Jurnal Varian
Vol. 8, No. 2, April 2025, pp 179–188
e-ISSN: 2581-2017

Panel Data Regression Modeling with Weighted Least Squares Method Using Fair Weights

Muhammad Ferdiansyah, Raupong Raupong, Siswanto Siswanto
Universitas Hasanuddin, Makassar, Indonesia

Article Info

Article history:

Received : 08-19-2024
Revised : 02-17-2025
Accepted : 07-23-2025

Keywords:

Fair Weighting;
Heteroscedasticity;
Life Expectancy;
Weighted Least Square.

ABSTRACT

Panel data regression is a robust method for analyzing relationships between dependent and independent variables by combining time-series and cross-sectional data. Its reliability hinges on key assumptions, particularly homoscedasticity. Violations, known as heteroscedasticity, lead to inefficient estimates and biased inference, as estimators fail to meet the best linear unbiased estimator criteria. The Weighted Least Squares (WLS) method addresses heteroscedasticity by weighting observations based on the inverse of their variance. WLS assumes prior knowledge of the heteroscedasticity structure, which is often impractical, creating a gap in evaluating its effectiveness compared to alternative methods. The purpose of this study is to examine life expectancy in South Sulawesi as the dependent variable, with expected years of schooling, per capita expenditure, and average years of schooling as independent variables. The research method used WLS with reasonable weighting, successfully addressing heteroscedasticity. The fixed-effects model was identified as the most appropriate, with an R-squared of 99.45%. The model explained life expectancy. Results show all variables positively and significantly influence life expectancy. In conclusion, the WLS method effectively overcomes heteroscedasticity in panel data regression, providing reliable estimators. This study emphasizes the significance of method selection in panel data analysis and provides guidance for policymakers seeking to enhance life expectancy in South Sulawesi.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v8i2.4392

Corresponding Author:

Raupong Raupong,
Department of Statistics, Universitas Hasanuddin
Email: raupong.stat.uh@gmail.com

How to Cite:
Ferdiansyah, M., Raupong, R., & Siswanto, S. 2025. Panel Data Regression Modeling with Weighted Least Squares Method Using Fair Weights. *Jurnal Varian*, *8*(2), 179-188.

A. INTRODUCTION

Regression analysis has become one of the widely used statistical methods in research. This development cannot be separated from the need for statistical methods capable of providing a variety of data types between units (cross-sections), between times (time series), and in combination, the two data are called panel data (Pramusinto & Daerobi, 2020). Panel data regression analysis explains regression analysis based on panel data to observe the relationship between a dependent variable and one or more independent variables. In panel data regression, there are three good deterministic models to use, which are the general effects model, fixed effects model, and random effects model (Du et al., 2025).

Some of these advantages include that panel data provides more informative data, reduces collinearity between variables, is more efficient, and can better detect the measurement of unobservable effects in cross-sections or time series with high efficiency (Abdussamad et al., 2024). The combination of cross-section and time series analysis will yield more accurate and reliable results than either method alone (Lv & Wu, 2019). By increasing the degrees of freedom and reducing collinearity between explanatory variables, the data is better able to improve estimation performance (Hsiao, 2022).

However, challenges arise when analyzing panel data, particularly concerning heteroskedasticity. When the classical assumptions are not met, it necessitates additional steps to address emerging issues. Various methods can be employed, including transformations (Rosalina et al., 2023). However, there is always the possibility that the data still does not meet the assumptions, so special methods are needed to overcome the problem. An example is when the data is heterogeneous. Previous research has largely discussed panel data regression to analyze socio-economic phenomena using fixed or random effects models without paying particular attention to the problem of heteroscedasticity which can affect the efficiency of parameter estimation. Several studies, such as those by Alamsyah et al. (2022) and Suhendra et al. (2020), employ panel regression to identify social factors, including poverty and income inequality, but have not explicitly implemented weighting methods to address violations of classical assumptions. Meanwhile, Nisa et al. (2020) and Rosalina et al. (2023) have begun to apply Weighted Least Squares (WLS) to address heteroscedasticity, but without a specific weighting approach such as fair weighting.

