656-5846*
P-ISSN: *2656-2286*

best estimation model among CEM, FEM, and REM. The selection of the model is based on statistical tests that have been validated in econometric literature. It is hoped that the findings of this study will contribute to data-driven and more targeted economic policy-making.

This study also aims to further analyze how economic growth reflected in GRDP can be used as a tool to address poverty issues, particularly in Medan City. Using a panel data regression approach, it is expected to identify which variables significantly influence GRDP and the direction and magnitude of their effects. Through this understanding, local governments can design more effective policies to enhance economic growth while reducing poverty rates sustainably.

As the provincial capital, Medan plays a vital role in driving economic growth in North Sumatra, so development strategies in this city will greatly determine the direction of regional development. This study is also expected to serve as a reference for other researchers in developing studies on regional economic dynamics using an econometric approach. By looking at the estimation results and the selection of the best model, this study can provide an empirical basis for inclusive and equitable economic development planning. The research findings can also be utilized by other stakeholders, such as academics and regional planning institutions, to support evidence-based policy-making. Therefore, this study holds strategic value in efforts to promote sustainable economic development in North Sumatra Province.

2. Method

2.1. Panel Data Regression

Panel data regression is a type of regression based on panel data used to observe the correlation between one dependent variable and one or more independent variables (Nguyen, 2023). The variables that have an impact are called independent variables, while the variables that are affected are called dependent variables. Panel data regression is an expansion of multiple linear regression. When panel data is the type of data, such as time series and cross-sectional data, panel data regression is used (Tohir & Utomo, 2023).

2.2. Panel Data Regression Model

In determining the panel data regression model, there are several alternative models that can be solved with panel data of Equation 1 (Elviera & Irawan, 2020),

$$Y_{it} = \beta_0 + \sum_{k=2}^{J} \beta_j X_{jit} + \varepsilon_{it} \tag{1}$$

Description:
i = $1,2,\dots,N$
t = $1,2,\dots,T$
N = number of cross-section units
T = number of time series data
Y_{it} = dependent variable for cross section i and time series t
x_{it} = independent variable j for cross section i and time series t
β_{it} = estimated parameter
ε_{it} = population disturbance element
J = number of estimated parameters

E-ISSN: 2656-5846
P-ISSN: 2656-2286

2.3. Common Effect Model (CEM)

The CEM method is the simplest approach in determining panel data regression model estimates, because this approach combines all data, both cross section data and time series data (Delmar, 2022). CEM assumes that the intercept and slope in the cross section and time series units are the same. Generally, the model equation (Hasibuan et al., 2022) is written as shown Equation 2 (Kosmaryati et al., 2019),

$$Y_{it} = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_j X_{jit} + \varepsilon_{it} \quad (2)$$

Explanation:

Y_{it} = dependent variable for cross section i and time series t
β_0 = intercept model
β_j = slope regresi to-j
X_{jit} = independent variable to-j for cross section i and time series t
J = number of independent variables to- j ; $j = 1,2, \dots, k$
i = regional unit to e- i ; $i = 1,2, \dots, n$
t = time period to t ; $t = 1,2, \dots, p$

2.4. Fixed Effect Model (FEM)

The FEM assumes that the slope coefficient is constant but the intercept is not constant. The method that can be used to estimate the model in FEM is the Least Square Dummy Variable method, often referred to as LSDV (Verlita, 2022). In the LSDV method, estimation is performed by inserting a dummy variable used to explain the different intercept values due to differences in unit values (Zebua, 2022). The regression model equation in FEM is written as Equation 3 and the explanation of this equation is the same as the Equation 2,

$$Y_{it} = \beta_{0it} + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_j X_{jit} + \varepsilon_{it} \quad (3)$$

2.5. Random Effect Model (REM)

There are two methods that can be used to estimate REM, namely the LSDV method and the Generalized Least Square (GLS) method (Heydari, 2021). Because the LSDV method adds variables, this results in a large number of variables in the equation compared to the amount of data. In addition, the degree of freedom is not fulfilled, so the LSDV method cannot be used. Therefore, it is necessary to perform estimation using the GLS method, as this method performs estimation directly without adding dummy variables (Silalahi et al., 2014). Several model equations in REM are described as Equation 4,

$$Y_{it} = \beta_{0it} + \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_j X_{jit} + (\mu_i + \varepsilon_{it}) \quad (4)$$

