CHAPTER II

THEORITICAL BACKGROUND

II.I Dividend Policy And Its Determinants

Dividend policy is a policy related to decision made by company whether to pay dividend or not, the amount to be distributed and the form of distribution (Gitmann, 2012). This policy is one of the most carefully monitored decisions by stockholder because it influences their wealth. The way it affects stockholders' wealth is when the dividend is distributed or when the stock price rises as the policy information enters stock market.

Dividend policy could be used by investor as a consideration of investment decision as it has informational content about company's future performance. As asymmetric information between company and investor exists, dividend can signal manager's superior information about company performance (Goddard, McMillan and Wilson, 2006). When a company decides to cut (maintain or raise) its dividend, it conveys information about its difficulty (good prospect) in the future. Therefore, information about dividend could reduce investor uncertainty about future. This certainty drives investor to pay higher price for the stock of dividend-paying-company than not-paying-dividend-company, thus maximizing the value of their investment.

Dividend policy would affect company's financing as well as investing decisions (Van Horne, 2008). Decision to pay out dividend could not be separated

from its financing needs. Therefore, there are many factors affecting decision encountered by company to pay out dividend to its shareholders. Some factors that might be the determinant of dividend policy are liquidity, contractual constraints, growth opportunities, future capital needed, investment, firm size as well as earnings. Frankfurter et al (2003) find out that only earnings, free cash flows, beta, and firm size that significantly influence dividend policy. In addition, leverage or debt constraints negatively significantly affect dividend policy (Brunzell et al, 2014)

II.II Multiple Linear Regression

Multiple linear regression is a commonly used method in forecasting or prediction of financial information data. This method is used to model the relationship of two or more independent variables and dependent variable based on observations or data used. In doing prediction, linear regression is used to find fit between dependent variables and independent variables. The strength of relationship among variables could be quantified through this method. The degree of strength is indicated by the coefficient $\boldsymbol{\alpha 2 \ldots \alpha n}$ in the following equation.

$$y_t = \alpha_1 + \alpha_2 x1_t + \alpha_3 x2_t + \ldots + \alpha_n xn_t + e$$

Common method in estimating the coefficient is using ordinary least square (OLS). This method aims at minimizing the residual sum of squared. To achieve this objective, some assumptions have to be fulfilled. Meeting these assumptions is important if one would to generate best coefficient to identify the

strength of relationship as proposed by Gauss and Markov. These assumptions is also known as BLUE (best linear unbiased estimator).

Gujarati (2008) summarizes the assumptions to be met so the estimator is BLUE. The criteria are first, it has to be a linear function. In other words, the estimator satisfies the following equation.

$$y = \beta 0 + \beta 1x1 + \beta 2x2 + ... + \beta k\, xk + u$$

Second, ***x*** value should be fixed and uncorrelated to the error term. Third, the error term ***u*** from the above equation should have and expected value of zero or the mean value of the disturbance ***u*** is zero.

$$E(u|x1, x2, ..., xk\ ) = 0\ ;\ E(u) = 0$$

Fourth, data has to fulfill homoskedasticity assumption. In other words, whatever the value of ***x,*** the variance of error term has to be 0. Fifth, the correlation of disturbance of independent variable should be zero. Under this assumptions, the estimator of OLS are said "BLUE". Other important assumption to consider regarding the OLS is sixth, there is no exact relationship among independent variable. In other words, no multicollinearity exists.

In order these criteria to be met, some violations should be checked. If potential violation exists, then treatment should be done. Otherwise, the unchecked and untreated violation will cause the coefficient to be unreliable or the MLR will provide wrong conclusion about the relationship among the variables. Some problem that may cause the relationship to be "unBLUE" is multicollinearity, heteroskedasticity and autocorrelation. If all the problem have been treated, then the goodness of fit (adjusted R-squared of the model) is reliable.

II.II.I Multicollinearity

Multicollinearity existence would not affect the goodness of fit of a model. Nevertheless, it would affect the strength of the estimator or coefficient. Multicollinearity exists when two or more independent variable high correlated. For example, from the following equation,

$$y_t = \alpha_1 + \alpha_2 x1_t + \alpha_3 x2_t + \ldots + \alpha_n xn_t + e$$

multicollinearity exists if

$$x2 = 2x1_t$$

If there is relationship between $x2$ and $x1$, then the coefficient α_2 and α_3 will change significantly if the data of the model change. Therefore it is important to find out the correlation among independent variable (x) to determine the existence of multicollinearity. The data is correlated if the correlation is greater than (lesser) 0.8 (-0.8). Should multicollinearity exists, then one of the highly correlated independent variable should be dropped from the equation or model.

