La GUBU, PhD Student
E-mail: lagubu2014@gmail.com
Department of Mathematics
Gadja Mada University, Yogyakarta Indonesia
Halu Oleo University, Kendari Indonesia
Professor Dedi ROSADI, Dr.rer.nat.
E-mail: dedirosadi@gadjahmada.edu
Department of Mathematics
Gadja Mada University, Yogyakarta Indonesia
Associate Professor ABDURAKHMAN, Dr. M.Si.
E-mail: rachmanstat@ugm.ac.id
Department of Mathematics
Gadja Mada University, Yogyakarta Indonesia

ROBUST MEAN-VARIANCE PORTFOLIO SELECTION WITH WARD AND COMPLETE LINKAGE CLUSTERING ALGORITHM

Abstract. In this paper, we present a robust mean-variance portfolio selection method with preprocessing data using cluster analysis. Using this proposed method, we obtain the best portfolio (i.e. portfolio with the highest Sharpe ratio) efficiently when there is a large number of stocks involved in the formulation of the portfolio. On the other hand, this procedure is also robust against the possibility of outliers existence in the data. Based on our empirical study, we find that the performance of portfolio produced using clustering with Ward algorithm is better than portfolio performance produced by the clustering with complete linkage algorithm for all risk aversion values γ. Besides, we also find that portfolio performance with robust FMCD estimation is better than portfolio performance with robust S estimation and classic MV portfolio for all risk aversion values γ, for both portfolios produced by cluster analysis with Ward algorithm and complete linkage algorithm.

Keywords: cluster analysis, Ward, complete linkage, Sharpe ratio, robust portfolio.

JEL Classification: C61, G11

1. Introduction

In the recent years, there are strong interest from practitioners who want time and cost efficiency in doing portfolio management. This aim can be reached by e.g. using clustering analysis as a first step in selecting securities that will form an optimum portfolio. The optimal portfolio selection using cluster analysis has been carried out by e.g., Guan and Jiang (2007), Tolaet. al. (2008), Chen and Huang (2009), Nanda et al. (2010) and Long et.al. (2014).The difference between these studies lies in the method of clustering analysis used and the method of selecting securities in the optimal portfolio. In all of the above studies, the optimal portfolio obtained by meaning of Mean-Variance (MV) Markowitz portfolio model. All of these studies state that the usage of cluster analysis in the formation of the optimum portfolio is very efficient when the number of securities involved in the portfolio formation is relatively large.

The main problem of the MV portfolio model is that the mean vectors and variance-covariance matrices must be estimated from data which can be highly volatile. Parameter estimation can be done with various choices of estimation techniques, which will inevitably contain estimation errors. As a very important input in the formation of MV portfolio models, estimation errors will significantly affect the results of optimal portfolio formation. Several studies related to estimation errors and their relationship with optimal portfolio formation have been carried out by Best and Grauer (1991), Broadie (1993), Chopra and Ziemba (1993), Bengtsson (2004), and Ceria and Stubbs (2006). Based on these studies conclude that although the MV model is supported by strong theory and has eased in computing, the MV model shows some weaknesses including the optimal portfolio produced by this model is not well diversified. The resulting portfolio tends to be concentrated in a small portion of assets. In addition, the MV model is also highly sensitive to changes in input parameters, which are the mean vector and the variance-covariance matrix.

Therefore some researchers have built a robust portfolio, a portfolio that can reduce the error of the estimated vector mean and the variance-covariance matrix in the MV portfolio model. One of the standard approaches in forming an optimal robust portfolio is through a robust estimation approach. Several studies on the formation of optimal portfolios using robust estimation have been carried out by Victoria-Feser (2000), Lauprete (2001), Vaz-de Melo and Camara (2003), Zhou (2006), Welsh and Zhou (2007), DeMiguel and Nogales (2008), Kusch (2012), Hu (2012), Supandi (2017).The difference between these studies lies in the use of robust estimates. All of the results of the study stated that the performance of the portfolio with robust estimation is better than the classical portfolio. However, none of the above literature has considered a combination of cluster analysis and robust estimation methods in the formation of an optimum portfolio.

In this paper, as our new contribution, we combine cluster analysis and robust estimation method in the formation of an optimum portfolio. We use the

clustering method with Ward and complete linkage algorithm in grouping of stocks
into clusters. For the mean and covariance of the data, we use Fast Minimum
Covariance Determinant (FMCD) and S robust estimation.

2. Material and Method

Based on the literature review, it can be seen that the portfolio selection
problem can be resolved more efficiently by grouping stocks into clusters and then
selecting stocks in clusters to form efficient portfolios. The problem framework in
this study is shown in Figure 1.
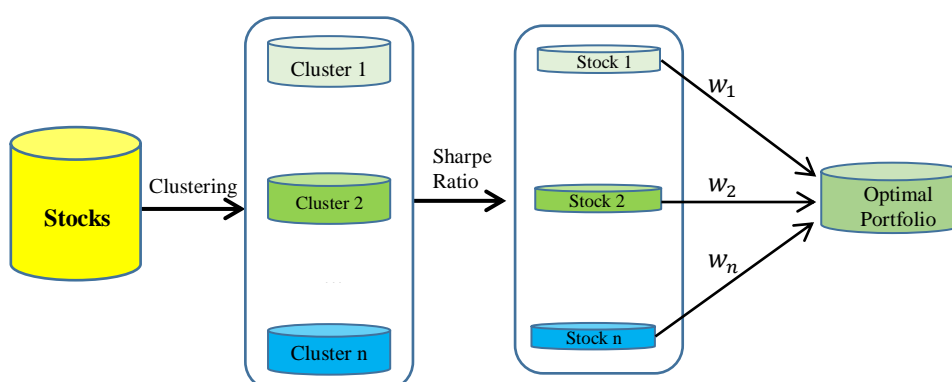
Figure 1. Optimal portfolio selection process

First, stocks are grouped into several clusters using the Ward algorithm and
complete linkage clustering which will be discussed in the next section.
Furthermore, the return and risk of historical stock data are calculated in each
cluster. From the calculation of return and risk, it can be determined the
performance of each stock in each cluster using the Sharpe ratio. The next step is to
choose stocks that will represent each cluster to form the optimum portfolio. The
stock chosen as representations of a cluster are stocks with the highest Sharpe ratio.
After the stocks that build the optimum portfolio are selected, the next step is to
determine the weight of each stock that builds the portfolio using robust FMCD
estimation method and robust S estimation method. To see the advantages of these
two methods, the performance of portfolios that formed then compared with
portfolio performance formed using the classic Mean-Variance (MV) method.

