
ROBUST PORTFOLIO SELECTION WITH CLUSTERING BASED ON BUSINESS SECTOR OF STOCKS

La Gubu^{1,2}, Dedi Rosadi², Abdurakhman²

¹ Jurusan Matematika FMIPA Universitas Halu Oleo

² Departemen Matematika FMIPA Universitas Gadjah Mada

e-mail: lagubu2014@gmail.com

DOI: 10.14710/medstat.14.1.33-43

Article Info:

Received: 30 August 2020

Accepted: 19 April 2021

Available Online: 30 June 2021

Keywords:

business sector; portfolio; Sharpe ratio; robust estimation, portfolio performance

Abstract: In recent years there have been numerous studies on portfolio selection using cluster analysis in conjunction with Markowitz model which used mean vectors and covariance matrix that are estimated from a highly volatile data. This study presents a more robust way of portfolio selection where stocks are grouped into clusters based on business sector of stocks. A representative from each cluster is selected from each cluster using Sharpe ratio to construct a portfolio and then optimized using robust FCMD and S-estimation. Calculation Sharpe ratio showed that this method works efficiently on large number of data while also robust against outlier in comparison to k-mean clustering. Implementation of this method on stocks listed on the Indonesia Stock Exchange, which included in the LQ-45 indexed for the period of August 2017 to July 2018 showed that portfolio performance obtained using clustering base on business sector of stocks combine with robust FMCD estimation is outperformed the other possible combination of the methods.

1. INTRODUCTION

Markowitz (1952) proposed a portfolio model using the mean and variance of asset returns to express the trade-off between return and risk of portfolio. Hence the Markowitz portfolio model is also called the Mean-Variance (MV) portfolio model. This model is an optimization problem with two opposing goals. This means that the expected return from the portfolio results needs to be maximized, on the other hand, the portfolio risk represented by the variance of returns from assets needs to be minimized.

Various studies have been conducted to solve and develop the Markowitz portfolio model. All of this is done to adapt the existing model to financial market conditions and the demands of capital market practitioners. One of the researches focuses in portfolio selection is the optimal portfolio selection time efficiency. This is understandable because the greater the number of securities involved in portfolio selection, the more likely portfolios can be formed. The number of securities involved in portfolio selection can be overcome by grouping the securities data using cluster analysis. Securities that have similar characteristics are grouped into the same cluster.

In recent years, many studies on portfolio selection have used cluster analysis, including (Guan & Jiang, 2007), Tola et al. (2008), Chen & Huang (2009), Nanda et al. (2010), and Long et al. (2014). Based on the results reported by those studies, the optimal

portfolio was obtained using the MV Markowitz portfolio model. The main problem of MV Markowitz portfolio model is that the mean vectors and the covariance matrix must be estimated from highly volatile data. Parameter estimation can be done using a variety of estimation technique, which will inevitably contain estimation errors. As a very important input in the formation of the MV model portfolio, the estimation error will significantly affect the results of the optimal portfolio formation. Best & Grauer (1991), Chopra & Ziemba (1993), and Ceria & Stubbs (2006) have conducted several studies related to estimation errors and their relationship to optimal portfolio formation. Based on these studies, it is concluded that although the MV model is supported by a strong theory and has ease of computation, the MV model shows several weaknesses. One of the weaknesses is that the optimal portfolio produced by this model is not well diversified and tends to be concentrated in a small proportion of assets (Fabozzi et al., 2007). In addition, the MV model is also sensitive to the changes of input parameters, namely the mean vector and the covariance matrix.

To overcome the weaknesses previously discussed, several researchers have built a robust portfolio, which is a portfolio that can reduce the error in estimating the mean vector and covariance matrix in the MV model portfolio. One of the standard approaches in forming the optimal robust portfolio is through the robust estimation approach. Several studies on optimal portfolio formation using robust estimates have been carried out by Lauprete (2001), Vaz-de Melo & Camara (2005), Welsch & Zhou (2007), DeMiguel & Nogales (2009), Supandi (2017), and Gubu et al. (2020). The difference between these studies lies in the robust estimation used in portfolio optimization. All the results of these studies indicate that the performance of the portfolio formed using robust estimation is better than the performance of the classic MV portfolio if there are outliers. However, the literature considers the combination of cluster analysis and robust estimation methods in the formation of an optimal portfolio is still limited. In addition, stock clustering as an initial analysis in portfolio formation uses cluster analysis that is well known in the literature.

