CHAPTER III

RESEARCH METHODS

3.1 Types of Research

This study uses a quantitative method approach, which is a research approach that focuses on collecting and processing data in numerical form and analyzed using statistical methods. According to Siroj et al. (2024), the quantitative method approach is a research approach that uses numerical data and is analyzed with statistical techniques to test hypotheses objectively and systematically. This approach emphasizes the measurement of research variables using numbers and drawing conclusions based on the results of statistical analysis, so that the research results obtained are measurable and generalizable. Furthermore, this research is included in comparative quantitative research. According to Aida, Hermina, and Norlaila (2025), comparative research is a type of quantitative research method that aims to compare two or more data groups to find out whether or not there is a statistically significant difference.

Based on this description, the comparative quantitative method approach was chosen because it is in accordance with the purpose of this study, which is to analyze the difference in the performance of the processing industry sector and the large trade and retail sector on economic growth. This approach allows researchers to test differences between sectors objectively through the differential test method, so that the research results obtained have a strong empirical basis and can be scientifically accounted for, and this method also provides a strong analytical picture to help the government carry out its own regional policies.

3.2 Nature of Research

This research is descriptive and comparative, Descriptive nature is used to provide a systematic and objective picture of the condition and development of the performance of the processing industry sector as well as the large trade and retail sector in Sidoarjo Regency based on the analyzed data (Sugiyono, 2019). According to (Setiawan and Prakoso, 2023), descriptive research aims to factually explain the characteristics of an economic phenomenon based on available quantitative data without treating the research variables. In addition to being descriptive, this research is also comparative, which is research conducted to compare two objects or data groups to find out whether or not there is a statistically significant difference. (Sari & Nugroho, 2024) states that comparative research in regional economic studies is used to analyze differences in performance between economic sectors based on certain indicators, so that it can be known which sectors have a more dominant contribution to Economic Growth.

The selection of the nature of comparative descriptive research is considered appropriate because it is in line with the purpose of the research, which is not only to describe the influence of each economic sector, but also to analyze the differences in the influence of the processing industry sector and the large and retail trade sector on the economic growth of Sidoarjo Regency. According to (Kuncoro and Wahyuni, 2023), a comparative approach in regional economic research is able to provide stronger empirical evidence in explaining differences in the performance of leading sectors as the basis for regional development policy formulation.

3.3 Location and Research Period

This research was carried out in Sidoarjo Regency, East Java Province. The selection of the research location was based on the consideration that Sidoarjo Regency is one of the regions with an economic structure dominated by the processing industry sector as well as the large trade and retail sectors. The data used in this study is secondary data in the form of a time series with an observation period of 2010–2024.

3.4 Population and Sample

3.4.1 Population

Population is the entire unit of analysis used in the study. In quantitative method research based on secondary data, the population is not defined as individuals or respondents, but rather as a collection of data that is available and relevant to the research objectives (Susanto et al., 2024). According to (Setiawan and Prakoso, 2023). Population data includes the value of sectoral production, the number of workers, and investment consisting of Domestic Investment (PMDN) and Foreign Investment (PMA).

3.4.2 Sample Size Determination Technique

In time-series secondary data-based research, the determination of samples is generally adjusted to the availability and completeness of data (Putri and Kurniawan, 2024). All data in the population were used as research samples (Pratama & Lestari, 2023). The sample includes all annual economic growth data measured by GDP, Investment, Labor and production value during the period 2010–2024.

3.5 Data Source

The data source used in this study is secondary data that is quantitative. Secondary data was chosen because this study focuses on the analysis of economic sectors and regional economic growth that requires consistent, standardized, and statistically analyzeable time-lapse data. The research data was obtained from the official publication of the Central Statistics Agency (BPS), especially the BPS of Sidoarjo Regency in the publication of Sidoarjo Regency in numbers and GDP of Sidorjo Regency according to the business field.

3.6 Data Collection Methods

This study uses quantitative secondary data, secondary data was chosen because this research focuses on sectoral economic analysis and regional economic growth which requires consistent and measurable time-lapse data (Siyoto, 2020). According to (Kuncoro and Wahyuni, 2023), secondary data is very relevant to be used in macroeconomic and regional research because it is able to represent economic conditions objectively and allows for systematic statistical analysis. The data that has been collected is then processed and analyzed according to the analysis method used in this study.