2.1 Mean-Variance portfolio

In investment management, there is a strong and linear relationship
between risk and return. If the risk is high, the return will also be high, whereas if
the return is low, the risk will also be low. Harry M. Markowitz developed a theory
in the 1950s called the Modern Portfolio Theory. This theory formulates the
existence of the elements of return and risk in an investment, where the risk

113

La Gubu, Dedi Rosadi, Abdurakhman

---

element can be minimized through diversification and combining various investment instruments into a portfolio. In 1952 the theory was widely published in the Journal of Finance.

Markowitz's portfolio theory is based on the mean and variance approach, where the mean is a measurement of the level of the expected return and variance is a measurement of the level of risk (Markowitz, 1952). Therefore, Markowitz's portfolio theory is also called the mean-variance model (MV). This model emphasizes efforts to maximize expected return and minimize risk to choose and build an optimal portfolio. According to Supandi (2017), the mean-variance portfolio can be formulated by solving the following optimization problems:

$$\max_{\boldsymbol{w}} \boldsymbol{w}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \tag{1}$$

$$\boldsymbol{w}'\boldsymbol{e} = 1 \tag{2}$$

Where $\boldsymbol{w}$ denotes the weight of the portfolio, $\boldsymbol{\mu}$ is the mean vector, $\boldsymbol{\Sigma}$ is covariance matrix, $\boldsymbol{e}$ is the column matrix where all the elements are 1 and $\gamma \geq 0$ are the risk aversion parameters, namely the relative measure of risk avoidance. The mean-variance portfolio model is chosen by investors using the utility function criteria. Assuming that investor attitudes are risk averse, that is investors will choose portfolios that provide the maximum expected utility E(U). So the mean-variance portfolio is equivalent to

$$\max E(U(\boldsymbol{w})) \tag{3}$$

$$\boldsymbol{w}'\boldsymbol{e} = 1 \tag{4}$$

Given initial capital W_0, under a portfolio with a weight vector $\boldsymbol{w}$ at the end of the period, the capital becomes $W_0(1 + R_p)$. Where R_p is a random variable of portfolio return. The utility function of $W_0(1 + R_p)$ is $U\big(W_0(1 + R_p)\big)$. The utility function can be expand using the Taylor second order approximation i.e.

$$U\big(W_0(1 + R_p)\big)$$
$$= U(W_0) + U'(W_0)W_0R_p + \frac{1}{2}U''(W_0)\big(W_0R_p\big)^2$$
$$+ \frac{1}{6}U'''(W_0)\big(W_0R_p\big)^3 + \cdots$$

$$= U(W_0) + W_0U'(W_0)R_p + \frac{1}{2}W_0^2U''(W_0)R_p^2 + O(R_p^3) \tag{5}$$

Taking the expected value of equation (5) we obtained

$$E\left(U\big(W_0(1 + R_p)\big)\right) = U(W_0) + W_0U'(W_0)\mu_p + \frac{1}{2}W_0^2U''(W_0)\sigma_p^2$$
$$= U(W_0) + W_0U'(W_0)\left(\mu_p + \frac{1}{2}W_0\frac{U''(W_0)}{U'(W_0)}\sigma_p^2\right) \tag{6}$$

Based on the approximation above, maximizing the expected utility function is equivalent to maximizing

$$\mu_p - \frac{1}{2}\gamma\sigma_p^2 \tag{7}$$

where $\gamma = -W_0 \frac{U''(W_0)}{U'(W_0)}$ express the relative size of risk aversion. It is known that $\mu_p = \boldsymbol{w}'\boldsymbol{\mu}$ and $\sigma_p^2 = \boldsymbol{w}'\Sigma\boldsymbol{w}$, so equation (7) can be rewritten in the form

$$\boldsymbol{w}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}'\Sigma\boldsymbol{w} \tag{8}$$

So equation (1) is equivalent to equation (8) plus the constraint $\boldsymbol{w}'\boldsymbol{e} = 1$.

Every investor wants a certain amount of risk to obtain a certain level of return. Because profits compensate with risks, investors should be balance the trade of between profits and risks by choosing the right γ.

There are two cases of extreme situations, where investors want to increase the rate of return (profit) and reduce risk (loss), that is when $\gamma = 0$, equation (1) provides the maximum rate of return without regard to the risk to be borne. Meanwhile, if $\gamma = \infty$, investors will choose the minimum risk regardless of the level of return.

The optimization problems in equations (1) and (2) can be solved using the following Lagrange method. Firstly, form the Lagrange function:

$$\boldsymbol{L} = \boldsymbol{w}'\boldsymbol{\mu} - \frac{1}{2}\gamma\boldsymbol{w}'\Sigma\boldsymbol{w} + \lambda(\boldsymbol{w}'\boldsymbol{e} - 1) \tag{9}$$

Based on the Kuhn-Tucker theorem, the necessary conditions for the optimal of equation (9) are:

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{w}} = 0 \tag{10}$$

$$\frac{\partial \boldsymbol{L}}{\partial \lambda} = 0 \tag{11}$$

From equations (9), (10), and (11) we obtain

$$\boldsymbol{w} = \frac{\Sigma^{-1}}{\gamma}(\boldsymbol{\mu} + \lambda\boldsymbol{e}) \tag{12}$$

and

$$\boldsymbol{e}'\boldsymbol{w} = 1 \tag{13}$$

Substitute equation (12) in to equation (13) we obtain

$$\lambda = \gamma(\boldsymbol{e}'\Sigma^{-1}\boldsymbol{e})^{-1} - (\boldsymbol{e}'\Sigma^{-1}\boldsymbol{e})^{-1}\boldsymbol{e}'\Sigma^{-1}\boldsymbol{\mu} \tag{14}$$