II.II.II Heteroskedasticity

Heteroskedasticity refers to different variabilities of the population or data sample used. To quantify variability, term variance is used. If heteroskedasticity exists, the coefficient is still efficient. Nevertheless, the significance of statistical test could not be validated. Heteroskedasticity would affect error terms used to produce t-statistic used to draw conclusion about the relationship of effect of one independent variable toward dependent variable.

II.II.III Autocorrelation

Autocorrelation refers to the correlation of current data with previous data. Its existence would violate because of correlation of data with itself. To meet Gauss-Markov assumptions, the following equation should hold

$$\text{cov}(u_i, u_j | x_i, x_j) = E(u_i u_j) = 0 \qquad i \neq j$$

However, if autocorrelation exists then the equation becomes

$$E(u_i u_j) \neq 0 \qquad i \neq j$$

The existence of autocorrelation would not affect the relationship or affect independent variable toward dependent variable. However, its presence will cause t and F statistic no longer valid. If it is invalid anymore, then the conclusion drawn from the statistical test will be wrong.

II.III Artificial Neural Network

Artificial Neural Network is a term that comes from the biological neuron model, used as an information processing system. This network contains several interconnected neurons, as the biological neurons, and it produces output pattern based on the input pattern given (Sivanandam& Paulraj,2009). Biological neurons that resides in human brain consists of mililions of neurons that communicate one another through electrochemicals signals.

There are several elements of an artificial neuron that models the biological neuron , these include: 1)inputs that models the synapses, 2)weights,

that models the strength of the signals, and 3)activation function that decides whether the neuron is to be activated or not, and output. Figure 2.1 describes the artificial neurons that modeled the biological neuron based on the work of McCulloch and Pitts (1943).

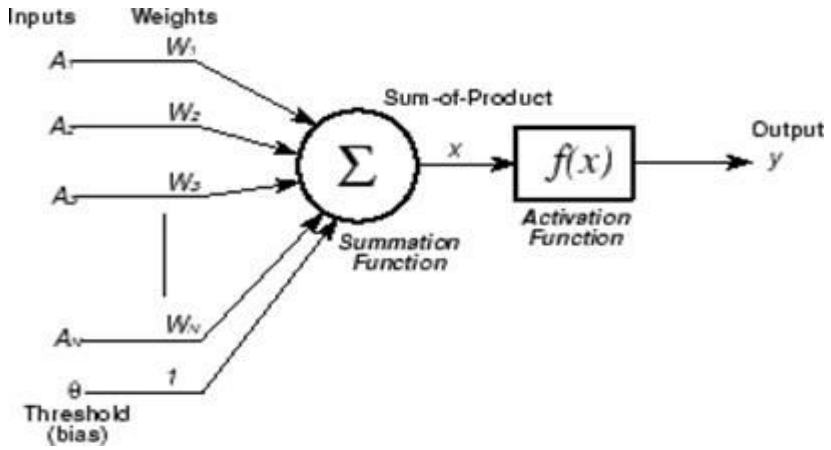

Figure 2.1 Artificial Neuron

Every inputs are associated with their weights, where all inputs will be multiplied with their weights . Summation Function will calculate all theses results using the following equation :

$$V_k = \sum_{j=1}^{k} w_j \, . \, x_j .$$

A single neuron as shown in figure 2.2 may not be useful unless it is connected with other neurons to form a network. This network describes the relationships or connections among neurons and shows how the neurons or nodes are being arranged. Multilayer neural network is one of the common neural network topology that consists of an input layer, hidden layer and an output layer, with the amount of hidden layer may be more than one. The input layer acts as the

sensor organ of a neural network, where the parameters of the environment is being set. The parameters describes the information that the neural network will use to produce output. The hidden layer in the multilayer network acts as the cognitive brain of the network, while the output layer consists of the neurons that represents the solution of the problem (Aamodt,2010). Multilayer is a feedforward network, in which the information in this network moves only in one direction and does not form a directed cycle. The information moves from the input layer to the hidden layer and to the output layer.

