Time Series Clustering for Robust Mean-Variance Portfolio Selection: Comparison of Several Dissimilarity Measures

La Gubu¹,²*, Dedi Rosadi², and Abdurakhman²
¹ Department of Mathematics Gadjah Mada University Yogyakarta, Indonesia
² Department of Mathematics Halu Oleo University Kendari, Indonesia
*Email: lagubu2014@gmail.com

Abstract. This paper shows how to create a robust portfolio selection with time series clustering by using some dissimilarity measure. Based on such dissimilarity measures, stocks are initially sorted into multiple clusters using the Partitioning Around Medoids (PAM) time series clustering approach. Following clustering, a portfolio is constructed by selecting one stock from each cluster. Stocks having the greatest Sharpe ratio are selected from each cluster. The optimum portfolio is then constructed using the robust Fast Minimum Covariance Determinant (FMCD) and robust S MV portfolio model. When there are a big number of stocks accessible for the portfolio formation process, we can use this approach to quickly generate the optimum portfolio. This approach is also resistant to the presence of any outliers in the data. The Sharpe ratio was used to evaluate the performance of the portfolios that were created. The daily closing price of stocks listed on the Indonesia Stock Exchange, which are included in the LQ-45 indexed from August 2017 to July 2018, was utilized as a case study. Empirical study revealed that portfolios constructed using PAM time series clustering with autocorrelation dissimilarity and a robust FMCD MV portfolio model outperformed portfolios created using other approaches.

Keywords: Time series, clustering, robust

1. Introduction

In the analysis of economic and financial modeling, portfolio selection is an important and challenging issue. The primary goal of portfolio selection is to maximize return while minimizing risk, which can be expressed as an optimization problem. It was first performed by Markowitz using statistical measures derived from historical stock price data. The mean of the data is used to represent the stock return, while the variance is used to represent the risk, giving rise to the name Mean-Variance (MV) portfolio model [1].

Recently, there are strong interest to improve on the effectivity and efficiency of existing methods using several methods, one of which is clustering analysis. The optimal portfolio selection using cluster analysis has been carried out by many researchers [2-4]. 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 previous studies, the optimal portfolio still build using MV portfolio model. All of these studies reported that the usage of clustering analysis in the formation of the optimum portfolio is very efficient when the number of securities involved in the portfolio formation is large.

Despite the fact that clustering analysis can make models perform better, the MV portfolio model still have several flaws. The MV portfolio model's main flaw is that the mean vectors and variance-covariance matrix must be estimated from highly fluctuating data, sometimes even when the data differs significantly from the majority of the data (outliers). Because estimation errors are a vital input in the formulation of MV portfolio models [5-8], they will have a significant impact on the results of optimal
portfolio design. Furthermore, changes in the mean vector and variance-covariance matrix are extremely sensitive to the MV model.

Therefore, some studies have developed a robust portfolio model that can minimize the effects of outliers on mean vector and variance-covariance matrix estimation in MV portfolio models. One of the most common ways for generating an optimal robust portfolio is to employ a robust estimation strategy [9-11]. The type of robust estimations used in portfolio optimization are the main difference between them. If there are outliers in the data, all studies revealed that the performance of a portfolio with robust estimation outperforms a classical portfolio.

Based on those previous study, we use Partitioning Around Medoids (PAM) time series clustering with some dissimilarity measure, namely: Euclidean distance, Dynamic Time Warping (DTW) distance, correlation distance, and autocorrelation (AC) distance to group stocks which are included in the LQ-45 index based on daily closing price. The Sharpe ratio for all stocks is the calculated and ones with highest Sharpe ratio are chosen from each cluster [12]. Using a robust FMCD and robust S MV portfolio model, the representative stocks are then used to form an optimum robust portfolio.

2. Mathematical formulation

2.1. Mean-variance portfolio

The MV model for constructing an optimum portfolio is based on a trade-off between portfolio return and risk, which is represented by the mean and variance of stocks. [1]. Therefore, this model can be expressed as the optimization problem below [10]:

$$\max_w w^\prime \mu - \frac{1}{2} w^\prime \Sigma w$$

(1)

$$w^\prime e = 1$$

(2)

where $w$ denotes the portfolio's weight, $\mu$ denotes the mean vector, $\Sigma$ denotes the covariance matrix, $e$ denotes the column matrix with all entries equal to 1, and $\gamma \geq 0$ denotes the risk aversion parameters, i.e. the relative measure of risk avoidance.