Substituting equation (14) in equation (12) produces:

$$\boldsymbol{w} = \frac{\Sigma^{-1}}{\gamma}\{\boldsymbol{\mu} + \boldsymbol{e}[\gamma(\boldsymbol{e}'\Sigma^{-1}\boldsymbol{e})^{-1} - (\boldsymbol{e}'\Sigma^{-1}\boldsymbol{e})^{-1}\boldsymbol{e}'\Sigma^{-1}\boldsymbol{\mu}]\}$$

$$= \frac{1}{\gamma}(\Sigma^{-1} - \Sigma^{-1}\boldsymbol{e}(\boldsymbol{e}'\Sigma^{-1}\boldsymbol{e})^{-1}\boldsymbol{e}'\Sigma^{-1})\boldsymbol{\mu} + \Sigma^{-1}\boldsymbol{e}(\boldsymbol{e}'\Sigma^{-1}\boldsymbol{e})^{-1} \tag{15}$$

115

La Gubu, Dedi Rosadi, Abdurakhman

Equation (15) shows that the optimal portfolio (w) depends on input $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and assumes multivariate normality.

The sufficient condition for optimal optimization is used to determine whether the optimal point w obtained from the necessary condition for optimal optimization is the minimum or maximum.

Suppose that function (9) is continuous and twice differentiable. The Hessian matrix $\boldsymbol{\mathcal{H}}$ for function (9) is

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \frac{\partial^2 L}{\partial w^2} & \frac{\partial^2 L}{\partial w \partial \lambda} \\ \frac{\partial^2 L}{\partial \lambda \partial w} & \frac{\partial^2 L}{\partial \lambda^2} \end{bmatrix} = \begin{bmatrix} -\gamma\boldsymbol{\Sigma} & \boldsymbol{e} \\ \boldsymbol{e}' & 0 \end{bmatrix} \tag{16}$$

Furthermore, based on the Hessian matrix (16) it can be determined whether the weight of the portfolio (15) is the minimum or maximum solution with the following criteria (Chong and Zak, 2001).
1. If $\boldsymbol{\mathcal{H}}$ is positive definite matrix, then w is the minimum solution;
2. If $\boldsymbol{\mathcal{H}}$ is negative definite matix, then w is the maximum solution.

According to Leon (2002), if a symmetric matrix A of size *pxp* is positive definite, it will have the following properties:
1. If A is a symmetric matrix positive definite, then $|A| > 0$.
2. If A is a symmetric matrix positive definite, then all major submatrices of $A_1, ... A_p$ are symmetric matrix positive definite.
3. If A is a symmetric matrix positive definite, then A can be reduced to upper triangular matrix using row elementary matrix operations and the pivot element will always be positive.

Otherwise, a symmetric matrix *A* of size *pxp* is negative definite, it will the following properties:
1. If A is a symmetric matrix negative definite, then $|A| < 0$.
2. If A is a symmetric matrix negative definite, then all major submatrices of $A_1, ... A_p$ are symmetric matrix negative definite.
3. If A is a symmetric matrix negative definite, then A can be reduced to upper triangular matrix using row elementary matrix operations and the pivot element will always be negative.

A square matrix $\boldsymbol{A}_{pxp}$ is called a negative matrix and is denoted by $\boldsymbol{A} \prec 0$ if A is symmetric and $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} < 0$ for every $\boldsymbol{x} \neq 0, \boldsymbol{x} \in \mathbb{R}^p$.

The covariance matrix of the variable $\boldsymbol{r}_i \in \mathbb{R}^p$, $i = 1, ..., N$ with the mean vector $\boldsymbol{\mu}$ is

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{r}_i - \boldsymbol{\mu})(\boldsymbol{r}_i - \boldsymbol{\mu})' \tag{17}$$

Suppose $\boldsymbol{A} = -\boldsymbol{\Sigma}$, for each vector $\boldsymbol{x} \in R^p$ where $\boldsymbol{x} \neq 0$, then

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = -\boldsymbol{x}'\left(\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{r}_i - \boldsymbol{\mu})(\boldsymbol{r}_i - \boldsymbol{\mu})'\right)\boldsymbol{x}$$

___

$$= -\frac{1}{N} \sum_{i=1}^{N} x'(r_i - \mu)(r_i - \mu)'x$$

$$= -\frac{1}{N} \sum_{i=1}^{N} ((r_i - \mu)x)'(r_i - \mu)'x$$

$$= -\frac{1}{N} \sum_{i=1}^{N} ((r_i - \mu)x)^2 < 0 \tag{18}$$

Therefore the matrix $A = -\Sigma$ is a matrix negative definite. Given the Hessian matrix $\mathcal{H}$ (16) and assumed investor is risk averse ($\gamma > 0$). Suppose the main submatric is $A_p = \gamma\Sigma$, i.e.

$$A_p = \begin{pmatrix} \sigma_{11}^* & \sigma_{12}^* & \cdots & \sigma_{1p}^* \\ \sigma_{21}^* & \sigma_{22}^* & \cdots & \sigma_{2p}^* \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_p^* & \sigma_{p2}^* & \cdots & \sigma_{pp}^* \end{pmatrix} \tag{19}$$

where $\sigma_{ij}^* = \gamma\sigma_{ij}$ for all $i, j = 1, \dots, p$.

