3. RESEARCH METHODOLOGY

This paper investigates the impact of corporate governance especially board diversity as the independent variable on a company's tax aggressiveness as the dependent variable in Indonesia's energy industry. Other variables include the firm size, leverage and return on assets which acts as the control variables. These variables have all been disclosed and discussed in the previous chapters, and so the research model incorporated in this study will be shown below:

3.1 Theoretical Model

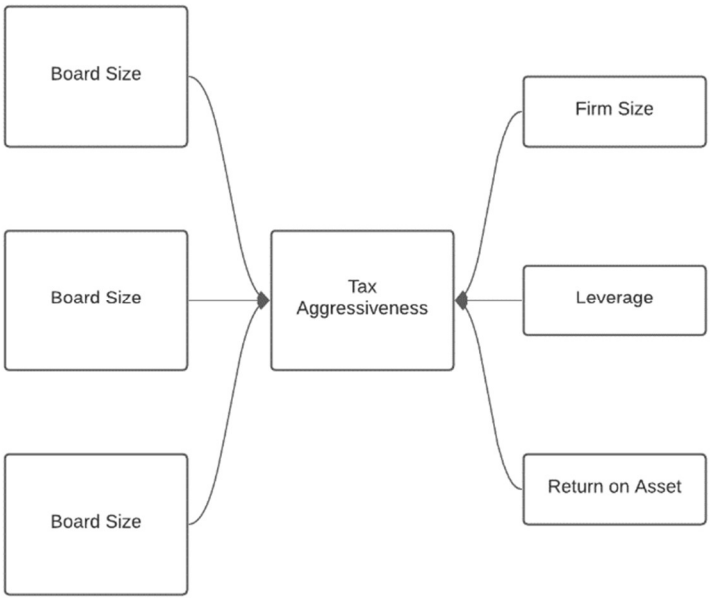

Figure 3.1 Research Model

According to the depiction of the figure above, this research will be focusing on the correlation between the structures and diversity of the board in a company and the influence it has on a company's tax aggressiveness. The independent variable is made up of board size, board independence, and board gender diversity, all of which make up the company's board structure and diversity. Meanwhile for the dependent variable, it will be the company's rate of tax avoidance, measured using ETR or commonly known as tax aggressiveness. The control variables are firm size, leverage and its profitability, measured using ROA.

The model of testing that will be used for testing the hypothesis is therefore:

$$ETR_{it} = SIZE_{it} + IND_{it} + BGD_{it} + FSIZE_{it} + FLEV_{it} + FROA_{it} \quad (3.1)$$

Where:

α = Intercept of the regression line in which it is constant

$SIZE_{it}$ = Board size of company *i* in time *t*

IND_{it} = Board Independence of company *i* in time *t*

BGD_{it} = Board Gender Diversity of company *i* in time *t*

$FSIZE_{it}$ = Firm Size of company *i* in time *t*

$FLEV_{it}$ = Leverage of company *i* in time *t*

$FROA_{it}$ = Return on Asset of company *i* in time *t*

ε = Error of the company *i* in time *t*

3.2. Operational Variable and Measurement

Moving deeper into the variables, this paper integrates three main ideas behind board diversity. Those are board's independence, board size, and percentage of women on board. The Tax aggressiveness as the independent variable will make use of the ETF formula or effective tax rate to measure a company's real tax rate. Finally, there are two more indicators acting as the control variable which are the firm's size and also its leverage. This will all be explained further in detail in the table below.

3.2.1. Independent Variable

Table 3.1

Independent Variable

Indicators	Variable Operationalization	Scale of Measurements
Board Size (SIZE)	One of the most general yet fundamental aspects of the board, the size of the board can indicate the intensity of decision making as well as intensity of the problems faced. It is calculated simply by the total	**Ratio (revisi)**

	amount of directors present on board.	
Board Independence (IND)	The level of independence inside a board can indicate the level of subjectiveness inside the decision-making process, which is very important in deciding tax aggressiveness. It is calculated by taking the number of independent directors as a proportion of total number of directors on the board room. **Board Independence** = Number of independent director / Total number of directors	Ratio
Board Gender Diversity (BGD)	Gender can be one of the key components in order to highlight the level of diversity present inside a board. It is therefore calculated as the number of female directors as a proportion of total number of directors on the board room and will represent the board's overall demographic diversity of the company. **Board Gender Diversity** = Number of female directors / Total number of directors	Ratio

Source: Author's compilation

3.2.2. Dependent Variable

Table 3.2

Dependent Variable

Indicators	Variable Operationalization	Scale of Measurements
Tax Aggressiveness	Companies usually take extensive measures	Ratio

	conducted by the management to reduce payments of tax to the government. To measure the level of tax avoidance present, the total tax expense of a company for the period is divided by the pre-tax book income for the period. This will result in a rate of taxes paid by a company in a given period, and will be compared to other periods to measure tax aggressiveness. **Tax Aggressiveness (ETR)** = Total Tax Expense / Pre-tax net income for the period	