The Lagrange method [13] can be used to solve the optimization problems in (1) and (2). First, construct the Lagrangian:

$$L = w^\prime \mu - \frac{1}{2} w^\prime \Sigma w + \lambda (w^\prime e - 1)$$

(3)

For (3) to be optimal, the following conditions must be satisfied:

$$\frac{\partial L}{\partial w} = 0 \quad \frac{\partial L}{\partial \lambda} = 0$$

(4)

From (3) and (4) we obtain

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

(5)

and

$$e^\prime w = 1$$

(6)

Substituting (5) in to (6) we obtain

$$\lambda = \gamma (e^\prime \Sigma^{-1} e)^{-1} - (e^\prime \Sigma^{-1} e)^{-1} e^\prime \Sigma^{-1} \mu$$

(7)

Substituting (7) in (5) resulting in:

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

(8)

2.2. Time series clustering

When we observe a collection of time series and try to organize them into distinct groups or clusters, we handle it by using time series clustering [14]. In this section we will give a brief overview of the time series clustering analysis that will be used in this paper.
2.2.1. Dissimilarity measure. The dissimilarity measure is the most important thing in conducting cluster analysis. There have been numerous proposed in the literature for time series dissimilarity measure. In this study the dissimilarity measure used are Euclidean distance, dynamic time warping (DTW) distance, correlation distance, and autocorrelation distance.

a) Euclidean distance
Let \( X_T = (X_1, ..., X_T)^T \) and \( Y_T = (Y_1, ..., Y_T)^T \) denote partial realizations from two real-valued processes \( X = \{X_t, \ t \in \mathbb{Z}\} \) and \( Y = \{Y_t, \ t \in \mathbb{Z}\} \), respectively. Euclidean distance between \( X_T \) and \( Y_T \) is defined by

\[
d_E(X_T, Y_T) = \left( \sum_{t=1}^{T} (X_t - Y_t)^2 \right)^{1/2}
\]

Signal transformations such as shifting or time scaling (stretching or shrinking of the time axis) are extremely sensitive to this metric [15].

b) Dynamic time warping distance
Dynamic Time Warping (DTW) distance is aimed to find a mapping \( r \) between the series so that a specific distance measure between the coupled observations \( (X_{a_i}, Y_{b_i}) \) is minimized [15]. The definition of the DTW distance is given by

\[
d_{DTW}(X_T, Y_T) = \min_{r \in M} \left( \sum_{i=1,...,m} |X_{a_i} - Y_{b_i}| \right)
\]

Even in the face of signal changes such as shifting and/or scaling, DTW distance allows for the recognition of similar forms. \( d_{DTW} \), like \( d_E \), ignore the temporal structure of the values because the proximity is based on the differences \( |X_{a_i} - Y_{b_i}| \) independent of how the values behave around them.

c) Correlation distance
The Pearson's correlation factor between \( X_T \) and \( Y_T \) is given by:

\[
COR(X_T, Y_T) = \frac{\sum_{t=1}^{T}(X_t \bar{X}_T)(Y_t \bar{Y}_T)}{\sqrt{\sum_{t=1}^{T}(X_t \bar{X}_T)^2 \sum_{t=1}^{T}(Y_t \bar{Y}_T)^2}}
\]

Where \( \bar{X}_T \) and \( \bar{Y}_T \) the average values of the serial realizations \( X_T \) and \( Y_T \) respectively.

d) Autocorrelation distance
Let \( \hat{\rho}_{X_T} = (\hat{\rho}_{1X_T}, ..., \hat{\rho}_{LX_T})^T \) and \( \hat{\rho}_{Y_T} = (\hat{\rho}_{1Y_T}, ..., \hat{\rho}_{LY_T})^T \) be the estimated autocorrelation function vectors of \( X_T = (\rho_{1X_T}, ..., \rho_{LX_T})^T \) and \( Y_T = (\rho_{1Y_T}, ..., \rho_{LY_T})^T \) respectively from lag-1 to lag-L such that \( \hat{\rho}_{1X_T} \approx 0 \) and \( \hat{\rho}_{1Y_T} \approx 0 \) for \( i > L \) [15] define distance between \( X_T \) and \( Y_T \) as follows.

\[
d_{ACF}(X_T, Y_T) = \sqrt{(\hat{\rho}_{X_T}, ..., \hat{\rho}_{Y_T})^T \Omega^{-1}(\hat{\rho}_{X_T}, ..., \hat{\rho}_{Y_T})}
\]

where \( \Omega \) is the weights matrix.