The Hessian matrix $\mathcal{H}$ (16) can be can be written as

$$\begin{pmatrix} \sigma_{11}^* & \sigma_{12}^* & \cdots & \sigma_{1p}^* & 1 \\ \sigma_{21}^* & \sigma_{22}^* & \cdots & \sigma_{2p}^* & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_p^* & \sigma_{p2}^* & \cdots & \sigma_{pp}^* & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \tag{20}$$

After the (p-1)-th row elementary operation, the Hessian matrix $\mathcal{H}$ (16) becomes:

$$\begin{pmatrix} \sigma_{11}^* & \sigma_{12}^* & \cdots & \sigma_{1p}^* & 1 \\ 0 & \sigma_{22}^{*(1)} & \cdots & \sigma_{2p}^{*(1)} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp}^{*(p-1)} & 1 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \tag{21}$$

Because $|A_p| < 0$, using the negative definite property then the pivot element is $\sigma_{pp}^{*(p-1)} < 0$. Furthermore, we perform p-th row elementary operation, we obtained the following matrix

$$\begin{pmatrix} \sigma_{11}^* & \sigma_{12}^* & \cdots & \sigma_{1p}^* & 1 \\ 0 & \sigma_{22}^{*(1)} & \cdots & \sigma_{2p}^{*(1)} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp}^{*(p-1)} & 1 \\ 0 & 0 & \cdots & 0 & -\frac{1}{\sigma_{pp}^{*(p-1)}} \end{pmatrix} \tag{22}$$

Based on the matrix (22), it is found that $|\mathcal{H}| = |A_p|(-\frac{1}{\sigma_{pp}^{*(p-1)}}) < 0$. Therefore, in this case the Hessian matrix (22) is a symmetric matrix negative definite. It can be concluded that the solution (15) is a solution that will maximize the objective function (9).

2.2 Clustering stocks

Because the number of stocks available in the capital market is quite large, it is very difficult to determine the proportion of investment for each stock. Therefore, it is necessary to use data mining techniques to deal with this. One of the data mining techniques that can be used is cluster analysis. Cluster analysis is a statistical analysis that aims to separate objects into several groups that have the same/different characteristics from one group to another group (Xu and Wunsch, 2009). In this analysis, each group is homogeneous between members in groups or variations of objects in groups that are formed as small as possible. Cluster analysis, also called segmentation, has various purposes. Everything is related to grouping or segmenting several objects into subsets or clusters. Objects in a cluster will have a closer relationship compared to objects in other clusters.

There are many cluster techniques in the literature. In this study, the clustering using Ward and complete linkage algorithm will be applied. Ward cluster algorithm is a clustering method based on the square of the distance in the cluster and the square of distance between the clusters. While the complete linkage clustering algorithm is a clustering method based on the farthest distance between objects. If two objects are separated by long distances, then the two objects will be combined into one cluster.

2.3 Sharpe ratio

After the clusters are formed, then the performance of each stock will be assessed in each cluster using the Sharpe ratio. Sharpe ratio or Sharpe index is a measure of excess return (or risk premium) per unit risk in an asset Sharpe (1994). Furthermore, Sharpe (1994) state that Sharpe ratio is used to characterize how well asset returns compensate investors for the risks taken. Sharpe ratio *(SR)* is calculated by comparing the difference between stock returns *(R)* and risk return free rate (R_f) with a standard deviation of stock return (σ) or can be written as follows:

$$SR = \frac{R - R_f}{\sigma} \tag{23}$$

In general, it can be said that the greater the value of the Sharpe ratio of a stock, the better the performance of the stock.

2.4 Robust estimation

There are many types of robust estimation for multivariate data. In this study robust estimation used is robust estimation which is included in the affine equivariant class.

Given a data set $(R = r_1, \ r_2, \dots, r_n)$ with $r_i \in \mathbb{R}^p$. The robust estimate for the mean vector is $\hat{\mu}(R) \in \mathbb{R}^p$ while the covariance matrix estimation is $\hat{\Sigma}(R) \in P_p$ (the set of all symmetric matrices positive definite of size *pxp*). According to (Marona et al, 2006), an estimation is included in the affine equivariant class if it has the properties described in the following definition.

Definition 2.1 (Marona et al, 2006)
Given Q is an invertible matrix of size pxp, vector $v \in \mathbb{R}^p$ and data sets $R \in \mathbb{R}^{pxn}$.
  i. *$\hat{\mu}$ is affine equivariant if $\hat{\mu}(QR + v) = Q\hat{\mu}(R) + v$*
  ii. *$\hat{\Sigma}$ is affine equivariant if $\hat{\Sigma}(QR + v) = Q\hat{\Sigma}(R) + v$*

Robust estimation included in the affine equivariant class that will be used in this study are Fast Minimum Covariance Determinant (FMCD) and S estimation. The following will be briefly presented the FMCD and S estimation. We also presented the procedure for determining $\hat{\mu}$ and $\hat{\Sigma}$ using the robust FMCD estimation method and the robust S estimation method.

2.4.1 Robust FMCD estimation

The minimum covariance determinant (MCD) estimation aims to find robust estimates based on the observations of total observations *(n)*, where the covariance matrix has the smallest determinant. The MCD estimation is a pair of $\hat{\mu} \in \mathbb{R}^p$ and $\hat{\Sigma}$ is a symmetric positive definite matrix with a dimension of *pxp* from a sample of *h* observation, where $\frac{(n+p+1)}{2} \leq h \leq n$ with

$$\hat{\mu} = \frac{1}{h}\sum_{i=1}^{h} r_i \tag{24}$$

The estimation of the covariance matrix can be obtained by solving the following equation:

$$\hat{\Sigma} = \frac{1}{h}\sum_{i=1}^{h}(r_i - \hat{\mu})(r_i - \hat{\mu})' \tag{25}$$

MCD calculations can be very complicated if the data dimensions are getting bigger, this is because this method must examine all possible subsets of *h* from a number of *n* data. Therefore Rousseeuw and Driessen (1999) found a faster calculation algorithm for calculating MCD called Fast MCD (FMCD). The FMCD method is based on the C-Step theorem described below.