The consequence of heterogeneity is that the obtained OLS estimation results are still linear and unbiased. Still, the obtained variance becomes inefficient, meaning that the variance tends to increase or is no longer the minimum variance, so the resulting estimate is no longer BLUE (Nisa et al., 2020). This inefficiency affects the calculation of standard errors in OLS methods, resulting in erratic t-tests that can bias conclusions drawn from the analysis. This study aims to construct a panel data regression model using the Fixed Effect Model approach, applying the Weighted Least Squares (WLS) method with fair weighting as a solution to address heteroscedasticity issues in panel data. One proposed solution is the weighted least squares (WLS) method, which aims to mitigate issues related to heterogeneity in panel data regression (Rosalina et al., 2023). Although WLS presents a viable approach to addressing these challenges, it is essential to evaluate its limitations and compare it with alternative methods critically. A more thorough discussion of unresolved issues or debates in panel data analysis can improve the understanding and application of this statistical technique. By addressing these knowledge gaps, researchers can better manage the complexities inherent in panel data regression analysis. There are several weights in panel data regression using the WLS method, including reasonable weights. Fair weighting is an approach used to assign different weights to each observation, treating each one based on its characteristics to produce a more accurate estimate.

B. RESEARCH METHOD

1. Panel Data Regression

Panel data can be divided into two types, namely balanced and unbalanced panel data. Balanced panel data occurs when each individual has the same number of time observations, while unbalanced panel data occurs when the number of time observations is different (Arumtiwi et al., 2025). Balanced panel data is also referred to as complete panel data, whereas unbalanced panel data is known as incomplete panel data (Rüttenauer & Ludwig, 2023). It can be expressed in Equation (1).

$$y_{it} = \beta_{0it} + \sum_{k=1}^{K} \beta_k x_{kit} + \varepsilon_{it} \tag{1}$$

Description; i: Observation location of the region i $(i = 1, 2, \ldots, N)$, t : Observation time of year t $(t = 1, 2, \ldots, N)$, y_{it} : Dependent variable for location-i and time-t, β_{0it} : Intercept coefficient for location-i and time-t, x_{kit} : Independent variable for location-i and time-t, β_k : Regression coefficient of the independent variable-k, ε_{it} : Error component for location-i and time-t. There are three types of models in panel data regression given as follows:

a. Common Effect Model

The least squares method can be used to estimate the parameters of the common effect model. The common effect model is constant or the same in each individual and each period. The general form of the common effect model can be written in Equation (2) as follows:

$$y_{it} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kit} + \varepsilon_{it} \tag{2}$$

DOI: https://doi.org/10.30812/varian.v8i2.4392

Description; β_0 : Intercept.

b. Fixed Effect Model

Panel data regression with a fixed effect model has a general form given Equation (3) as follows:

$$y_{it} = \beta_{0i} + \sum_{k=1}^{K} \beta_k x_{kit} + \varepsilon_{it} \tag{3}$$

Description; β_{0i} : Intercept for location-i

The fixed effect model assumes that the regression coefficients are constant, but the intercept is not (Rüttenauer & Ludwig, 2023). The assumptions of the individual effects model and the fixed effects model are used in conjunction with the within-group estimator approach. This approach is done by eliminating the unit location effect (β_{0i}), then the value of the dependent and independent variables from each location is averaged over time (Jangphanish et al., 2025).

The model with fixed regression coefficients but different intercepts between individuals (Hsiao, 2022). Equation (4) and Equation (5) are given as:

$$y_{it} = \beta_{0i} + \sum_{k=1}^{K} \beta_k x_{kit} \; + \; \varepsilon_{it} \tag{4}$$

$$\bar{y}_i = \beta_{0i} + \sum_{k=1}^{K} \beta_k \bar{x}_{ki} \; + \; \bar{\varepsilon}_i \tag{5}$$

Description; $\bar{y}_i$: Average of y_{it}, $\bar{x}_{ki}$: Average of x_{kit}, and $\bar{\varepsilon}_i$: Average of ε_{it}.
Within the group, the estimator is obtained by subtracting Equation (4) from Equation (5) to obtain Equation (6):

$$\widetilde{y}_{it} = \beta_1 \widetilde{x}_{1it} + \ldots + \beta_K \widetilde{x}_{Kit} + \widetilde{\varepsilon}_{it} \tag{6}$$

then to estimate the parameter β, assuming the value of ε is distributed $\varepsilon \sim N(\mu, \sigma^2)$, using maximum likelihood, the parameter estimates for the individual effect model as Equation (6) to obtain Equation (7):

$$\hat{\boldsymbol{\beta}} = \left[\left(\widetilde{\boldsymbol{x}}^T \widetilde{\boldsymbol{x}} \right)^{-1} \left(\widetilde{\boldsymbol{x}}^T \widetilde{\boldsymbol{y}} \right) \right] = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_K \end{bmatrix} \tag{7}$$

c. Random Effect Model

The random effect model assumes that individual effects for all cross-section units are random (Bell et al., 2019). Random effect model has a general form given in Equation (8) as follows:

$$y_{it} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kit} + \mu_{it} + \varepsilon_{it} \tag{8}$$

