CHAPTER III

RESEARCH METHODOLOGY

A. Operational Variable Definition

The research variables according to Sugiyono (2011) everything that shaped as well as certain traits or activities that have certain variations set by the researcher to be examined and then drawn conclusions. It can be said as the object of research that will be a point of attention to be investigated by researchers.

Objects in this study include 5 districts/cities of D.I Yogyakarta, namely Kulonprogo district, Bantul district, Gunung Kidul district, Sleman district and Yogyakarta city. Some operational definitions of variables in this research, as follows:

1. Dependent Variable.

The data on economic growth based on economic growth in 5 districts/cities of D.I Yogyakarta which is calculated by GRDP at constant market prices and stated in millions of rupiah. The data of this research calculated the value of GRDP at 2000 constant market prices period 1996-2015.

2. Independent Variables.

a. Local Original Revenue (X1).

Local original revenue is one of the values of the budget plan of government's receipt that expressed in millions of rupiah. This research uses data of local original revenue in 5 district/cities of D.I Yogyakarta period 1996-2015.

b. Government's Expenditures (X2).

Government's expenditure is the value of budget plan of government's expenditure that expressed in millions of rupiah. This research uses data of government's expenditure in 5 district/cities of D.I Yogyakarta period 1996-2015.

c. Population (X3).

The population is the number o individuals who registered officially and resides on a specific area within a specific in 5 district/cities of D.I Yogyakarta period 1996-2015.

d. Fiscal Decentralization (X4).

The fiscal decentralization in this research is proportional to the realization of the local original revenues compared to the total receipts expressed in the formula as follows:

$$DF = \frac{PAD}{TPD} x\ 100\%$$

Where:

DF : Fiscal Decentralization (%)

PAD : Local Original Revenue (million rupiahs)

TPD : Total receipts (million rupiahs)

B. Type and Sources of Data

Type of data use in this research is quantitative data that is using secondary data with time-series and cross-section. The data comes from Central Bureau of Statistics (BPS) in 5 district/cities of D.I Yogyakarta period 1996-2015. In this research data used Gross Regional Domestic Product (GRDP) at 2000 constant market price, Local Original Revenue, Government's Expenditures, Population and Fiscal Decentralization.

C. Methods of Analysis

Data analysis methods to be used in this research is a method of quantitative analysis, it is analysis technique that can be used to estimate the parameter. Data analysis is done by testing statistics on the variables that have been collected with the help of software EViews 7. The result is expected to be used to determine the magnitude of the impact of the independent variable on the dependent variable.

The econometric model is used in this research to determine the inter-relationships between theory formulations, tests, and empirical estimates. In econometric theory, panel data is a combination of cross-section data and time-series data. Thus, the amount of observation data in panel data is time series observation data (t>1) and cross-section observation data (n>1). The basic model to be used in this research is as follows:

$$Y = \beta_0 + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 X_{3it} + \beta_4 X_{4it} + u$$

The regression model in this research is based on models that have been used by Pujiati (2007) and Sodik (2007). It is transforming the equation of regression into logarithmic form. Then the formulation transformed in logarithmic form with the following equation:

$$LogY = \beta_0 + \beta_1 Log\ X_{1it} + \beta_2\ Log\ X_{2it} + \beta_3\ LogX_{3it} + \beta_4 X_{4it} + u$$

Where:

Y : GRDP at constant market prices (in log)

β_0 ,β_1 ,β_2 ,β_3 ,β_4 ,β_5 : Coefficient

X_1 : Local original revenue (in log)

X_2 : Government's expenditures (in log)

X_3 : Population (in log)

X_4 : Fiscal Decentralization (percent)

i : Districts/Cities

t : Year

u : Error term

D. Estimation of Regression Model

According to Ghozali (2006) some advantages of using panel data in research: a) Panel data is combining the data time series and cross-section, the panel data provides data more informative, more varied, low multicollinearity between variables, a greater degree of freedom and more efficient; b) Panel data relate to individuals, firms, cities, countries at all times, it will be heterogeneous within the unit. Techniques for estimating panel data can explicitly enter heterogeneity for each individual variable; c) Panel data is able to detect and measure unobservable effects through time-series data or cross-section data; d) Panel data allows to learning more complex behavioral models.