119

Theorem 2. 1 (Rousseeuw and Driessen, 1999)

If H_1 is the set of size h taken from data of size n, the sample statistics are:

$$\hat{\mu}^1 = \frac{1}{h}\sum_{i \in H_1} r_i \tag{26}$$

$$\hat{\Sigma}^1 = \frac{1}{h}\sum_{i \in H_1}(r_i - \hat{\mu}^1)(r_i - \hat{\mu}^1)' \tag{27}$$

If $\left|\hat{\Sigma}^1\right| > 0$ than distance $d_i = (r_i; \hat{\mu}^1, \hat{\Sigma}^1)$. Next, specify H_2 is subset consist of the observation with the smallest distance d_i, namely $\{d_1(i)|i \in H_2\} = \{(d_1)_1, \dots, (d_1)_h\}$ where $(d_1)_1 \le (d_1)_2 \le \cdots \le (d_1)_n$ is a sequential distance. Based on H_2, calculate $\hat{\mu}^2$ and $\hat{\Sigma}^2$ using equations (19) and (20), so that

$$\left|\hat{\Sigma}^2\right| \le \left|\hat{\Sigma}^1\right| \tag{28}$$

Equation (28) will be the same if $\hat{\mu}^1 = \hat{\mu}^2$ and $\hat{\Sigma}^1 = \hat{\Sigma}^2$.

C-Step theorem is done repeatedly until $\left|\hat{\Sigma}_{new}\right| = 0$ or $\left|\hat{\Sigma}_{new}\right| = \left|\hat{\Sigma}_{old}\right|$. However, there is no guarantee that the end of the iteration process will produce a new $\hat{\Sigma}_{new}$ that has global minimum of objective function the MCD estimation. Therefore, the MCD solution approach can be done by selecting a number of initial sets of H_1, then using C-Step for each set to get the smallest determinant. Rousseeuw and Driessen (1999) created an algorithm to overcome this problem known as the Fast Minimum Covariant Determinant (FMCD) algorithm. The process of calculating FMCD estimation can be explained in the following algorithm:

1. Take subset of the matrix **R** consist of $h = (n + p + 1)/2$ observations and denoted by H_1.
2. Calculate mean vector ($\hat{\mu}^1$) and covariance matrix ($\hat{\Sigma}^1$)
3. Calculate mahalanobis distance $d_1(i) = (r_i - \hat{\mu}_1)'\Sigma_1^{-1}(r_i - \hat{\mu}_1)$
4. Sort $d_1(i)$ from the smallest to the largest value
5. Define the new subset with H_2, such that
   $$\{d_1(i); i \in H_2\}\{(d_1)_{1:n}, (d_1)_{2:n}, \dots, (d_1)_{h:n}\}$$
   where di mana $(d_1)_{1:n} \le (d_1)_{2:n} \le \cdots \le (d_1)_{h:n}$
6. Calculate mean vector ($\hat{\mu}_2$) and covariance matrix ($\hat{\Sigma}^2$), and $d_2(i)$
7. Repeat steps 1 through steps 6 until we get $\left|\hat{\Sigma}^2\right| \le \left|\hat{\Sigma}^1\right|$

2.4.2 Robust S estimation

This estimation was first introduced by Rosseeuw and Yohai (1984) which was later developed again by Lopuhaa (1989) and Davies (1987).

Definition 2.2 (Davies, 1987)

Given $\{r_i, i = 1, \dots, n\}$ is data set in $\mathbb{R}^p$ and P_p is set of symmetric matrices positive definite with size $p \times p$. S estimation for measure of location $\hat{\mu} \in \mathbb{R}^p$ and dispersion $\hat{\Sigma}(R) \in P_p$ is a pair of $\hat{\mu}$ and $\hat{\Sigma}(R)$ that minimized $|\Sigma|$ with condition

$$\frac{1}{n}\sum_{i=1}^{n}\rho[(\boldsymbol{r}_i - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{r}_i - \boldsymbol{\mu})]^{1/2} = b_0 \tag{29}$$

where ρ *is loss function and* b_0 *is constant. This constant must be determined precisely because this value affects the result of estimation. If the data distribution is unknown then we choose* $b_0 = E\{\rho\|r\|\}$.

The S estimator can be obtained by solving the following equation:

$$\frac{1}{n}\sum_{i=1}^{n}u(d_i)(\boldsymbol{r}_i - \boldsymbol{\mu}) = 0 \tag{30}$$

$$\frac{1}{n}\sum_{i=1}^{n}pu(d_i)(\boldsymbol{r}_i - \boldsymbol{\mu})(\boldsymbol{r}_i - \boldsymbol{\mu})' - v(d_i)\boldsymbol{\Sigma} = 0 \tag{31}$$

where $d_i = (\boldsymbol{r}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{r}_i - \boldsymbol{\mu})$, $\psi(d_i) = \frac{\partial\rho}{\partial d}$, $u(d_i) = \psi(d_i)/d_i$, while $v(d_i) = \psi(d_i)d_i - \rho(d_i) + b_0$.

Calculation of S estimation is done iteratively using equations (30) and (31). According to [30], the algorithm for S estimation is:

1. Determine the initial estimation of mean vector and covariance matrix, $\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{\Sigma}}_0$

2. Calculate $d_i = (\boldsymbol{r}_i - \hat{\boldsymbol{\mu}}_0)'\hat{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{r}_i - \hat{\boldsymbol{\mu}}_0)$

3. Determine k_0 so that $\frac{\sum\rho(d_i/k_0)}{n} = b_0$

4. Calculate $\tilde{d}_i = \frac{d_i}{k_0}$

5. Determine $\hat{\boldsymbol{\mu}} = \frac{\sum\psi(\tilde{d}_i)r_i}{\sum\psi(\tilde{d}_i)}$ and $\hat{\boldsymbol{\Sigma}} = \frac{p\sum\psi(\tilde{d}_i)(r_i-\mu)(r_i-\mu)'}{\sum\psi(\tilde{d}_i)}$

6. Repeat steps 2-3 until $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ convergent

3. Empirical Study

3.1 Clustering and stocks representation of clusters

In this study, we used the daily return of all stocks which are included in the LQ-45 index, listed on the Indonesia Stock Exchange for the period of August 2017-July 2018, which is accessed online through the website https://finance.yahoo.com. The cluster analysis used were Ward and complete linkage algorithm. By using **varclus function** in R package, we found that clustering with Ward algorithm and complete linkage algorithm, stocks LQ-45 indexed can be grouped into 7 clusters, with members of each cluster as presented in Table 1 and Table 2.

