ZIBELINE INTERNATIONAL™ PUBLISHING

ISSN: 2521-0831 (Print)
ISSN: 2521-084X(Online)
CODEN: MSMAD

Matrix Science Mathematic (MSMK)

DOI: http://doi.org/10.26480/msmk.01.2023.07.15

CrossMark

RESEARCH ARTICLE

APPLICATION OF ROBUST REGRESSION FOR PORTFOLIO OPTIMIZATION

Ezra Putranda Setiawan[a*], Dedi Rosadi[b]

[a]*Universitas Negeri Yogyakarta Fakultas Matematika dan Ilmu Pengetahuan Alam*
[b]*Universitas Gadjah Mada*
**Corresponding Author Email: ezra.ps@uny.ac.id*

ARTICLE DETAILS

Article History:

Received 25 November 2022
Revised 05 December 2022
Accepted 09 January 2023
Available online 13 January 2023

ABSTRACT

The single-index model is a portfolio optimization method that uses each asset's beta'. In general, the beta is estimated using the return data by the least square method. However, the return data frequently contains several outliers, so the estimation resulting from the least square method is inaccurate. This study examines several beta estimators from robust regression methods, namely the least absolute value estimator, M-estimator, LMS-estimator, LTS-estimator, MM-estimator, and Tau estimator to estimate the beta of each asset and make an optimal portfolio based on this estimated value. We also evaluate the effect of robust beta estimators on the stability and performance of each portfolio. We present the Sharpe ratio and some turnover measures, namely the l-period portfolio turnover, maximum turnover, lower bound single-asset turnover, and lower bound multiple-asset turnover. Among various estimators used here, the Tau estimator is the best estimator to replace the OLS for estimating the beta.

KEYWORDS

Portfolio Optimization; Robust Regression; Asset Turnover; Stock

1. INTRODUCTION

The mean-variance model is one of the most important foundations of modern investment portfolio selection (Markowitz, 1952). In this model, the objective is to minimize the portfolio risk, measured by the variance of the portfolio's return, for a specific level of expected return. One should estimate the variance-covariance matrix based on assets' return data to calculate the variance. It means that one should estimate the variance of each asset's return and the covariance of each asset pair. Consequently, the number of parameters that must be estimated would be increased swiftly when more assets are incorporated into the portfolio.

To simplify the calculation of a portfolio's risk, introduces the single-index model (Sharpe, 1963). In this model, it is assumed that the dynamic movement of each asset's return has a linear relationship with the dynamic movement of the market return. The 'beta' of each asset is defined as the slope of a regression line, where the asset's return becomes the response or dependent variable and the market return becomes the predictor variable. Therefore, the 'beta' could be used to measure investment risk. A higher beta value indicates that an asset's price movement is similar to the market index. In contrast, a more negligible or negative beta indicates that the asset's price movement is unrelated to the market index movement.

The idea of a single index model underlies several developments in portfolio theory. It Has a introduced the ratio of excess return to beta as a volatility measure of an asset, similar to the ratio of excess return to standard deviation as the risk measure (Treynor, 1973). Later, found that the ratio of excess return to beta could be used to calculate an optimal portfolio, which was then known as the single-index portfolio optimization model (Elton et al.,1976). Using this model, the investor only needs to estimate the beta and the standard deviation for each asset and does not need to calculate the covariance between each asset. Compared to Markowitz's mean-variance portfolio optimization model, the single index model is more straightforward and uses fewer estimators. Although less popular than the Mean-Variance model, the single index model is still widely used, for example, by (Nandan and Sritavasta., 2017; Ahuja, 2017; Yuwono and Ramdhani, 2017; Parton et al.,2017; Murthy, 2018).

The 'beta' of an asset is the slope of the regression line between the asset return and the market return. Therefore the ordinary least square (OLS) method is frequently used to estimate the beta. On the other hand, show that the empirical return data frequently inhibit fat tails and contains one or more outliers (Lauprete, 2002). The single index model used the return data as a dependent and independent variable. Hence, the estimated single index model may contain outliers or leverage points. It is stated that the presence of outliers or leverage points would significantly affect and offset the result of the OLS estimator (Rousseeuw and Leroy, 1987).

In dealing with outliers, an estimation method known as a robust estimator could be used to get an estimated value that is not so strongly affected by the presence of outliers. According to robust estimators provide an excellent fit to the bulk of the data when the data contain outliers and when the data are free of them (Maronna et al.,2006).