As a new contribution in this paper, stock clustering is carried out based on business sector of stock. The stock representation of each cluster then combined with a robust estimation to form an optimal portfolio. For the mean vector and covariance matrix, it is estimated using the robust Fast Minimum Covariance Determinant (FMCD) estimation and the robust S estimation, because these two estimators have high breakdown points (Supandi, 2017).

2. LITERATURE REVIEW

2.1. Mean-Variance Portfolio

Markowitz's portfolio theory is based on the mean and variance approach, where the mean is a measure of the expected rate of return and variance is a measure of the level of risk (Markowitz, 1952). Therefore, Markowitz's portfolio theory is also called the Mean-Variance (MV) portfolio model. This model emphasizes efforts to maximize expected returns and minimize risks to form an optimum portfolio. According to Supandi (2017), the mean-variance portfolio can be formulated as the following optimization problem:

$$\max_w w' \mu - \frac{\gamma}{2} w' \Sigma w \quad (1)$$

$$w' e = 1 \quad (2)$$

where w represents the portfolio weight, μ is the mean vector, Σ is the covariance matrix, e is the column matrix where all elements are 1 and $\gamma \geq 0$ is the risk aversion parameter, that is the relative measure of risk avoidance.

The optimization problem in Equations (1) and (2) can be solved using the Lagrange method (Winston & Goldberg, 2004). First, form the Lagrange function:

$$L = w' \mu - \frac{1}{2} \gamma w' \Sigma w + \lambda (w' e - 1) \quad (3)$$

Based on the Kuhn-Tucker theorem (Winston & Goldberg, 2004), the necessary conditions to find the optimal value of Equation (3) are:

$$\frac{\partial L}{\partial w} = 0 \quad \text{and} \quad \frac{\partial L}{\partial \lambda} = 0 \quad (4)$$

From Equations (3) and (4) we have

$$w = \frac{\Sigma^{-1}}{\gamma} (\mu + \lambda e) \quad \text{and} \quad e' w = 1 \quad (5)$$

From Equation (5) it is obtained

$$\lambda = \gamma (e' \Sigma^{-1} e)^{-1} - (e' \Sigma^{-1} e)^{-1} e' \Sigma^{-1} \mu \quad (6)$$

Substitution of Equation (6) to Equation (5) gives:

$$w = \frac{1}{\gamma} (\Sigma^{-1} - \Sigma^{-1} e (e' \Sigma^{-1} e)^{-1} e' \Sigma^{-1}) \mu + \Sigma^{-1} e (e' \Sigma^{-1} e)^{-1} \quad (7)$$

Equation (7) shows that the optimal portfolio weight (w) depends on input μ and Σ .

2.2. Sharpe Ratio

After the clusters are formed, an assessment of the performance of each stock in each cluster is carried out using the Sharpe ratio (SR). Sharpe ratio or Sharpe index is a measurement of excess return (or risk premium) per unit risk in assets (Sharpe, 1994). Sharpe ratio is used to characterize how well the return on assets compensates investors for the risk taken. Furthermore, Sharpe (1994) states that SR is calculated by comparing the difference between stock return (R) and return risk-free rate (R_f) with the standard deviation of stock return (σ) or it can be written as follows:

$$SR = \frac{R - R_f}{\sigma} \quad (8)$$

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

2.3. Optimum Portfolio Selection Using Robust Estimation

The classic MV portfolio becomes ineffective when faced the conditions of return data that do not meet the assumptions of a multivariate normal distribution, because classical estimates of mean and variance are not robust and are strongly influenced by observations that deviate (outliers), (Maronna et al., 2006). Robust statistics are concerned with establishing stable statistical procedures when there are parts of the data that do not fit the assumed distribution or there are deviations from the model.

In this study, the weight of the selected stocks to form the optimum portfolio is determined using the robust FMCD estimation method and the robust S estimation method.

The following will briefly present the FMCD robust estimation method procedure and the robust S estimation method.

2.3.1 Robust FMCD Estimation Method

Minimum Covariance Determinant (MCD) estimation aims to find robust estimation based on h observations from total observations (n), where the covariance matrix has the smallest determinant. The MCD estimation is a pair of $\hat{\boldsymbol{\mu}} \in \mathbb{R}^p$ and $\hat{\boldsymbol{\Sigma}}$, which is a symmetrical positive definite matrix of dimensions $p \times p$ from a sample subspace of size h , where $\frac{(n+p+1)}{2} \leq h \leq n$,

$$\hat{\boldsymbol{\mu}} = \frac{1}{h} \sum_{i=1}^h \mathbf{r}_i \quad (9)$$

where \mathbf{r}_i is stock return of i -th stock, $i = 1, \dots, h$

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

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{h} \sum_{i=1}^h (\mathbf{r}_i - \hat{\boldsymbol{\mu}})(\mathbf{r}_i - \hat{\boldsymbol{\mu}})' \quad (10)$$

The calculation of MCD can become complicated as the data dimensions get bigger, this is because this method has to examine all possible subsets h of a number of n data. Therefore, Rousseeuw & Van Driessen (1999) found a faster calculation algorithm for calculating MCD called Fast MCD (FMCD). The FMCD method is based on the C-Step theorem which is explained below.

