Jurnal Matematika dan Statistika serta Aplikasinya, Vol. 13, No. 2, 172-183, 2025
DOI:https://doi.org/10.24252/msa.v13i2.60197

Autoregressive Distributed Lag (ARDL) Method for Estimating Poverty Levels in Polewali Mandar Regency

Andi Tenri Abeng[1]*, Wahidah Alwi[2], Adnan Sauddin[3], Sri Dewi Anugrawati[4], Nur Aeni[5]
[1,2,3,4,5]*Mathematics Study Program, Universitas Islam Negeri Alauddin Makassar, Gowa, Indonesia*
***Corresponding author**: anditenri1811@gmail.com

***Submission date**: 29 July 2025, **Revision**: 03 August 20205 2025, **Accepted**: 27 November 2025

ABSTRACT
Polewali Mandar Regency is the region with the highest poverty rate in West Sulawesi. According to a publication by the Central Bureau of Statistics in March 2022, the percentage of the poor population was 11.75%, an increase compared to March 2021. The forecasting method used in this study is the Autoregressive Distributed Lag (ARDL) method. This study aims to determine the Autoregressive Distributed Lag (ARDL) model, which is then used to forecast the number of poor people in Polewali Mandar Regency. The results of the study using the ARDL method yielded the best estimation model, namely ARDL (3, 3, 2, 2). The forecast results for the percentage of the poor population using the ARDL (3, 3, 2, 2) model for the following semesters are 21.79%, 10.15%, and 16.52%, respectively. The forecasting accuracy test using the Mean Absolute Percentage Error (MAPE) yielded a value of 12.18%, indicating that the ARDL model produced in this study is suitable for forecasting the percentage of the poor population in Polewali Mandar Regency.

KEYWORDS
Autoregressive Distributed Lag(ARDL), Forecasting Accuracy, Poor Population, Poverty.

1. INTRODUCTION

Poverty is a major issue in developing countries such as Indonesia. In West Sulawesi, individuals are classified as poor if their monthly expenditure falls below IDR 405,187 (as of 2022). Polewali Mandar Regency recorded the highest poverty rate at 11.75% in March 2022, an increase from the previous year. Accurate poverty data is essential for policy evaluation, regional and temporal comparisons, and the effective distribution of social assistance. One approach that can be employed is forecasting based on historical data [1].

Autoregressive Distributed Lag (ARDL) model was selected due to its ability to analyze both short-term and long-term relationships among variables. By combining autoregressive and distributed lag components, ARDL is particularly relevant for modeling the number of people living in poverty in Polewali Mandar. A previous study by Permata et al. (2017) demonstrated that the ARDL model outperforms ARIMAX in forecasting economic variables, yielding lower forecast errors [2]. Rahmasari et al. (2019), in a study conducted in West Nusa Tenggara Province, also found ARDL to be highly accurate in predicting poverty, with a MAPE of only 3% [3]. Therefore, with its high level of accuracy and capacity to capture temporal dynamics, ARDL is considered an appropriate method for forecasting poverty in Polewali Mandar Regency

Traditional poverty measurement relies heavily on household surveys and census data, which provide detailed consumption information necessary for establishing monetary poverty lines. However, these conventional data collection methods face significant operational constraints in developing countries. Survey rounds are often infrequent and costly, creating temporal gaps that limit real-time policy responses and program monitoring [4]. Furthermore, changes in questionnaire design across survey rounds can hinder temporal comparability, while small sample sizes at the district or village level produce unreliable estimates for localized policy interventions [5]. These limitations underscore the importance of developing robust forecasting models that can leverage historical data to generate timely and spatially disaggregated poverty estimates for policy planning and resource allocation.

The success of poverty forecasting models depends critically on the selection of appropriate socioeconomic indicators as predictor variables. Empirical studies across developing countries have identified several key determinants that consistently influence poverty dynamics. In the Indonesian context, research by Sugiharti et al. (2022) found that household demographic structure, education level, employment type (particularly casual agricultural work), and access to basic services are significant predictors of chronic poverty [6]. Similarly, studies in other developing countries demonstrate that household food expenditure, asset ownership, and utility expenditures substantially improve model prediction accuracy [7, 8]. The integration of these socioeconomic variables into time-series models such as ARDL enables more accurate forecasting of poverty trends, which in turn supports better targeting of social assistance programs and more effective evaluation of poverty reduction policies [9].

2. LITERATURE REVIEW

2.1 Poverty

Poverty is a global issue resulting from disparities in capabilities and access to resources. Although difficult to eradicate, poverty must not be overlooked due to its significant social impacts. It was previously defined as the inability to meet basic needs, but it now encompasses education, health, social participation, and access to information [10].