Literature surveys show that usage of the robust estimator in portfolio optimization is dominated by robust covariance matrix estimation for mean-variance portfolio (Lauprete, 2002; Lauprete et al., 2003; Perret-Gentil and Victoria Fesser, 2003; Mendes and Leal, 2005; Welsch and Zhou, 2007; DeMiguel and Nogales, 2009; Supandi et al., 2014, 2015; Huo et al., 2012). In most cases, the robust covariance estimator yields a better portfolio than the classical estimator. On the other hand, there needs to be more information about the usage of the robust estimator in the single index portfolio optimization model. Through an empirical study using the Indonesian stock market dataset, implement the Bisquare-M estimation, Least Trimmed Squares (LTS), and MM estimation in Single Index Model (Rosadi and Setiawan, 2017). They stated that usage of robust estimator yield portfolios with higher Sharpe Ratio.

Quick Response Code	Access this article online	
	Website: www.matrixsmathematic.com	**DOI:** 10.26480/msmk.01.2023.07.15

Cite The Article: Ezra Putranda Setiawan, Dedi Rosadi (2023). Application of Robust Regression for Portfolio Optimization. *Matrix Science Mathematic, 7(1): 07-15.*

This study aims to extend and validate studies by implementing various robust regression approaches for the single index portfolio optimization model and performing a simulation procedure (Rosadi and Setiawan's, 2017). We compare more methods not considered in that paper, i.e., Least Absolute Value (LAV) regression, various types of M-regression, MM-regression, Least Median of Squares (LMS), and τ-regression. We also evaluate the estimated value of the asset's beta, the portfolio's performance, and the behaviour of the asset's weight using the rolling horizon method proposed by (DeMiguel and Nogales, 2009). In addition, we propose several new measures of portfolio turnover that can reflect the investor's preference for keeping or changing the portfolio over time.

The rest of this paper is organized as follows. Section 2 is dedicated to the foundations of single-index portfolio optimization and robust regression. In section 3, we explain the simulation and empirical study procedures. Sections 4 and 5 present the simulation and empirical study, respectively. The last section is devoted to the conclusion and suggestions for future research.

2. Single-Index Portfolio and Robust Regression

2.1 Single Index Portfolio Optimization

Suppose that R_{it} and R_{mt} denote the return of the i^{th} asset and the return of the market index at time t, respectively. According to there is a linear relationship between the asset return and the market index return, so we could write (Elton et al.,1976).

$$R_{it} = a_i + \beta_i R_{Mt}. \tag{1}$$

In the formula (1), a_i is a random variable that represent the component of the asset return that unexplained by the fluctuation of the market price. This component could be breaked into two component, namely the expected value α_i and a random element denoted by e_i. The beta or β_i is a constant that measure the expected movement in the stock return R_i for every change of market return R_m. Therefore, we can restate the equation (1) as

$$R_{it} = \alpha_i + \beta_i R_{Mt} + e_i. \tag{2}$$

When examine two asset together, we assume that the random error of the i-th asset, e_i, is uncorrelated to the random error of the j-th asset, e_j. This means that the fluctuation of asset return is only related to the fluctuation of market return, and no effects outside the market return.

$$\sigma_i^2 = \beta_i^2 \sigma_M^2 + \sigma_{ei}^2. \tag{3}$$

Following the single index model, the variance as a risk measure of each security could be broken down into two parts: systematic risk and unsystematic risk. Unsystematic risk (σ_{ei}) is the risk that can be zero by adding more assets to the portfolio. In contrast, systematic risk is not affected by the number of assets but corresponds to the market risk. Since the market risk (σ_M) is constant for all assets in the market, the beta β_i could be used to measure the contribution of market risk to the systematic risk of an asset.

Suppose a risk-free asset is traded in the stock market with a return equal to Rf. Since investing in stock is not risk-free, a rational investor will choose an asset that can yield a higher return than the risk-free asset or has a positive excess return. Recall that the beta of an asset is a measure of systematic risk, so a rational investor should choose an asset with a smaller beta. Therefore, in the single index portfolio optimization model, the first step is to calculate the excess return to beta (ERB) ratio for each asset, that is

$$ERB_i = \frac{\bar{R}_i - R_f}{\beta_i}. \tag{4}$$

A rational investor will choose an asset with a larger ERB ratio to be included in the portfolio since a larger ERB represent a more significant excess return and lower beta (lower risk). The ordered list of stocks based on the ERB ratio is the basis to determine the order of stock that will be included in the optimal portfolio. If an asset with a particular ERB is included in the optimal portfolio, the other assets with larger ERB also must be included in the optimal portfolio.