3.7 Variable Operational Definition

The operational definition of variables is used to concretely explain the meaning of research variables so that they can be measured empirically and do not cause differences in interpretation. According to (Zafira, 2023), an operational definition is needed to translate conceptual variables into indicators that can be observed and measured scientifically. The dependent variable in this study is economic growth which is measured by the constant price GDP indicator which reflects the dynamics of increasing the influence of economic fluctuations in each sector analyzed.

3.7.1 Variable Dependency

The dependent variable in this study is Economic Growth which is measured by the constant price GDP indicator (Y), which is the gross added value produced by the processing industry sector as well as the large trade and retail sectors in a certain period. GDP is used as the main indicator to describe the ability of an economic sector to produce the output of goods and services in a region. Sectoral GDP reflects the level of economic activity that occurs in each sector and its contribution to the regional economy.

In regional economic analysis, changes in the value of GDP over time reflect sectoral economic growth. Economic growth is defined as the process of increasing real output that is sustainable in the long term. According to (Mankiw, 2021), economic growth reflects an increase in an economy's ability to produce goods and services measured based on an increase in real output, not just a nominal increase due to price changes. Therefore, economic growth is an important indicator in assessing the development of a sector or region.

In this study, sectoral GDP is used to represent economic growth in the processing industry sector and the trade sector large and retail. The measurement of GDP is carried out on a constant price basis, as this measure is able to describe the increase in the real output of the sector without being affected by changes in price levels or inflation. The use of constant prices aims to eliminate the inflation bias so that the observed changes in the value of GDP are truly reflective increase in real economic activity in related sectors. This approach is in line with the sectoral

GDP calculation method applied by the Central Statistics Agency in the publication of regional economic statistics (Central Statistics Agency, 2023). Operationally, the economic growth variables in this study are measured using the GDP value of the processing industry sector and the large and retail trade sector on a constant price basis. The sectoral GDP data is then analyzed to find out the difference in GDP value between the two economic sectors studied.

3.7.2 Independent Variables

Independent variables are variables that act as explanatory elements that can affect changes in dependent variables. In economic studies at the regional level, independent variables usually reflect the economic sectors as well as the elements of production that contribute to the performance of economic growth. According to (Pratama and Hidayat, 2023), the analysis of independent variables based on economic sectors is important to understand differences in economic growth performance through the structure of production, investment, and sectoral labor absorption. In this study, independent variables consist of two economic sectors, namely the processing industry sector and the large trade and retail sector. Each sector is measured using three indicators, namely the value of production, investment, and labor. Operationally, the processing industry sector and the large and retail trade sector in this study are measured through three indicators, namely:

Production value

The value of production reflects the amount of output produced by the processing industry sector in a given period measured in billions of rupiah. The production value of the processing industry sector in this study is defined operationally as the total real output value or nominal value of all physical and chemical transformation results from raw materials or raw materials to semi-finished goods or finished goods in the Sidoarjo Regency area. This variable includes the accumulation of output from processing industry business commodities on a micro, small, medium, and large manufacturing scale. In order to reflect the real value of production, this variable is measured through the indicator of the output value of the processing industry on a constant price basis (ADHK). The annual production value of the processing industry sector in Sidoarjo Regency throughout the 2010–2024 time period is presented in units of Rupiah (Billion Rp) sourced from periodic publications of the Central Statistics Agency (BPS Sidoarjo Regency, 2024).

Investment

Investment in the processing industry sector in this study is operationally defined as real capital or gross capital accumulation that is invested and realized, both through Domestic Investment (PMDN) and Foreign Investment (PMA) schemes, for the procurement of capital goods, machinery, factory buildings, and physical supporting infrastructure in the manufacturing industry business field in the Sidoarjo Regency area. This operational definition refers to an empirical study by (Sari, 2023) which confirms that the realization of secondary sector investment reflects the expansion of long-term real capital capacity that is free from short-term price fluctuation distortions. The indicator used to measure this variable is the total value of annual real investment realization in the the processing industry sector expressed in units of Rupiah (Billion Rp). The source of this variable data is obtained periodically from the official reports of the Investment and One-Stop Integrated Services Office (DPMPTSP) and the Central Statistics Agency (BPS

Sidoarjo Regency, 2024).

