PEMBENTUKAN PORTOFOLIO SAHAM MENGGUNAKAN KLASTERING TIME SERIES K-MEDOID DENGAN UKURAN JARAK DYNAMIC TIME WARPING

La Gubu[1], Dedi Rosadi[2], and Abdurakhman[3]

[1,2,3]Departemen Matematika Universitas Gadjah Mada, Yogyakarta-Indonesia
[1]Jurusan Matematika Universitas Halu Oleo, Kendari-Indonesia
e-mail: [1]lagubu2014@gmail.com, [2]dedirosadi@gadjahmada.edu, [3]rachmanstat@ugm.ac.id

Abstrak

Pada penelitian ini akan disajikan pembentukan portofolio saham dengan *preprocessing* data menggunakan klastering *time series* dengan ukuran jarak *Dynamic Time Warping* (DTW). Pertama-tama saham-saham dikelompokkan ke dalam beberapa klaster menggunakan klastering *time series Partitioning Around Medoids* (PAM) berdasarkan ukuran jarak DTW. Setelah proses klastering, saham dipilih untuk mewakili masing-masing klaster untuk membangun portofolio optimum. Saham yang dipilih dari masing-masing klaster merupakan saham yang memiliki *Sharpe ratio* tertinggi. Portofolio optimal ditentukan dengan menggunakan tiga model portofolio, yaitu: model portofolio *Mean-Variance* (MV) klasik, model portofolio MV *robust Fast Minimum Covariance Determinant* (FMCD) dan model portofolio *robust Scale* (S). Dengan menggunakan prosedur ini, dapat diperoleh portofolio optimum secara efisien bila ada banyak saham yang terlibat dalam proses pembentukan portofolio. Untuk mengukur kinerja portofolio yang terbentuk digunakan *Sharpe ratio*. Hasil kajian empiris menunjukkan bahwa kinerja portofolio yang dihasilkan dengan menggunakan klastering *time series* PAM dengan ukuran disimilaritas jarak DTW yang dikombinasikan dengan model portofolio MV klasik mengungguli kinerja portofolio yang dihasilkan kombinasi dengan model yang lain.

Kata kunci: klastering, *k-medoids*, portofolio saham, estimasi *robust*, kinerja portofolio

Abstract

This research will present the formation of stock portfolios by preprocessing data using time-series clustering with a distance measure of Dynamic Time Warping (DTW). First, stocks are grouped into several clusters using the Partitioning Around Medoids (PAM) time series cluster based on the DTW distance measure. After the clustering process, stocks are selected to represent each cluster to build the optimum portfolio. The stock selected from each cluster is the one with the highest Sharpe ratio. The optimal portfolio is determined using three portfolio models, namely: the classic Mean-Variance (MV) portfolio model, the FMCD robust MV portfolio model, and the S robust portfolio model. Using this procedure, an optimum portfolio can be obtained efficiently if there are many stocks involved in the portfolio formation process. Sharpe ratio is used to measure the performance of the portfolios. The results of the empirical study show that the portfolio performance generated using the PAM time series clustering with DTW distance dissimilarity measure combined with the classic MV portfolio model outperforms the resulting portfolio performance in combination with other models.

Keywords*: clustering, k-medoid, stock portfolio, robust estimation, portfolio performance*

PENDAHULUAN

Hakikat pembentukan portofolio adalah mengalokasikan modal pada berbagai alternatif investasi, sehingga risiko investasi dapat diminimumkan (Elton, dkk., 2014). Dalam memilih portofolio, *investor* berusaha memaksimalkan keuntungan dari investasi dengan tingkat risiko tertentu yang diterima. Atau dalam pengertian lain, *investor* berusaha meminimalkan risiko yang dihadapi pada tingkat keuntungan tertentu. Portofolio yang dapat mencapai tujuan tersebut disebut portofolio efisien (Reilly dan Brown, 2012). Apabila seorang *investor* memiliki beberapa pilihan portofolio yang efisien, maka portofolio optimal yang akan dipilihnya. Portofolio yang optimal adalah portofolio yang dapat memaksimalkan preferensi *investor* sehubungan dengan tingkat keuntungan dan risiko.

Pembentukan portofolio merupakan masalah yang menarik dan menantang dalam studi pemodelan ekonomi dan keuangan. Masalah utama dalam pembentukan portofolio adalah memaksimalkan return sekaligus meminimalkan risiko, sehingga dapat dirumuskan menjadi masalah optimasi (Supandi, 2017). Upaya pertama dalam memodelkan masalah optimasi ini telah dilakukan oleh Markowitz pada tahun 1952. Markowitz (1952) menggunakan ukuran-ukuran statistik dari data historis harga saham untuk membangun portofolio optimal. Dalam model ini, return saham direpresentasikan sebagai mean dari data sedangkan risiko direpresentasikan sebagai variansi, sehingga model ini disebut model portofolio *Mean-Variance* (MV).

Studi untuk mengembangkan dan memperluas model MV lebih lanjut telah dilakukan oleh banyak peneliti. Pengembangan dan perluasan model tersebut dilakukan untuk menyesuaikan model agar lebih adaptif dan *robust* terhadap kondisi pasar keuangan yang terus berubah, serta membangun cara yang lebih efektif dan efisien untuk membangun portofolio yang optimal. Salah satu cara yang dilakukan adalah dengan metode klastering. Penelitian tentang pembentukan portofolio dengan menggunakan klastering telah dilakukan oleh beberapa peneliti, antara lain oleh Guan dan Jiang (2007), Tola dkk (2008), Chen dan Huang (2009), Nanda dkk (2010), Long dkk (2014), dan Gubu, dkk (2019). Dalam pendekatan ini, saham dipisahkan menjadi beberapa kelompok (klaster), kemudian dipilih perwakilan dari masing-masing kelompok untuk membentuk portofolio optimal dengan menggunakan model portofolio MV.