Table 1. Clustering of stocks with Ward algorithm

Cluster	Stocks									
1	ADHI	PTPP	WIKA	WSKT						
2	ASII	BBCA	BBTN	BMRI	BBNI	BBRI				
3	GGRM	HMSP	TLKM	UNVR	ICBP	KLBF	SCMA	PWON	INTP	INDF
	JSMR									
4	ANTM	INCO								
5	ADRO	UNTR	BUMI	PTBA						
6	AALI	LSIP								
7	AKRA	MYRX	BMTR	SRIL	BJBR	SSMS	PGAS	BRPT	EXCL	SMRA
	BSDE	PPRO	LKPR	SMRG	LPPF	MNCN				

Table 2: Clustering of stocks with complete linkage algortihm

Cluster	Stocks									
1	GGRM	HMSP	ASII	BBCA	ICBP	KLBF	TLKM	UNVR	SCMA	PWON
	INDF	JSMR								
2	LPPF	MNCN	BRPT	EXCL	PGAS	BBTN	BMRI	BBNI	BBRI	
3	BUMI	PTBA	ANTM	INCO	UNTR					
4	BJBR	SSMS								
5	MYRX	BMTR	SRIL							
6	AALI	LSIP								
7	INTP	LKPR	SMGR	SMRA	BSDE	PPRO	AKRA	ADHI	PTPP	WIKA
	WSKT									

After clusters are formed the next step is the calculation of the Sharpe ratio of each stock in each cluster produced by two clustering methods to select stocks representation of clusters. In Sharpe ratio calculations, the risk return free rate used is the latest Bank Indonesia rate that is 5.25% per year. A Stock that represents a cluster is a stock that has the highest Sharpe ratio in the cluster. Based on the calculation of the Sharpe ratio of each stock in each cluster in two methods of clustering, we obtained stocks that represent each cluster to form the optimum portfolio as presented in Table 3 and Table 4.

Table 3. Stocks representation of clusters with Ward algorithm

Cluster	Cluster Representation	Return	Risk	Sharpe Ratio
1	WSKT	-0.00006	0.02545	-0.01033
2	BBCA	0.00089	0.01285	0.05233
3	HMSP	0.00066	0.02146	0.02137
4	INCO	0.00258	0.02673	0.08885
5	PTBA	0.00233	0.02601	0.08158
6	LSIP	-0.00111	0.01942	-0.06800
7	BRPT	0.00042	0.02296	0.00954

Table 4. Stocks representation of clusters with complete linkage algorithm

Cluster	Cluster Representation	Return	Risk	Sharpe Ratio
1	BBCA	0.00089	0.01285	0.05233
2	BRPT	0.00042	0.02296	0.00954
3	INCO	0.00258	0.02673	0.08885
4	BJBR	0.00023	0.02246	0.00087
5	MYRX	0.00043	0.03291	0.00675
6	LSIP	-0.00111	0.01942	-0.06800
7	WSKT	-0.00006	0.02545	-0.01033

3.2 Comparison of portfolios performance

In this study, the optimum portfolio was determined using the MV portfolio model with robust FMCD estimation (MV_{FMCD}) and MV portfolio model with robust S estimation (MV_S). The first step is determining the portfolio weights

of the two models for various risk aversion values γ using **function CovSest** and
CovMcd in R packages. As a comparison, we also determining the portfolio
weights formed using the classic MV portfolio model ($MV_{Classic}$). The stocks used
are stocks that represent each cluster for two clustering methods as presented in
Table 1 and Table 2. The portfolio weights generated for the two clustering
methods are presented in Table 5 and Table 6.

Table 5.Portfolio weight of MV classic and robust portfolio with Ward algorithm

Model	γ	WSKT	BBCA	HMSP	INCO	PTBA	LSIP	BRPT
$MV_{Classic}$	0.5	-2.41372	4.07947	0.567023	6.03144	4.36919	-9.98716	-1.64626
	1	-1.19807	2.27208	0.325333	3.04801	2.23501	-4.90548	-0.77688
	2	-0.59025	1.36839	0.204489	1.55630	1.16792	-2.36465	-0.34219
	5	-0.22556	0.82617	0.131983	0.66127	0.52766	-0.84014	-0.08138
	10	-0.10399	0.64542	0.107814	0.36293	0.31424	-0.33198	0.00555
MV_{FMCD}	0.5	-11.09913	3.01825	5.134420	2.64043	14.09999	-13.02744	0.23346
	1	-5.51484	1.78887	2.603222	1.32007	7.08539	-6.41399	0.13128
	2	-2.72269	1.17417	1.337622	0.65989	3.57809	-3.10727	0.08019
	5	-1.04741	0.80536	0.578263	0.26378	1.47371	-1.12323	0.04954
	10	-0.48899	0.68242	0.325143	0.13174	0.77225	-0.46189	0.03932
MV_S	0.5	-6.21418	1.74467	4.597106	3.50689	9.68051	-12.74687	0.43186
	1	-3.06731	1.10568	2.340481	1.76421	4.88307	-6.28318	0.25705
	2	-1.49388	0.78619	1.212169	0.89286	2.48435	-3.05133	0.16964
	5	-0.54982	0.59449	0.535181	0.37005	1.04512	-1.11222	0.11719
	10	-0.23513	0.53059	0.309519	0.19579	0.56538	-0.46585	0.09971

Table 6.Portfolio weight of MV classic and robust portfolio with complete linkage algorithm

