Risk analysis of stock investment using the value at risk methods with the Bayesian normal mixture approach

**S Inayati1, Sahid2 and R Kusumawati3**

1, 2, 3Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

E-mail: syarifah.inayati@uny.ac.id

**Abstract.** In a stock investment, the greater the desired profits, the greater risks will be implied. The big changes in the stock market prices encourage us to measure the financial risks. The value at risk (VaR), a parametric method under the assumption of normally distributed data, is one of the most popular and accurate risk measurement methods. If the stock data does not match a normal distribution, then the normal mixture distribution can be implemented. In this study, we calculate the risk data of the shares of three companies registered in the Jakarta Islamic Index (JII) using the VaR methods through the normal mixture approach. Those companies are PT. Astra International Tbk (ASII), PT. Telekomunikasi Indonesia Tbk (TLKM) and PT. Unilever Indonesia Tbk (UNVR). The data were taken in 2019. The parameter estimation was done using the Bayesian Markov Chain Monte Carlo (MCMC) approach. Based on the obtained VaR values, the highest risk is 0.124272 for TLKM, then 0.02533735 for ASII, and the lowest one is 0.02298288 for UNVR.

1. Introduction

Nowadays, one kind of financial investment being developed by the Indonesian capital market is Islamic stocks, which are based on sharia/islamic principles, like the one on Jakarta Islamic Index (JII) [1]. Stock investment is such a fascinating thing in the business sector because investing will get a lot of returns. However, it is at risk of loss indicated by the uncertainty of the return that will be received. The higher the return desired, the greater the risk. Before purchasing shares, an investor needs to investigate the possible risks, therefore. In addition, as stock exchange prices largely change gradually, it is necessary to measure financial risk.

 A popular and widely used technique in measuring risks is value at risk (VaR) [2], [3], a parametric method based on the assumption of a normal distribution. In reality, many financial data do not follow a normal distribution, so normal mixture models can be implemented into the data to solve it [4].

 The mixture model is a special model suitable for multimodal data. This robust and highly flexible approach has been used in a variety of ways, in both finite and infinite mixture frameworks. It can be used to facilitate complicated identification of multimodal data distribution [5], [6], [7]. When the constituent components of the mixture model are normal distribution, it is called normal mixture model.

 This study aims to analyze the risks of investing in Islamic stocks in three companies with the largest market capitalization registered in JII .i.e PT Telekomunikasi Indonesia Tbk (TLKM), PT Unilever Indonesia Tbk (UNVR), and PT Astra International Tbk (ASII), through the normal mixture model approach. Parameter estimation is performed using the Bayesian Markov Chain Monte Carlo (MCMC) approach.

1. Literature
	1. Normal mixture model

The mixture model is used in multimodal data, the data originating from sub-populations or groups which are constituent components of the mixture model with different proportions. It is called a special model because of its ability to combine data while maintaining the characteristics of the original data [8], [9], [10]. It is, furthermore, able to combine a number of components originating from different distributions so that it can provide a more realistic picture of the characteristics of the original data.

 The mixture density function of an observational data taken from the number of *k* subpopulations, which is referred to as the finite mixture model, is defined as follows [5], [8]:

where is the mixture density function of data with the model parameter vector and the weighting vector **,**  the density function to-, j = 1,2,…, k with parameters which is a vector parameter whose characteristics depend on the distribution form of each component of the mixture model, and is the proportion (weighting) parameter vector of the mixture model with elements where and .

 If it is known that the *k* sub-populations of each sub-population have a normal distribution, then equation (1) is the normal mixture distribution, which can be formulated as follows:

where is a normal mixture density function of data with weighting vector **,** averageparameter and standard deviation parameter . Parameter as a weight has elements and is normal density function to- of the number of of constituent component of mixture model mixture with parameter .

* 1. Bayesian mixture model

The Bayesian model is developed from the Bayes method and has high flexibility in various application problems [11], [12]. Bayesian analysis on the mixture model views that all the parameters in the model are random variables that have a certain prior distribution, so this analysis requires a prior distribution specification for each parameter in the model. This forms a hierarchical arrangement that forms the model including the number of *k* components in the model,, weighing proportion where and , and the specific parameters of components .

 According to Richardson and Green [13], the preparation of the mixture model is conducted with the view that every observation will be a member of one of the unknown sub-populations. If the allocation of each observation in each sub-population of the mixture models in equation (1) is denoted by, then the allocation proportion of each observation is determined by the following distribution:

 According to equation (3), if value is given, then according to equation (1) observation data comes from the sub-population distributing as follows:

Thus, the combined posterior distribution produced by the mixture models will be in the form of a mixed distribution of all the variables in the following mixture model:

where is the combined posterior distribution of mixture model with component *k* from data with model parameter , weighting and allocation parameter **,**  is the prior distribution of the parameter of the number of mixture components as much as , is the prior distribution of the weighting parameter , on one condition, *k* has a certain value, is the prior distribution of the allocation parameter , on one condition, and *k* have a certain value, is the prior distribution of the model parameters , on one condition, and have certain values, and is the likelihood function of the data x knowinglyand have certain values.