Masalah utama model portofolio MV adalah bahwa vektor mean dan matriks variansi-kovariansi harus diestimasi dari data yang dapat sangat volatil. Estimasi parameter dapat dilakukan dengan menggunakan berbagai pilihan teknik estimasi, yang tentunya akan mengandung kesalahan estimasi. Sebagai input yang sangat penting dalam pembentukan model portofolio MV, kesalahan estimasi akan berpengaruh signifikan terhadap hasil pembentukan portofolio optimal. Beberapa penelitian terkait dengan kesalahan estimasi dan hubungannya dengan pembentukan portofolio optimal telah dilakukan oleh Best dan Grauer (1991), Chopra dan Ziemba (1993), dan Ceria dan Stubbs (2006). Studi-studi tersebut menyimpulkan bahwa meskipun model MV didukung oleh teori yang kuat dan relatif mudah dalam komputasi, model MV menunjukkan beberapa kelemahan, diantaranya portofolio optimal yang dihasilkan oleh model ini belum terdiversifikasi dengan baik. Selain itu, model MV juga sangat sensitif terhadap perubahan parameter input yaitu vektor mean dan matriks variansi-kovariansi.

Oleh karena itu beberapa peneliti telah membangun suatu portofolio *robust* (Demiguel dan Nogales, 2009), yaitu portofolio yang dapat mengurangi kesalahan estimasi vektor mean dan matriks variansi-kovariansi pada model portofolio MV. Salah satu pendekatan standar dalam membentuk portofolio *robust* optimal adalah melalui pendekatan estimasi *robust*. Beberapa studi tentang pembentukan portofolio optimal menggunakan estimasi *robust* telah dilakukan oleh Lauprete (2001), Vaz-de Melo dan Camara (2005), Welsh dan Zhou (2007), DeMiguel dan Nogales (2008), dan Supandi (2017). Perbedaan utama antara penelitian-penelitian ini terletak pada estimasi *robust* yang digunakan dalam pembentukan portofolio optimal.

Berdasarkan penelitian-penelitian sebelumnya, disimpulkan bahwa, untuk meningkatkan efisiensi proses pembentukan portofolio, perlu mengelompokkan saham-saham ke dalam klaster-klaster, kemudian memilih saham representasi dari masing-masing klaster dan menentukan bobot masing-masing saham representasi.

Pada penelitian ini, saham-saham dikelompokkan ke dalam klaster-klaster menggunakan klastering *time series Partitioning Around Medoids* (PAM) dengan ukuran jarak *Dynamic Time Warping* (DTW).

Pemilihan PAM sebagai analisis klaster yang digunakan dengan pertimbangan bahwa analisis klaster PAM lebih *robust* dibandingkan dengan analisis klaster yang lain seperti *k-mean*. Pada analisis klaster PAM, pusat klaster adalah *medoid* dari data dalam suatu klaster, sehingga jika ada *outlier* dalam data maka analisis klaster PAM akan lebih *robust* dibandingkan dengan analisis klaster *k-mean*.

Ukuran jarak DTW juga disebut sebagai *non-linear sequence alignment*, sehingga ukuran jarak ini lebih realistis untuk digunakan dalam mengukur disimilaritas suatu pola dibandingkan menggunakan ukuran jarak yang lain seperti jarak Euclidean, Manhattan, Minkowski dan lain-lain (Berndt dan Clifford, 1994). Data yang diolah selalu berada dalam kawasan waktu, sehingga rentetan data yang kita miliki dianggap bervariasi terhadap waktu. Atas dasar ini maka ukuran jarak DTW dipilih untuk digunakan dalam penelitian ini.

Bobot portofolio ditentukan dengan menggunakan tiga model portofolio, yaitu model portofolio MV klasik, model portofolio MV *robust* FMCD dan model portofoli MV *robust* S. Untuk mengukur kinerja portofolio yang terbentuk dengan menggunakan tiga model portofolio digunakana *Sharpe ratio*.

METODOLOGI

Model Portofolio *Mean-Variance*

Model portofolio *Mean-Variance* (MV) untuk membentuk portofolio optimal didasarkan pada *trade-off* antara return dan risiko portofolio yang direpresentasikan sebagai mean dan variansi saham (Markowitz, 1952). Oleh karena itu, model ini dapat dirumuskan sebagai masalah optimasi berikut (Supandi, 2017):

$$\max_{\boldsymbol{w}} \boldsymbol{w}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \qquad (1)$$

$$\boldsymbol{w}'\boldsymbol{e} = 1 \qquad (2)$$

dengan $\boldsymbol{w}$ menyatakan bobot portofolio, $\boldsymbol{\mu}$ adalah vektor mean, $\boldsymbol{\Sigma}$ adalah matriks variansi-kovariansi, $\boldsymbol{e}$ adalah matriks kolom dengan semua elemennya adalah 1 dan $\gamma \geq 0$ adalah parameter penghindaran risiko (*risk aversion*), yaitu ukuran relatif penghindaran risiko. Model portofolio MV dipilih oleh *investor* dengan menggunakan kriteria fungsi utilitas. Dengan asumsi bahwa sikap *investor* adalah *risk averse*, *investor* akan memilih portofolio yang memberikan utilitas maksimum yang diharapkan $E(U)$. Jadi, portofolio *mean-variance* setara dengan

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

$$\boldsymbol{w}'\boldsymbol{e} = 1 \qquad (4)$$

Diberikan modal awal W_0, di bawah portofolio dengan vektor bobot $\boldsymbol{w}$, pada akhir periode, modal menjadi $W_0(1+R_p)$, di mana R_p adalah variabel random return portofolio (Supandi, 2017). Fungsi utilitas $W_0(1+R_p)$ adalah $U\big(W_0(1+R_p)\big)$. Selanjutnya, Supandi (2017) menyatakan bahwa fungsi utilitas dapat diperluas dengan menggunakan pendekatan deret Taylor orde dua, yaitu