2.6. Gross Regional Domestic Product (GRDP) at Constant Prices

Gross Regional Domestic Product (GRDP) reflects the total value of goods and services produced in a region, both by local residents and foreigners living in that regio (Szwacka-Mokrzycka, 2020) and is used to measure the economic contribution of a region to the national economy. In this study, RDP based on constant prices (ADHK) is used because this method eliminates the influence of inflation, thereby reflecting real economic growth more accurately (Kosmaryati dkk., 2019 ; Brzozowska-Rup, 2020). GRDP is calculated using three approaches:

E-ISSN: 2656-5846
P-ISSN: 2656-2286

income, expenditure, and production, and serves as the primary indicator for assessing the economic progress of a region (Kosmaryati et al., 2019).

This study uses PDRB data for North Sumatra Province from 2019 to 2023, covering 33 districts/cities with unbalanced characteristics, as some regions have a much more dominant economic contribution than others, such as Medan City. This sample representation imbalance has the potential to influence the results of the model implementation, making a focus on Medan City particularly important to identify the main factors influencing GRDP in the context of poverty (Priyatno, 2023).

3. Results and Discussion

3.1. Descriptive Statistical Analysis

An overview of the variables used in this study, including minimum, maximum, average, and standard deviation values. The data show significant variations in variables such as GDP, population, number of poor people, human development index (HDI), open unemployment rate (OUR), and labor force participation rate (LFPR) in Medan City during the period 2019-2023, as presented in Table 1.

Table 1. Descriptive Statistical Analysis

Variable	Minimum	Maximum	Mean	Standar Deviasi
PDRB (Y)	850.79	173445.69	17.175,03	29.070,25
JP (X1)	48935.00	2494512,00	452.752,0	510.947,6
JPM (X2)	4.01	193,03	38.89176	34.2143
IPM (X3)	61.14	82,19	71,31921	4.509589
TPT (X4)	0.19	11,50	4.814061	2.739589
TPAK (X5)	51.83	88,95	72.93885	7.9737008

3.2. Multicollinearity Test

The multicollinearity test was conducted by examining the Variance Inflation Factor (VIF) value, where a VIF value greater than 10 indicates multicollinearity in the model. The results of the multicollinearity test are shown in Table 2.

Table 2. Multicollinearity Test

Variable	VIF
X_1	7.773710
X_2	7.006006
X_3	1.607456
X_4	2.327282
X_5	1.882673

Table 2 informs that the variable X_1, X_2, X_3, X_4, and X_5 have a VIF value < 10. This indicates that there is no multicollinearity problem between the independent variables in the Regional Domestic Product (RDP) in North Sumatra in 2019-2023. These results indicate that the data used can provide reliable estimates without distortion due to multicollinearity.

3.3. First Stage Panel Data Regression Estimation

Panel data regression estimation will be estimated using three models, namely CEM, FEM, and/or REM. The parameter estimation process for each model has

E-ISSN: 2656-5846
P-ISSN: 2656-2286

been described previously.

3.3.1. *Common Effect Model (CEM)*

The panel data regression model for CEM that produces estimates is presented in Table 3.

Table 3. Common Effect Model (CEM)

Variable	Parameter Estimates	Standard Deviation	t-value	p-value
C	-81.72	16.54	-4.93	0.00
X_1	0.02	0.00	6.02	0.00
X_2	427.11	54.75	7.80	0.00
X_3	1142.3	198.99	5.74	0.00
X_4	-646.56	394.14	-1.64	0.10
X_5	-90.58	121.79	-0.74	0.45

The panel data regression model estimation for CEM of Table 3 can be written in Equation 5,

$$\hat{Y}_{it} = -81.725 + 0.023283X_{1it} + 427.11_{2it} + 1142.3X_{3it} + -646.56X_{4it} + -90.585X_{5it} \quad (5)$$

3.3.2. *Fixed Effect Model (FEM)*

Panel data regression model for FEM that produces estimates as shown in Table 4.