Theorem 1 (Rousseeuw & Van Driessen, 1999) *If H_1 is the set of size h taken from data of size n , the sample statistics are:*

$$\hat{\boldsymbol{\mu}}^1 = \frac{1}{h} \sum_{i \in H_1} \mathbf{r}_i \quad (11)$$

$$\hat{\boldsymbol{\Sigma}}^1 = \frac{1}{h} \sum_{i \in H_1} (\mathbf{r}_i - \hat{\boldsymbol{\mu}}^1)(\mathbf{r}_i - \hat{\boldsymbol{\mu}}^1)' \quad (12)$$

If $|\hat{\boldsymbol{\Sigma}}^1| > 0$ than distance $d_i = (\mathbf{r}_i; \hat{\boldsymbol{\mu}}^1, \hat{\boldsymbol{\Sigma}}^1)$. 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 $(d_1)_1 \leq (d_1)_2 \leq \dots \leq (d_1)_n$ is a sequential distance. Based on H_2 , calculate $\hat{\boldsymbol{\mu}}^2$ and $\hat{\boldsymbol{\Sigma}}^2$ using equations (11) and (12), so that

$$|\hat{\boldsymbol{\Sigma}}^2| \leq |\hat{\boldsymbol{\Sigma}}^1| \quad (13)$$

Equation (13) will be the same if $\hat{\boldsymbol{\mu}}^1 = \hat{\boldsymbol{\mu}}^2$ and $\hat{\boldsymbol{\Sigma}}^1 = \hat{\boldsymbol{\Sigma}}^2$.

C-Step theorem is done repeatedly until $|\hat{\boldsymbol{\Sigma}}_{new}| = 0$ or $|\hat{\boldsymbol{\Sigma}}_{new}| = |\hat{\boldsymbol{\Sigma}}_{old}|$.

2.3.2 Robust S Estimation Method

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

Definition 2.1 (Davies, 1987) Given $\{\mathbf{r}_i, i = 1, \dots, n\}$ 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 $\hat{\boldsymbol{\mu}} \in \mathbb{R}^p$ and dispersion $\hat{\boldsymbol{\Sigma}}(R) \in P_p$ is a pair of $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}(R)$ that minimized $|\boldsymbol{\Sigma}|$ with condition

$$\frac{1}{n} \sum_{i=1}^n \rho[(\mathbf{r}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{r}_i - \boldsymbol{\mu})]^{1/2} = b_0 \quad (14)$$

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 $b_0 = E\{\rho\|\mathbf{r}\|\}$.

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

$$\frac{1}{n} \sum_{i=1}^n \mathbf{u}(d_i) (\mathbf{r}_i - \boldsymbol{\mu}) = \mathbf{0} \quad (15)$$

$$\frac{1}{n} \sum_{i=1}^n p \mathbf{u}(d_i) (\mathbf{r}_i - \boldsymbol{\mu}) (\mathbf{r}_i - \boldsymbol{\mu})' - v(d_i) \boldsymbol{\Sigma} = \mathbf{0} \quad (16)$$

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

Calculation of S estimation is done iteratively using equations (15) and (16). According to Hardin (2000), 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 $d_i = (\mathbf{r}_i - \hat{\boldsymbol{\mu}}_0)' \hat{\boldsymbol{\Sigma}}_0^{-1} (\mathbf{r}_i - \hat{\boldsymbol{\mu}}_0)$
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 $\hat{\boldsymbol{\mu}} = \frac{\sum \psi(\tilde{d}_i) \mathbf{r}_i}{\sum \psi(\tilde{d}_i)}$ and $\hat{\boldsymbol{\Sigma}} = \frac{p \sum \psi(\tilde{d}_i) (\mathbf{r}_i - \boldsymbol{\mu}) (\mathbf{r}_i - \boldsymbol{\mu})'}{\sum \psi(\tilde{d}_i)}$
6. Repeat steps 2-3 until $\hat{\boldsymbol{\mu}}$ dan $\hat{\boldsymbol{\Sigma}}$ convergent

3. METHODOLOGY

3.1. Data

There are several data used in this research, namely:

1. Data of stocks included in the LQ-45 group for the period August 2017 - January 2018.
2. Data grouping of stocks on Indonesia Stock Exchange based on the business sector of stocks by the Indonesia Stock Exchange.
3. Data of daily closing prices for LQ-45 stocks for the period of August 2017 to January 2018 which was later expanded to July 2018 obtained from www.yahoo.finance.com.