 If equation (5) is applied to the normal mixture distribution with model parameters in the form of  **μ**, then the combined posterior distribution will be obtained for all parameters in the normal mixture model as follows:

* 1. Estimating parameters

The parameter is estimated for equation (6) to form a full conditional distribution for each parameter in the model. To form a fully conditional distribution, it is necessary to know the combined likelihood function and prior distribution of each model parameter that will be estimated. According to [8] the likelihood function for normal mixture models (2) can be formulated as:

where and is the number of components of the mixture model.

 Prior distribution in normal mixture modeling given as follow [13]:

where prior to is where is midpoint of and is a precision number whose value is equal to , where is range of data Meanwhile, prior to is with parameter , where constant has range and 𝛽 is hyperparameter for that has distribution of gamma with scale parameter that can aslo be given value and constant parameter . Furthermore is parameter of distribution of Dirichlet which is constant value. Every observation data is taken independently from unknown subpopulation with and for . Then, is determined sequentially as follows: where . Meantime, weighing parameter of every mixture component is determined as .

 The full conditional distribution can be obtained from the combined distribution of all parameters in the model provided that the other parameters are considered constant. The full conditional distribution of the normal mixture model can be determined by manipulating the combined distribution form of all parameters in the normal mixture model, as presented in equation (6). The full conditional distribution for each parameter in the model can be seen in [14], [15].

 Markov Chain Monte Carlo (MCMC) can be implemented to obtain the posterior distribution of Bayesian through a simulation method, which is a combination of Monte Carlo and Markov Chain properties to obtain sample data based on certain sampling scenarios [16]. Estimation of normal mixture model parameters can be done using a numerical MCMC approach, especially the Gibbs Sampler [17]. In implementing it into real problems, to facilitate the estimation process of the normal mixture modeling parameters, we use the WinBUGS 1.4 program.

* 1. Distribution Multiplication Structure (DMS)

Carlin and Chib [18] have developed the way how to select the model by generating it through the bayesian MCMC at the posterior likelihood conditional probability to estimate the value of the bayes factor. Distribution Multiplication Structure (DMS) is a method of selecting the best model formed through the combined distribution of several models using the principle of multiplication and not paying attention to the assumption of normality in the residuals.

 For example there are two different mixture models, for 2 components and for 3-components which are independent, which will be compared to obtain the correct model. Also, parameter λis given as an indicator of the new density model dominance . The new density formed from the two densities is formulated as follows:

* 1. Bayes Factor

Hypothesis testing to select the best model that fits the data used the Bayesian method, where the solution used the Bayes Factor, a summary of all data evidence to make it more suitable for one distribution or model from a number of distributions or models compared to DMS.

 The Bayes factor value can be obtained from MDS using the MCMC method. The calculation method is given in the following equation:

1. For the two constituent distributions in DMS
2. For more than two constituent distributions in DMS

where 𝜆𝑗 (𝑔) is the number of distributions or models to − where which can be generated from as much as 𝑁 iteration of MCMC for MDS where the constituent distribution.

 Kass and Raftery [19] describe Bayes working principle as follows. For example, there are two models that are suitable for data .i.e. dan , then the most suitable model for data is determined by testing the two models using the hypothesis:

(Model one in accordance with data )

(Model two in accoradance with data )

 A category for simplifying interpretations that can be made from the results of the Bayes factor calculation is stated in table 1.

**Table 1.** Interpretation of Bayes factor values to select hypotheses

|  |  |  |
| --- | --- | --- |
|  |  | Reliability proof of and  |
|  |  | Negative |
|  |  | Zero |
|  |  | Positive |
|  |  | Firm |
|  |  | Very firm |

* 1. Return

Return of an asset is the change in price from the initial price, and it is one of the factors motivating investors to invest [20]. The formula used to determine return [21]:

where is the stock price return on day t, is the price on day t, and is the stock price on day t .

* 1. Value at Risk (VaR) through normal mixture approach

Value at Risk (VaR) is a concept to measure risk in risk management. It is used to measure the estimated minimum loss that may occur in a certain period of time under normal market conditions with a certain confidence level

 Bodie, Kane and Marcus [22] state that VaR is another name for the quantile of a distribution. is expressed as the quantile form of the return (R) gain or loss distribution, where is the density function of and is the cumulative distribution function. If return follows a normal function, with mean 𝜇 and standard deviation 𝜎, then the function can be written in the following equation:

Simply, VaR of at the level of trust can be written:

where . Therefore, equation (19) can also be written as follows:

where average return, quantile- and return distribution, and standard return deviation.