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

Mengambil nilai harapan persamaan (5) diperoleh

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

Berdasarkan pendekatan tersebut, memaksimalkan fungsi utilitas yang diharapkan sama dengan memaksimalkan

$$A = \mu_p - \frac{1}{2}\gamma\sigma_p^2 \quad (7)$$

dengan $\gamma = -W_0 \frac{U''(W_0)}{U'(W_0)}$ menyatakan ukuran relatif penghindaran risiko. Telah diketahui bahwa $\mu_p = \boldsymbol{w}'\boldsymbol{\mu}$ dan $\sigma_p^2 = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}$, sehingga persamaan (7) dapat ditulis dalam bentuk

$$B = \boldsymbol{w}'\boldsymbol{\mu} - \frac{\gamma}{2}\boldsymbol{w}'\Sigma\boldsymbol{w} \quad (8)$$

Sehingga persamaan (1) ekuivalen dengan persamaan (8) ditambah kendala $\boldsymbol{w}'\boldsymbol{e} = 1$.

Permasalahan optimasi pada persamaan (1) dan (2) dapat diselesaikan dengan menggunakan metode Lagrange. Pertama, bentuk fungsi Lagrange:

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

Berdasarkan teorema Kuhn-Tucker (Winston dan Goldberg, 2004), syarat perlu persamaan (9) mencapai nilai optimum adalah:

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{w}} = 0 \quad (10)$$

$$\frac{\partial \boldsymbol{L}}{\partial \lambda} = 0 \quad (11)$$

Dari persamaan (9), (10) dan (11) diperoleh

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

dan

$$\boldsymbol{e}'\boldsymbol{w} = 1 \quad (13)$$

Substitusi persamaan (12) ke persamaan (13) diperoleh

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

Substitusi persamaan (14) pada persamaan (12) menghasilkan:

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

Persamaan (15) menunjukkan bahwa bobot portofolio ($\boldsymbol{w}$) bergantung pada input vektor mean $\boldsymbol{\mu}$ dan matiks variansi-kovariansi $\boldsymbol{\Sigma}$.

Klastering *Time Series*

Masalah klastering *time series* terjadi ketika kita mengamati himpunan *time series* dan kita ingin mengaturnya ke dalam kelompok atau kluster terpisah (Alonso, 2006). Pada bagian ini akan disajikan secara singkat tentang hal-hal yang terkait dengan analisis klaster *time series* yang akan digunakan dalam penelitian ini.

Ukuran Disimilaritas

Ukuran disimilaritas *(dissimilarity measure)* merupakan hal terpenting dalam melakukan analisis klaster. Banyak ukuran disimilaritas antara *time series* telah diajukan dalam literatur. Dalam penelitian ini ukuran disimilaritas yang digunakan adalah jarak *Dynamic Time Warping* (DTW).

Jarak DTW telah dipelajari secara mendalam oleh Sankoff dan Kruskal (1983) dan digunakan untuk menemukan pola dalam deret waktu oleh Berndt dan Clifford (1994). Jarak DTW bertujuan untuk menemukan pemetaan r antara deret waktu tersebut sehingga ukuran jarak antara pengamatan berpasangan $\left(X_{a_i}, Y_{b_i}\right)$ adalah minimum. Montero dan Vilar (2014) memberikan definisi jarak DTW sebagai berikut.

$$d_{DTW}(\boldsymbol{X}_T, \boldsymbol{Y}_T) = \min_{r \epsilon M}\left(\sum_{i=1,\dots,m}\left|X_{a_i} - Y_{b_i}\right|\right) \quad (16)$$

Lebih lanjut, Montero dan Vilar (2014) menyatakan bahwa jarak DTW memungkinkan untuk mengenali bentuk yang serupa, bahkan dengan adanya transformasi sinyal seperti pergeseran dan/atau penskalaan.

k-Medoids Clustering

Klastering *k-medoids* adalah teknik klastering yang mirip dengan teknik klastering *k-means* untuk mempartisi kumpulan data *(data set)* menjadi *k* kelompok atau klaster. Dalam pengelompokan *k-medoids*, setiap klaster diwakili oleh salah satu titik data dalam klaster. Titik-titik ini disebut *medoids* dari klaster. Istilah *medoid* mengacu pada entitas dalam sebuah klaster yang memiliki perbedaan rata-rata minimal antara titik tersebut dan semua anggota klaster lainnya. Sebagai catatan, dalam klastering k-means, pusat kluster ditentukan sebagai nilai rata-rata dari semua titik data kluster. *K-medoids* adalah alternatif yang *robust* untuk klastering *k*-means. Ini berarti bahwa klastering *k-medoids* kurang sensitif terhadap pencilan dibandingkan dengan klastering *k-means*, karena metode klastering tersebut menggunakan *medoid* sebagai pusat cluster. *Partitioning Around Medoids* (PAM) adalah pendekatan klastering yang paling umum untuk *k-medoids* (Kaufman dan Rousseeuw, 1990).

Algoritma PAM didasarkan pada pencarian *k* objek perwakilan *(representative)* atau *medoid* di antara pengamatan kumpulan data *(data set)*. Setelah menemukan satu set *k medoid*, klaster dibangun dengan menetapkan setiap observasi ke *medoid* terdekat. Selanjutnya, setiap *medoid* yang dipilih dan setiap titik data *non-medoid* ditukar *(swapped)* dan fungsi tujuan dihitung. Fungsi objektif adalah jumlah disimilaritas semua objek ke *medoid* terdekatnya.

Langkah pertukaran *(swap)* mencoba meningkatkan kualitas pengelompokan dengan menukar objek yang dipilih sebagai *medoid* dan objek *non-medoid*. Jika nilai fungsi objektif dapat dikurangi dengan menukar objek yang dipilih sebagai *medoid* dengan objek *non-medoid*, maka pertukaran dilakukan. Hal ini terus dilakukan hingga fungsi objektif tidak dapat diturunkan lagi. Tujuannya adalah untuk menemukan objek perwakilan yang meminimalkan jumlah disimilaritas pengamatan ke objek perwakilan terdekat mereka.