3.2. Procedures

This research was conducted with the following procedures:

1. LQ-45 stocks are grouped into several clusters based on business sector of stock.

2. Calculate the return and risk of historical stock data in each cluster. From the calculation of return and risk, the performance of each stock in each cluster can be determined using the Sharpe ratio.
3. Select stocks that represent each cluster to build the optimum portfolio. Stocks that are chosen as a representation of a cluster are stocks with the highest Sharpe ratio.
4. Determine the weight of each stock to build a portfolio using the robust FMCD estimation method, the robust S estimation method and the classic MV method.
5. To see the advantages of the proposed method, the performance of the portfolios formed is then compared with the performance of the portfolios formed using the familiar clustering method, namely the k-means clustering method as used by Rifa et al. (2020). The number of clusters formed by the k-mean method were 9 and 4 clusters. The choice of 9 clusters corresponds to the number of clusters based on the sector business of stock, while the choice of 4 clusters is the optimum number of clusters for the data used.

4. RESULTS AND DISCUSSION

4.1. Clustering Results

In this study, the stock clustering was carried out based on the business sector of the stock. Based on the grouping criteria issued by the Indonesia Stock Exchange, LQ-45 stocks are grouped into nine clusters. As a comparison, clustering is also carried out using the k-means method. The number of clusters formed by the k-mean method were 9 and 4 clusters.

After the clusters are formed, the next step is to calculate the Sharpe ratio of each stock in each cluster. In calculation of Sharpe ratio, the return risk-free rate used is the Bank Indonesia rate at the time of data collection, namely 5.25% per year. Based on the Sharpe ratio calculation for each stock in each cluster, stocks are obtained which represent each cluster to build the optimum portfolio for two clustering methods presented in Table 1.

Table 1. Stocks Representation of Clusters

Cluster	Clustering Method					
	Based on Business Sector		k-mean (9 cluster)		k-mean (4 cluster)	
	Representation	Sharpe Ratio	Representation	Sharpe Ratio	Representation	Sharpe Ratio
1	SSMS	-0.04139	UNTR	0.02962	BBCA	0.05713
2	INCO	0.09116	ANTM	0.05068	INCO	0.09116
3	BRPT	0.01222	UNVR	-0.03521	ICBP	0.00958
4	SRIL	-0.00583	LPPF	-0.03123	GGRM	0.00941
5	HMSP	0.02424	ADRO	0.01236		
6	MYRX	0.00862	GGRM	0.00941		
7	PGAS	-0.01165	BBCA	0.05713		
8	BBCA	0.05713	INCO	0.09116		
9	UNTR	0.02962	ICBP	0.00958		

Using clustering based on the business sector of stock, cluster 1 consists of 3 stocks and SSMS stock has the best performance compared to other stocks in the cluster which are marked by the highest Sharpe ratio value in the cluster, which is -0.04139. So that SSMS stock is chosen as a representation of cluster 1. Furthermore, in cluster 2 which consists of 5 stock and INCO stock with Sharpe ratio of 0.09116 is a representation of cluster 2. And so on, the stock BRPT, SRIL, HMSP, MYRX, PGAS, BBCA and UNTR are respectively representation of clusters 3, 4,5, 6, 7, 8, and 9.

Using the k-mean clustering method with 9 clusters, in cluster 1 there is only 1 stock, namely UNTR, so that UNTR stock become a representation of cluster 1, then in cluster 2 there are 11 stocks and ANTM stock is the stocks with the highest Sharpe ratio in the cluster. In the same way, the stock UNVR, LPPF, ADRO, GGRM, BBKA, INCO and ICBP are representations of clusters 3, 4, 5, 6, 7, 8, and 9. While the clustering using the optimal k-mean (4 clusters) is obtained BBKA, INCO, ICBP and GGRM stock as a representation of clusters 1, 2, 3, and 4.