II.III.I Training in Neural Network

The process of learning in Artificial Neural Network is done through a training. As soon as the network has been structured, it is ready for training. During the process of training, the weights of a neuron are being adjusted, in order for a set of inputs to produces the desired outputs. (Slowik&Bialco, 2008).

Training of a network can be done either using a supervised training method, unsupervised training or reinforced method. Supervised training is a process of learning based on the comparison of the actual output with the targeted output. While unsupervised training is the process of learning merely based on the correlations of the input data without the present of the targeted or desired output. Reinforced method is a supervised learning, where the supervisor is present in the learning process,without the targeted output . This method will indicate whether the computed output is correct or incorrect. There are various learning algorithms that can be used to train the network, but the most widely

used in supervised training, is the backpropagation algorithm(Slowik& Bialco,2008).

The process of learning or training can be terminate by evaluating its Mean Square Error(MSE) value. Usually, this value is used as the primary termination procedure for the training because it describes the performance function of the training. It is the goal of the training to achieved the smallest value of error. On each iteration, the network will calculate the MSE in that pass, and it will be compared with mse_{min} . If $mse \leq mse_{min}$ then the training will terminate. In order to achieve a minimum value of error, then the network do an adjustment on the weights. This is the reason why, on every iteration, the network should adjust its weights.

The MSE can be expressed with the following equation :

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(t_i - \alpha_i)^2$$

Backpropagation algorithm is one of the most widely used training algorithm for ANN because it produces a balance ability in memory and generalization. Eventhough backpropagation algorithm has the power of generalization, but it has some limitations, and one of the greatest limitation of backpropagation is the local minima. Backpropagation algorithm is a gradient method that gives no warranty in achieving the global minimum of error, which is the smallest error value. This algorithm may get stuck in a local minima, which is the other minimum value of error (but not the smallest value) (Yuceturk,et al, 1999).

Several studies shows that in order to solve the backpropagation problem of local minima is by using Genetics Algorithms.

II.IV Genetic Algorithm

Genetics Algorithm (GA) is an evolutionary algorithm that uses iterative method to search the optimal solution. This algorithm is inspired by the biological evolution that have the nature to select the best, and discard the rest. The basic process of GA is shown in figure 2.2 .

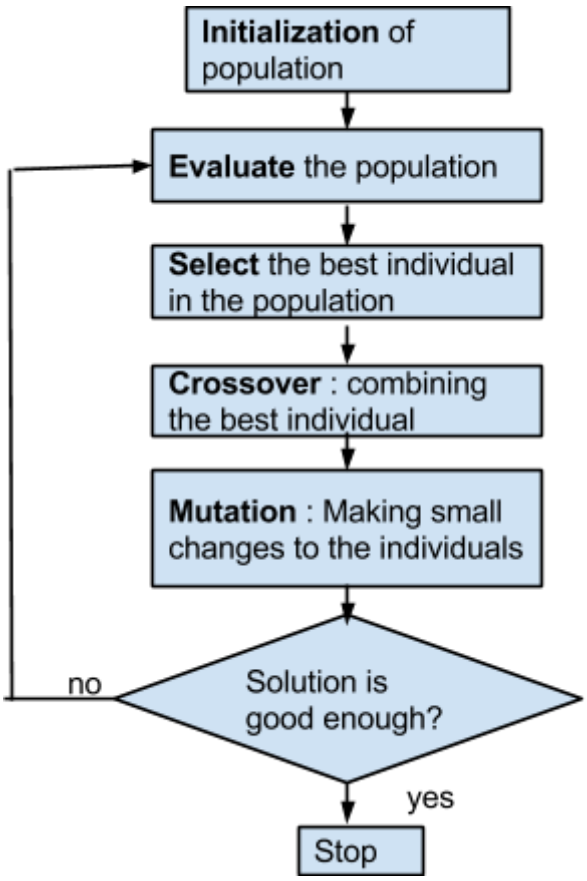

Figure 2.2 The basic process of Genetic Algorithm

The details of the GA process will be discussed below.