Poverty is increasingly understood as a multidimensional problem that hinders overall well-being. Governments address this issue through adaptive policies in line with contemporary developments. The poverty line serves as a key indicator in determining poverty status; therefore, accurate data on the proportion of the poor population is essential for formulating effective policies [11].

2.2 Autoregressive Distributed Lag (ARDL)

Autoregressive Distributed Lag (ARDL) model is an econometric approach that combines the elements of autoregression (AR) and distributed lag (DL) to analyze both short-term and long-term relationships among variables. By incorporating the time dimension, ARDL transforms economic theory from a static to a dynamic framework [12].

The main advantage of the ARDL model lies in its ability to examine the short-run and long-run effects of variables by accounting for lag lengths, thus providing a more comprehensive analysis compared to conventional regression models [13]. The following are the formulas commonly used in the ARDL method:

Single-variable formula

$$X = \beta_0 + \emptyset_1 X_{t-1} \tag{1}$$

Two-variable formula

$$X = \beta_0 + \emptyset_1 X_{t-1} + \ldots + \emptyset_p X_{t-p} + \beta_0 Y_t + \ldots + \emptyset_q Y_{t-q} \tag{2}$$

where:
X: level-stationary or dependent variable
β_0 : constant
$\emptyset_1$: coefficient of the dependent variable
X_{t-1}: lagged value of variable X (previous time period)
$t-1$: previous time period
t: current time period
Y_{t-1}: lagged value of the second variable Y

Autoregressive Distributed Lag (ARDL) model offers several advantages, including: ARDL model does not require uniform stationarity or integration levels of the data. It can be applied even when the variables exhibit different degrees of stationarity—whether at the level, first difference, or second difference. The use of ARDL does not necessitate that all variables share the same order of integration. In other words, the model remains applicable even when the variables are integrated at different levels.

ARDL model is not heavily dependent on large sample sizes. It can still be effectively employed with relatively small datasets, unlike many other statistical methods that typically require large amounts of data to produce valid results. ARDL can also be used to test for reciprocal or causal relationships among variables in time series data, both in the short run and the long run

2.3 Stationarity Test

Stationarity testing is essential in the ARDL method to ensure the consistency of time series data [14]. ARDL can only be applied when the data is stationary at the level or first difference, but not at the second difference. One common method to test for stationarity is differencing. The differencing process is formulated as follows:

$$Z_t' = Z_t - Z_{t-1} \tag{3}$$

where Z_t is the value of variable Z at time t and Z_{t-1} is the value of Z at time $t-1$.

In this study, the Augmented Dickey-Fuller (ADF) test is used to examine the presence of a unit root in the time series data. If the data is not stationary at level [I(0)], the test is continued to higher orders such as the first difference [I(1)] or second difference [I(2)], until stationarity is achieved [15]. The ADF test is based on the following equations:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_P X_{Pt} \tag{4}$$

$$\Delta Y_t = \beta_0 + \beta_1 X_{1t} - X_{1t-1} + \beta_2 X_{2t} - X_{2t-2} + \ldots + \beta_P X_{Pt} - X_{Pt-1} \tag{5}$$

$$\Delta Y_{t-1} = \beta_0 + \beta_1 X_{1t} - X_{1t-1} - X_{1t-2} + \beta_2 X_{2t} - X_{2t-2} - X_{2t-2} + \ldots + \beta_P X_{Pt} - X_{Pt-1} - X_{Pt-2} \tag{6}$$

where:
Y : level-stationary variable
ΔY_t: first difference of Y
ΔY_{t-1}: second difference of Y
β_0: constant or intercept
β_1: regression coefficient for trend
t : time
Hypotheses:
H_0: $\delta = 0$ (There is a unit root; the variable Y is non-stationary)
H_1: $\delta \neq 0$ (There is no unit root; the variable Y is stationary)
Test statistic:

$$t_\delta = \frac{\hat{\delta} - \delta_0}{se\left(\hat{\delta}\right)} \tag{7}$$

If the value of t_δ is greater than the ADF critical value, the data is non-stationary. Conversely, if it is smaller, the data is stationary. If the data is not stationary at order zero [I(0)], it must be tested again at a higher order, such as I(1) or I(2), until the degree of stationarity is determined.

2.4 Cointegration Test

Cointegration indicates the existence of a long-term relationship among non-stationary variables. If a combination of these variables is stationary, it implies a stable long-term relationship; otherwise, the relationship is not statistically significant [16].