Before performing regression analysis, the step taken is to test the model estimations to obtain the most appropriate model estimation used. Then the next step is to test the classical assumption to test the research hypothesis.

Regression analysis with panel data can be done with three estimation method:

1. Common Effect Model.

Common effect model estimation is an estimation of panel data that only combines time series and cross-section data using Ordinary Least Square (OLS) method. This approach does not take into account individual dimensions or time. In this model, there is an assumption that intercepts and regression coefficients are fixed for each research object and time.

2. Fixed Effect Model.

Fixed effect model assume that each object has a different intercept but has the same coefficients. To distinguish between one object with another then used dummy variables or pseudo variables so this method is also called Least Square Dummy Variables (LSDV). The use of a dummy in this research is using dummy area. By inserting the dummy variable region into the equation, the equation model is as follows:

$$LogY = \beta_0 + \beta_1 Log\ X_{1it} + \beta_2\ Log\ X_{2it} + \beta_3\ LogX_{3it} + \beta_4 LogX_{4it} + \beta\ 1D1 + \beta\ 2D2 + \beta\ 3D3 + \beta\ 4D4 + \beta\ 5D5 + u$$

Where:

Y : GRDP at constant market prices (in log)

β_0 ,β_1 ,β_2 ,β_3 ,β_4 ,β_5 : Coefficient

X_1 : Local original revenue (in log)

X_2 : Government’s expenditures (in log)

X_3 : Population (in log)

X_4 : Fiscal Decentralization (percent)

D1 : Dummy Kulonprogo

D2 : Dummy Bantul

D3 : Dummy Gunung Kidul

D4 : Dummy Sleman

D5 : Dummy Yogyakarta

i : Districts/Cities

t : Year

u : error term

3. Random Effect Model.

This method does not use dummy variables such as those used in the fixed effect method. This method uses residuals that are suspected to have inter-time and inter-object relationships. Random effect Model assumes each variable has different intercept but the intercept is random or stochastic. Thus the model equation becomes:

$$Y_{it} = \beta_{0i} + \beta_1 I_{it} + v_{it} \text{ where } v_{it} = e_{it} + u_i$$

In this method, residual v_{it} consists of two components, namely (1) residual e_{it} which is the residual overall as well as a combination of time series and cross-section; (2) the individual residuals represented by u_i. In this case, each object has a different residual u_i but remains inter-time. Generalized Least Square Method (GLS) is used to estimate this regression model as a substitute for the OLS method.

E. The Selection of Panel Data Model

To select the model estimation which is considered the most appropriate for the three types of models, then to do a series of test.

1. Chow Test (F-Test).

Chow test is used to know between two models to be selected for data estimation, between Pooled Least Square (PLS) or Fixed Effect Model (FEM). Testing this test to determine the model of common effect or the model of fixed effect is most appropriate used in panel data estimation. The hypothesis in Chow test:

H_0 : Common effect model/PLS

H_1 : Fixed effect model/FEM

To find out is used Chow test is formulated as follows:

$$\text{Chow} = \frac{(\text{RRSS} - \text{URSS})/|\ (\text{n} - 1)}{\text{URSS}/(\text{nT} - \text{n} - \text{k})}$$

Where:

RRSS : Restricted Residual Sum Square (*Sum of Square Residual* that obtained from PLS model)

URSS : Unrestricted Residual Sum Square (*Sum of Square Residual* that obtained from FEM)

n : Number of *cross section* data

T : Number of *time series* data

k : Number of independent variables

While F-table is formulated as follows:

$$\text{F-table} = \{\alpha : df\,(n - 1,\ nt - n - k)\}$$

where:

α : Level of significance (alfa)

n : Number of cross section data

nt : Number of cross section data x number of time series data

This test uses the F statistical distribution. If the value of F-stat > F-table then the model to be used is the FEM model. Whereas if F-stat < F-table then the model to be used is the PLS model.