4.2. Comparison of the Performance of Portfolios

In this study, the optimum portfolio is determined using the MV portfolio model with robust FMCD estimation (MV_{FMCD}), the MV portfolio model with robust estimation S (MV_S) and the classic MV portfolio model ($MV_{classic}$). The first step is to determine the portfolio weights of the three models for various risk aversion values γ using the **CovMcd** and **CovSest** functions on R packages (Würtz et al., 2009). As a comparison, the optimum portfolio is also determined using the k-mean clustering method (using the **kmeans** function in R packages) with 9 and 4 clusters. The stocks used are stocks which represent each cluster as presented in Table 1. The portfolio weights are presented in Table 2, Table 3, and Table 4.

Table 2. Portfolio Weight with Clustering Based on Business Sector of Stock

Model	γ	SSMS	INCO	BRPT	SRIL	HMSP	MYRX	PGAS	BBKA	UNTR
$MV_{classic}$	0.5	-7.18415	6.32096	-1.03784	-2.17993	1.06934	0.11039	-1.77714	6.17780	-0.49941
	1	-3.46418	3.19607	-0.48011	-1.02109	0.55483	0.07446	-0.89853	3.26714	-0.22858
	2	-1.60420	1.63363	-0.20125	-0.44166	0.29758	0.05650	-0.45923	1.81181	-0.09317
	5	-0.48820	0.69616	-0.03393	-0.09401	0.14322	0.04572	-0.19565	0.93861	-0.01192
	10	-0.11621	0.38367	0.02184	0.02188	0.09177	0.04212	-0.10779	0.64754	0.01517
	15	0.00779	0.27951	0.04043	0.06050	0.07462	0.04093	-0.07850	0.55052	0.02420
	20	0.06979	0.22743	0.04973	0.07982	0.06605	0.04033	-0.06385	0.50201	0.02871
MV_{FMCD}	0.5	-9.00513	5.72357	1.25498	11.09480	4.13869	-6.34818	-16.39009	6.43109	4.10026
	1	-4.27551	2.86477	0.65309	5.61774	2.07552	-3.15460	-8.18224	3.35394	2.04730
	2	-1.91071	1.43536	0.35214	2.87920	1.04393	-1.55780	-4.07831	1.81537	1.02082
	5	-0.49182	0.57772	0.17157	1.23608	0.42498	-0.59973	-1.61595	0.89223	0.40493
	10	-0.01886	0.29184	0.11138	0.68838	0.21866	-0.28037	-0.79517	0.58451	0.19963
	15	0.13880	0.19655	0.09132	0.50581	0.14989	-0.17392	-0.52157	0.48194	0.13120
	20	0.21762	0.14890	0.08129	0.41452	0.11551	-0.12069	-0.38478	0.43065	0.09698
MV_S	0.5	-3.30986	6.92331	1.10496	1.70397	3.40532	-9.70056	-8.32274	6.82135	2.37426
	1	-1.47976	3.46773	0.58224	0.91513	1.71654	-4.81706	-4.13967	3.56525	1.18961
	2	-0.56471	1.73994	0.32088	0.52070	0.87215	-2.37532	-2.04813	1.93720	0.59728
	5	-0.01568	0.70326	0.16407	0.28405	0.36551	-0.91027	-0.79321	0.96037	0.24189
	10	0.16733	0.35771	0.11180	0.20517	0.19664	-0.42192	-0.37490	0.63476	0.12342
	15	0.22833	0.24252	0.09437	0.17887	0.14034	-0.25914	-0.23547	0.52623	0.08393
	20	0.25883	0.18493	0.08566	0.16573	0.11220	-0.17774	-0.16575	0.47196	0.06419

Table 3. Portfolio Weight with Clustering k-mean (9 Cluster)

Model	γ	UNTR	ANTM	UNVR	LPPF	ADRO	GGRM	BBKA	INCO	ICBP
$MV_{classic}$	0.5	0.35038	0.69091	-11.72962	-3.20853	-1.16961	0.65909	8.25952	6.60250	0.54535
	1	0.21030	0.35707	-5.76372	-1.58951	-0.59901	0.36861	4.28176	3.33585	0.39866
	2	0.14026	0.19015	-2.78077	-0.78001	-0.31372	0.22337	2.29288	1.70253	0.32531
	5	0.09824	0.09000	-0.99100	-0.29430	-0.14254	0.13623	1.09955	0.72253	0.28130
	10	0.08423	0.05661	-0.39441	-0.13240	-0.08548	0.10718	0.70177	0.39587	0.26663
	15	0.07956	0.04548	-0.19555	-0.07843	-0.06646	0.09750	0.56918	0.28698	0.26174
	20	0.07722	0.03992	-0.09611	-0.05145	-0.05695	0.09266	0.50288	0.23253	0.25930
MV_{FMCD}	0.5	12.33172	-10.89482	6.57445	-6.53360	5.07325	-5.04100	4.66632	3.68058	-8.85689
	1	6.17115	-5.41833	3.44028	-3.25310	2.51785	-2.49957	2.48513	1.86089	-4.30431
	2	3.09087	-2.68009	1.87320	-1.61286	1.24016	-1.22885	1.39453	0.95105	-2.02802
	5	1.24270	-1.03714	0.93295	-0.62871	0.47354	-0.46642	0.74017	0.40515	-0.66224
	10	0.62664	-0.48949	0.61954	-0.30066	0.21800	-0.21228	0.52205	0.22318	-0.20698
	15	0.42129	-0.30694	0.51506	-0.19131	0.13282	-0.12756	0.44934	0.16253	-0.05523
	20	0.31861	-0.21567	0.46283	-0.13663	0.09023	-0.08521	0.41299	0.13220	0.02065