The cointegration test employs the Bound Test, which is suitable for short time series data and variables integrated at order I(0) or I(1). This method complements ARDL in producing consistent and asymptotically normal estimates of long-run coefficients. The hypotheses for the cointegration test are as follows:
H_0: $\lambda_1 = \lambda_2 = \lambda_n = 0$ (no cointegration)
H_a: $\lambda_1 \neq \lambda_2 \neq \lambda_n \neq 0$ (cointegration exists)

The decision in the Bound Test is based on the F-statistic value. If it is less than the lower bound, there is no cointegration (fail to reject H_0). If it exceeds the upper bound, cointegration exists (reject H_0). If it falls between the two bounds, the result is inconclusive [17].

2.5 Assumption Testing

A regression model estimated using the Ordinary Least Squares (OLS) method has the advantage of being a Best Linear Unbiased Estimator (BLUE) that is, a linear and unbiased estimator with the smallest possible variance. However, this advantage holds only if the classical linear regression assumptions are met. Therefore, it is necessary to test for possible violations of these assumptions.

2.6 Optimal Lag Selection

The purpose of determining the optimal lag is to address the problem of autocorrelation in Vector Autoregression (VAR) systems. This process involves the use of several criteria, such as the Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC), Hannan-Quinn Criterion (HQ), and Final Prediction Error (FPE). The model with the smallest criterion value is considered the best [11]. AIC is widely used because it penalizes overly complex models, making the lag with the lowest AIC value the most optimal. The formula for AIC is:

$$AIC = \log\left(\sum \frac{\varepsilon_t^2}{n}\right) + \frac{2k}{n} \tag{8}$$

where:
$\sum \frac{\varepsilon_t^2}{n}$: sum of squared residuals
n: sample size
k: number of parameters (variables)

2.7 Forecasting Accuracy

The primary objective of forecasting is to produce optimal and accurate predictions with minimal error. The error is calculated as the difference between the actual value and the forecasted value:

$$Error\,(E) = Y_t - \hat{Y}_t \tag{9}$$

where Y_t is the actual value at time t and $\hat{Y}_t$ is the forecasted value at time t.

To measure the accuracy of the forecast, one commonly used metric is the Mean Absolute Percentage Error (MAPE). MAPE is calculated as the average percentage of the absolute difference between the predicted and actual values, regardless of whether the difference is positive or negative. The formula for calculating MAPE is as follows:

$$MAPE = \ 100\ \times \frac{1}{n}\sum_{i=t}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \tag{10}$$

where Y_t is an actual value at time t, $\hat{Y}_t$ is a forecasted value at time t and n is the number of data points. A lower MAPE value indicates better and more accurate forecasting model performance. More detailed interpretation of forecasting accuracy levels was depicted in the **Table 1**.

Table 1. MAPE Value Interval

MAPE	Level of Accuracy
$< 10\%$	Very Good
$10\% - 20\%$	Good
$21\% - 50\%$	Fair
$> 50\%$	Poor

3. METHODOLOGY

The type of research used in this study is applied research. The data utilized in this study includes the percentage of the poor population (Y), Human Development Index (X_1), Open Unemployment Rate (X_2), and Economic Growth (X_3) on a semi-annual basis, ranging from 2006 to the first semester of 2023. The data were obtained from the official website `https://polewalimandarkab.bps.go.id`.

3.1 Data Analysis Process:

The research procedure consists of the following steps:

1. Descriptive Data Analysis
2. Modeling using the Autoregressive Distributed Lag (ARDL) method, with the following stages:
   (a) Stationarity Test: Performed using a unit root test, specifically the Augmented Dickey–Fuller (ADF) test. If the data is non-stationary, differencing is applied until stationarity is achieved.
   (b) Cointegration Test: Used to determine the presence of a long-term relationship among variables using the Bound Test.
   (c) Model Selection: The best ARDL model is selected by determining the optimal lag based on the Akaike Information Criterion (AIC), after which the selected model is estimated.
   (d) Parameter Testing: Conducted to ensure that there is a statistically significant relationship among the variables and that the model meets the requirements for validity.
   (e) Assumption Testing: Ensures that the estimations meet the criteria of the Best Linear Unbiased Estimator (BLUE).
3. Forecasting using the Autoregressive Distributed Lag (ARDL) method, with the following steps:
   (a) Perform forecasting based on the selected ARDL model.
   (b) The final step is to evaluate forecast accuracy by calculating the Mean Absolute Percentage Error (MAPE)

4. RESULT & DISCUSSION

4.1 Data Description

The data used in this study is secondary time series data in the form of semi-annual observations, collected from the first semester of 2006 to the first semester of 2023. The dependent variable in this study is the number of people living in poverty (Y) in Polewali Mandar Regency. The independent variables include the Human Development Index (X_1), Open Unemployment Rate (X_2), and Economic Growth (X_3). An overview of the dynamics of each variable from 2006 to 2023 is presented in **Table 2**.