Source: Author's compilation

3.2.3. Control Variable

Table 3.3

Control Variable

Indicators	Variable Operationalization	Scale of Measurements
Firm Size (FSIZE)	Natural logarithm of total assets which the company possess and have control on. Regarded as the control variable as it may affect the tendency of management to be more tax aggressive. *Firm Size = ln(total assets)*	Ratio
Firm Leverage (FLEV)	The amount of debt which the company has to finance assets and represent as control variable. This may affect the tendency of management to be more tax aggressive, and so will be controlled during this research.	Ratio

	Firm Leverage = Total Debt / Total Asset	
Firm Return on Asset (FROA)	To control for profitability, as the amount of profit can directly impact the tendencies to commit more aggressive tax avoidance practices. It is calculated as the pre-tax income of a company in a given period divided by the total assets in the same period. *Firm ROA = Pre-tax net income / Total Asset*	Ratio

Source: Author's compilation

3.3. Type and Source of Data

This research will be assessing quantitative data. Quantitative data itself is the data more commonly expressed in numbers and are prevalently used for most statistical methods. This type of approach is deemed to be more suitable in examining cause and effect relationship (Ahmad, 2019). This paper will then utilize a quantitative approach and so will present data and results using statistical methods expressed in a nominal form.

This paper will be using secondary data extracted from the annual reports and financial statements of Indonesia's energy industry listed in Indonesia's Stock Exchange. Further details of the type and source of data will be shown below:

Table 3.4

Types and Sources of Data

Data	Type of Data	Source of Data
Board Size	Quantitative	Annual Report
Board Independence	Quantitative	Annual Report
Board Gender Diversity	Quantitative	Annual Report
Tax aggressiveness	Quantitative	Annual Report / Financial Statements
Firm Size	Quantitative	Bloomberg database

Firm Leverage	Quantitative	Annual Report / Financial Statements
Firm Return on Asset	Quantitative	Annual Report / Financial Statements

Source: Author's Compilation

3.4. Instrument for Data Collection

The method of data collection used to compile data in this paper is through the documentation method. This method of data gathering combines data from many different sources, mainly secondary sources. This means that the data used are not for the purpose of this specific research, as in indirect. This research will therefore use mainly historical data of board size, board independence, and also percentage of female directors on board obtained from the company's annual report and also financial statements. To sum up, these reports will be obtained from the IDX website, Bloomberg, Tax Justice Network, and other reliable sources. Other supporting reports from a trusted publisher might also be used for reference to assist the accuracy and reliability of the research.

3.5. Population

Population is defined as the total number of samples possessing similar characteristics to be further observed for this research. One specific objective to use a population is to ensure the data tested are relevant and reliable to avoid obtaining distorted results obtained from irrelevant data. For this research, this paper will be targeting a population of 36 listed companies from the energy industry for a period of 5 years / 60 months between years 2015-2019

3.6. Sample Technique

A Sample is a fraction of the population selected to be tested and is deemed to represent the entire population data. From the two sampling methods commonly known which are probability sampling and non-probability sampling techniques, this paper will use the latter. This means that the sample does not always need to be representative of the whole population, although a clear reasoning is needed as the basis of the selection (Taherdoost, 2016). To be specific, this research will utilize purposive sampling as its main technique. Purposive as in the sample selected will be derived from a list of criteria handpicked by the

author's judgements. From the 71 listed companies selected as the population earlier, 36 will be selected as they have met the following purposive sampling criteria:

1. Companies listed in IDX consistently from years 2015-2019
2. Companies with a published annual report consistently from years 2015-2019
3. Companies within the energy sector
4. Companies with sufficient financial and taxation disclosure of information in the annual report consistently
5. Companies with a positive net income value

3.7. Unit Analysis

To sum up, the unit analysis in this research will be the annual report of listed companies in the energy industry between years 2015-2019, along with other disclosures of relevant data relating to taxation. The total firm years would be 139 firm years derived from a total of 36 companies in 5 years.

3.8. Data Analysis Technique

To complete this research methodology plan, the final consideration and perhaps the most important one to disclose is regarding the technique applied in analysing the interrelationship between the independent and dependent variables mentioned above. As the sample data for this research includes both time series and cross-sectional data, this research will attempt to use the GRETL statistical programs to assist in obtaining the results. Time series itself refers to the number of years / periods being tested, whereas cross-sectional data refers to the different companies being tested.

The steps to operate the application of the program will be further explained in detail below.

1. Data Gathering

The most fundamental step before trying to analyse any data is to gather and collect the correct and relevant data for this research. As mentioned previously, data for this research will be mostly compiled of secondary data obtained from reliable sources such as the company's annual report as well as other global economic research centre such as Bloomberg and Tax Justice Network. The variables will then be calculated, all within the period from 2015-2019 using measurements obtained from 3.2

2. Determining the descriptive statistics for the entire set of data

To assist in creating a general format for the application to read, the variables present in this paper will have its descriptive statistics computed. The descriptive statistics will range from the maximum, minimum, and median value, as well as its standard deviations. This itself will present a more useful information when used for further research

3. Run the Diagnostic Tests for Panel Regressions.

Regression analysis is a common set of methods to analyse the relationships between one dependent variable – also known as the outcome variable, and one or more independent variables. In general, there are three main variants of the regression model suitable for panel data analysis, and those are the Random Effect Model, the Fixed Effect Model, and the Ordinary Least Squared model. With the options in place, diagnostic tests are required to determine which type of model is the most appropriate for this intended research. In addition, this research will be utilizing a balanced panel data approach, meaning every member of the sample will be tested for every year without exception. Three tests will be conducted, each to compare the respective models, and those are the Hausman test, the Fixed Effect estimator, and finally the Breusch-Pagan Test.

a. Hausman test

The Hausman test is conducted to compare the Fixed Effect model and the Random Effect Model. Under this test, the Random Effect model is generally preferred due to its flexibility and efficiency and so will be the null hypothesis. The Fixed effect model will then be the alternative hypothesis.