Table 4 Fixed Effect Model (FEM)

Variable	Parameter Estimates	Standard Deviation	t-value	p-value
C	-25.90	15.55	-1.66	0.09
X_1	0.01	0.00	3.36	0.00
X_2	442.7	82.55	-0.53	0.59
X_3	669.7	184.8	3.62	0.00
X_4	-85.58	154.2	-0.55	0.57
X_5	-0.22	29.97	-0.00	0.99

The panel data regression model estimation for FEM gives Equation 6,

$$\hat{Y}_{it} = -25.900_i + 0.01271X_{1it} + 442.7_{2it} + 669.7X_{3it} + -85.58X_{4it} + -0.2276X_{5it} \quad (6)$$

3.3.3. *Random Effect Model (REM)*

Panel data regression model for REM that produces estimates as Table 5.

Table 5. Random Effect Model (REM)

Variable	Parameter Estimates	Standard Deviation	t-value	p-value
C	-70.22	13.17	-5.33	0.00
X_1	0.02	0.00	6.59	0.00
X_2	312.22	56.51	5.52	0.00
X_3	909.88	180.58	5.03	0.00
X_4	-89.57	162.68	-0.55	0.58
X_5	7.77	32.214	0.24	0.80

The panel data regression model estimation for REM has mathematical equation

 E-ISSN: 2656-5846
P-ISSN: 2656-2286

as Equation 7,

$$\hat{Y}_{it} = -70.220 + 0.022583X_{1it} + 312.22_{2it} + 909.88X_{3it} + -89.576X_{4it} + 77759X_{5it} \quad (7)$$

3.4. Best Model Approach Phase One

After obtaining three different approach models, the next step is to determine which model is most suitable for the data among the three models. Model selection tests are conducted using the Chow test, Hausman test, and Lagrange Multiplier test.

3.4.1. Chow Test

The Chow test is used to select the best model between FEM and CEM. After obtaining an initial estimate by estimating FEM, the next step is to perform the Chow test. The hypothesis is as follows,

H_0 : $\beta_{01} = \beta_{02} = \cdots = \beta_{0n} = 0$ (the best model is CEM)
H_1 : at least one $\beta_{0i} \neq 0$ (the best model is FEM)

The test results obtained F_{count} is 177.73. Meanwhile, the value of F_{table} with degrees of freedom N1 = 32 and degrees of freedom N2 = 127 at the actual level (a = 0.05), is 1.628. Therefore, reject H_0 because $F_{count} > F_{table}$. Therefore, the Chow test results prove that the best model selected is Fixed Effect Model (FEM). With this, the testing process continues to the Hausman Test.

3.4.2. Hausman Test

The Hausman test is used to select the best model between FEM and REM. This test is performed if the previous Chow test results indicate that FEM is the best model. The hypothesis uses,

H_0 : $corr\left(X_{it,}\varepsilon_{it}\right) = 0$ (REM are better),
H_1 : $corr\left(X_{it,}\varepsilon_{it}\right) \neq 0$ (FEM are better)

The results of the Hausman test conduces that the value of W is 33.398. To obtain the table value X^2 can be seen in the Chi-Square table with degrees of freedom at the significant level (a = 0.05), then the value obtained is X^2_{table} = 11.070. Thus, it is obtained that $W > X^2_{table}$ therefore reject H_0 which means that the FEM model is better.

3.4.3. Lagrange Multiplier Test

The Lagrange Multiplier test was used to select the best model between CEM and REM. With the help of R software, this study used the Lagrange Multiplier test (LM_{lag}). The hypothesis for the Lagrange Multiplier test is as follows,

H_0 : $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (CEM are better)
H_1 : there is at least one $\sigma_i^2 \neq \sigma^2$(REM are better), $i = 1,2,\ldots,n$

The result of the Lagrange Multiplier Test has $p - value$ which is 0.0000. This result obtained $p - value < a$ so that reject H_0 which means that the better

E-ISSN: 2656-5846
P-ISSN: 2656-2286

estimation model is the Random Effect Model (REM). According to the Chow Test, Hausman Test, and Lagrange Multiplier Test, it was obtained that the best model is the Fixed Effect Model (FEM). The best equation obtained is,

$$\hat{Y}_{it} = -25.900_i + 0.01271X_{1it} + -44.27_{2it} + 669.7X_{3it} + -85.58X_{4it} + -0.2276X_{5it} \quad (8)$$