Table 2. Descriptive Statistics of the Research Data

Description	Y	X_1	X_2	X_3
Minimum	10.87	56.73	1.29	-7.36
Maximum	20.74	67.89	6.45	14.81
Mean	13.267714	62.538571	3.232571	7.248
Q_1	11.395	60	2.335	5.21
Median	12.23	62.46	3.04	7.61
Q_3	13.79	65.655	3.72	10.115
Stdev	2.662037	3.297985	1.197908	4.294855
Skewness	1.425982	-0.140751	0.805405	-0.99402
Kurtosis	0.985304	-1.292498	0.070229	1.957629

Based on **Table 2**, the minimum value is found in variable X_3 at -7.36% while the maximum value occurs in X_1 at 67.89%. The highest average is also observed in X_1 at 62.54%, whereas the lowest average is in X_2 at 3.23%. The skewness and kurtosis values fall within the normal range (-1.96 to +1.96), indicating that the data distribution is normal and symmetrical. All standard deviation values are smaller than their respective means, suggesting the absence of extreme outliers. A plot of the data was generated to observe the trends, as shown in **Figure 1**.

As depicted on Figure 1, the movement of the poverty rate (Y) from the first semester of 2006 to the first semester of 2023 is characterized by short-term fluctuations but shows a long-term downward trend. The Human Development Index (X_1)

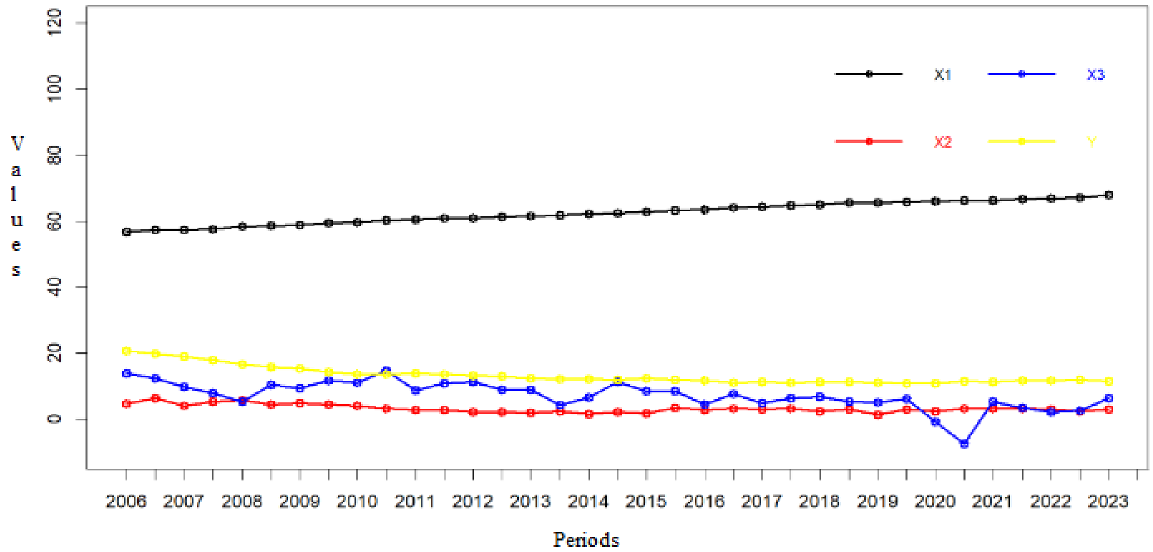

Figure 1. Trends of the data

exhibits an upward trend, reflecting improvements in the population's quality of life. The Open Unemployment Rate (X_2) is also fluctuating but tends to decline in the long run. Meanwhile, Economic Growth (X_3) displays a cyclical pattern—peaking in the second semester of 2010, sharply declining until the second semester of 2020, and stabilizing from 2021 to 2023.

Poverty Estimation Modeling Using the Autoregressive Distributed Lag (ARDL) Method Forecasting the number of people living in poverty using the Autoregressive Distributed Lag (ARDL) method is carried out through the following stages:

1. Stationarity Test

   In this study, the stationarity of the data is tested using the Unit Root Test through the Augmented Dickey-Fuller (ADF). The results of the stationarity test using the ADF test are presented in **Table 3**. This table shows the variable X_3 is

Table 3. Unit Root Test Result

ADF Variables	p-value	Critical Value ($\alpha = 5\%$)
Y	0.0018	0.05
$\mathbf{X_1}$	0.0000	0.05
$\mathbf{X_2}$	0.0117	0.05
$\mathbf{X_3}$	0.0736	0.05

non-stationary because the p-value is greater than 0.05. Since it is not yet stationary, differencing is applied. The test results at the first difference level are presented in **Table 4**. Based on the test results, the p-value for variable X_3 is less

Table 4. Unit Root Test Results at First Difference

ADF Variable	p-value	Critical Value ($\alpha = 5\%$)
$\mathbf{X_3}$	0.0000	0.05

than 0.05, indicating that the data is stationary at the first difference. All variables have met the stationarity requirements, and the analysis can proceed.