The workforce in this study refers to the capacity of the labor market absorption by the processing industry (X_3) sector and the large trade and retail sector (X_3) in the Sidoarjo Regency area. This variable is operationally defined as the number of the population of the workforce aged 15 years and above who are actively absorbed and depend on their main source of income in these two business fields during the observation year. In accordance with the empirical foundation of (Saputra, 2022), the absorption rate of sectoral labor indicates the flexibility and ability of the secondary and tertiary sectors to accommodate the abundance of local labor force to encourage regional productivity. This variable is measured quantitatively based on the absolute number of people working in each business field using the unit of Soul (People). Secondary data for these two variables are collected periodically from the annual report of the National Labor Force Survey (Sakernas) published by the Central Statistics Agency (BPS Sidoarjo Regency, 2024).

3.8 Data Analysis

3.8.1 Classic Assumption Test

The normality test aims to find out whether the residual of the regression model follows the normal distribution, since the t-test and the F-test assume residual normality in order for the statistical test to be valid. This study uses *Jarque–Bera* which measures the deviation of the residual distribution from normal based on skewness:

$$JB = n\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right)$$

In practice, residual is categorized as normal if *the p-value* > 0.05, and is categorized as abnormal if the *p-value* < 0.05. This criterion indicates the extent to which the residual distribution corresponds to normal assumptions. This category is important because the normal residual distribution ensures the validity of the t- and F tests, reduces the risk of type I and type II errors, and keeps the interpretation of the regression coefficient unbiased. If the residual is abnormal, the study should consider data transformations, such as logs or differentiation, to improve the residual distribution. The main objective of this test is to ensure that the regression model provides valid estimates and reliable predictions, as the Jarque–Bera test combines information about skewness and kurtosis to provide a thorough evaluation of the residual distribution

3.8.2 Heteroscedasticity test

The heteroscedasticity test aims to find out whether the residual variance is constant (homoscedasticity) or not (heteroscedasticity). Unconstant residual variance can reduce the efficiency of standard error estimation, resulting in statistical tests being biased and research conclusions less accurate. This study uses the Breusch–Pagan–Godfrey Test, which tests whether residual squares are affected by independent variables. The hypotheses tested were H_0: constant residual variance (no heteroscedasticity) and H_1: non-constant residual variance (heteroscedasticity occur). The LM statistical formula for the Breusch–Pagan–

Godfrey test is as follows:

Description:

a. n = number of observations
b. R^2_{aux} = the coefficient of determination of the residual quadratic regression with respect to all independent variables
c. k = number of independent variables Test criteria:
d. If *the p-value* > 0.05, the residual is considered homoskedastic → stable model, efficient error standards, and reliable statistical tests.
e. If *the p-value* < 0.05, heteroscedasticity occurs → variable residual variance, the standard error is inefficient, so the model needs to be improved (e.g., log transformation, weighted least squares, or robust regression).

   Test Result Categories:

f. Homoskedastic: residual variance is stable, the regression coefficient can be interpreted accurately, the t and F tests are valid.
g. Heteroskedastic: residual variance is unstable, standard error bias, so that the interpretation of the coefficient becomes unreliable and the model's prediction is invalid.

The main purpose of this heteroscedasticity test is to ensure the efficiency of the regression coefficient estimation and maintain the validity of the statistical test. Thus, the regression results obtained can be used as the basis for economic policy analysis or GDP predictions with a high level of reliability. The Breusch–Pagan–Godfrey Test is an option because it effectively detects changes in residual variance as an independent variable function, and provides more accurate information than classical tests such as Glejser or White.

3.8.3 Multicollinearity test

The multicollinearity test aims to detect whether there is a high linear relationship between independent variables in a multiple linear regression model. Multicollinearity can make the regression coefficient unstable, increase the standard of error, and make it difficult to interpret the influence of each variable. The type of multicollinearity test used in this study is the Variance Inflation Factor (VIF), which measures the extent to which the estimated variance of the regression coefficient increases due to the correlation between independent variables.

The VIF formula for the th independent variable is:

$$VIF_i = \frac{1}{1 - R_i^2}$$

Hipotesis:

a. H_0: no significant multicollinearity occurs
b. H_1: significant multicollinearity occurs

VIF criteria and categories:

a. VIF < 5 → there is no multicollinearity, independent variables can be interpreted clearly
b. ≤ VIF ≤ 10 → of moderate multicollinearity, interpretation should be careful as some of the information overlaps
c. VIF > 10 → high multicollinearity, the model needs to be improved (e.g. by removing redundant variables)

VIF criteria and categories:

a. VIF < 5 → there is no multicollinearity, independent variables can be interpreted clearly
b. ≤ VIF ≤ 10 → of moderate multicollinearity, interpretation should be careful as some of the information overlaps
c. VIF > 10 → high multicollinearity, the model needs to be improved (e.g. by removing redundant variables)

The purpose of the multicollinearity test is to ensure that each independent variable exerts a unique influence on the dependent variable, maintains the stability of the regression coefficient, and allows the interpretation of the influence of the variables to be carried out accurately. With the fulfillment of this test, regression models can produce reliable and valid estimates for data-driven decision-making.