Description; μ_{it} : Error component for the time series, ε_{it} : Error component for the cross-section.

2. Weighted Least Squares

If the variance or square of the standard deviation (σ_ε^2) is known or can be estimated, the easiest way to overcome heteroscedasticity is the weighted least squares method using weights that provide the best linear unbiased estimator (Ulfa, 2019). The fixed effect model regression model in Equation (6) will be transformed into Equation (9) as follows:

$$\widetilde{y}_{it}\left(\frac{1}{\hat{s}_i}\right) = \beta_1 \widetilde{x}_{1it}\left(\frac{1}{\hat{s}_i}\right) + \beta_2 \widetilde{x}_{2it}\left(\frac{1}{\hat{s}_i}\right) + \ldots + \beta_K \widetilde{x}_{Kit}\left(\frac{1}{\hat{s}_i}\right) + \widetilde{\varepsilon}_{it}\left(\frac{1}{\hat{s}_i}\right) \tag{9}$$

Description; $\hat{s}_i$ = Standard deviation of the error of the i region. If $v_i = \frac{1}{\hat{s}_i}$, then Equation (9) is substituted into Equation (10):

$$\widetilde{y}_{it}v_i = \beta_1\widetilde{x}_{1it}v_i + \beta_2\widetilde{x}_{2it}v_i + \ldots + \beta_K\widetilde{x}_{Kit}v_i + \widetilde{\varepsilon}_{it}v_i \quad (10)$$

to estimate the parameter β^* assuming the value of ε is distributed $\varepsilon \sim N(\mu, \sigma^2)$, using maximum likelihood, the parameter estimates for the individual effect model are defined in Equation (11) as follows:

$$\widetilde{\boldsymbol{\varepsilon}}^T\widetilde{\boldsymbol{\varepsilon}} = \widetilde{\boldsymbol{y}}^T\boldsymbol{w}\widetilde{\boldsymbol{y}} - \boldsymbol{2}\boldsymbol{\beta^*}^T\widetilde{\boldsymbol{x}}^T\boldsymbol{w}\widetilde{\boldsymbol{y}} + \boldsymbol{\beta^*}^T\widetilde{\boldsymbol{x}}^T\boldsymbol{w}\widetilde{\boldsymbol{x}}\boldsymbol{\beta}^* \quad (11)$$

so that the parameter estimation results for the WLS model with $\mathbf{w}$ weights are obtained using maximum likelihood as obtained Equation (12):

$$\hat{\boldsymbol{\beta}}^* = \left(\widetilde{\boldsymbol{x}}^T\boldsymbol{w}\widetilde{\boldsymbol{x}}\right)^{-1}\widetilde{\boldsymbol{x}}^T\boldsymbol{w}\widetilde{\boldsymbol{y}} = \begin{bmatrix}\hat{\beta}_1^* \\ \hat{\beta}_2^* \\ \vdots \\ \hat{\beta}_K^*\end{bmatrix} \quad (12)$$

3. Data

The data used in this study are secondary data obtained from the Central Bureau of Statistics of South Sulawesi Province through the official website of CBS South Sulawesi. This dataset consists of 120 time-series data points for 5 years and 24 districts/cities as a cross-section. The dependent variable is life expectancy, while the independent variables are expected years of schooling, per capita expenditure, and average years of schooling.