2. Cointegration Test

   The purpose of the cointegration test is to determine the existence of a long-term relationship between the independent and dependent variables. In this study, the cointegration test is conducted using the Bound Test The results based on **Table**

Table 5. Cointegration Test Results

Test Statistic	Value	Significant Level	I(0)	I(1)
F-statistic	9.18			
		10%	2.17	3.19
		5%	2.72	3.83
		1%	3.88	5.30

4 shows an F-statistic value of 9.18, which exceeds the upper bound value at $\alpha = 5\%$ (3.83), leading to the rejection of H_0. This indicates the presence of a long-term relationship (cointegration) in the model.

3. Determination of the Best ARDL Model

   The determination of the optimal lag length in this study uses the Akaike Information Criterion (AIC), with the results presented in **Table 6**. Out of the 18 models identified in **Table 6**, ARDL(3,3,2,2) is selected as the best model because it

Table 6. Akaike Information Criteria (Top Models)

No.	Y	X_1	X_2	X_3	AIC
1	3	3	2	2	12.71923
2	3	3	3	2	12.98647
3	1	0	3	2	13.15186
4	3	3	3	3	13.32779
5	3	3	2	3	13.65325
6	3	2	3	2	13.77842
7	2	3	2	2	14.07628
8	1	0	3	3	14.40687
9	3	2	3	3	14.87668
10	2	3	3	2	15.19567
11	1	0	2	2	15.40236
12	2	3	2	3	15.93325
13	2	4	2	2	16.32823
14	1	0	2	1	17.42600
15	2	2	2	2	18.71100
16	4	4	4	4	21.74792
17	1	0	1	1	23.04391
18	1	1	1	1	24.73237

has the smallest error compared to the other models. The parameter coefficients for each lagged variable, based on the optimal lag test results, are presented in **Table 7**. The ARDL(3, 3, 2, 2) estimation model for forecasting the number of people living in poverty in Polewali Mandar Regency for future periods is as follows:

$$\begin{aligned} Y_t = {} & 16.385769 + 1.025371Y_{t-1} - 0.034436Y_{t-2} - 0.260198Y_{t-3} + 0.493227X_{1t} - 0.643018X_{1t-1} + 0.388965X_{1t-2} \\ & - 0.440616X_{1t-3} - 0.224025X_{2t} - 0.004062X_{2t-1} + 0.244984X_{2t-2} - 0.079095X_{3t} + 0.010822X_{3t-1} \\ & - 0.044381X_{3t-2} \end{aligned}$$

4. Parameter Testing

   This stage involves several tests, including the coefficient of determination (R^2), simultaneous test (f-test), and partial test (t-test).

   (a) Coefficient of Determination

Table 7. Optimal Lag Test

Variable	Coefficient	Std. Error	t-Statistic	Prob.*
$\mathbf{Y(-1)}$	1.025371	0.183062	3.220	0.00475
$\mathbf{Y(-2)}$	-0.034436	0.242460	5.601	0.0000258
$\mathbf{Y(-3)}$	-0.260198	0.184405	-0.142	0.88864
$\mathbf{X_1}$	0.493227	0.358251	1.377	0.18547
$\mathbf{X_1(-1)}$	-0.643018	0.441808	-1.455	0.16277
$\mathbf{X_1(-2)}$	0.388965	0.397647	0.978	0.34096
$\mathbf{X_1(-3)}$	-0.440616	0.386864	-1.139	0.26966
$\mathbf{X_2}$	-0.224025	0.097000	-2.776	0.03298
$\mathbf{X_2(-1)}$	-0.004062	0.079441	-0.051	0.95978
$\mathbf{X_2(-2)}$	0.244980	0.088245	2.776	0.01246
$\mathbf{X_3}$	-0.079095	0.020987	-3.769	0.00141
$\mathbf{X_3(-1)}$	0.010822	0.0191949	0.565	0.57895
$\mathbf{X_3(-2)}$	-0.044381	0.018701	-2.373	0.02897
C	16.385769	5.089531	3.220	0.004

Coefficient of determination indicates an R^2 value of 0.9886 and an adjusted R^2 of 0.9804, meaning that 98.04% of the variation in the dependent variable (number of people living in poverty) can be explained by the independent variables: Human Development Index (X_1), Open Unemployment Rate (X_2), and Economic Growth (X_3). The remaining 1.96% is explained by factors outside the model.