Secara terperinci, Kaufman dan Rousseeuw (1990) menguraikan langkah-langkah algoritma analisis klaster PAM sebagai berikut:

1. Pilih *k* objek untuk menjadi *medoid*;
2. Hitung matriks disimilaritas menggunakan persamaan (16);
3. Tetapkan setiap objek ke *medoid* terdekatnya;
4. Untuk setiap pencarian klaster jika salah satu objek klaster menurunkan koefisien ketidaksamaan rata-rata, pilih entitas yang paling banyak menurunkan koefisien ini sebagai *medoid* untuk klaster ini;
5. Jika setidaknya satu *medoid* telah berubah, lanjutkan ke (3), jika tidak, akhiri algoritma.

Sharpe Ratio

Setelah klaster-klaster terbentuk, selanjutnya dilakukan penilaian terhadap kinerja masing-masing saham dalam setiap klaster dengan menggunakan Sharpe *ratio* (SR). Sharpe *ratio* atau indeks Sharpe adalah ukuran dari *excess return* (atau *risk premium*) per unit risiko dalam aset (Sharpe, 1994). Sharpe *ratio* digunakan untuk mengkarakterisasi seberapa baik return aset mengkompensasi *investor* untuk risiko yang diambil. Lebih lanjut Sharpe (1994) menyatakan bahwa SR dihitung dengan membandingkan selisih antara return saham (R) dan *return risk-free rate* (R_f) dengan standar deviasi return saham (σ) atau dapat ditulis sebagai berikut:

$$SR = \frac{R-R_f}{\sigma} \tag{17}$$

Secara umum dapat dikatakan bahwa semakin besar nilai Sharpe *ratio* suatu saham maka kinerja saham semakin baik. Sharpe *ratio* juga dapat digunakan untuk mengukur kinerja portofolio. Jika Sharpe *ratio* digunakan untuk mengukur kinerja portofolio maka return dan risiko yang digunakan adalah return dan risiko portofolio.

Pemilihan Portofolio Optimum dengan Menggunakan Estimasi *Robust*

Ada dua terminologi penting dalam pembentukan portofolio, yaitu portofolio efisien dan portofolio optimum. Portofolio efisien adalah portofolio dengan return tertinggi pada risiko tertentu, atau portofolio dengan risiko terendah pada return tertentu (Elton, dkk., 2014). Investor perlu mempertimbangkan dan menentukan sekuritas apa saja yang membentuk portofolio dan dapat mencapai efisiensi maksimal. Indikator portofolio efisien adalah:

1. Mampu memberikan *expected* return terbesar dengan risiko yang sama. \
2. Mampu memberi risiko terkecil dengan *expected* return yang sama.

Penentuan portofolio efisien dilakukan dengan cara memilih tingkat *expected* return tertentu dan meminimumkan risikonya, atau menentukan tingkat risiko tertentu dan kemudian memaksimumkan *expected* returnnya. Portofolio efisien merupakan portofolio yang baik, tetapi bukan yang terbaik. Portofolio yang terbaik adalah portofolio yang optimal. Portofolio efisien hanya mempunyai satu dari faktor terbaik, yaitu faktor *expected* return atau faktor risikonya. Sementara, portofolio yang optimal adalah portofolio yang memiliki kombinasi *expected* return dan risiko yang terbaik (Elton, dkk., 2014).

Portofolio MV klasik menjadi tidak efektif ketika dihadapkan pada kondisi data return yang tidak memenuhi asumsi distribusi normal multivariat, karena estimasi klasik dari mean dan variansi tidak *robust* dan sangat dipengaruhi oleh pengamatan-pengamatan yang menyimpang (*outlier*), (Maronna dkk, 2006). Statistika *robust* berkaitan dengan membangun prosedur statistik yang stabil ketika ada bagian dari data yang tidak sesuai dengan distribusi yang diasumsikan atau terdapat penyimpangan dari model.

Pada penelitian ini, vektor mean dan matriks variansi-kovariansi diestimasi dengan menggunakan metode estimasi *robust Fast Minimum Covariance Determinant* (FMCD) dan metode estimasi *robust Scale* (S). Berikut akan disajikan secara sekilas prosedur metode estimasi *robust* FMCD dan metode estimasi *robust* S.

Metode Estimasi *Robust FMCD*

Estimasi *Minimum Covariance Determinant* (MCD) bertujuan untuk mencari estimasi *robust* berdasarkan h pengamatan dari total pengamatan *(n)*, dimana matriks kovariansi mempunyai determinan paling kecil. Estimasi MCD merupakan pasangan $\hat{\boldsymbol{\mu}} \in \mathbb{R}^p$ dan $\hat{\boldsymbol{\Sigma}}$ adalah matriks definit positif simetris berdimensi $p \times p$ dari suatu sub ruang sampel berukuran h pengamatan, di mana $\frac{(n+p+1)}{2} \le h \le n$ dengan

$$\hat{\boldsymbol{\mu}} = \frac{1}{h}\sum_{i=1}^{h} \boldsymbol{r_i} \qquad (18)$$

dengan $\boldsymbol{r_i}$ return saham ke-i, $i = 1, \dots, h$.

Estimasi matriks kovariansi dapat diperoleh dengan menyelesaikan persamaan berikut ini:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{h}\sum_{i=1}^{h} (\boldsymbol{r}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{r}_i - \hat{\boldsymbol{\mu}})' \qquad (19)$$

Perhitungan MCD bisa menjadi sangat rumit apabila dimensi data semakin besar, hal ini dikarenakan pada metode ini harus memeriksa semua kemungkinan himpunan bagian *h* dari sejumah *n* data. Oleh karena itu Roesseeuw dan Van Driessen (1999) menemukan algoritma perhitungan yang lebih cepat untuk menghitung MCD yang disebut *Fast* MCD (FMCD). Metode FMCD berdasarkan teorema C-Step yang dijelaskan berikut ini.