1. Initialization of population

The first process in GA is the initialization process. The initialization process includes creating the population of solution. This population should be represented in a form that can be prosessed by the computer. The common way to represent the population is by using binary strings, or array of integers or decimal numbers. The position of each strings or numbers in the population represent the aspect of the solution. In order to create members of the population, a random technique may be used.

2. Evaluate the population

Each member of the population will then be evaluated for their fitness in the desired requirements. The goal of the evaluation is to determine which member of the population fits well for the solution. In order to evaluate the population, a fitness function is used to perform the evaluation. Fitness function will return a fitness value that depicts how well an individual or member in the population fits the solution. The potential member will then proceed to the next generation, while the others will be deleted.

3. Select the best individual in the population

The potential member of population that have the highest fitness value will then be selected. The selection process is meant to improve the overall fitness of the population, by discarding all the suboptimal members of the population, therefore

leaving only the optimal members. There are several selection methods that can be used in selecting the best individual in the population,and one of the popular methods used in the selection method is the tournament selection (TOS). According to Jebari&Madiafi (2013), there is no common guidelines or theoritical assistance on how to choose the best selection method.

TOS is a method that selects the best individuals using ranking method. At first, this method will randomly select *k* set of individuals. These individuals will be ranked according to their fitness. Selection is made by choosing the most fitted individuals. This process is repeated *n* times for the whole population.

4. Crossover

Crossover is a recombination process aims to create new individuals, by combining a pair of the selected individuals from the previous step. The purpose of this process is to create a better individual based on the assumption that if two optimal individuals are combined will result in an even better offspring.

5. Mutation

Mutation is a process where a very small change occur on the individuals code. After the crossover process, a single point of the individuals code will be mutated randomly.The mutation is depicted using a rate. Mutation rate could be 1%, or 0.1% or any rate described in the GA algorithm. 1% mutation rate means the probability for every individual code to be mutated is 1%. Mutation process is an optional process. Some problems do not have mutation process.

GA will terminate if the algorithm have found the best solution which meets the criteria.

II.V Related Research

II.V.I Statistical Method

Several research in financial management sector related to dividend have used different statistical method to reveal the determinant of dividend policy. Kapoor, Anil and Mishra (2010) conduct a research about dividend policy determinant of Indian Service Sector using factorial analysis and regression analysis. Having used these two method, they find out that only firm size, solvency and earnings as the determinant of dividend policy.

Another research to find out dividend's determinant is conducted by Frankfurter et al. (2003). They use Vector Autoregression Model (VAR) to analyze the significance of earnings, debt, investment, debt, free cash flows, firm size, beta, investment and industry classification. By using VAR, earnings, free cash flows, beta, and firm size found to be the determinant of dividend policy.

Maldjian and Khoury (2014) find that firm size, risk and dividend last year are positive estimator of dividend, while growth opportunity and company's profitability are negative predictor for dividend. This result is generated by employing multiple linear regression and conducted over Lebanese listed banks.

II.V.II Comparison of Multiple Linear Regression(MLR) and ANN on Financial Data

Most recent publication related to this research is done by Mohamad, Ibrahim and Massoud (2013). They used financial data taken from financial statement as well as external data to forecast one output variable (firm's net profit). In making prediction, they compared two method called Neural Network Method (NNM) and Multiple Regression Model (MRM). They use 23 variables as input to determine one output which is net profit of the company. The object of their study is 20 Egyptian construction companies listed on Egyptian Stock Exchange from 2000-2010. Having excluded incomplete data,t he result of their research found that both method could be used to forecast construction company's net income. However, MRM has higher accuracy in forecasting compared to NNM. Likewise, Altay and Satman (2005) compare forecasting performance of ANN and linear regression method in forecasting stock price of companies listed in Istambul Stock Exchange. They found out that linear regression method outperformed ANN in forecasting accuracy the chosen sample of their research.

However, other research found the opposite. Abdou, et al (2012) found that Generalized Regression Neural Network is better in minimizing error of prediction of capital structure of 100 UK retail companies. Sample used in their research covers period from 2000-2006. In addition, Longinis and Symeonidis (2013) recommended to use NN (backpropagation algorithm) as method of predicting dividend payment as their research found NN is more accurate than linear regression in making prediction. Another study by Saiful (2011) , shows

that ANN outperform MLR with 94.62% accuracy in predicting depositor return of Islamic Bank in Indonesia.