The Data was analyzed using the R programming language with a data panel regression model, employing fair weighting to overcome heteroscedasticity problems that occur in the data. Figure 1 is the flowchart analysis and the detailed stages of the data processing process in this study:

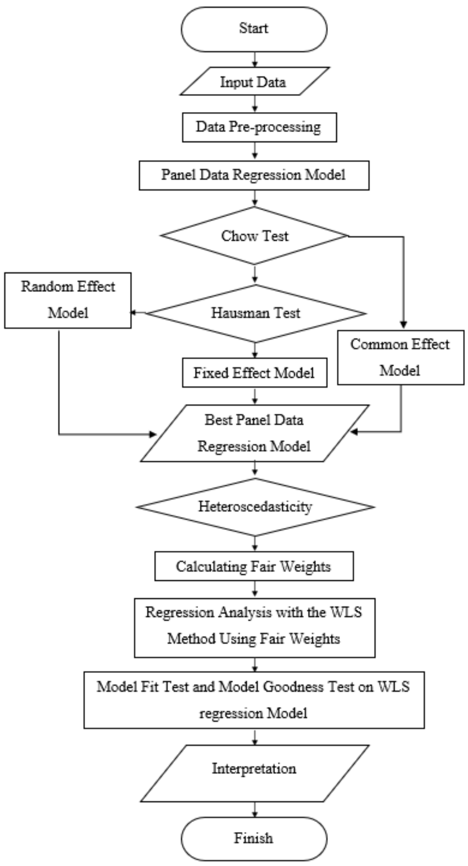

Figure 1. Research flowchart: Data analysis process with a weighted least square method using fair weights

Figure 1 illustrates the flow of research implementation in the data processing section, which addresses heteroscedasticity issues in data using the Weighted Least Squares method with fair weights. The following is a more detailed explanation of the flow chart above:

a. Exploring descriptive data to see the initial picture of research data.
b. Estimating regression models with panel data with three approaches, namely the common effect model, fixed effect model, and random effect model.
c. The Chow test performed with fixed effect model testing similar to the F test is given in Equation (13) as follows:

$$F = \frac{(SSE_{CEM} - SSE_{FEM})/(N-1)}{SSE_{FEM}/(NT-N-K)} \tag{13}$$

Description; SSE_{CEM} : Sum square error of the common effect model, SSE_{FEM} : Sum square error of the fixed effect model, N : Number of cross-section units, T : Number of time series units, and K : Number of independent variables.
If F_{count} is more than F_{table} with the decision that the most appropriate model to use is the fixed-effect model (Novák & Truong, 2023).

d. The Hausman test works by testing whether there is a relationship between the errors in the model (composite errors) and one or more independent variables in the model (Suhendra et al., 2020), the following equation is given in Equation (14) as follows:

$$W = (\hat{\beta}_{FEM} - \hat{\beta}_{REM})\prime \left[var\left(\hat{\beta}_{FEM}\right) - var\left(\hat{\beta}_{REM}\right)\right]^{-1} (\hat{\beta}_{FEM} - \hat{\beta}_{REM}) \tag{14}$$

Description: $\hat{\beta}_{FEM}$ = Intercept coefficient of fixed effect model, $\hat{\beta}_{REM}$ = Intercept coefficient of random effect model. If W is more than $x^2_{a,K}$ or p-value is less than α, then the correct model is the fixed effect model.

e. Heteroscedasticity testing is carried out to determine the data distribution pattern that supports each research variable. According in (Ahzar et al., 2021), the Lagrange multiplier (LM) test or breach-pagan test is used to see whether there are symptoms of heteroscedasticity or not with the test statistics used in the breach-pagan test. The following equation is given in Equation (15) as follows:

$$LM = \frac{NT}{2(T-1)} \left[\frac{\Sigma_{i=1}^{N} \left[\Sigma_{t=1}^{T} \varepsilon_{it}\right]^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \varepsilon_{it}^2} - 1 \right]^2 \tag{15}$$

if the LM value is more than $\chi^2_{\alpha,N-1}$ then reject H_0 so that the error variance structure has a heteroscedasticity.

f. Calculating fair weights, which will be used as weights in the regression equation later on.
g. Test the fit of the weighted least square model using the F-test and T-test and test the goodness of fit using the coefficient of determination test.
h. Displays the interpretation results of the panel data regression analysis obtained.