3.5. Significance Test of First Stage Model Parameters

3.5.1. Simultaneous Regression Coefficient Significance Test (F test)

The F test is used to determine the linear relationship between all independent variables and the dependent variable. The hypothesis uses,

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_k = 0$,
H_1 : at least one $\beta_p \neq 0$ with $p = 1,2, \ldots, k$

The test results were obtained $F_{count} = 30.909$ as large as and for $F_{table} = F_{(a,k-1,n-k)} = F_{0,05;5,127} = 2.29$. The FEM output has that $p - value(0.0000) <$ actual level ($a = 0.05$). The value obtained is $F_{count} = 30.909 > F_{table} = 2.29$ therefore it is concluded to reject H_0 which means that there is at least one independent variable that has a significant effect on the dependent variable.

3.5.2. Partial Regression Significance Test (t-test)

The t-test is used to determine the causal relationship between independent variables individually and the dependent variable. The results of the partial model significance test are shown in Table 6.

Table 6. Result of t-test

Variable	Parameter Estimates	Standard Deviation	t-value	p-value
C	-25.90	15.55	-1.66	0.09
X_1	0.01*	0.00	3.36	0.00
X_2	- 442.7	82.55	-0.53	0.59
X_3	669.7*	184.8	3.62	0.00
X_4	-85.58	154.2	-0.55	0.57
X_5	-0.22	29.97	-0.00	0.99

The t-test of Table 4 shows that variable X_1 (Number of Residents), X_3 (Human Development Index) have a value $p < 0,05$, so that both variables have a significant effect on GRDP partially. The variable X_2 (Number of Poor People) X_4 (Open Unemployment Rate) and X_5 (The Labor Force Participation Rate has a value of $p > 0.05$, so that it does not significantly affect the dependent variable. Therefore, the variable X_2, X_4, and X_5 can be considered for exclusion from the advanced regression model.

3.6. Second Stage Panel Data Regression Estimation

Similar to the first stage panel data regression estimation, the second stage follows the same process as the first stage.

 E-ISSN: 2656-5846
P-ISSN: 2656-2286

3.6.1. *Common Effect Model (CEM)*

The panel data regression model for CEM is presented in Table 7.

Table 7. Common Effect Model (CEM)

Variable	Parameter Estimates	Standard Deviation	t-value	p-value
C	-46.82	14.27	-3.28	0.00
X_1	0.05	0.00	28.01	0.00
X_3	576.30	204.58	2.81	0.00

The panel data regression model estimation for CEM stage two can be written in Equation 9,

$$\hat{Y}_{it} = -46.828 + 0.050584X_{1it} + 576.30_{2it} \tag{9}$$

3.6.2. *Fixed Effect Model (FEM)*

The panel data regression model for CEM is presented in Table 8.

Table 8. Fixed Effect Model (FEM)

Variable	Parameter Estimates	Standard Deviation	t-value	p-value
C	-33.190	11.150	-2.873	0.004753
X_1	0.01286	0.003735	3.443	0.000775
X_3	721.5	168.8	4.276	0.0000366

The panel data regression model estimation for the second stage of the FEM, as derived from the FEM results presented above, is expressed in Equation 10.

$$\hat{Y}_{it} = -33.190_i + 0.01286X_{1it} + 721.5_{2it} \tag{10}$$

3.6.3. *Random Effect Model (REM)*

The panel data regression model for REM is presented in Table 9.

Tabel 9. Random Effect Model (REM)

Variable	Parameter Estimates	Standard Deviation	t-value	p-value
C	-43.967	12.808	-3.3547	0.0007946
X_1	0.031356	0.0030856	10.1620	0.00000000000000022
X_3	644.23	181.55	3.5484	0.0003875

The panel data regression model estimation for the second stage of REM, as shown in Table 9, is expressed in Equation 11.

$$\hat{Y}_{it} = -43.967 + 0.031356X_{1it} + 644.23_{2it} \tag{11}$$

3.7. *Best Model Approach Phase Two*

After obtaining three different approach models, the next step is to determine which model is suitable for the data among the three models. The model selection test is conducted using the Chow test, Hausman test, and Lagrange Multiplier test.