2. Hausman Test.

According to Nachrowi et al (2005) to select the fixed effect model (FEM) or random effect model (REM) as an appropriate model there are several ways to determine:

a. If T (number of time-series data)> N (number of cross-section data), then use fixed effect model (FEM).

b. If N (number of cross-section data)> T (amount of time-series data), then use random effect model (REM).

c. If the cross-sectional effect is correlated with one or more variable X, then the FEM estimator is unbiased and appropriate.

d. Hypothesis test which can be used to more convince decision in choosing best model is by using Hausman test.

Hausman test is used to determine the FEM or REM model to be selected. This test is based on the idea that both OLS and GLS methods are consistent but OLS is not efficient in Ho. Following Wald's criterion, this Hausman test will follow the following chi-squares distribution.

$$m = \hat{q}\,'\, var(\hat{q})^{-1}\, \hat{q}$$

$$\text{where } \hat{q} = [\beta_{OLS} - \beta_{GLS}]$$

$$\text{and } var(\hat{q}) = var(\beta_{OLS}) - var(\beta_{GLS})$$

This statistic follows the statistical distribution of chi-squares with df as much as k, where k is the number of independent variables. If the value of Hausman stat > the critical value then the exact model is the FEM model and vice versa.

3. Lagrange Multiplier Test.

The Lagrange Multiplier test is used to determine between a random effect model (REM) or a PLS model. The test was developed by Bruesch-Pagan in 1980. This LM test is based on the residual value of the PLS model. The LM statistic value is calculated based on the following formula:

$$LM = \frac{nT}{2(T-1)}\left(\frac{\sum_{i=1}^{n}(Te_{it}^{\pi})^2}{\sum_{i-1}^{n}\sum_{t=1}^{T}\hat{e}_{it}^{2}} - 1\right)^2$$

Where:

n : number of individuals

T : number of period time

ê : residual PLS method

The LM test is based on the distribution of chi-squares with a df value (degrees of freedom) as much as the number of independent variables. If the value of LM stat > stat values of chi-squares then the selected model is REM model and vice versa.

F. Classical Assumption Test

The need for classical assumption testing in this research depends on the results of the estimation method selection. According to Gujarati and Porter (2009) equations that satisfy classical assumptions, in estimation model estimation using GLS method only Random Effect Model, while Fixed Effect Model and Common Effect Model using Ordinary Least Square (OLS). When based on the selection of an appropriate estimation method for regression equation is a random effect, then no need for a classical assumption test. Conversely, if the regression equation is more suitable to use the common effect model or fixed effect model (OLS) then it is necessary to test the classical assumption.

1. Normality Test.

Normality test in regression analysis was conducted to test whether the data to be researched has a residual variable that normally distributed. In this study used Jarque-Bera test statistics contained in the EViews program. If the Jarque-Bera probability value is greater than the alpha value ($p > \alpha$) then the data is normally distributed whereas if the probability value is less than the Alpha value ($p < \alpha$) then the data is not normally distributed.

2. **Multicollinearity Test.**

The linear relationship between independent variables in multiple regressions is called multicollinearity. Models with large standard errors and low t-statistical values are an early indication of muticollinearity problems. In this research, multicollinearity test is done by testing the correlation coefficient (r) between independent variables. If the correlation coefficient is high enough that > 0.9 then it can be concluded the existence of multicollinearity problem. But if the correlation coefficient is less that <0.9 then there is no problem multicollinearity.