C. RESULT AND DISCUSSION

1. Data Description

The panel data regression model was obtained in Equation (16), Equation (17), and Equation (19) as follows:

Common effect model,

$$\hat{y}_{it} = 58.0057 + 0.4991x_{1it} - 0.4071x_{2it} + 1.1257x_{3it} \tag{16}$$

Fixed effect model,

$$\hat{y}_{it} = 0.5971x_{1it} + 0.7858x_{2it} + 0.2820x_{3it} \tag{17}$$

In fixed-effects model parameter estimation, the model includes an individual effect value. The following are the calculation results using Equation (7), and we obtained the value in Table 1 as follows:

DOI: https://doi.org/10.30812/varian.v8i2.4392

Table 1. The Value of Individual Effects of Regions in The Province of South Sulawesi

Index (i)	Region	Individual Effect
1	Bantaeng	524.586
2	Barru	502.940
3	Bone	507.173
4	Bulukumba	498.350
5	Enrekang	519.554
6	Gowa	531.619
7	Jeneponto	504.619
8	Kepulauan Selayar	520.574
9	Makassar City	461.051
10	Palopo City	487.899
11	Parepare City	489.489
12	Luwu	524.690
13	East Luwu	503.392
14	North Luwu	497.005
15	Maros	507.954
16	Pangkep	480.296
17	Pinrang	502.152
18	Sidenreng Rappang	503.260
19	Sinjai	503.530
20	Soppeng	525.828
21	Takalar	497.133
22	Tana Toraja	576.974
23	North Toraja	574.006
24	Wajo	478.666

We observed that it is used as a substitute for intercept because, in the fixed effect model estimation, there is an assumption that the regression coefficient is fixed, but the intercept varies between individuals. For example, the Bantaeng region has an individual effect value of 52.4586, so it can be known that the regression model in the Bantaeng region FEM model is as follows:

$$\hat{y}_{it} = 52.4586 + 0.5971x_{1it} + 0.7858x_{2it} + 0.2820x_{3it} \tag{18}$$

Random effect model,

$$\hat{y}_{it} = 51.2114 + 0.5882x_{1it} + 0.6531x_{2it} + 0.4046x_{3it} \tag{19}$$

We first determine which panel data model is best to use in this case by using the Chow test and the Hausman test.

2. Chow Test

The results of testing the Chow test using R software obtained the value of F_{count} 752.6100 with $p-value = 2.20 \times 10^{-16}$ less than $\alpha = 0.05$. Therefore, the decision is that the better model chosen is the fixed effects model.

3. Hausman Test

The results of testing the Hausman test using R software obtained the results of the W value of 27.5350 with a $p-value = 4.55 \times 10^{-06}$ less than $\alpha = 0.05$. Therefore, the conclusion is that the best model is the fixed effects model. Next, we conducted a heteroscedasticity test to determine if heteroscedasticity was present in the data.

4. Heteroscedasticity Test

In this study, the heteroscedasticity test was carried out using the Breusch-Pagan test with R software, and the results of the BP value of 9.8750 with a $p-value = 0.0196$ less than $\alpha = 0.05$, so the conclusion is that heteroscedasticity occurs in the selected model, because the data in this case is heteroscedasticity, it is necessary to do weighting in the weighted least square method, with the weights given in Table 2 as follows:

Table 2. Fair Weighting Value in Each Region in The Province of South Sulawesi

Region	s_i^2	s_i	Fair Weighting (v_i)
Barru	0.0000	0.0032	3.162.278
Bone	0.0025	0.0502	199.007
Bulukumba	0.0037	0.0608	164.377
Enrekang	0.0004	0.0202	495.074
Gowa	0.0018	0.0421	237.289
Jeneponto	0.0000	0.0020	5.000.000
Kepulauan Selayar	0.0001	0.0095	1.048.285
Makassar City	0.0107	0.1037	96.462
Palopo City	0.0024	0.0491	203.785
Parepare City	0.0014	0.0371	269.386
Luwu	0.0007	0.0259	386.046
East Luwu	0.0025	0.0496	201.743
North Luwu	0.0099	0.0997	100.347
Maros	0.0010	0.0322	310.835
Pangkep	0.0032	0.0563	177.499
Pinrang	0.0005	0.0214	466.252
Sidenreng Rappang	0.0010	0.0313	319.928
Sinjai	0.0000	0.0070	1.428.571
Soppeng	0.0064	0.0801	124.815
Takalar	0.0000	0.0045	2.236.068
Tana Toraja	0.0111	0.1052	95.079
North Toraja	0.0006	0.0237	421.450
Wajo	0.0012	0.0344	290.865

We observed that fair weighting value (v_i) in each region in South Sulawesi Province. Using the weighted least square method, this value will later be used as weights in the panel data regression model. For example, the Bantaeng region has a fair weighting value of 23.307, indicating that the weight used in this region is 23.307. This also applies to other regions.