MV_S	0.5	4.61673	-3.69464	-3.83908	-2.89185	2.65530	-2.31557	7.17107	4.56815	-5.27012
	1	2.32529	-1.82747	-1.76909	-1.43322	1.32556	-1.12770	3.72289	2.30413	-2.52039
	2	1.17956	-0.89388	-0.73409	-0.70391	0.66069	-0.53376	1.99880	1.17212	-1.14553
	5	0.49213	-0.33373	-0.11309	-0.26632	0.26177	-0.17740	0.96434	0.49291	-0.32061
	10	0.26299	-0.14701	0.09391	-0.12046	0.12879	-0.05861	0.61952	0.26651	-0.04564
	15	0.18660	-0.08477	0.16291	-0.07184	0.08447	-0.01902	0.50458	0.19104	0.04602
	20	0.14841	-0.05365	0.19741	-0.04753	0.06230	0.00078	0.44711	0.15331	0.09185

Table 4. Portfolio Weight with Clustering k-mean (4 Cluster)

Model	γ	BBCA	INCO	ICBP	GGRM
$MV_{Classic}$	0.5	2.92203	5.00731	-4.57171	-2.35762
	1	1.66200	2.56234	-2.11519	-1.10915
	2	1.03199	1.33985	-0.88693	-0.48491
	5	0.65398	0.60636	-0.14997	-0.11037
	10	0.52798	0.36187	0.09568	0.01448
	15	0.48598	0.28037	0.17756	0.05609
	20	0.46498	0.23962	0.21851	0.07690
MV_{FMCD}	0.5	0.91641	6.70728	-4.74234	-1.88135
	1	0.73525	3.36042	-2.21482	-0.88085
	2	0.64467	1.68700	-0.95106	-0.38060
	5	0.59032	0.68294	-0.19280	-0.08046
	10	0.57220	0.34825	0.05995	0.01959
	15	0.56616	0.23669	0.14420	0.05294
	20	0.56314	0.18091	0.18633	0.06962
MV_S	0.5	4.03509	4.11326	-4.40308	-2.74527
	1	2.27367	2.08270	-2.05765	-1.29872
	2	1.39296	1.06742	-0.88494	-0.57544
	5	0.86454	0.45825	-0.18131	-0.14148
	10	0.68839	0.25520	0.05323	0.00318
	15	0.62968	0.18751	0.13141	0.05139
	20	0.60032	0.15367	0.17050	0.07550

From Table 2 it can be seen that stocks with negative returns, namely SSMS and PGAS stocks, have a negative weight (short selling) for almost all risk aversion values γ in the three portfolio models. On the other hand, stocks with large returns, namely INCO, HMSP and BBCA stocks always have positive weights on the three portfolio models. From Table 2 it can also be seen that the greater the value of γ , the smaller the weight of the stock with a positive return, conversely, for stocks with a negative return, the stock weight will get bigger along with the increase value of γ . The same thing also occurs in Table 3, in this case UNVR and LPPF have negative returns and BBCA and INCO stocks have large returns.

Based on the portfolio weightings, as well as the mean vectors and covariance matrices, the Sharpe ratio was calculated for the three portfolio models, as presented in Table 5, Table 6, and Table 7.