Teorema 1 (Rousseeuw dan Driessen,1999) *Jika H_1 adalah himpunan bagian berukuran h yang diambil dari data berukuran n, maka statistik sampel adalah:*

$$\hat{\boldsymbol{\mu}}^1 = \frac{1}{h}\sum_{i \in H_1} \boldsymbol{r}_i \qquad (20)$$

$$\hat{\boldsymbol{\Sigma}}^1 = \frac{1}{h}\sum_{i \in H_1} (\boldsymbol{r}_i - \hat{\boldsymbol{\mu}}^1)(\boldsymbol{r}_i - \hat{\boldsymbol{\mu}}^1)' \qquad (21)$$

Jika $\left|\hat{\boldsymbol{\Sigma}}^1\right| > 0$ *maka jarak* $d_i = \left(\boldsymbol{r}_i; \hat{\boldsymbol{\mu}}^1, \hat{\boldsymbol{\Sigma}}^1\right)$. *Selanjutnya tentukan H_2 himpunan bagian yang terdiri dari dari h pengamatan dengan jarak terkecil d_i, yaitu* $\{d_1(i) | i \in H_2\} = \{(d_1)_1, \dots, (d_1)_h\}$ *dengan* $(d_1)_1 \le (d_1)_2 \le \dots \le (d_1)_n$ *merupakan jarak yang berurutan. Berdasarkan H_2 dengan menggunakan persamaan (20) dan (21), diperoleh*

$$\left|\hat{\boldsymbol{\Sigma}}^2\right| \le \left|\hat{\boldsymbol{\Sigma}}^1\right| \qquad (22)$$

Ekspresi (22) akan sama apabila $\hat{\boldsymbol{\mu}}^1 = \hat{\boldsymbol{\mu}}^2$ *dan* $\hat{\boldsymbol{\Sigma}}^1 = \hat{\boldsymbol{\Sigma}}^2$.

Teorema C-Step dilakukan berulang kali sampai $\left|\hat{\boldsymbol{\Sigma}}_{baru}\right| > 0$ atau $\left|\hat{\boldsymbol{\Sigma}}_{baru}\right| = \left|\hat{\boldsymbol{\Sigma}}_{lama}\right|$.

Metode Estimasi Robust S

Estimasi ini diperkenalkan pertama kali oleh Rousseeuw and Yohai (1984) yang kemudian dikembangkan lagi oleh Lopuhaa (1989) dan Davies (1987).

Definisi 1 (Davies,1987) *Diberikan* $\{\boldsymbol{r}_i,\ i = 1, \dots, n\}$ *adalah himpunan data di* $\mathbb{R}^p$ *dan diberikan* $\mathcal{P}_p$ *adalah himpunan semua matriks simetrik definit positif berukuran* $p \times p$. *Estimasi S untuk ukuran lokasi* $\hat{\boldsymbol{\mu}} \in \mathbb{R}^p$ *dan dispersi* $\hat{\boldsymbol{\Sigma}}(R) \in \mathcal{P}_p$ *adalah setiap pasangan yang meminimumkan* $|\boldsymbol{\Sigma}|$ *dengan kendala*

$$\frac{1}{n}\sum_{i=1}^{n} \rho[(\boldsymbol{r}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{r}_i - \boldsymbol{\mu})]^{1/2} = b_0 \qquad (23)$$

dengan ρ adalah fungsi kerugian (loss function), dan b_0 adalah suatu *konstanta*. Konstanta ini harus ditentukan dengan tepat karena nilai ini berpengaruh terhadap hasil estimasi. Apabila distribusi data tidak diketahui maka dipilih $b_0 = E\{\rho\|\mathbf{r}\|\}$.

Estimator S dapat diperoleh dengan menyelesaikan persamaan berikut ini:

$$\frac{1}{n}\sum_{i=1}^{n} u(d_i)(\boldsymbol{r}_i - \boldsymbol{\mu}) = 0 \tag{24}$$

$$\frac{1}{n}\sum_{i=1}^{n} pu(d_i)(\boldsymbol{r}_i - \boldsymbol{\mu})(\boldsymbol{r}_i - \boldsymbol{\mu})' - v(d_i)\boldsymbol{\Sigma} = 0 \tag{25}$$

dengan $d_i = (\boldsymbol{r}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{r}_i - \boldsymbol{\mu})$, $\psi(d_i) = \frac{\partial \rho}{\partial d}$, $u(d_i) = \psi(d_i)/d_i$, sedangkan $v(d_i) = \psi(d_i)d_i - \rho(d_i) + b_0$. Estimasi S ditentukan oleh pemilihan fungsi kerugian (*loss function*).

Perhitungan estimasi S dilakukan secara iteratif dengan menggunakan persamaan (24) dan (25). Menurut Hardin (2000) algoritma perhitungan estimasi S adalah sebagai berikut:

1. Tentukan estimasi awal vektor mean dan matriks kovariansi, $\widehat{\boldsymbol{\mu}}_0$ dan $\widehat{\boldsymbol{\Sigma}}_0$
2. Hitung $d_i = (\boldsymbol{r}_i - \widehat{\boldsymbol{\mu}}_0)'\widehat{\boldsymbol{\Sigma}}_0^{-1}(\boldsymbol{r}_i - \widehat{\boldsymbol{\mu}}_0)$
3. Tentukan k_0 sedemikian hingga $\frac{\sum \rho(d_i/k_0)}{n} = b_0$
4. Hitung $\widetilde{d_i} = \frac{d_i}{k_0}$
5. Tentukan $\widehat{\boldsymbol{\mu}} = \frac{\sum \psi(\widetilde{d_i})\boldsymbol{r}_i}{\sum \psi(\widetilde{d_i})}$ dan $\widehat{\Sigma} = \frac{p\sum \psi(\widetilde{d_i})(\boldsymbol{r}_i - \boldsymbol{\mu})(\boldsymbol{r}_i - \boldsymbol{\mu})'}{\sum \psi(\widetilde{d_i})}$
6. Ulangi langkah 2-3 sampai $\widehat{\boldsymbol{\mu}}$ dan $\widehat{\boldsymbol{\Sigma}}$ konvergen