5. Weighted Least Squares

Panel data regression modeling employs the weighted least squares method to mitigate heteroscedasticity. The results of the weighted least square model using Equation (10) with fair weights from Table 2 and individual effects based on Equation (7) are listed in Table 1, so we can get the WLS equation as in Table 3 as follows.

Table 3. Weighted Least Square Model with Fair Weighting on 24 Regions in South Sulawesi Province

Region	Weighted Least Square Models
Bantaeng	$\hat{y}_{1t} = 52.4586 + 0.4630x_{1,1t} + 0.8873x_{2,1t} + 0.3317x_{3,1t}$
Barru	$\hat{y}_{2t} = 50.2940 + 0.4630x_{1,2t} + 0.8873x_{2,2t} + 0.3317x_{3,2t}$
Bone	$\hat{y}_{3t} = 50.7173 + 0.4630x_{1,3t} + 0.8873x_{2,3t} + 0.3317x_{3,3t}$
Bulukumba	$\hat{y}_{4t} = 49.8350 + 0.4630x_{1,4t} + 0.8873x_{2,4t} + 0.3317x_{3,4t}$
Enrekang	$\hat{y}_{5t} = 51.9554 + 0.4630x_{1,5t} + 0.8873x_{2,5t} + 0.3317x_{3,5t}$
Gowa	$\hat{y}_{6t} = 53.1619 + 0.4630x_{1,6t} + 0.8873x_{2,6t} + 0.3317x_{3,6t}$
Jeneponto	$\hat{y}_{7t} = 50.4619 + 0.4630x_{1,7t} + 0.8873x_{2,7t} + 0.3317x_{3,7t}$
Kepulauan Selayar	$\hat{y}_{8t} = 52.0574 + 0.4630x_{1,8t} + 0.8873x_{2,8t} + 0.3317x_{3,8t}$
Makassar City	$\hat{y}_{9t} = 46.1051 + 0.4630x_{1,9t} + 0.8873x_{2,9t} + 0.3317x_{3,9t}$
Palopo City	$\hat{y}_{10t} = 48.7899 + 0.4630x_{1,10t} + 0.8873x_{2,10t} + 0.3317x_{3,10t}$
Parepare City	$\hat{y}_{11t} = 48.9489 + 0.4630x_{1,11t} + 0.8873x_{2,11t} + 0.3317x_{3,11t}$
Luwu	$\hat{y}_{12t} = 52.4690 + 0.4630x_{1,12t} + 0.8873x_{2,12t} + 0.3317x_{3,12t}$
East Luwu	$\hat{y}_{13t} = 50.3392 + 0.4630x_{1,13t} + 0.8873x_{2,13t} + 0.3317x_{3,13t}$
North Luwu	$\hat{y}_{14t} = 49.7005 + 0.4630x_{1,14t} + 0.8873x_{2,14t} + 0.3317x_{3,14t}$
Maros	$\hat{y}_{15t} = 50.7954 + 0.4630x_{1,15t} + 0.8873x_{2,15t} + 0.3317x_{3,15t}$

DOI: https://doi.org/10.30812/varian.v8i2.4392

Region	Weighted Least Square Models
Pangkep	$\hat{y}_{16t} = 48.0296 + 0.4630x_{1,16t} + 0.8873x_{2,16t} + 0.3317x_{3,16t}$
Pinrang	$\hat{y}_{17t} = 50.2152 + 0.4630x_{1,17t} + 0.8873x_{2,17t} + 0.3317x_{3,17t}$
Sidenreng Rappang	$\hat{y}_{18t} = 50.3260 + 0.4630x_{1,18t} + 0.8873x_{2,18t} + 0.3317x_{3,18t}$
Sinjai	$\hat{y}_{19t} = 50.3530 + 0.4630x_{1,19t} + 0.8873x_{2,19t} + 0.3317x_{3,19t}$
Soppeng	$\hat{y}_{20t} = 52.5828 + 0.4630x_{1,20t} + 0.8873x_{2,20t} + 0.3317x_{3,20t}$
Takalar	$\hat{y}_{21t} = 49.7133 + 0.4630x_{1,21t} + 0.8873x_{2,21t} + 0.3317x_{3,21t}$
Tana Toraja	$\hat{y}_{22t} = 57.6974 + 0.4630x_{1,22t} + 0.8873x_{2,22t} + 0.3317x_{3,22t}$
North Toraja	$\hat{y}_{23t} = 57.4006 + 0.4630x_{1,23t} + 0.8873x_{2,23t} + 0.3317x_{3,23t}$
Wajo	$\hat{y}_{24t} = 47.8666 + 0.4630x_{1,24t} + 0.8873x_{2,24t} + 0.3317x_{3,24t}$

Next, the f- and t-tests are carried out to determine whether the model obtained significantly affects the weighted least squares model.