(b) Simultaneous Test

F-test is used to examine the combined influence of all independent variables on the dependent variable. If the p-value is less than 5%, it can be concluded that the independent variables simultaneously affect the dependent variable. If n represents the number of observations, k is the number of independent variables, and R^2 is the coefficient of determination obtained from the R^2 test, then the calculated F-value is obtained as follows.

$$F_{value} = \frac{\frac{R^2}{(k-1)}}{\frac{(1-R^2)}{n-k}} = \frac{\frac{0.9804}{(3-1)}}{\frac{(1-0.9804)}{35-3}} = 800.326$$

The degrees of freedom for the numerator are calculated as $N1 = k-1$, where k is the number of independent variables. Meanwhile, the degrees of freedom for the denominator are $N2 = n-k-1$, where n is the number of observations. The resulting F_{table} in **Table 8** as follows:

Table 8. Critical Values from the F_{table}

$\mathbf{N_1}$ $(k-1)$	$\mathbf{N_2}$ $(n-k-1)$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
$\mathbf{3-1=2}$	$35-3=32$	5.34	3.29	2.48

Based on **Table 8**, the calculated F_{value} (800.326) is greater than the critical F_{tabel} (3.29) at $\alpha = 5\%$, thus $_0$ is rejected. This indicates that the independent variables (X_1,X_2, X_3) have a significant simultaneous effect on the number of people living in poverty in Polewali Mandar Regency.

(c) Partial Test

Partial test (t-test) is used to analyze the individual effect of each independent variable on the dependent variable. If the p-value $>$ 5%, the variable does not have a statistically significant partial effect. Given n = number of observations and k = number of independent variables, the t_{tabel} value can be determined as shown in **Table 9**.

The calculated t_{value} along with the results of the t-test for all independent variables in the ARDL(3,3,2,2) estimation model, are presented in **Table 10**. It is shows that the variables which have a significant effect on the number of people living in poverty (Y) are $Y\,(-1)$, X_2, $X_2\ (-2)$, X_3 and $X_3(-2)$.

Table 9. t_{table}

Degree of Freedom (df) $(n-k)$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
$\mathbf{35 - 3 = 32}$	2.448678	1.693889	1.308573

Table 10. Partial Test at $\alpha = 5\%$

Variable	Coefficient	t_{value}	H_0	Decision
$\mathbf{Y}(-1)$	1.025371	5.601	Rejected	Significant
$\mathbf{Y}(-2)$	-0.034436	-0.142	Accepted	Not Significant
$\mathbf{Y}(-3)$	-0.260198	-1.411	Accepted	Not Significant
$\mathbf{X_1}$	0.493227	1.377	Accepted	Not Significant
$\mathbf{X_1}(-1)$	-0.643018	-1.455	Accepted	Not Significant
$\mathbf{X_1}(-2)$	0.388965	0.978	Accepted	Not Significant
$\mathbf{X_1}(-3)$	-0.440616	-1.139	Accepted	Not Significant
$\mathbf{X_2}$	-0.224025	-2.310	Rejected	Significant
$\mathbf{X_2}(-1)$	-0.004062	-0.051	Accepted	Not Significant
$\mathbf{X_2}(-2)$	0.244984	2.776	Rejected	Significant
$\mathbf{X_3}$	-0.079095	-3.769	Rejected	Significant
$\mathbf{X_3}(-1)$	0.010822	0.567	Accepted	Not Significant
$\mathbf{X_3}(-2)$	-0.044381	-2.373	Rejected	Significant

5. Assumption Testing

Assumption testing is conducted to ensure that the estimation satisfies the criteria of the Best Linear Unbiased Estimator (BLUE). The assumption tests that must be fulfilled are as follows:

(a) Normality Test

Normality test is performed to determine whether the data used are normally distributed. One method to detect normality issues is the Jarque-Bera test, with the following hypotheses:
H_0: the residuals are not normally distributed
H_1: the residuals are normally distributed
The result of the Jarque-Bera test shows a value of $0.07756 > \alpha = 0,05$, thus H_0 is rejected, and it is concluded that the data are normally distributed.

(b) Homoscedasticity Test

Homoscedasticity test aims to ensure that the variance of the residuals is constant. One method used for this test is the Breusch-Pagan-Godfrey test. The hypotheses used in this test are as follows:
H_0 : the residuals are not homoscedastic
H_1 : the residuals are homoscedastic
The test result shows a p-value of $0.09774 > \alpha = 0,05$, thus H_0 is rejected. This indicates that the model satisfies the assumption of homoscedasticity, meaning the residuals have constant variance.