H0: Random Effect Model

H1: Fixed Effect Model

During the test, if the results indicates that the overall probability of the Hausman test is less than the alpha (5%), this means that the H0 is rejected, meaning that it is more suitable to use the Fixed Effect Model

b. Fixed Effect estimator

The Fixed Effect estimator test is conducted to compare if the panel data is more suitable to be processed using the Ordinary least squared model or the Fixed effect model. Fixed effect models are generally more suitable for data with massive variances over time, and will be the alternative hypothesis for this test. The Ordinary Least Square model will then be the null hypothesis for this test.

H0: Ordinary Least Squared model

H1: Fixed Effect Model

If the results of the test indicates that the fixed model is less than the alpha (5%), then the H0 is rejected, meaning that it is more suitable for the research to use the Fixed Effect model

c. Breusch-Pagan Test

Breusch-Pagan test is conducted to compare between Ordinary Least Squared model or Random Effect model. The testing will look at the Chi-square goodness of fit with a significant value of alpha (5%). The hypothesis will be shown as follows:

H0: Ordinary Least Squared model

H1: Random Effect Model

If the results indicate that the Chi-square is less than the alpha (5%), then the H0 will be rejected, thus meaning that a Random Effect model is more appropriate.

4. Perform a general assumption test in the regression model

a. Heteroscedasticity test

Heteroscedasticity is essentially a regression problem in which the variance of errors is different within the observations. Where most studies assume that the variance would be homogenous, this problem has potential to misinterpret the data and thus produce misleading results (Heteroscedasticity and Homoscedasticity, 2007). For this cause, a heteroscedasticity test should then be performed in order to ensure that the tests took account for the difference in variance, assuming that a different variance exists. To test for heteroscedasticity, the following hypothesis is produced.

H0: the variables are constant or homoscedastic

H1: the variables are not constant or heteroscedastic

A significant rate of less than alpha (5%) would imply that the H0 is rejected, meaning the variables are not constant. A weighted OLS must then be conducted to prevent the research from being altered.

b. Multicollinearity test

Where the heteroscedasticity test addresses the issue of different variance of errors, the multicollinearity test addresses the issue of over – intercorrelations. Similar to the issue found in heteroscedasticity, multicollinearity can alter the results into inaccurate findings due to the oversaturated levels of dependence between the independent variables (Wilcox, 2018). In most common situations, this issue can lead to wider confidence intervals, producing less reliable probabilities. To measure multicollinearity level, a Variance Inflation Factor (VIF) can be used. Generally, the inflation factor will be at 1 if no correlations exist between the

independent variables. As the number increases, it can be assumed that multicollinearity exists. The rule of thumb would be to limit the number at 10. Anything above this number can be subject to more scrutiny as it poses as a sign that a multicollinear issue exists.

5. Testing the model

Having chosen the most suitable method to conduct this research, the author will proceed with statistical testing. For a multiple regression model, a probability value would be used to test the significance. An alpha of 10% would be the limit to determine whether the regression could be indicated as significant. Further testing to be applied to this model are the F test and the T test in order to determine the determinant coefficient of the variables.

a. Determinant Coefficient Test

A determinant coefficient test is conducted essentially to analyse the impact of the independent variable towards the dependent variable. It can also be viewed as the proportion of the variance in the dependent variable that is predictable from the variance in the independent variable. The coefficient of determination itself is the square of the correlation (r) between the independent and dependent variable, and so it ranges from 0 to 1. A value closer to 1 will indicate a stronger correlation between the variables.

b. F test

The F test is conducted to compare two variances. In this case, between the independent and dependent variable. It is based on Snedecor F-distribution under null hypothesis, and is performed to confirm that a data conforms to a regression model. With a 90% confidence level, the null hypothesis will be rejected when the P-value is less than 10%.

c. T test

Similar to the F test, the T test is conducted to test a single variable against another. It is mainly used when standard deviation is not known. With a similar 90% confidence level being used, the null hypothesis will be rejected if the P-value is less than 10%

6. Conduct hypothesis test

The finalizing steps to complete this research is to analyse the results obtained from the previous tests, and compare it against the hypothesis mentioned in the early stages. It is essentially to find out whether the hypothesis is accepted or rejected.

7. Analyse and conclude each hypothesis

The final result of this research paper will then be further analysed and concluded by the author. Justifications and disclosures regarding the results will also be included, as well as supporting evidence from previous studies.