2.2.2. K-Medoids clustering. The clustering of k-medoids is a clustering approach for partitioning a data collection into k groups or clusters, comparable to the clustering of k-means [16]. In k-medoids clustering, one of the cluster data points represents each cluster. These points are called medoids of the cluster. The term medoid refers to an item within a cluster that has the least average dissimilarity from the other members of the cluster. Partitioning Around Medoids (PAM) is the most common clustering approach for k-medoids [16].

In more detail, [16] describes the steps of the PAM cluster analysis algorithm as follows:
1. Select k objects to be medoid, or if these objects are provided use them as medoid.
2. Calculate the dissimilarity matrix.
3. Assign each object to its nearest medoid;
4. For each cluster, find if one of the cluster objects decreases the average dissimilarity value; if yes, select the entity that reduces the most as the medoid for this cluster.
5. If at least one medoid has changed, continue to (3), otherwise, end the algorithm.

2.3. Robust estimation for portfolio selection
In this study, robust FMCD estimation and robust S estimation methods are used to estimate the weight of the selected stocks that make up the optimal portfolio. The approach for obtaining portfolio weights using the robust FMCD and robust S estimate methods is briefly described below.

2.3.1. Robust FMCD estimation. The minimum covariant determinant (MCD) estimation seeks to find robust estimates based on total observations (n), with the smallest of covariance matrix. The Minimum Covariant Determinant (MCD) estimation tries to find robust estimates with the smallest determinant covariance matrix based on total observations (n). From a sample of h observations, 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 $p \times p$, where $\frac{(n+p+1)}{2} \leq h \leq n$ with

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

(13)

The following equation is used to estimate the covariance matrix:

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

(14)

Because this method evaluates all possible subsets of h from n data, MCD computations become increasingly sophisticated as the data dimensions are larger. Therefore, [17] construct Fast MCD (FMCD) algorithm, an algorithm for quicker calculation of MCD. The C-Step theorem, which is discussed below, underpins the FMCD approach.

**Theorem 1** ([17])

If $H_1$ is a set of size h constructed from data of size n, then the sample statistics are as follows:

$$\hat{\mu}_1 = \frac{1}{h} \sum_{i \in H_1} r_i$$

(15)

$$\hat{\Sigma}_1 = \frac{1}{h} \sum_{i \in H_1} (r_i - \hat{\mu}_1)(r_i - \hat{\mu}_1)'$$

(16)

if $|\hat{\Sigma}_1| > 0$ then distance compute $d_i = (r_i; \hat{\mu}_1, \hat{\Sigma}_1)$. Next, specify $H_2$, which is a subset consist of the observation with the smallest distance $d_i$, namely $\{d_i(i)|i \in H_2\} = \{(d_1)_1, \ldots, (d_1)_h\}$ where $(d_1)_1 \leq (d_1)_2 \leq \cdots \leq (d_1)_h$ is a sequential distance. Based on $H_2$, using equations (15) and (16), we obtain

$$|\hat{\Sigma}_2| \leq |\hat{\Sigma}_1|$$

(17)

If $\hat{\mu}_2 = \hat{\mu}_1$ and $\hat{\Sigma}_2 = \hat{\Sigma}_1$, then expression (17) reach equality. Repeat C-Step theorem until $|\hat{\Sigma}_{new}| > 0$ or $|\hat{\Sigma}_{new}| = |\hat{\Sigma}_{old}|$.

2.3.2. Robust S estimation. [18] was the first who introduce this estimation, which was then expanded upon by [19].

**Definition 1** ([19])

Given $\{r_i, i = 1, \ldots, n\}$ 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

$$\frac{1}{n} \sum_{i=1}^{n} \rho[(r_i - \mu)\Sigma^{-1}(r_i - \mu)]^{1/2} = b_0$$

(18)

where $\rho$ denotes the loss function and $b_0$ denotes the constant. This constant must be determined precisely because this value affects the result of estimation. By solving the following equation, we can get the S estimator.
\[ \frac{1}{n} \sum_{i=1}^{n} u(d_i)(r_i - \mu) = 0 \]  
\[ \frac{1}{n} \sum_{i=1}^{n} p u(d_i)(r_i - \mu)(r_i - \mu)' - v(d_i) \Sigma = 0 \]

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

The S estimate is calculated iteratively using equations (19) and (20). The S estimate algorithm is as follows [2]:

1. Calculate the initial estimates of the mean vector and covariance matrix, \( \hat{\mu}_0 \) and \( \hat{\Sigma}_0 \) respectively.
2. Determine \( d_i = (r_i - \hat{\mu}_0)'\hat{\Sigma}_0^{-1}(r_i - \hat{\mu}_0) \)
3. Determine \( k_0 \) such that \( \frac{\sum p(d_{ij}/k_0)}{n} = b_0 \)
4. Calculate \( \tilde{d}_i = \frac{d_i}{k_0} \)
5. Determine \( \hat{\mu} = \frac{\sum \psi(\tilde{d}_i)r_i}{\sum \psi(\tilde{d}_i)} \) and \( \hat{\Sigma} = \frac{p \sum \psi(\tilde{d}_i)(r_i - \mu)(r_i - \mu)'}{\sum \psi(\tilde{d}_i)} \)
6. Steps 2–5 should be repeated until the convergent is achieved.