6. F-test

The test results using Equation (9) using R software obtained the value of F_{count} 7114.51 greater than F_{table} 2.6821 with $p - value = 0.0000$ less than $\alpha = 0.05$. Therefore, it can be concluded that the variables of expected years of schooling, per capita expenditure, and average years of schooling simultaneously have a significant effect on life expectancy in the model formed.

7. T-test

In the T-test, we use the following formula in Equation (20):

$$T_{count} = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \tag{20}$$

Description; $\hat{\beta}_k$: Intercept of each independent variable, and SE : Standard Error of each independent variable.

The test results using Equation (20) obtained are in Table 4 below:

Table 4. T-test Result on the Weighted Least Square Model

Variable	T_{count}	$p - value$
X_1	2.5500	0.0120
X_2	68.0600	0.0000
X_3	15.4300	0.0000

Based on Table 4, it can be seen that the independent variables of expected years of schooling, per capita expenditure, and average years of schooling have a p-value of less than $\alpha = 0.05$ with a T_{count} value greater than the T_{table} value of 1.6580 so it can be concluded that the variables of expected years of schooling, per capita expenditure, and average years of schooling partially significantly affect life expectancy in the regression model formed.

8. Goodness of Fit Using the Coefficient of Determination

Testing the goodness of the model used is the coefficient of determination or R^2. This is used to assess the goodness of the weighted least squares panel data regression model applied to life expectancy data in South Sulawesi Province. In this case, the R^2 result is 0.9945. This shows that the value of life expectancy in South Sulawesi Province is influenced by the expected length of schooling, per capita expenditure, and the average length of schooling by 99.45%, and the remaining 0.55% is influenced by other variables not included in this study.

This study develops a panel data regression model using the Weighted Least Squares (WLS) method, previously applied by Nisa et al. (2020), with fair weighting. This approach allows each observation unit to be treated proportionally based on its characteristics. In a previous study by Nisa et al. (2020), which aimed to estimate the parameters of the WLS method to address heteroscedasticity, obtained analysis results with an accuracy level of 95.45%. Based on these research results, we have developed

DOI: https://doi.org/10.30812/varian.v8i2.4392

a method for using weighting to obtain more optimal analysis results. This has proven effective in addressing heteroscedasticity problems and improving the quality of model estimation, as indicated by the very high coefficient of determination (R^2) value of 99.45%. Thus, there is a 4.00% increase in accuracy, which means that the accuracy level using fair weighting is better at explaining WLS regression analysis with heteroscedasticity characteristics. Thus, this study not only strengthens the application of applied statistical methods but also enriches strategies for handling heteroscedasticity in the context of panel regression.

D. CONCLUSION AND SUGGESTION

This study concluded that the best model used for the data, in this case, is the fixed effect model according to $\hat{y}_{it} = 0.5971x_{1it} + 0.7858x_{2it} + 0.2820x_{3it}$. The weighted least squares regression equation, using the fixed effect model approach with fair weights, shows that the independent variables of expected years of schooling, per capita expenditure, and average years of schooling have a positive and significant influence on life expectancy. The results of this study indicate that the regression model has a very high coefficient of determination, which is 0.9945, suggesting that these variables influence life expectancy in South Sulawesi Province to the extent of 99.45%. The remaining 0.55% is influenced by other variables not included in this study's model. Further research is expected to use other methods that can overcome the problem of heteroscedasticity.

ACKNOWLEDGEMENT

We would like to thank Universitas Hasanuddin and the parties involved in this research for their support.

DECLARATIONS

AUTHOR CONTIBUTION

All authors contributed to this manuscript, from exploring ideas to writing this article.