3.7.1. *Chow Test*

The Chow test is used to select the best model between FEM and CEM. After obtaining an initial estimate by estimating FEM, the next step is to perform the

E-ISSN: *2656-5846*
P-ISSN: *2656-2286*

Chow test. The hypothesis uses,

H_0 : $\beta_{01} = \beta_{02} = \cdots = \beta_{0n} = 0$ (the best model is CEM)
H_1 : at least one $\beta_{0i} \neq 0$ (the best model is FEM

The test results indicate that F_{count} is 253.31. Meanwhile, the value F_{table} with degrees of freedom N1= 32 with degrees of freedom N2=130 pada actual level ($a =$ 0.05), is 1.533. Thus, reject H_0 because $F_{count} > F_{table}$. Therefore, the Chow test results prove that the best model selected is Fixed Effect Model (FEM). With this, the testing process continues to the Hausman Test.

3.7.2. Hausman Test

The Hausman test is used to select the best model between FEM and REM. This test is performed if the previous Chow test results indicate that FEM is the best model. The hypothesis uses,

H_0 : $corr\,(X_{it,}\varepsilon_{it}) = 0$ (REM are better),
H_1 : $corr\,(X_{it,}\varepsilon_{it}) \neq 0$ (FEM are better)

The Hausman test results yield a value of $W = 76.283$. To obtain the critical chi-square value, the Chi-Square table is consulted with degrees of freedom equal to 2 and a significance level of $a = 0.05$. The value obtained is $X^2_{table} = 5{,}991$. Thus, we obtain $W > X^2_{table}$ so that reject H_0 which means that the FEM model is better.

3.7.3. Lagrange Multiplier Test

The Lagrange Multiplier test was used to select the best model between CEM and REM. With the help of R software, this study used the Lagrange Multiplier test (LM_{lag}). The hypothesis uses,

H_0 : $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$ (CEM are better)
H_1 : there is at least one $\sigma_i^2 \neq \sigma^2$ (REM are better), $i = 1,2,\ldots,n$

The Lagrange Multiplier Test yields a value of 0.0000, indicating that the p-value is less than α, leading to the rejection of the null hypothesis in favor of the REM. According to the Chow Test, Hausman Test, and Lagrange Multiplier Test, it was found that the best model was the FEM. The best equation obtained was,

$$\hat{Y}_{it} = -33.190_i + 0.01286X_{1it} + 721.5_{2it} \quad (12)$$

3.8. Significance Test of Model Parameters in Stage Two

3.8.1. Simultaneous Regression Coefficient Significance Test (F test)

The F test is used to determine the linear relationship between all independent variables and the dependent variable. The hypothesis is as follows,

H_0 : $\beta_1 = \beta_2 = \cdots = \beta_k = 0$,
H_1 : at least one $\beta_p \neq 0$ with $p = 1,2,\ldots,k$

E-ISSN: 2656-5846
P-ISSN: 2656-2286

The Lagrange Multiplier Test result is 0.0000, indicating that the p-value is less than α. This leads to the rejection of the null hypothesis, suggesting that the REM is not the preferred specification. Considering the outcomes of the Chow Test, Hausman Test, and Lagrange Multiplier Test collectively, the FEM is identified as the most appropriate estimation model. The best equation obtained was,

$$\hat{Y}_{it} = -33{,}190_i + 0.01286X_{1it} + 721{,}5_{2it} \tag{13}$$

3.8.2. *Partial Regression Significance Test (t-test)*

The t-test is used to determine the causal relationship between independent variables individually and the dependent variable. The results of the partial model significance test can be shown in Table 10.

Table 10 Partial Regression Significance Test (t-test)

Variable	Parameter Estimates	Standard Deviation	t-value	p-value
C	-33.190	11.150	-2.873	0.004753
X_1	0.01286*	0.003735	3.443	0.000775
X_3	721.5*	168.8	4.276	0.0000366

The actual level used is 0.05. The partial regression significance test (t-test) from Table 10 shows that variable X_1 (Population), X_3 (Human Development Index) has a value of $p - value < 0.05$. It can be said that each of these independent variables has a significant effect on GRDP.