(c) Autocorrelation Test

Autocorrelation refers to the presence of correlation among the error terms in a regression model. To detect violations of classical assumptions related to autocorrelation, the Breusch-Godfrey method is commonly used, with the following hypotheses:
H_0: no autocorrelation is present
H_1: autocorrelation is present
Based on the test results, the p-value obtained is $0.3562 > \alpha\ (0,05)$, which means $H_0$is accepted. Therefore, it can be concluded that the regression model does not indicate the presence of autocorrelation.

4.2 Calculating Forecasting Results Using the Autoregressive Distributed Lag (ARDL) Method

1. Estimation of the Number of People Living in Poverty

   The forecasting results for the number of people living in poverty using the Autoregressive Distributed Lag (ARDL) method are obtained as follows: The ARDL forecasting results in **Table 10** indicated the fluctuations in the number of

Table 11. Forecasting Results for Future Periods

Year	Semester	Forecasted Results
2023	2	21,79
2024	1	10,15
2024	2	16,52

people living in poverty in Polewali Mandar Regency: an increase of 21.79% in the second semester of 2023, a decrease of 10.15% in the first semester of 2024, followed by another increase of 16.52% in the second semester of 2024.

The following graph in **Figure 2** comparing the actual and predicted poverty levels. Based on the figure, there is no significant difference between the forecasted results and the actual data.

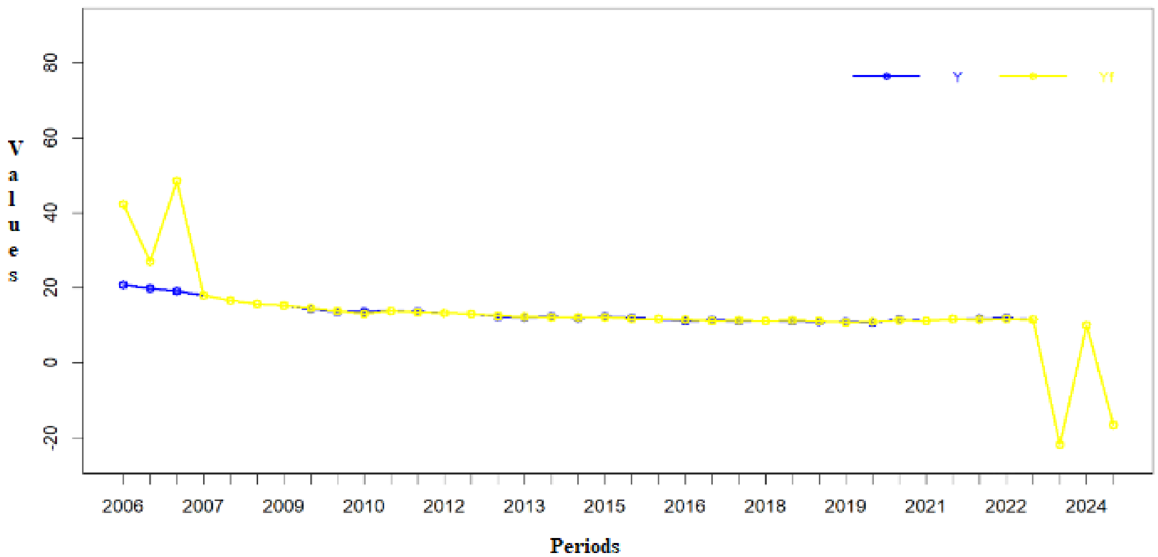

Figure 2. Actual Data and Predicted Data

2. Forecast Accuracy Test

   The final step in assessing the quality of the forecasting results is to conduct a forecast accuracy test. In this study, the method used to measure the level of forecasting accuracy is the Mean Absolute Percentage Error (MAPE), which serves as an indicator of forecasting accuracy.

$$
\begin{aligned}
MAPE &= 100\% \times \frac{1}{n}\sum_{i=t}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \\
&= 100\% \times \frac{1}{35}\left(\frac{20.74 - 42.22448}{20.74} + \cdots + \frac{11.49 - 11,55183}{11.49}\right) \\
&= 12.18\%
\end{aligned}
$$

Based on the MAPE value obtained using the ARDL method, which is 12.18%, it indicates that the ARDL model generated using R Studio is suitable for forecasting the number of people living in poverty in Polewali Mandar Regency for future periods

5. CONCLUSION

Based on the analysis of poverty data in Polewali Mandar Regency from the first semester of 2006 to the first semester of 2023 using the Autoregressive Distributed Lag (ARDL) model, several conclusions can be drawn as follows:

1. The estimation model generated in this study using the ARDL method is as follows:

$$Y_t = 16.385769 + 1.025371Y_{t-1} - 0.034436Y_{t-2} - 0.260198Y_{t-3} + 0.493227X_{1t} - 0.643018X_{1t-1} + 0.388965X_{1t-2} - 0.440616X_{1t-3} - 0.224025X_{2t} - 0.004062X_{2t-1} + 0.244984X_{2t-2} - 0.079095X_{3t} + 0.010822X_{3t-1} - 0.044381X_{3t-2}$$

2. The forecasting results for the number of people living in poverty in Polewali Mandar Regency, based on the ARDL model analysis, indicate percentage variations for upcoming periods. The forecast covers the second semester of 2023 through the second semester of 2024, with the projected figures being 21.79%, 10.15%, and 16.52%, respectively.