3.8.4 Autocorrelation test

The autocorrelation test aims to find out whether the residual of the regression model correlates between periods, especially in *time series* data. Autocorrelation can make regression coefficients inefficient and standard errors biased, so the interpretation of the influence of independent variables on dependents is inaccurate. The type of autocorrelation test used is the Breusch–Godfrey Serial Correlation LM Test, which is capable of detecting autocorrelation up to the order of high and in models that have dependent variable lag.

The autocorrelation test aims to find out whether the residual of the regression model correlates between periods, especially in *time series* data. Autocorrelation can make regression coefficients inefficient and standard errors biased, so the interpretation of the influence of independent variables on dependents is inaccurate. The type of autocorrelation test used is the Breusch–Godfrey Serial Correlation LM Test, which is capable of detecting autocorrelation up to the order of high and in models that have dependent variable lag.

Formula *Breusch–Godfrey LM Test:*

$$LM = n \cdot R^2_{aux} \sim \chi^2_p$$

Description:

a. n = number of observations
b. R^2_{aux} = the coefficient of determination of the residual regression to the residual lag and independent variables

Hipotesis:

a. H_0: no autocorrelation occurs (independent residual)
b. H_1: autocorrelation occurs (correlated residuals)

Criteria and categories:

a. If the p-value > 0.05 → independent residual, the model is autocorrelated free
b. If the p-value < 0.05 → autocorrelation occurs, the model needs to be corrected, for example by adding a lag of dependent variables or using the Generalized Least Squares (GLS) method.

The purpose of this test is to ensure that the residual is independent between periods, maintain the efficiency of the regression coefficient, and improve the accuracy of the model's predictions. This test is important because unaddressed autocorrelations can degrade the reliability of models in economic analysis, including GDP predictions.

3.8.5 Multiple Linear Regression Analysis

This model describes the linear relationship between one bound variable and more than one independent variable (Damodar N. Gujarati, 2021). In a quantitative method study that compares two independent groups, namely the processing industry sector and the large trade and retail sector, regression models are prepared separately for each sector.

Mathematically, the common forms of multiple linear regression are (Gujarati, 2021):

$$Y = \alpha + \beta 1X1 + \beta 2X2 + \beta 3X3 + \varepsilon$$

Based on this general form, the model in this study is formulated as follows:

Processing Industry Sector Model

$$Y_A = \alpha_A + \beta_{1A}X_{1A} + \beta_{2A}X_{2A} + \beta_{3A}X_{3A} + \varepsilon_A$$

Description:

- Y_A = Economic growth of the processing industry sector
- X_1_A = Production value X_2_A = Investment X_3_A = Labor α_A = Constant
- β_1A, β_2A, β_3A = Coefisien regressions
- ε_A = Error term

Large Trade and Retail Sector Model

$$Y_B = \alpha_B + \beta_{1B}X_{1B} + \beta_{2B}X_{2B} + \beta_{3B}X_{3B} + \varepsilon_B$$

Description:

- Y_B = Economic growth of the large trade and retail sector
- X_1_B = Production value
- X_2_B = Investment
- X_3_B = Labor α_B = Constant
- $\beta_1 B$, $\beta_2 B$, $\beta_3 B$ = Coefficient regressions
- ε_B = Error term

With the use of these two separate regression models, the study was able to compare the magnitude of the influence of each independent variable on economic growth in both sectors more comprehensively and identify which sectors have a more dominant contribution to regional economic growth (Gujarati, 2021). This separation of models also allows for a more in-depth analysis of the differences in the economic structure and characteristics of each sector, so that the resulting parameter estimation becomes more specific and unbiased due to the incorporation of cross-sector data. In addition, the estimated results of each model can be compared through the value of the determination coefficient (R^2), the significance of the parameters, and the magnitude of the regression coefficient to determine the sectors that have a stronger influence on regional economic growth (Imam Ghozali, 2021).

3.8.6 Hypothesis Test

In multiple linear regression, hypothesis testing includes t-test (partial), F-test (simultaneous), and Coefficient of Determination (R^2) test.