3. Result and discussions

3.1. Stocks clustering

The daily stock prices of all stocks included in the LQ-45 index from August 2017 to July 2018 which is traded on the Indonesia Stock Exchange (see https://finance.yahoo.com), were used in this study. We used PAM time series clustering with four dissimilarity which are: 1) Euclidean distance, 2) Dynamic Time Warping (DTW) distance, correlation distance, and autocorrelation (AC) distance. Utilizing TSclust function in R package, we discovered that using clustering PAM time series clustering with those dissimilarity measures, LQ-45 stocks can be clustered into six clusters, as shown in Tables 1, 2, 3, and 4.

**Table 1.** The cluster of using PAM time series clustering with Euclidean distance dissimilarity

| Cluster | Stocks |
|---------|--------|
| 1       | AALI   |
| 2       | ADHI   |
| 3       | ANTM   |
| 4       | BBCA   |
| 5       | GGRM   |
| 6       | UNVR   |

**Table 2.** The Cluster of using PAM time series clustering with DTW distance dissimilarity

| Cluster | Stocks |
|---------|--------|
| 1       | AALI   |
| 2       | ADHI   |
| 3       | ANTM   |
| 4       | BBCA   |
| 5       | GGRM   |
| 6       | UNVR   |
Table 3. The Cluster of using PAM time series clustering with correlation distance dissimilarity

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

Table 4. The Cluster of using PAM time series clustering with AC dissimilarity

| Cluster | Stocks |
|---------|--------|
| 1       | AALI   | BBTN   | GGRM   | HMSP   | INTP   | LSIP   | PWON   | SCMA   | SMGR   | SRIL   |
|         | UNVR   |        |        |        |        |        |        |        |        |        |
| 2       | ADHI   | BUMI   | ICBP   | SMRA   | UNTR   |        |        |        |        |        |
| 3       | ADRO   | BMTR   | BRPT   | INCO   | LPPF   | MNCN   | MYRX   | PGAS   | WSKT   |        |
| 4       | AKRA   | EXCL   | JSMR   | LPKR   |        |        |        |        |        |        |
| 5       | ANTM   | ASII   | BBCA   | BNBI   | BBRI   | BJBR   | BMRI   | INDF   | KLBF   | PPRO   |
| 6       | ICBP   | JSRR   | SRIL   |        |        |        |        |        |        |        |

3.2. Stock representation of cluster

After the clusters have been constructed, the Sharpe ratio for each stock in each cluster is calculated. The Sharpe ratio is calculated using the most recent Bank Indonesia rate of 5.25 percent per year. To establish the best portfolio, we selected stocks that represent each cluster based on the calculations of each stock's Sharpe ratio in each cluster, as shown in Table 5.

Table 5. Stocks representation of clusters

| Dissimilarity | Clusters Representation |
|---------------|-------------------------|
| Euclidean     | ICBP | INCO | ANTM | BBCA | GGRM | UNVR |
| DTW           | ICBP | INCO | ANTM | BBCA | GGRM | UNVR |
| Correlation   | BJBR | INTP | ADRO | INCO | UNTR | ICBP |
| AC            | HMSP | UNTR | INCO | EXCL | PTBA | PTPP |

3.3. Portfolios weight and comparison of portfolios performance

Robust FMCD MV portfolio model ($MV_{FMCD}$) and robust S MV portfolio model ($MV_{S}$) were used to establish the optimum portfolio. The first step is using CovMed and CovSest function in R packages to calculate the portfolio weights of the models for varying risk aversion values $\gamma$. Tables 6, 7, and 8 show the portfolio weights derived using PAM clustering with four levels of dissimilarity.
3.4. Discussions

As shown in Table 5, we use the Sharpe ratio for each stock in each cluster to establish cluster representations for four dissimilarities. The clusters representation of clustering with dissimilarity
Euclidean distance and DTW distance are identical, as shown in Table 5. Cluster 1 consists of ten stocks based on these dissimilarities (Euclidean/DTW), and ICBP has the best performance among them, as seen by the highest Sharpe ratio in cluster 1, which is 0.0052. Furthermore, INCO with a Sharpe ratio of 0.0888 represents cluster 2, whereas ANTM, BBCA, GGRM, and UNVR are representatives of clusters 3, 4, 5, and 6.