FUNDING STATEMENT

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

REFERENCES

Abdussamad, S. N., Adityaningrum, A., & Payu, M. R. F. (2024). Panel Data Regression Analysis for Modeling the Human Development Index in North Sulawesi Province. *Parameter: Journal of Statistics*, *4*(1), 46–53. https://doi.org/10.22487/27765660.2024.v4.i1.17138

Ahzar, F. A., Rosadi, S., & Wati, A. (2021). Corporate Governance, Karakteristik Dewan Pengawas Syariah dan Maqoshid Sharia Index pada Perbankan Syariah di Indonesia. *AKTSAR: Jurnal Akuntansi Syariah*, *4*(2), 241. https://doi.org/10.21043/aktsar.v4i2.12730

Alamsyah, I. F., Esra, R., Awalia, S., & Nohe, D. A. (2022). Analisis Regresi Data Panel untuk Mengetahui Faktor yang Memengaruhi Jumlah Penduduk Miskin di Kalimantan Timur. *Prosiding Seminar Nasional Matematika dan Statistika*, *2*.

Arumtiwi, A. A., Raupong, R., & Siswanto, S. (2025). Minimum Variance Quadratic Unbiased Estimation Approach in Panel Data Regression Analysis with Two-Way Error Component Estimation Using Biggers Method. *Jurnal Matematika, Statistika dan Komputasi*, *21*(2), 439–455. https://doi.org/10.20956/j.v21i2.37325

Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: Making an informed choice. *Quality & Quantity*, *53*(2), 1051–1074. https://doi.org/10.1007/s11135-018-0802-x

Du, S., Wilke, R. A., & Homrighausen, P. (2025). On Omitted Variables, Proxies, and Unobserved Effects in Empirical Regression Analysis. *Journal of Official Statistics*, 0282423X241312644. https://doi.org/10.1177/0282423X241312644

Hsiao, C. (2022). *Analysis of Panel Data* (4th ed.). Cambridge University Press.

Jangphanish, K., Boonyanet, W., & Tongkong, S. (2025). Analyzing robust dividend payout policy with dynamic panel regression: Application of speed of adjustment to half-life (M. Alexandre, Ed.). *PLOS ONE*, *20*(1), e0316478. https://doi.org/10.1371/journal.pone.0316478

Lv, T., & Wu, X. (2019). Using Panel Data to Evaluate the Factors Affecting Transport Energy Consumption in China's Three Regions. *International Journal of Environmental Research and Public Health*, *16*(4), 555. https://doi.org/10.3390/ijerph16040555

Nisa, H., Kusnandar, D., & Martha, S. (2020). Estimasi Parameter Metode Weighted Least Square Dalam Mengatasi Masalah Heteroskedastisitas. *Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya*, *9*(1), 65–70. https://doi.org/10.26418/bbimst.v9i1.38586

Novák, V., & Truong, T. T. P. (2023). A Combination of Fuzzy Techniques and Chow Test to Detect Structural Breaks in Time Series. *Axioms*, *12*(2), 103. https://doi.org/10.3390/axioms12020103

Pramusinto, N. D., & Daerobi, A. (2020). Labor Absorption of the Manufacturing Industry Sector in Indonesia. *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, *3*(1), 549–561.

Rosalina, M., Martha, S., & Imro'ah, N. (2023). Pemodelan Regresi Nonparametrik Birespon Spline pada Persentase Penduduk Miskin dan Indeks Kedalaman Kemiskinan. *BIMASTER : Buletin Ilmiah Matematika, Statistika dan Terapannya*, *12*(1), 69–78. https://doi.org/10.26418/bbimst.v12i1.62849

Rüttenauer, T., & Ludwig, V. (2023). Fixed Effects Individual Slopes: Accounting and Testing for Heterogeneous Effects in Panel Data or Other Multilevel Models. *Sociological Methods & Research*, *52*(1), 43–84. https://doi.org/10.1177/0049124120926211

Suhendra, I., Istikomah, N., Ginanjar, R. A. F., & Anwar, C. J. (2020). Human Capital, Income Inequality and Economic Variables: A Panel Data Estimation from a Region in Indonesia. *The Journal of Asian Finance, Economics and Business*, *7*(10), 571–579. https://doi.org/10.13106/JAFEB.2020.VOL7.NO10.571

Ulfa, N. L. (2019, December). *Penanganan Kasus Heteroskedastisitas dengan Menggunakan Metode WLS (Weighted Least Square)* [Bachelor Thesis]. UIN Sunan Kalijaga.