3. **Heteroscedasticity Test.**

Heteroscedasticity is a deviation of OLS assumptions in the form of impaired variance estimation, which is generated by OLS estimation that not constant. Formally homoscedasticity can be expressed:

$$\text{as Var}(u|X_1,X_2,\ldots\ldots..,X_k)=\delta^2$$

If such assumption is violated then heteroscedasticity may occur as follows:

$$\text{Var }(u|X_1,X_2,\ldots\ldots..,X_k)=\delta_i^2$$

Where the notation i states that the variance is changed from each research object.

Heteroscedasticity test is not only by looking at the Plot Scatter or in the disturbance pattern. Some statistical methods can be used to determine whether a model is free of heteroscedasticity problems such as White Test, Park Test, Gletjer Test, and others. The hypothesis whether heteroscedasticity or not is as follows:

H_0 : There is no heteroscedasticity

H_1 : There is heteroscedasticity

In this study the heteroscedasticity test was performed with the EViews program through Park Test .if the *p*-value of probability is greater than *Alpha value* ($p > \alpha$) then the variance error is homoscedasticity, whereas if the *p*-value of probability value is less than the Alpha value ($p < \alpha$) then the variance error is heteroscedasticity.

G. Statistical Test

1. **Partial Test (t-test).**

The t-test is used to know the magnitude of the impact the independent variable on the dependent variable. the value of the t-statistic was known, then compared with the value of the t-table by using two directions on a certain confidence level.

The independent variable is said to be significant to the dependent variable if the t-statistic value of the independent variable is located within the critical area or in other words the statistical t-value is greater than the value of t-table (t-statistic > t-table). This means that there is impact of the independent variable on dependent variable. Vice versa if the value of t-statistics is smaller than the value of t-table (t-statistic < t-table), then it can be said there is no impact of independent variables to the dependent.

2. **Significance of Stimulant F (F-test).**

The f-statistic test basically indicates whether all independent variables included in the model have the impact of the dependent variable simultaneously. To know the impact of independent variable on the dependents variable simultaneously, use F-test by making the hypothesis as follows:

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

All independent variables have no an impact on dependent variable simultaneously.

$H_1 : \beta_1 \neq \beta_2 \neq \beta_3 \neq \beta_4 \neq 0$

All independent variables have an impact on dependent variable simultaneously.

This means that all independent variables are simultaneously a significant explanation of the dependent variable. To test this hypothesis used F-statistic with decision-making criteria are comparing the value of F-statistic with 95% confidence level or comparing the value of F-calculation with the value of F-table. F-statistics can be fulfilled with the following formula:

$$\mathrm{F-statistic} = \frac{\mathrm{R2}/(\mathrm{k}-1)}{1-\mathrm{R2}/(\mathrm{n}-\mathrm{k})}$$

Where:

R2 : Coefficient of determination

k : Number of independent variables included constants

n : Number of cross section data

How to perform the F-test as follows:

a. Quick look: if the F-value is greater than 4 then Ho is rejected with 95% confidence degree H1 accepted, which means that all independent variables are stimulant and significant influence on dependent variable.
b. Compare the value of F-statistics with F-table. If the F-statistic value is greater than F table then the alternative hypothesis is accepted.

If the probability value F stat < 95% confidence level then H_0 rejected and accept H_1, it means there is influence of independent variables together with the dependent variable. Conversely if, F statistic > 95% confidence level then H_0 is accepted and H_1 is rejected, it means there is no relation between the dependent variable and the independent variable. In addition to this method, hypothesis testing can also be done by comparing the F-count value with F-table. If the F-count value is greater than F-table then H_0 is rejected and vice versa.

3. Coefficient of Determination (R^2).

The coefficient of determination (R^2) basically measures how far the ability of a model in explaining the dependent variable. Where the coefficient that measures how much variation of the dependent variable can be explained by the variation of the independent variable, where the value of R^2 has a range of values 0 to 1. The closer to 1 is better.