REFERENCES

[1] K. Nurfadilah, F. R. C, and I. Kasse, "Peramalan Tingkat Suku Bunga Pasar Uang Antar Bank (Puab) Dengan Vector Autoregressive Exogenous (VARX)," *msa*, vol. 6, no. 1, p. 51, Jun. 2018.

[2] W. F. Permata, M. Rahmi, and F. I. Yusuf, "Perbandingan model arimax dan ardl untuk peramalan data (aplikasi pada banyaknya uang beredar di indonesia)," *Transformasi: Jurnal Pendidikan Matematika dan Matematika*, vol. 1, no. 2, 2017.

[3] A. Rahmasari, E. H. Sunani, M. Jannah, F. Fathulaili, L. Kurnia, and A. Satria, "ARDL Method: Forecasting Data Kemiskinan di NTB," *JTAM*, vol. 3, no. 1, p. 52, Apr. 2019.

[4] T. Kilic, H.-A. H. Dang, C. Carletto, K. Abanokova, and K. Abanokova, *Poverty Imputation in Contexts without Consumption Data: A Revisit with Further Refinements*. World Bank, Washington, DC, Nov. 2021.

[5] H. Dang, D. Jolliffe, and C. Carletto, "Data gaps, data incomparability, and data imputation: A review of poverty measurement methods for data-scarce environments," *Journal of Economic Surveys*, vol. 33, no. 3, pp. 757–797, Jul. 2019.

[6] L. Sugiharti, R. Purwono, M. A. Esquivias, and A. D. Jayanti, "Poverty Dynamics in Indonesia: The Prevalence and Causes of Chronic Poverty," *JPSS*, vol. 30, pp. 423–447, Feb. 2022.

[7] F. Nkurunziza, R. Kabanda, and P. McSharry, "Enhancing poverty classification in developing countries through machine learning: a case study of household consumption prediction in Rwanda," *Cogent Economics & Finance*, vol. 13, no. 1, p. 2444374, Dec. 2025.

[8] H. H. Dang, T. Kilic, K. Abanokova, and C. Carletto, "Poverty Imputation in Contexts Without Consumption Data: A Revisit With Further Refinements," *Review of Income and Wealth*, vol. 71, no. 1, p. e12714, Feb. 2025.

[9] M. Gualavisi and D. Newhouse, "Integrating Survey and Geospatial Data for Geographical Targeting of the Poor and Vulnerable: Evidence from Malawi," *The World Bank Economic Review*, vol. 39, no. 2, pp. 377–409, May 2025.

[10] I. Maipita *et al.*, *Mengukur kemiskinan & distribusi pendapatan*. Upp Stim Ykpn, 2014.

[11] N. M. W. Satyawati, I. M. Candiasa, and N. M. S. Mertasari, "Prediksi Penduduk Miskin di Indonesia Menggunakan Analisis Dekomposisi," *del.jur.il.pen.mat.*, vol. 9, no. 1, p. 77, Jan. 2021.

[12] D. N. Gujarati and D. C. Porter, *Dasar-Dasar Ekonometrika*. Salemba Empat, 2012.

[13] D. N. Gujarati and D. C. Porter, *Basic econometrics*, 5th ed. Boston: McGraw-Hill Irwin, 2009.

[14] B. Audina, M. Fatekurohman, and A. Riski, "Peramalan Arus Kas dengan Pendekatan Time Series Menggunakan Support Vector Machine," *IJAS*, vol. 4, no. 1, p. 34, May 2021.

[15] S. Kusumadewi and H. Purnomo, *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan Edisi 2*. Graha Ilmu, 2010.

[16] E. W. D. Dhewanty, Evy Sulistianingsih, E. Sulistianingsih, and S. Martha, "Analisis Kointegrasi dan Error Correction Model Indeks Harga Konsumen Kota Pontianak dan Singkawang," *Bimaster*, vol. 8, no. 1, Jan. 2019.

[17] M. H. Pesaran, Y. Shin, and R. J. Smith, "Bounds testing approaches to the analysis of level relationships," *Journal of Applied Econometrics*, vol. 16, no. 3, pp. 289–326, 2001.