In this research, we focus on the portfolio with only long positions. Therefore, the next step is to determine the cut-off point C^* so that all assets with $ERB > C^*$ are included in the optimal portfolio, and the other asset with $ERB < C^*$ must be discarded. Following an optimal C^* is defined as the value of C^* so that all assets that are used to calculate them have $ERB > C^*$, while the other asset has $ERB < C^*$. To obtain the C^*, we must calculate the candidate of the cut-off point, namely, (Elton et al.,1976)

$$C_i = \frac{\sigma_M^2 \sum_{j=1}^{i} \left(\frac{\beta_j \left[\bar{R}_j - R_F \right]}{\sigma_{ej}^2} \right)}{1 + \sigma_M^2 \sum_{j=1}^{i} \left(\frac{\beta_j^2}{\sigma_{ej}^2} \right)} \tag{5}$$

for $i = 1, 2, \ldots, n$, since adding or removing an asset from the portfolio will affect the cut-off point. After obtaining the optimal cut-off point C^*, we could determine which assets should be included in the optimal portfolio. Suppose that there are K assets with ERB ratio greater than the optimal cut-off point C^*, the weight of each of the K asset in the optimal portfolio equal to

$$X_i = \frac{Z_i}{\sum_{j \in K} Z_j}, i \in K \tag{6}$$

where

$$Z_i = \frac{\beta_i}{\sigma_{ei}^2} \left(\frac{\bar{R}_i - R_F}{\beta_i} - C* \right), i \in K. \tag{7}$$

From the above explanations, it can be inferred that the estimated value of beta plays a vital role in determining which asset should be included in the portfolio and how many portions of the capital that be used to buy the asset. In daily practice, an asset's beta is estimated using the historical data for the asset's and market's returns. As mentioned above, a simple and popular method to estimate the asset's beta is by minimising the square distance between the predicted asset's return from the estimated single index model and the actual value of the asset's return, known as the Ordinary Least Square (OLS) method. In this research, we examine the usage of several robust estimators to get the estimated value of beta, replacing the ordinary least square that had widely used. The estimated beta value in the single index model will be used to determine the portfolio's weight.

2.2 Least Absolute Value (LAV) Regression

Least absolute value (LAV) estimation probably the oldest estimation procedure known for linear regression (Dielmann, 2005). For the single-index model stated in eq. (2), the estimation value is chosen in order to minimize the sum of absolute value of the residual, that is

$$min \sum_{t=1}^{T} \left| R_{it} - (\alpha_i + \beta_i R_{mt}) \right| \tag{8}$$

According to the LAV estimation is less influenced by outlier data in the dependent variable. However, this method would not solve the problem when the outliers are present in both the dependent and independent variable (Andersen et al.,2008).

2.3 M-Regression

Table 1: Several kinds of $\rho(u)$ function for M-estimator (Draper and Smith, 1999).

Function name	$\boldsymbol{\rho(u)}$	$\boldsymbol{\psi(u)}$
Huber	$\begin{cases} \frac{1}{2}u^2 & , \lvert u \rvert \le k \\ k\lvert u \rvert - \frac{1}{2}u^2 & , \lvert u \rvert > k \end{cases}$	$\begin{cases} k & , u \ge k \\ u & , -k < u < k \\ -k & , E \le -k \end{cases}$
Tukey's Bisquare	$\begin{cases} \frac{u^2}{2}\left(1 - \frac{1}{2}\left(\frac{u}{a}\right)^2\right) & , \lvert u \rvert \le a \\ \frac{1}{2}a^2 & , \lvert u \rvert > a \end{cases}$	$\begin{cases} u\left(1 - \left(\frac{u}{a}\right)^2\right) & , \lvert u \rvert \le a \\ 0 & , \lvert u \rvert > a \end{cases}$
Hampel	$\begin{cases} \frac{1}{2}u^2 & , \lvert u \rvert \le a \\ a\lvert u \rvert - \frac{1}{2}a^2 & , a \le \lvert u \rvert \le b \\ \frac{a}{c-b}\left(c\lvert u \rvert - \frac{1}{2}u^2\right) - \frac{7}{6}a^2 & , b \le \lvert u \rvert \le c \\ a(b+c-a) & \lvert u \rvert \ge c \end{cases}$	$\begin{cases} u & , \lvert u \rvert \le a \\ a \operatorname{sign}(u) & , a \le \lvert u \rvert \le b \\ \frac{c \operatorname{sign}(u) - u}{c-b} & , b \le \lvert u \rvert \le c \\ 0 & \lvert u \rvert \ge c \end{cases}$