Table 5. Return, Risk, and Sharpe ratio of Portfolio Using Clustering Based on Business Sector of Stocks

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	0.02581	0.09309	0.06616	0.05073	0.18663	0.13281	0.11395	0.21516	0.18116
1	0.01315	0.04645	0.03297	0.01274	0.04668	0.03324	0.11520	0.21430	0.18007
2	0.00682	0.02312	0.01638	0.00324	0.01169	0.00834	0.11714	0.21249	0.17771
5	0.00302	0.00913	0.00642	0.00059	0.00190	0.00137	0.11880	0.20633	0.16930
10	0.00175	0.00446	0.00310	0.00021	0.00050	0.00038	0.11223	0.19398	0.15201

15	0.00133	0.00291	0.00199	0.00014	0.00024	0.00019	0.10210	0.17977	0.13287
20	0.00112	0.00213	0.00144	0.00011	0.00015	0.00013	0.09283	0.16463	0.11403

Table 6. Return, Risk and Sharpe Ratio of Portfolio Using Clustering k-means (9 Cluster)

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	0.03208	0.06095	0.02281	0.06319	0.12325	0.04565	0.12702	0.17320	0.10608
1	0.01630	0.03015	0.01141	0.01586	0.03085	0.01146	0.12826	0.17084	0.10526
2	0.00841	0.01475	0.00571	0.00403	0.00775	0.00291	0.13021	0.16589	0.10319
5	0.00368	0.00551	0.00230	0.00072	0.00129	0.00052	0.13199	0.14969	0.09428
10	0.00210	0.00243	0.00116	0.00024	0.00036	0.00018	0.12545	0.12027	0.07564
15	0.00157	0.00141	0.00078	0.00016	0.00019	0.00012	0.11472	0.09131	0.05877
20	0.00131	0.00089	0.00059	0.00012	0.00013	0.00009	0.10451	0.06537	0.04573

Table 7. Return, Risk and Sharpe Ratio of Portfolio Using Clustering k-means (4 Cluster)

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	0.01347	0.01729	0.00884	0.02545	0.03561	0.01798	0.08355	0.09083	0.06484
1	0.00713	0.00840	0.00436	0.00644	0.00895	0.00454	0.08714	0.08725	0.06253
2	0.00396	0.00395	0.00212	0.00168	0.00228	0.00118	0.09318	0.07974	0.05740
5	0.00206	0.00129	0.00077	0.00035	0.00041	0.00024	0.10261	0.05608	0.04048
10	0.00143	0.00040	0.00033	0.00016	0.00015	0.00011	0.10174	0.02080	0.01756
15	0.00122	0.00010	0.00018	0.00012	0.00010	0.00004	0.09630	-0.00438	0.00369
20	0.00111	-0.00005	0.00010	0.00011	0.00008	0.00008	0.09148	-0.02127	-0.00472

Measuring portfolio performance not only be seen from the return, but also must pay attention to the risks that will be borne by investors. There are several measurements that can be used to measure portfolio performance, one of which is Sharpe ratio. Table 5, Table 6, and Table 7 show the portfolio return, risk and Share ratio of portfolio formed using clustering based on business sector of stock and k-mean combined with the classic MV portfolio model, MV_{FMCD} model and MV_S model. From Table 5, Table 6 and Table 7, in general it can be seen that the portfolio performance formed by combining clustering results based on the stock business sector with robust FMCD estimation outperforms other portfolio performance for all risk aversion values. These results are in line with the results of research conducted by Gubu et al. (2020).

CONCLUSION

This paper discusses how to group stocks into clusters based on business sector of stock and then uses robust estimates to obtain an optimum portfolio. This can reduce a lot of time in stock selection because stocks from the same sector can be easily grouped into one cluster. The best performing stocks from each cluster are then selected to represent the cluster to form a portfolio. To see the advantages of the proposed method, a portfolio formation is also carried out using k-means clustering. The results showed that the portfolio performance formed by combining clustering results based on the business sector of stock with FMCD robust estimation outperformed the portfolio performance with other combinations.

ACKNOWLEDGMENTS

The authors thank the Ministry of Education and Culture Republic of Indonesia for financial support via Doctoral Research Grant (Penelitian Disertasi Doktor, PDD) 2020. The

first oauthor also thank to Indonesia Endowment Fund for Education (Lembaga Pengelola Dana Pendidikan, LPDP) Ministry of Finance of the Republic of Indonesia for the scholarship funds given for his Doctoral Program at the Mathematics Department of Gadjah Mada University.