3.9. *Coefficient of Determination (R^2)*

The panel data regression model for the multiple R-squared values is presented in Table 11. The coefficient of determination (R^2) is shown in Table 11.

Tabel 11. Coefficient of Determination (R^2)

multiple R-squared	
CEM	0.86
FEM	0.9974
REM	0.47103

The test results yield a multiple R-squared value of 0.9974, indicating that the independent variables explain 99.74% of the variation in the dependent variable, while the remaining 0.26% is attributed to factors not included in the model. The best model is selected based on the highest R-squared value, the FEM is chosen as the most appropriate specification.

In applying models to simulation processes, a number of substantive challenges often arise, particularly those related to data imbalance. This imbalance has the potential to cause bias in parameter estimation, which can ultimately reduce the accuracy of the analysis results. Therefore, selecting the right analysis method is essential, given that not all statistical approaches are capable of handling data complexity and dynamics optimally. For example, panel data regression models may have limitations in fully representing temporal variation. Additionally, hardware and software limitations can also pose significant challenges, especially when dealing with large volumes of data and complex data structures. Therefore, a comprehensive methodological and technical evaluation is necessary before simulation implementation to ensure the validity and reliability

E-ISSN: 2656-5846
P-ISSN: 2656-2286

of the results obtained.

4. Conclusion

The Chow, Hausman, and Lagrange Multiplier test results collectively indicate that the FEM is the most appropriate model for this study. The panel data regression model with FEM estimation is,

$$\hat{Y}_{it} = -33{,}190_i + 0.01286X_{1it} + 721{,}5_{2it}$$

In the panel data regression modeling stage, it was found that several variables had a significant effect on Gross Regional Domestic Product (GRDP), namely: Population X_1, Number of Poor People X_2, and Human Development Index X_3. Meanwhile, the Open Participation Rate (OPR) variable X_4, and Labor Force Participation Rate (LFPR) X_5, not significant to GRDP in the model used.

Contributions

The first author is responsible for the process of collecting, processing, analyzing data, and compiling the results into a scientific paper. Meanwhile, the second and third authors acted as advisors who provided guidance during the writing of this article.

Acknowledgments

We would like to express our gratitude to all parties who have provided support in the preparation, development, and writing of this article.

Conflict of Interest Statement

As the author, I declare that there are no conflicts of interest or other interests that conflict with this research. All research results are compiled objectively and free from the influence of any party that could affect the validity or independence of the findings.

References

Adityaningrum, A. (2024). Robust Least Median of Square Modelling Using Seemingly Unrelated Regression With Generalized Least Square on Panel Data for Tuberculosis Cases. *Barekeng*, *18*(4), 2293–2306. https://doi.org/10.30598/barekengvol18iss4pp2293-2306

Amaliah, E. N., Darnah, D., & Sifriyani, S. (2020). Regresi data panel dengan pendekatan common effect model (CEM), fixed effect model (FEM) dan random effect model (REM)(Studi kasus: persentase penduduk miskin menurut Kabupaten/Kota di Kalimantan Timur tahun 2015-2018). *ESTIMASI: Journal of Statistics and Its Application*, *1*(2), 106–115.

Brzozowska-Rup, K. (2020). Modelling road traffic safety indices by means of regression with panel data. *Engineering Management in Production and Services*, *12*(4), 40–51. https://doi.org/10.2478/emj-2020-0026

Castro, J. R. (2024). Active Modeling of Solar Panel Performance using K-Means and Multiple Linear Regression with Environmental Variables and Historical Data. In *Iop Conference Series Earth and Environmental Science* (Vol. 1434, Issue 1). https://doi.org/10.1088/1755-1315/1434/1/012010

E-ISSN: 2656-5846
P-ISSN: 2656-2286

Delmar, F. (2022). Modeling new-firm growth and survival with panel data using event magnitude regression. *Journal of Business Venturing*, *37*(5). https://doi.org/10.1016/j.jbusvent.2022.106245

Elviera, C. D., & Irawan, D. (2020). Pengaruh Indeks Pembangunan Manusia Dan Produk Domestik Bruto (PDRB) Terhadap Kemiskinan Provinsi Sumatera Utara. *Bisnis-Net Jurnal Ekonomi Dan Bisnis*, *3*(2), 103–109.