3.8.7 F Test

The simultaneous test, commonly referred to as the F-test, is employed in multiple linear regression analysis to evaluate the joint performance of the structural model. The primary objective of this statistical test is to determine whether all the independent variables included in the model simultaneously exert a statistically significant effect on the dependent variable. In academic research, this test serves as a crucial indicator to verify whether the combination of predictors can reliably predict the outcome variable, ensuring that the overall model structure is valid and meaningful rather than a result of mere coincidence.

For the purpose of this analysis, the testing is strictly conducted using a 10 percent significance level (alpha = 0.10). This threshold implies that the study tolerates a maximum 10 percent probability of committing a Type I error, which is the risk of incorrectly rejecting a true null hypothesis. To conduct the simultaneous evaluation, the statistical hypotheses are formulated as follows:

1. H0 (Null Hypothesis) The independent variables simultaneously do not have a significant effect on the dependent variable. This indicates that the regression model lacks overall explanatory power, meaning all partial regression coefficients are jointly equal to zero.
2. H1 (Alternative Hypothesis) The independent variables simultaneously have a significant effect on the dependent variable. This indicates that the regression model possesses sufficient explanatory power, meaning at least one of the predictor variables contributes significantly to explaining the variance in the dependent variable.

The systematic criteria for decision-making regarding the acceptance or rejection of the null hypothesis are established based on the comparison of either the probability values or the empirical test statistics:

1. Rejection of H0 If the observed empirical Significance value (Sig.) is less than the established alpha level of 0.10 (Sig. less than 0.10), or if the calculated empirical F-value exceeds the critical value obtained from the statistical distribution table (F-statistic greater than F-table), then the null hypothesis (H0) is definitively rejected and the alternative hypothesis (H1) is accepted. This statistical outcome demonstrates that the independent variables collectively and simultaneously exert a significant influence on the dependent variable.
2. Acceptance of H0 Conversely, if the observed empirical Significance value (Sig.) is greater than the established alpha level of 0.10 (Sig. greater than 0.10), or if the calculated empirical F-value is lower than the critical statistical value (F-statistic less than F-table), the null hypothesis (H0) cannot be rejected, meaning H1 is rejected. This outcome confirms that the independent variables simultaneously do not have a significant effect on the dependent variable, indicating that the combined predictors fail to explain the shifts in the model's outcome.

3.8.8 T test

The t-test is used to test the influence of each independent variable partially on the dependent variable. This test determines whether an independent variable significantly affects the dependent variable by taking into account the other variables in the model. The formula of the t-test is:

$$t = \frac{b_i}{SE(b_i)}$$

Description:

a. *b_i* = regression coefficient for independent variable i

b. *SE(b_i)*= standard error regression coefficient i Hypothesis

c. H_0: bi = 0→ independent variables have no significant effect

$$F = \frac{(SSR/k)}{(SSE/(n-k-1))}$$

Description:

a. *SSR* = sum of squares regression (sum of the regression squares)

b. *SSE* = sum of squares error (number of squares of error)

c. *k* = number of independent variables

d. *n* = number of observations

Criteria and categories:

a. If |*t-calculate*| > *t-table* or *p-value* < *0.05*, the independent variable has a significant effect on the dependent variable

b. If |*t-calculate*| < *t-table* or *p-value* > *0.05*, independent variables have no significant effect

The purpose of the t-test is to find out which variables make significant contributions in the model. The results of this test allow researchers to assess the specific influence of each factor, for example whether investment, production value, or labor individually affect the GDP of Sidoarjo Regency.

3.8.9 Coefficient of Determination (R^2)

The coefficient of determination (R^2) serves to assess how well the regression model can explain changes in dependent variables. R^2 describes the part of the change in dependent variables that can be explained by independent variables in the model. The formula for calculating R^2 is:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Description:

- SSR = sum of regression squares
- SSE = sum of squares of error
- SST = total number of quadrarates

Criteria: $0 \leq R^2 \leq 1$

The closer it is to 1, the more the model is able to explain the variation of dependent variables. The aim of R^2 is to assess the overall quality of the regression model. The high R^2 value indicates that independent variables in this study, such as the value of production, investment, and labor, are able to explain most of the variation in the economic growth of Sidoarjo Regency. And of course, there are also other factors that can explain the rise and fall of this value so that a balance of results is created in independent variables, namely the value of production, investment and labor to economic growth measured by GDP.