According to Table 3, BJBR has the best performance in cluster 1 when using dissimilarity correlation distance, as seen by the highest Sharpe ratio in cluster 1, which is 0.0009. As a result, BJBR was chosen as the representation for Cluster 1. Furthermore, INTP, which had a Sharpe ratio of -0.0225, represented cluster 2, and ADRO, INCO, UNTR, and ICBP, respectively, represented clusters 3, 4, 5, and 6.

Finally, utilizing dissimilarity based on AC distance, it was discovered that HMSP, UNTR, INCO, EXCL, PTBA, and PTPP are representations of clusters 1, 2, 3, 4, 5, and 6, respectively.

Stocks having a small Sharpe ratio or negative returns, such as ICBP and UNVR in Table 6 (portfolio weight using clustering with Euclidean/DWT dissimilarity), have a negative weight (short selling) for all risk aversion values. Stocks with high returns, such as ANTM and BBCA, on the other hand, always have positive weights in three portfolio models. Table 6 further shows that when the value of γ grows, the weight of a stock with a positive return will drop, and vice versa for stocks with negative returns. As demonstrated in Tables 7 and 8, the same thing happened when portfolios were built using stock representations derived using correlation and AC distance dissimilarity.

Table 9 display the returns, risks, and Sharpe ratio of portfolios formed using PAM clustering with four dissimilarity in combination with robust FMCD MV portfolio model and robust S MV portfolio model. From Table 9 it can be seen that the performance of the portfolios produced by PAM clustering with AC dissimilarity combined with robust FMCD MV portfolio model are outperforming those produced by others methods.

4. Conclusions
Portfolio performance must be measured not only in terms of return, but also in terms of the risks that the investor will be borne. The Sharpe ratio is one of several indicators that can be used to assess portfolio performance. The results of this study showed that the performance of portfolio generated by using PAM time series clustering with autocorrelation dissimilarity combined with robust FMCD MV portfolio model outperformed those created by other methods.

Acknowledgment
We acknowledge the financial support from Directorate General of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology Republic of Indonesia via Doctoral Research Grant (Penelitian Disertasi Doktor, PDD) 2021.

References
[1] Markowitz H M 1952 Journal of Finance 7 77–91
[2] Tola V, Lillo F, Gallegati M and Mantegna R N 2008 J. Econ. Dyn. Control 32 235–258
[3] Chen L H and Huang L 2009 Expert Systems with Applications 36 3720–3727
[4] Nanda R, Mahanty B and Tiwari M K 2010 Expert Systems with Applications 8793–8798
[5] Best M J and Grauer R R 1991 Review of Financial Studies 4(2) 315–342
[6] Broadie M 1993 Annals of Operations Research 45 21–58
[7] Chopra V K and Ziembba W T 1993 Journal of Portfolio Management 19(2) 6–11
[8] Ceria S and Stubbs R A 2006 Journal of Asset Management 7(2) 109–127
[9] DeMiguel V and Nogales F J 2008 Journal of Operation Research 57 560–577
[10] Supandi E D 2017 Developing of mean-variance portfolio modeling using robust estimatin and robust optimization method PhD Dissertation Gadjah Mada University
[11] Gubu L, Rosadi D and Abdurakhman 2020 Economic Computation and Economic Cybernetics Studies and Research 54(3) 111–127
[12] Sharpe W F 1994 The Journal of Portfolio Management 21 49–58
[13] Winston W L and Goldberg J B 2004 *Operations Research: Applications and Algorithms* (Toronto: Thomson Learning Inc.)

[14] Alonso A M, Barrendero J R, Hernandez A and Justel 2006 *A Computational Statistics & Data Analysis* **51** 762–776

[15] Montero P and Vilar J A 2014 *Journal of Statistical Software* **62**(1) 1–43

[16] Kaufman L and Rousseeuw P J 1990 *Finding Groups in Data: An Introduction* (New Jersey: John Wiley & Sons)

[17] Rousseeuw P J and Van Driessen K 1999 *Technometrics* **41** 212–223

[18] Rousseeuw P J and V. Yohai V 1984 *Lecture Notes in Statistics* **26** 256–272

[19] Davies P 1987 *The Annals of Statistics* **15** 1269–1292

[20] Hardin J S 200 Multivariate Outlier Detection and Robust Clustering with Minimum Covariance Determinant Estimation and S-Estimation PhD Dissertation University of California