REFERENCES

- Best, M. J., & 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. <https://doi.org/10.1093/rfs/4.2.315>
- Ceria, S., & Stubbs, R. A. (2006). Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management*, 7(2), 109–127. https://doi.org/10.1007/978-3-319-30794-7_12
- Chen, L. H., & Huang, L. (2009). Portfolio optimization of equity mutual funds with fuzzy return rates and risks. *Expert Systems with Applications*, 36(2 PART 2), 3720–3727. <https://doi.org/10.1016/j.eswa.2008.02.027>
- Chopra, V. K., & Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2), 6–11. https://doi.org/10.1142/9789813144385_0002
- Davies, P. L. (1987). Asymptotic Behaviour of S-Estimates of Multivariate Location Parameters and Dispersion Matrices. *The Annals of Statistics*, 15(3), 1269–1292. <https://doi.org/10.1214/aos/1176350505>
- DeMiguel, V., & Nogales, F. J. (2009). Portfolio selection with robust estimation. *Operations Research*, 57(3), 560–577. <https://doi.org/10.1287/opre.1080.0566>
- Fabozzi, F. J., Kolm, P. N., Pachamanova, D. A., & Focardi, S. M. (2007). *Robust portfolio optimization*. John Wiley and Sons, Inc.
- Guan, H. S., & Jiang, Q. S. (2007). Cluster financial time series for portfolio. *Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR '07*, 2, 851–856. <https://doi.org/10.1109/ICWAPR.2007.4420788>
- Gubu, L., Rosadi, D., & Abdurakhman. (2020). Robust mean-variance portfolio selection with ward and complete linkage clustering algorithm. *Economic Computation and Economic Cybernetics Studies and Research*, 54(3), 111–127. https://doi.org/10.24818/18423264/54.3.20.07_117
- Hardin, J. S. (2000). *Multivariate outlier detection and robust clustering with minimum covariance determinant estimation and S-estimation*. Ph.D. Dissertation University of California.
- Lauprete, G. J. (2001). *Portfolio risk minimization under departures from normality*. Ph.D. Dissertation Massachusetts Institute of Technology.
- Long, N. C., Wisitpongphan, N., Meesad, P., & Unger, H. (2014). Clustering stock data for multi-objective portfolio optimization. *International Journal of Computational Intelligence and Applications*, 13(2), 1–13. <https://doi.org/10.1142/S1469026814500114>
- Lopuhaa, H. P. (1989). On the Relation between S-Estimators and M-Estimators of Multivariate Location and Covariance. *The Annals of Statistics*, 17(4), 1662–1683.

- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). Robust Statistics: Theory and Methods. In *Robust Statistics: Theory and Methods*. John Wiley & Sons, Ltd.
- Nanda, S. R., Mahanty, B., & Tiwari, M. K. (2010). Clustering indian stock market data for portfolio management. *Expert Systems with Applications*, 37(12), 8793–8798. <https://doi.org/10.1016/j.eswa.2010.06.026>
- Rifa, I. H., Pratiwi, H., & Respatiwan, R. (2020). Clustering of Earthquake Risk in Indonesia Using K-Medoids and K-Means Algorithms. *Media Statistika*, 13(2), 194–205. <https://doi.org/10.14710/medstat.13.2.194-205>
- Rousseeuw, P. J., & Van Driessen, K. (1999). Fast Algorithm For Minimum Covariance Determinant Estimator. *Technometrics*, 41(3), 212–223.
- Rousseeuw, P., & Yohai, V. (1984). Robust Regression By Means of S estimators. *Lecture Notes in Statistics: Robust and Nonlinear Time Series Analysis*, 26, 256–272.
- Sharpe, W. F. (1994). The Sharpe Ratio. *The Journal of Portfolio Management*, 21, 49–58.
- Supandi, E. D. (2017). *Developing of Mean-Variance Portfolio Modeling Using Robust Estimation and Robust Optimization Method*. Ph.D. Dissertation Gadjah Mada University.
- Tola, V., Lillo, F., Gallegati, M., & Mantegna, R. N. (2008). Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control*, 32(1), 235–258. <https://doi.org/10.1016/j.jedc.2007.01.034>
- Vaz-de Melo, B., & Camara, R. P. (2005). Robust multivariate modeling in finance. *International Journal of Managerial Finance*, 1(2), 95–107. <https://doi.org/10.1108/17439130510600811>
- Welsch, R., & Zhou, X. (2007). Application of robust statistics to asset allocation models. *REVSTAT–Statistical Journal*, 5(1), 97–114.
- Winston, W. L., & Goldberg, J. B. (2004). Operations research. In *Operations Research: Applications and Algorithms* (4th ed.). Thomson Learning Inc.
- Würtz, D., Chalabi, Y., Chen, W., & Ellis, A. (2009). *Portfolio Optimization with R / Rmetrics*. 419. <https://finance.e-bookshelf.ch/portfolio-optimization-with-r-rmetrics-2329769.html>