HASIL DAN PEMBAHASAN

Hasil Klastering

Pada penelitian ini, kami menggunakan harga penutupan harian saham harian dari seluruh saham yang termasuk dalam indeks LQ-45 yang terdaftar di Bursa Efek Indonesia periode Agustus 2017 – Februari 2018 yang kemudian diperluas hingga Juli 2018 yang diakses melalui website https://finance. yahoo.com. Analisis klaster yang digunakan adalah analisis klaster *time series* PAM dengan ukuran disimilaritas jarak DTW. Dengan menggunakan fungsi **TSclust** pada *packages* R, diperoleh bahwa saham-saham LQ-45 dapat dikelompokkan menjadi 6 klaster, sebagaimana disajikan pada Tabel 1.

Dari Tabel 1 dapat dilihat bahwa dengan menggunakan klastering *time series* PAM dengan ukuran disimilaritas DTW, saham-saham LQ-45 tersebar ke dalam enam klaster dengan jumlah anggota masing-masing klaster cukup variatif. Klaster 1 terdiri dari 10 saham, klaster 2 terdiri 19 saham, klaster 3 terdiri dari 11 saham, klaster 4 terdiri dari 3 saham, sedangkan klaster 5 dan klaster 6 masing-masing beranggotakan 1 saham.

Saham Representasi Klaster

Setelah klaster-klaster terbentuk langkah selanjutnya adalah menghitung Sharpe *ratio* masing-masing saham pada setiap klaster. Dalam perhitungan Sharpe *ratio* ini, *return risk-free rate* yang digunakan adalah Bank Indonesia rate pada waktu pengambilan data, yaitu 5.25 % per tahun. Berdasarkan perhitungan Sharpe *ratio* setiap saham pada masing-masing klaster diperoleh saham-saham yang merupakan representasi masing-masing klaster untuk menyusun portofolio optimum sebagaimana disajikan pada Tabel 2.

Tabel 1. Klaster Saham

Klaster	Anggota Klaster									
1	AALI	AKRA	ASII	BBNI	BMRI	ICBP	INDF	JSMR	LPPF	SMC
2	ADHI	ADRO	BBRI	BBTN	BJBR	BRPT	BSDE	EXCL	HMSP	INC
	KLBF	PGAS	PTBA	PTPP	SCMA	SSMS	TLKM	WIKA	WSKT	
3	ANTM	BMTR	BUMI	LPKR	LSIP	MNCN	MYRX	PPRO	PWON	SMF
	SRIL									
4	BBCA	INTP	UNTR							
5	GGRM									
6	UNVR									

Tabel 2. Saham-saham Representasi Klaster

Klaster	Representasi	Return	Risiko	Sharpe Ratio
1	ICBP	0,00028	0,01395	0,00958
2	INCO	0,00258	0,02673	0,09116
3	ANTM	0,00138	0,02431	0,05068
4	BBCA	0,00088	0,01285	0,05713
5	GGRM	0,00032	0,01830	0,00941
6	UNVR	-0,00034	0,01383	-0,03521

Pada klaster 1 yang terdiri 10 saham dan saham ICBP dengan Sharpe *ratio* sebesar 0,00958 merupakan saham dengan Sharpe *ratio* tertinggi di klaster 1, sehingga saham ICBP terpilih sebagai representasi dari klaster 1. Selanjutnya pada klaster 2 yang terdiri dari 19 saham dan saham INCO dengan Sharpe *ratio* sebesar 0,09116 merupakan saham dengan Sharpe *ratio* tertinggi di klaster 2, sehingga saham INCO terpilih sebagai representasi dari klaster 2. Demikian seterusnya dengan cara yang sama, dipilih saham-saham ANTM, BBCA, GGRM, dan UNVR secara berturut-turut sebagai representasi dari klaster 3, 4, 5, dan 6.

Perbandingan Kinerja Portofolio yang Terbentuk

Pada penelitian ini portofolio optimum ditentukan dengan menggunakan tiga model portofolio, yaitu: model portofolio MV klasik (MV_{klasik}), model portofolio MV *robust* FMCD (MV_{FMCD}) dan model portofolio MV *robust* S (MV_S). Langkah pertama yang dilakukan adalah menentukan bobot portofolio ketiga model untuk berbagai nilai *risk aversion* γ menggunakan fungsi **CovMcd** dan **CovSest** pada *packages* R. Saham-saham yang digunakan adalah saham-saham yang merupakan representasi setiap klaster sebagaimana disajikan pada Tabel 1. Bobot portofolio yang dihasilkan diberikan pada Tabel 3.

Dari Tabel 3 dapat dilihat bahwa, untuk model portofolio MV klasik, saham dengan return negatif yaitu saham UNVR memiliki bobot negatif *(short selling)* untuk semua nilai *risk aversion* γ. Sebaliknya saham-saham yang lain selalu memiliki bobot positif, kecuali saham ICBP untuk $\gamma = 0{,}5$ dan $\gamma = 1$.