Thus, VaR with Mixture approach can be calculated through the following equation.

where is the proportion of components to-i, is average return of component to-i, is the quantile of return distribution and is holding of investment period [23].

1. Result
	1. Normal mixture modeling on Islamic stock investments

In this study, the risk analysis of stock investment was conducted based on sharia through Value at Risk (VaR) method with a normal mixture approach. The data used were daily close price data of Islamic stocks from PT Telekomunikasi Indonesia Tbk (TLKM), PT Unilever Indonesia Tbk (UNVR), and PT Astra International Tbk (ASII). The data were the ones in 2019 which can be accessed at [www.yahoo.finance.com](http://www.yahoo.finance.com).

The characteristics of stock returns can be viewed through descriptive analysis in table 1. Meanwhile, q-q plot of stock returns observed can be seen in figure 1.Based on the characteristics of stock return data and the normality test on each stock data, it has been found that TLKM, UNVR, and ASII stock return data do not normally distribute nor is feasible to predict univariate multimodal distribution, so that the modeling can be approached through the normal mixture model.

**Tabel 1.** Stock return statistics Description of TLKM, UNVR, and ASII

|  |  |  |  |
| --- | --- | --- | --- |
| Asset | Descriptive statistics |  |  |
| N | Mean | ST Dev | Skewness | Kurtosis |
| TLKM | 250 | 0.0003338138 | 0.01457042 | -0.03046413 | 2.888669 |
| UNVR | 250 | -0.0002122131 | 0.01410451 | 0.01889289 | 3.978136 |
| ASII  | 250 | -0.0005550948 | 0.01632887 | 0.05896569 | 3.769645 |

|  |  |  |
| --- | --- | --- |
| TLKM | UNVR | ASII |
|  |  |  |
| **Figure 1.** Normal q-q plot of stock return data |

The shape of the mixture distribution can be identified through the histogram data, and the return histogram for the three observed stocks is presented in figure 2. Based on figure 2, the histogram of return data for the three stocks shows a tendency that has different variability. In this study, many components of the normal mixture distribution are determined as 2 and 3 components. Based on the analysis results, the normal histogram of the 2 and 3 components of each stock data is given in Figure 3.In this study the determination of 2 and 3 components used the clustering technique with the R program through the caret function (), then selected so that the data were obtained in each of the components normally distributed. The normal histogram of stock return of the 2-3 constituent components is shown in Figure 3.

|  |  |  |
| --- | --- | --- |
|  |  |  |

**Figure 2**. Stock return histogram

|  |  |
| --- | --- |
|   |  |
|  |  |
|  | Histogram of Komponen 1, Komponen 2, Komponen 3 |

Figure 3. Normal histogram of stock returns of the 2 and 3 constituent components

* 1. Selection of the best model and parameter estimation

Distribution Multiplication Structure (DMS) was used to select the best normal mixture model among several components of the model constituent by adding 10000 samples. λ is the weight of each component of the normal mixture model, and DMS function is formed according to equation (14). After DMS function had been obtained, the best model was selected using the Bayes factor value in equation (16). These steps were conducted by compiling a program through WinBUGS 1.4 software. The results obtained for each stock data are as follows:

1. The DMS function formed for TLKM shares is stated in the following equation:

where and

1. The DMS function formed for UNVR shares is stated in the following equation:

where

 and

1. The DMS function formed for ASII shares is expressed in the following equation:

where and

 The selection of the best model using the Bayes Factor value for DMS return for the three stocks shows the following results:

**Tabel 2.** Bayes factor calculation results for DMS stock returns

|  |  |  |
| --- | --- | --- |
| STOCKS | Component number | Bayes Factor Value |
| 2 Components | 3 Components |
| TLKM | 2 | 1 | 0.6553 |
| TLKM | 3 | 1.5258 | 1 |
| UNVR | 2 | 1 | 0.658 |
| UNVR | 3 | 1.5195 | 1 |
| ASII | 2 | 1 | 0.65755 |
| ASII | 3 | 1.52079 | 1 |

 Based on the Bayes factor value in table 2, all in TLKM, UNVR, and ASII stocks, it can be viewed that the normal mixture model with 3 constituent components to the normal mixture model with 2 constituent components is in the interval 1 to 3 meaning that no model dominates. Meanwhile, based on the program output on the three stocks, it has been found that the Bayes factor value of 2 constituent components to the 3 constituent components is negative because the Bayes factor value <1, meaning that the model with 3 constituent components is better than that with 2 constituent components. The best model used, therefore, is the one with 3 constituent components

 Based on parameter estimation and analysis of the stock normal return mixture model, the following results are obtained:

1. The estimation result of the density function of the Mixture model of TLKM stock return is

 where

and = = =

1. The estimation result of the density function of Mixture model of UNVR stock return is

where

and =

1. The estimation result of the density function of Mixture model of ASII stock return is

where

and =,=

 The mixture model above can also interpret the magnitude of the proportion variation in each group of the stock return data constituents.