Gujarati, D. N. (2021). *Essentials of econometrics*. Sage Publications.

Hasibuan, R. R. A., Kartika, A., Suwito, F. A., & Agustin, L. (2022). Pengaruh Produk Domestik Regional Bruto (PDRB) terhadap Tingkat Kemiskinan Kota Medan. *Reslaj: Religion Education Social Laa Roiba Journal*, *4*(3), 683–693.

Heydari, A. (2021). Application of nonlinear mathematical modeling in panel data analyses of the pecking order theory in capital structure using multiple regression: Evidence from the tehran stock exchange. *International Journal of Nonlinear Analysis and Applications*, *12*(2), 2383–2401. https://doi.org/10.22075/ijnaa.2021.5376

Hidaka, K. (2024). Regression modeling of subjective probability of river inundation with epistemic uncertainty: a short-term panel data analysis. *Natural Hazards*, *120*(7), 6343–6375. https://doi.org/10.1007/s11069-024-06446-y

Hutagalung, I. P. (2022). Analisis Regresi Data Panel Dengan Pendekatan Common Effect Model (CEM), Fixed Effect Model (FEM) dan Random Effect Model (REM)(Studi Kasus: IPM Sumatera Utara Periode 2014–2020). *FARABI: Jurnal Matematika Dan Pendidikan Matematika*, *5*(2), 217–226.

Kosmaryati, K., Handayani, C. A., Isfahani, R. N., & Widodo, E. (2019). Faktor-faktor yang mempengaruhi kriminalitas di Indonesia tahun 2011-2016 dengan regresi data panel. *Indonesian Journal of Applied Statistics*, *2*(1), 10–20.

Musolesi, A. (2025). Modelling green knowledge production and environmental policies with semiparametric panel data regression models. *Empirical Economics*, *68*(1), 327–352. https://doi.org/10.1007/s00181-024-02634-8

Nandita, D. A., Alamsyah, L. B., Jati, E. P., & Widodo, E. (2019). Regresi data panel untuk mengetahui faktor-faktor yang mempengaruhi PDRB di Provinsi DIY tahun 2011-2015. *Indonesian Journal of Applied Statistics*, *2*(1), 42–52.

Nguyen, M. C. (2023). Estimation of Systems of Regression Equations with Unbalanced Panel Data and Random Effects: Modelling Food-Expenditure Shares in Vietnamese Households. In *Contributions to Economics* (pp. 609–625). https://doi.org/10.1007/978-981-99-4902-1_20

Priyatno, D. (2023). *Olah data sendiri analisis regresi linier dengan SPSS dan analisis regresi data panel dengan Eviews*. Penerbit Andi.

Ratnasari, V. (2023). Statistical modeling to analyze factors affecting the middle-income trap in Indonesia using panel data regression. In *Methodsx* (Vol. 11). https://doi.org/10.1016/j.mex.2023.102379

Silalahi, D., Sitepu, R., & Tarigan, G. (2014). Analisis ketahanan pangan provinsi Sumatera Utara dengan metode regresi data panel. *Saintia Matematika*, *2*(3), 237–251.

Szwacka-Mokrzycka, J. (2020). The panel data regression concept in consumption modelling. *Economic Annals Xxi*, *185*(9), 61–69. https://doi.org/10.21003/EA.V185-06

Tohir, S. S., & Utomo, Y. P. (2023). Analisis Faktor-Faktor Yang Mempengaruhi Investasi Di Provinsi Banten 2016-2020. *Eprints.Ums.Ac.Id*.

E-ISSN: *2656-5846*
P-ISSN: *2656-2286*

Verlita, T. (2022). Panel data regression modeling of net enrollment rate at senior high school/equal levels in NTT Province 2015-2019. In *Aip Conference Proceedings* (Vol. 2641). https://doi.org/10.1063/5.0117382

Zebua, T. N. (2022). Pengaruh Produk Domestik Regional Bruto (PDRB) Terhadap Tingkat Kemiskinan Di Provinsi Sumatera Utara (2011-2020). *Jurnal Syntax Fusion*, *2*(03), 439–453.