Dari Tabel 3, dapat dilihat pula bahwa untuk model portofolio MV *robust* FMCD, saham ICBP, INCO dan UNVR memiliki bobot negatif untuk hampir semua nilai *risk aversion* γ. Namun demikian bobot negatif tersebut semakin mengecil seiring dengan bertambahnya nilai *risk aversion* γ. Di sisi lain, saham ANTM, BBCA dan GGRM selalu memiliki bobot positif. Dengan

Tabel 3. Bobot Portofolio

Model	γ	ICBP	INCO	ANTM	BBCA	GGRM	UNVR
MV_{klasik}	0,5	-0,50075	6,01759	0,01913	7,49371	0,35268	-12,38236
	1	-0,11771	3,05234	0,02227	3,90435	0,22244	-6,08368
	2	0,07380	1,56971	0,02384	2,10966	0,15733	-2,93434
	5	0,18871	0,68014	0,02478	1,03285	0,11826	-1,04474
	10	0,22702	0,38361	0,02509	0,67392	0,10523	-0,41487
MV_{FMCD}	0,5	-11,64984	-2,47218	3,38129	15,32727	0,41117	-3,99770
	1	-5,72835	-1,21254	1,70727	7,80734	0,24466	-1,81838
	2	-2,76760	-0,58271	0,87026	4,04737	0,16140	-0,72872
	5	-0,99116	-0,20482	0,36806	1,79139	0,11145	-0,07492
	10	-0,39901	-0,07886	0,20066	1,03940	0,09480	0,14301
MV_S	0,5	-7,93013	0,21814	1,74590	11,37744	-0,87301	-3,53834
	1	-3,84454	0,13009	0,89565	5,83364	-0,40021	-1,61463
	2	-1,80174	0,08607	0,47052	3,06174	-0,16381	-0,65277
	5	-0,57606	0,05965	0,21545	1,39860	-0,02197	-0,07566
	10	-0,16751	0,05085	0,13042	0,84421	0,02531	0,11671

Tabel 4. Return, Risiko dan *Sharpe Ratio* Portofolio

γ	Return			Risiko			Sharpe Ratio		
	MV_{Klasik}	MV_{FMCD}	MV_S	MV_{Klasik}	MV_{FMCD}	MV_S	MV_{Klasik}	MV_{FMCD}	MV_S
0,5	0,02636	0,01959	0,01090	0,05167	0,04043	0,02222	0,11531	0,09669	0,07218
1	0,01346	0,00949	0,00536	0,01298	0,01015	0,00560	0,11687	0,09279	0,06976
2	0,00701	0,00444	0,00260	0,00331	0,00258	0,00145	0,11936	0,08470	0,06446
5	0,00314	0,00142	0,00093	0,00060	0,00046	0,00028	0,12213	0,05945	0,04695
10	0,00186	0,00041	0,00038	0,00022	0,00016	0,00012	0,11622	0,02111	0,02182

membesarnya nilai *risk aversion* γ bobot positif tersebut akan semakin menurun untuk mengimbangi saham-saham dengan bobot negatif.

Untuk model portofolio MV *robust* S, saham ICBP, GGRM dan UNVR memiliki bobot negatif untuk hampir semua nilai γ. Namun demikian bobot negatif tersebut semakin mengecil seiring dengan bertambahnya nilai *risk aversion* γ. Di sisi lain, saham INCO, ANTM dan BBCA selalu memiliki bobot positif semua nilai *risk aversion* γ. Sama dengan dua model yang lain, dengan membesarnya nilai *risk aversion* γ, bobot positif tersebut akan semakin menurun untuk mengimbangi saham-saham dengan bobot negatif.

Berdasarkan bobot portofolio, vektor mean dan matriks kovariansi, selanjutnya dapat ditentukan return, risiko dan Sharpe *ratio* ketiga portofolio yang dibentuk yang hasilnya sebagaimana disajikan pada Tabel 4.

Mengukur kinerja portofolio tidak bisa hanya dilihat dari returnnya saja tetapi juga harus memperhatikan risiko yang akan ditanggung *investor*. Ada beberapa ukuran yang dapat digunakan untuk mengukur kinerja portofolio, salah satunya adalah dengan menggunakan *Sharpe ratio*. Tabel 4 memperlihatkan return, risiko dan Sharpe *ratio* portofolio yang dibentuk dengan menggunakan klastering *time series* PAM dengan ukuran jarak DTW dikombinasikan dengan model portofolio MV klasik, model MV_{FMCD} dan model MV_S. Dari Tabel 4, secara umum dapat dilihat bahwa kinerja portofolio yang dibentuk dengan menggunakan saham representasi hasil klastering menggunakan PAM dengan ukuran DTW yang dikombinasikan dengan model portofolio MV klasik mengungguli kinerja portofolio yang dibentuk dengan kombinasi model portofolio yang lain.

KESIMPULAN DAN SARAN

Kesimpulan

Paper ini membahas bagaimana teknik klastering dapat digunakan dalam manajemen portofolio saham dan membangun sistem untuk mendapatkan portofolio optimum. Hal ini dapat mengurangi waktu pemilihan saham karena saham dengan kategori serupa dapat dengan mudah dikelompokkan ke dalam satu klaster. Setelah klaster-klaster terbentuk, kemudian dipilih saham berkinerja terbaik dari masing-masing klaster sebagai representasi klaster untuk membentuk portofolio optimum. Studi empiris menunjukkan bahwa dengan menggunakan klastering *time series* PAM dengan ukuran jarak DTW, 45 saham Bursa Efek Indonesia yang termasuk dalam indeks LQ-45 dapat dikelompokkan menjadi 6 klaster. Saham representasi masing-masing klaster kemudian digunakan untuk membentuk portofolio optimum menggunakan model portofolio MV klasik, model portofolio MV *robust* FMCD dan model portofolio MV *robust* S. Hasil penelitian ini menunjukkan bahwa kinerja portofolio yang dihasilkan dengan menggunakan klastering *time series* PAM dengan ukuran jarak DTW yang dikombinasikan dengan model portofolio MV klasik mengungguli kinerja portofolio yang dihasilkan kombinasi dengan model yang lain.

Saran

Untuk penelitian selanjutnya, akan menarik untuk menggunakan teknik klastering *time series* dengan ukuran disimilaritas yang lain seperti periodogram *based distance*, Piccolo *distance*, Maharaj *distance* dan mengkombinasikannya dengan model-model portofolio yang sudah ada, kemudian membandingkan kinerja portofolio yang dihasilkan.

UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih atas dukungan finansial dari Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia melalui Hibah Penelitian Disertasi Doktor (PDD) Tahun 2021. Penulis pertama juga mengucapkan terima kasih kepada Lembaga Pengelola Dana Pendidikan (LPDP) Kementerian Keuangan Republik Indonesia yang telah memberikan beasiswa untuk Program Doktor di Departemen Matematika Universitas Gadjah Mada.

DAFTAR PUSTAKA

Alonso, A. M., Barrendero, J. R., Hernandez, A., dan Justel, A. 2006. Time Series Clustering Based on Forecast Densities. *Computational Satistics & Data Analysis*, Vol. 51, 762–776.

Best, Michael J. dan Grauer, Robert R. 1991. On the Sensitivity of Mean-Variance Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results. *Review of Financial Studies*, Vol. 4(2), 315–342.

Berndt, Donald J. dan Clifford, James. 1994. Using Dynamic Time Warping to Find Patterns in Time Series. In *Workshop on Knowledge Discovery in Databases*, 359–370.

Ceria, Sabastian dan Stubbs, Robert A. 2006. Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management*, Vol. 7(2), 109–127.

Chen, Liang H. dan Huang, Lindsay. 2009. Portfolio optimization of equity mutual funds with fuzzy return rates and risks. *Expert Systems with Applications*, Vol. 36, 3720–3727.

Chopra, Vijay K. dan Ziemba, William T. 1993. The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice. *Journal of Portfolio Management*, Vol. 19(2), 6–11.

Davies, P. L. 1987. Asymptotic Behaviour of S-Estimates of Multivariate Location Parameters and Dispersion Matrices. *The Annals of Statistics*, Vol. 15(3), 1269–1292.

DeMiguel, Victor dan Nogales, Francisco J. 2009. Portfolio Selection With Robust Estimation. *Operations Research*, Vol. 57(3), 560–577.

Elton, E. J., Gruber, M. J., Brown, S. J., dan Geotzmann, W. N. 2014. *Modern Portfolio Theory and Investment Analysis*, 9th Edition, John Wiley and Sons, Inc., New York.

Guan, He S. dan Jiang, Qing S. 2007. Cluster financial time series for portfolio. Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR 2007, Vol. 2, 851–856.

Gubu, L. Rosadi, D., dan Abdurakhman. 2019. Classical portfolio selection with cluster analysis: Comparison between hierarchical complete linkage and Ward algorithm, in International Conference on Mathematics and its Applications 2019, *AIP Conference Proceedings 2192*, 090004-1–0960004-7.

Hardin, Johana S. 2000. Multivariate Outlier Detection and Robust Clustering with Minimum Covariance Determinant Estimation and S-estimation. *Disertasi*, University of California, California.

Kaufman, Leonard dan Rousseeuw, Peter J. 1990. *Finding Groups in Data: An Introduction*, John Wiley & Sons, Hoboken, New Jersey.

Lauprete, Geoffrey J. 2001. Portfolio Risk Minimization Under Departures from Normality. *Disertasi*, Massachusetts Institue of Technology.

Long, N. C., Wisitpongphan, N., Meesad, P., dan Unger, H. 2014. Clustering stock data for multi-objective portfolio optimization. *International Journal of Computational*

Intelligence and Applications, Vol. 13(2), 1–13.

Lopuhaa, Hendrik P. 1989. On the Relation between S-Estimators and M-Estimators of Multivariate Location and Covariance. *The Annals of Statistics,* Vol. 17(4), 1662–1683.

Markowitz, Harry. 1952. Portfolio Selection. *The Journal of Finance,* Vol. 7(1), 77–91.

Maronna, R. A., Martin, R. D. dan Yohai, V. J. 2006. *Robust Statistics: Theory and Methods.* John Wiley & Sons, Ltd.

Montero, Pablo dan Vilar, Jose A. 2014. TSclust: an R Package for Time Series Clustering. *Journal of Statistical Software,* Vol. 62(1), 1–43.

Nanda, S. R., Mahanty, B., dan Tiwari, M. K. 2010. Clustering Indian Stock Market Data for Portfolio Management. *Expert Systems with Applications,* Vol. 37(12), 8793–8798.

Reilly, Frank K. dan Brown, Keith C. 2012. *Investment Analysis and Portfolio Management. 10th Edition,* Thomson South-Western. USA

Rousseeuw, Peter J. dan Van Driessen, Katrien. 1999. Fast Algorithm For Minimum Covariance Determinant Estimator. *Technometrics,* Vol. 41(3), 212–223.

Rousseeuw, Peter J. dan Yohai, Victor. 1984. Robust Regression By Means of S estimators, *Lecture Notes in Statistics: Robust and* Nonlinear Time Series Analysis, Vol. 26, 256–272.

Sankoff, David dan Kruskal, Joseph. 1983. Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison, Addison Wesley.

Sharpe, William F. 1994. The Sharpe Ratio, *The Jornal of Portfolio Management,* Vol. 21, 49–58.

Supandi, Epha D. 2017. Pengembangan Model Portofolio Mean-Variance Melalui Metode Estimasi Robust dan Optimasi Robust. *Disertasi*, Departemen Matematika FMIPA Universitas Gadjah Mada, Yogyakarta

Tola, V., Lillo, F., Gallegati, M., dan Mantegna, R. N. 2008. Cluster analysis for portfolio optimization. *Journal of Economic Dynamics and Control*, Vol. 32(1), 235–258.

Vaz-de Melo, Beatriz dan Camara, Ricardo P. 2005. Robust Multivariate Modeling in Finance, *International Journal of Managerial Finance,* Vol. 1(2), 95–107.

Welsch, Roy E. dan Zhou, Xinfeng. 2007. Application of robust statistics to asset allocation models. *REVSTAT–Statistical Journal*, Vol. 5(1), 97–114.

Winston, Wayne L. dan Goldberg, Jeffrey B. 2004. *Operations Research: Applications and Algorithms, Fourth Edition,* Thomson Learning Inc.