Model	γ	BBCA	BRPT	INCO	BJBR	MYRX	LSIP	WSKT
$MV_{Classic}$	0.5	6.41036	-1.14087	6.98125	-0.51028	0.46109	-9.32584	-1.87571
	1	3.44013	-0.52416	3.52275	-0.17502	0.25239	-4.59449	-0.92157
	2	1.95501	-0.21581	1.79349	-0.00739	0.14803	-2.22883	-0.44451
	5	1.06394	-0.03079	0.75595	0.09318	0.08542	-0.80943	-0.15826
	10	0.76692	0.0308	0.41009	0.12671	0.06455	-0.33629	-0.06285
MV_{FMCD}	0.5	19.71313	4.05346	4.39266	-2.45391	-17.40476	-5.14132	-2.15927
	1	10.06745	2.07328	2.18078	-1.09881	-8.69087	-2.47936	-1.05248
	2	5.24461	1.08319	1.07485	-0.42127	-4.33392	-1.14838	-0.49908
	5	2.35091	0.48913	0.41129	-0.01474	-1.71975	-0.34979	-0.16704
	10	1.38634	0.29111	0.19009	0.12077	-0.84836	-0.08360	-0.05636
MV_S	0.5	16.85351	3.09968	6.33932	-7.27564	-8.60099	-7.27430	-2.14157
	1	8.61765	1.58682	3.17373	-3.51781	-4.25517	-3.55818	-1.04704
	2	4.49972	0.83039	1.59093	-1.63889	-2.08226	-1.70012	-0.49977
	5	2.02896	0.37653	0.64126	-0.51151	-0.77851	-0.58528	-0.17141
	10	1.20537	0.22525	0.32469	-0.13576	-0.34393	-0.21367	-0.06196

Based on portfolio weights, mean vectors and covariance matrices, we can then determine the return, risk, and Sharpe ratio of the three portfolio models as presented in Table 7 and Table 8.

Table 7: Returns, risks and Sharpe ratio of classic MV portfolio and robust portfolios with Ward algorithm

γ	Return			Risk			Sharpe Ratio		
	$MV_{Classic}$	MV_{FMCD}	MV_S	$MV_{Classic}$	MV_{FMCD}	MV_S	$MV_{Classic}$	MV_{FMCD}	MV_S
0.5	0.04027	0.08594	0.05637	0.07922	0.17208	0.11268	0.14255	0.20681	0.16751
1	0.02049	0.04293	0.02822	0.01988	0.04307	0.02823	0.14428	0.20617	0.16713
2	0.01059	0.02143	0.01415	0.00505	0.01082	0.00712	0.14718	0.20465	0.16602
5	0.00466	0.00853	0.00570	0.00089	0.00179	0.00120	0.15142	0.19831	0.16020
10	0.00269	0.00423	0.00289	0.00029	0.00049	0.00036	0.14738	0.18311	0.14469

Table 8: Returns, risks, and Sharpe ratio of classic MV portfolio and robust portfolios with complete linkage algorithm

γ	Return			Risk			Sharpe Ratio		
	$MV_{Classic}$	MV_{FMCD}	MV_S	$MV_{Classic}$	MV_{FMCD}	MV_S	$MV_{Classic}$	MV_{FMCD}	MV_S
0.5	0.03375	0.07947	0.06715	0.06655	0.16069	0.13560	0.13026	0.19789	0.18195
1	0.01713	0.03931	0.03326	0.01671	0.04021	0.03395	0.13143	0.19533	0.17974
2	0.00883	0.01923	0.01632	0.00425	0.01009	0.00853	0.13323	0.19004	0.17510
5	0.00384	0.00718	0.00615	0.00076	0.00165	0.00142	0.13436	0.17308	0.15967
10	0.00218	0.00317	0.00276	0.00026	0.00045	0.00039	0.12656	0.14260	0.13109

2. Discussions

The results obtained through cluster analysis with Ward algorithms and complete linkage algorithms are that stocks included in the LQ-45 index can be grouped into 7 clusters as shown in Table 1 and Table 2. With clustering using the Ward algorithm, in cluster 1 WSKT has the best performance compared to other stocks in cluster 1 which are marked by the highest Sharpe ratio in cluster 1, which is -0.01033. So that WSKT was chosen as a representation of cluster 1. Furthermore, in cluster 2, BBCA with a Sharpe ratio of 0.05233 was representations of cluster 2, and so on BRPT with Sharpe ratio 0.00954 was a representation of cluster 7. On the other hand, clustering using complete linkage algorithms is obtained that in cluster 1, BBCA has the best performance compared to other stocks in cluster 1 which are marked by the highest Sharpe ratio in cluster 1, which is equal to 0.0523. So that BBCA is taken as a representation of cluster 1. Furthermore, in cluster 2, BRPT with Sharpe ratio 0.00954 is representations of cluster 2, and so on WSKT with Sharpe ratio -0.01033 is representations of cluster 7. Therefore, it is enough to consider stocks as presented in Table 1 (Ward algorithm) or Table 2 (complete linkage algorithm) for investment decisions.

From Table 5 (with Ward algorithm), it can be seen that stocks with negative returns, namely WSKT and LSIP have a negative weight (short selling) for all risk aversion values γ in three portfolio models. Conversely, stocks with large returns, namely BBCA, HMSP, INCO and PTBA always have positive

124

weights on three portfolio models. From Table 5, it can also be seen that the greater the value of γ, the weight of the stock with a positive return will be smaller, as well as for stocks with negative returns the weight of stocks for short selling also decreases as the value of γ increases. The same thing also happened in the formation of portfolios using representation stocks generated by using clusters with the complete linkage algorithm, as shown in Table 6.

Measuring portfolio performance not only be seen from the return but also must pay attention to the risks that will be borne by the investor. Several measures can be used to measure portfolio performance, one of which is using the Sharpe ratio. Table 7 and Table 8 show the returns, risks, and Sharpe ratio of portfolios formed using cluster analysis with Ward algorithm and complete linkage combined with the classic MV portfolio model, MV_{FMCD} model and MV_S model. From Table 7 and Table 8, in general, it can be concluded that the performance of the portfolio generated by using clustering with Ward algorithm is better than the portfolio performance produced by the complete linkage algorithm for all risk aversion values γ.