* 1. Value at Risk calculation based on the best mixture model

To calculate the amount of risks an investor runs when investing in ASII, TLKM and UNVR, the Value at Risk (VaR) method is used in equation (21). Within one day and with a confidence level of 95%, based on the VaR value obtained, the highest risk occurs in TLKM by 0.124272, followed by ASII by 0.02533735, and UNVR by 0.02298288, the lowest.

1. Conclusion

Risk calculation using VaR through normal mixture approach can be applied to analyze the risks of investing in Shari'ah/ Islamic stocks in PT. Astra International Tbk (ASII), PT. Telekomunikasi Indonesia Tbk (TLKM) and PT. Unilever Indonesia Tbk (UNVR). This parametric method can be used to solve VaR calculations for abnormal data in financial risk analysis. Based on the VaR value obtained, the highest risk occurs in TLKM stock, which is 0.124272, followed by ASII by 0.02533735, and UNVR by 0.02298288, the lowest. To sum up, the greater the obtained VaR value, the greater the risks by investors.

References

[1] Huda N and Nasution M E 2007 *Investasi pada Pasar Modal Syariah* (Jakarta: Kencana)

[2] Jorion P 2007 *Value at Risk: The New Benchmark for Managing Financial Risk, third edition*  (New York: McGraw-Hill)

[3] Liao G Pan T-H, Chang L-F, Chang L-F, Huang S-C and Wu C-F 2012 *J. Stat. Manag. Syst.* **15** 345

[4] Kocak K, Calis N and Unal D 2013 *J. Business, Econ. Financ.* **2** 13

[5] Malsiner-Walli G, Frühwirth-Schnatter S and Grün B 2017 *J. Comput. Graph. Stat.* **26** 285

[6] Miftahurrohmah B, Iriawan N and Fithriasari K 2017 *J. Phys.: Conf. Ser.*  **855** 012026

[7] Putri U M 2016 *Analisis Risiko Investasi Saham Syariah Menggunakan Value at Risk dengan Pendekatan Bayesian Mixture Normal Autoregressive* (Surabaya: Institut Teknologi Sepuluh Nopember).

[8] McLachlan G J and Basford K E 1988 *Mixture Models Inference and Applications to Clustering* (New York: Marcel Dekker)

[9] Escobar M D and West M 1995 *J. Am. Stat. Assoc.* **90** 577

[10] Gelman A, Carlin J B, Stern H S and Rubin D B 2004 *Bayesian Data Analysis* (London: Chapman & Hall/CRC)

[11] Dunson D B 2001 *Am J Epidemiol* **153** 1222

[12] Fuadi M Z, Kusumawati R, Inayati S and Sahid 2020 *J. Phys.: Conf. Ser.*  **1581** 012016

[13] Richardson S and Green P J 1997 *J. R. Stat. Soc.* **59** 731

[14] Astuti A B 2017 *Bayesian Mixture Model Averaging untuk Mengidentifikasi Perbedaan Ekspresi Gen Percobaan Microarray* (Surabaya: Institut Teknologi Sepuluh Nopember)

[15] Gilks W R and Richardson S 1995 *Markov Chain Monte Carlo in Practise* (London: Chapman & Hall/CRC) pp 75–88

[16] Greyserman A, Jones D H and Strawderman W E 2006 *J. Bank. Financ.* **30** 669

[17] Casella G and George E I 1992 *Am. Stat.* **46** 167

[18] Carlin B P and Chib S 1995 *J. R. Stat. Soc. Ser. B* **57** 473

[19] Kass R E and Raftery A E 1995 *J. Am. Stat. Assoc.* **90** 773

[20] Ruppert D 2004 *Statistics and Finance: An Introduction* (New York: Springer)

[21] Bogdan S, Baresa S and Ivanovic Z 2015 *UTMS J. Econ.* **6** 165

[22] Bodie Z, Kane A and Marcus A J 2004 *Essentials of Investements* (New York: McGraw-Hill)

[23] Valecký J 2012 *Manag. Model. Financ. risks* **6** 653