4. Conclusions and Future Research

This paper shows how to integrate clustering techniques into portfolio management and building systems to get an efficient portfolio. This can reduce a lot of time in stocks selection because stocks of similar categories can be easily grouped into one cluster. The best performing stocks from each cluster are then chosen as representations of the cluster to built the portfolio. The results showed that by using cluster analysis with Ward algorithm and complete linkage algorithm, 45 stocks of the Indonesia Stock Exchange included in the LQ-45 index can be grouped into 7 clusters. Stocks as a representation of each cluster are then used to form a portfolio using the MV model with robust FMCD estimation, robust S estimation, and the classic MV model. Portfolio performance that formed by combining the result of two clustering methods and robust estimation techniques then compared. The results showed that portfolio performance with robust FMCD estimation was better than portfolio performance with robust S estimation as well as classic MV portfolio for all risk aversion values γ, both for portfolios produced by cluster analysis with Ward algorithm and complete linkage algorithm.

There are many future research potentials in portfolio selection using clustering and robust estimation. It will be interesting to use nonclassical clustering methods and combine them with robust estimation methods and compare the performance of generated portfolios.

125

REFERENCES

[1] **Bengtsson,C. (2004),** *The Impact of Estimation Error on Portfolio Selection for Investor with Constant Relative Risk Aversion.* Working Paper, *Department of Economics Lund University;*

[2] **Best, M.J. and Grauer, R.R. (1991),** *On the sensitivity of Mean-Variance Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results.* Review of Financial Studies, 4(2): 315-342;

[3] **Broadie,M. (1993),** *Computing Efficient Frontiers Using Estimated Parameters.* Annals of Operations Research, 45: 21-58;

[4] **Ceria, S. and Stubbs, R.A. (2006),** *Incorporating Estimation Errors into Portfolio Selection: Robust Portfolio Construction.* Journal of Asset Management, 7(2): 109-127;

[5] **Chen, L.H., and L. Huang. (2009),** *Portfolio Optimization of Equity Mutual Funds with Fuzzy Return Rates and Risks.* Expert Systems with Applications, 36: 3720-3727;

[6] **Chong, E.D.P and Zak, S.H. (2001),** *An Introduction to Optimization. Second Edition. John Wiley and Sons, Inc.;*

[7] **Chopra, V. K. and Ziemba,W. T. (1993),** *The Effects of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice.* Journal of Portfolio Management, 19(2): 6-11;

[8] **DeMiguel V. and Nogales,F.J. (2008),** *Portfolio Selection with Robust Estimation.* Journal of Operation Research, 57: 560-577;

[9] **Davies, P. (1987),** *Asymptotic Behavior of S-estimators of Multivariate Location Parameters and Dispersion Matrices.* The Annals of Statistics, 15: 1269-1292;

[10] **Guan,H.S. and Jiang,Q.S. (2007),** *Cluster Financial Time Series for Portfolio.* Proceedings of International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR, Beijing, China, 2: 851–856;

[11] **Hardin,J.S. (2000),** *Multivariate Outlier Detection and Robust Clustering with Minimum Covariance Determinant Estimation and S-Estimation.* PhD Dissertation, University of California;

[12] **Hu,J. (2012),** *An Empirical Comparison of Different Approaches in Portfolio Selection.* Project Report, Uppsala Universitet;

[13] **Kusch,P. (2012),** *Portfolio Optimization with Robust Mean-Variance and Mean-Conditional Value at Risk Strategies.* Project Report, Department of Economics Umea University;

[14] **Lauprete (2001),** *Portfolio Risk Minimization under Departures from Normality.* PhD Disertation, Massachusetts Institute of Technology;

[15] **Leon, S.J. (2002),** *Linear Algebra with Applications.* Sixth Edition, Prentice-Hall, New Jersey;

[16] **Long, N.C., Wisitponghan, N., Meesad, P. (2014),** *Clustering Stock Data for Multi-Objective Portfolio Optimization.*International Journal of Computational Intelligence and Applications, 13(2): 1-13;

[17] **Lopuhaa,H.P. (1989),** *On The Relation between S-estimators and M-estimators of Multivariate Location and Covariance*. The Annals of Statistics, 17: 1662-1683;

[18] **Markowitz,H.M. (1952),** *Portfolio Selection*. Journal of Finance, 7: 77-91;

[19] **Marona,R.A., Martin,R.D., Yohai,V.J. (2006),** *Robust Statistics: Theory and Methods*. John Wiley and Sons;

[20] **Nanda,R., Mahanty, B., Tiwari, M.K. (2010),***Clustering Indian Stock Market Data for Portfolio Management.*Expert Syst. Appl., 37(12): 8793–8798;

[21] **Rousseeuw, P. and Yohai,V. (1984),** *Robust Regression by Means of S-estimators, Appeared in Robust and Nonlinear Time Series Analysis*. Lecture Notes in Statistics, 26: 256 – 272;

[22] **Rousseeuw, P.J. and Van Driessen, K. (1999),** *A Fast Algorithm for the Minimum Covariance Determinant Estimator*. Technometrics, 41: 212 - 223.

[23] **Sharpe,W.F. (1994),** *The Sharpe Ratio*. The Journal of Portfolio Management, 21: 49–58;

[24] **Supandi,E.D. (2017),** *Developing of Mean-Variance Portfolio Modeling Using Robust Estimation and Robust Optimization Method*. PhD Dissertation, Mathematics Department Gadjah Mada University, Indonesia;

[25] **Tola, V., Lillo, F., Gallegati, M., Mantegna, R.N. (2008),** *Cluster Analysis for Portfolio Optimization*. J. Econ. Dyn. Control, 32(1): pp. 235–258;

[26] **Vaz de Melo, B. and Camara,R.P. (2003),** *Robust Multivariate Modelling in Finance*. Working Paper Series 355, Federal University at Rio de Janeiro, Rio de Janeiro, Brazil;

[27] **Victoria-Feser,M.P. (2000),** *Robust Portfolio Selection*. Research Report, Universite de Geneve;

[28] **Welsh, R.Y. and Zhou (2007),** *Application of Robust Statistics to Asset Allocation Models*. Statistical Journal, 5(1): 97-114;

[29] **Xu, R. and Wunsch, D.C. (2009),** *Clustering*. John Wiley and Sons Inc., Hoboken, New Jersey;

[30] **Zhou,X. (2006),** *Application of Robust Statistics to Asset Allocation Models.* MasterThesis, Massachusetts Institute of Technology.