Since outlier data can be identified as observations with relatively large residual than the others, the main idea of M-regression is giving smaller weight to the observations with large residual and larger weight to the observations which has residual located in the middle of the distribution. This approach, introduced by, is done by creating a weight function that symmetric, continuous, and strictly increasing in positive real number domain (Andrews, 2008). Several type of weight functions have been proposed, as mentioned in table 1 below.

With these weight function, an M-estimate of asset's beta according to equation (2) is obtained by solving

$$min \sum_{t=1}^{T} \rho \left(\frac{R_{it}-(\alpha_i+\beta_i R_{mt})}{s_i}\right) \quad (9)$$

where ρ is the weight function.

This is show that the Iteratively Reweighted Least Square (IRLS) could be used to estimate the regression parameter using M-estimator (Maronna et al.,2006). The M-estimator could deal with a more significant number of an outlier than the classical least (sum of) square and the least absolute value method.

2.4 Least Median of Squares (LMS) Regression

It is widely known that as a central tendency measure of numerical data, the median is more robust than the mean or average. Based on this idea, suggests using the median instead of the sum of the squared residuals, later known as the median least square method (Rousseeuw, 1984). Mathematically, the least median of squares (LMS) estimation for the asset's beta is the solution to the minimization problem

$$min \left[median\left(R_{it}-(\alpha_i+\beta_i R_{mt})\right)^2\right] \quad (10)$$

According to Massart *et al.* (1986), this estimation method is very robust concerning the outliers in the data of dependent and independent variables. The estimated value obtained by this method will not be significantly affected when the number of outliers in the data is as large as 50%. Nevertheless, the relative efficiency of this estimation model is 37% only, relative to the efficiency of the classical least square estimation method. These conditions make the least median square method rarely used today.

2.5 Least Trimmed Square (LTS) Regression

Proposed by, Least Trimmed Square (LTS) estimation for the beta of an asset is given by (Rousseeuw, 1984)

$$min \left[\sum_{t=1}^{h} \left(R_{i(t)}-\left(\alpha_i+\beta_i R_{m(t)}\right)\right)^2\right]. \quad (11)$$

To obtain this estimator, one first needs to estimate a preliminary model and get the residual based on this preliminary estimation. Calculate each residual's square, order them increasingly, and then discard some observations with the most significant residual (which probably are outliers). In other words, from *T* available data, only *h* smallest residual were taken to calculate the sum of square, which are minimized. In the best robustness of this estimator is achieved when half of the available data are used. In other words, take $h = T/2$ (Rousseeuw and Leroy, 1987).

Compared to the classical least square estimator, least absolute value estimator, and M-estimator, the Least Median of Square (LMS) and Least Trimmed Square (LTS) estimator could manage the outlier on the independent variable as well as a large number of outliers in the data.

2.6 MM-Regression

Despite the higher robustness exhibited by both LMS and LTS estimation methods, these estimators have lower efficiency than the ordinary least square. To increase the efficiency, introduce a new estimation method that yields a higher breakdown point (can deal with more outliers data) and higher efficiency (up to 95% relative to the ordinary least square method) (Yohai and Zamar, 1987). This method consists of a two-step M-estimation procedure, hence known as MM-estimation. Using this estimation method, the estimated beta of an asset can be calculated as follows:

- Obtain an initial parameter estimation of regression coefficient in equation (2), using a high-breakdown estimation method, i.e. least median of square (LMS) or least trim med square (LTS). The estimation method must be consistent but not need to be efficient.
- Calculate the residual from the estimated regression obtained in the step 1,

$$e_{it} = R_{it} - \left(\hat{\alpha}_i + \hat{\beta}_i R_{mt}\right), \quad (12)$$

and use these residuals to calculate an M-estimation of scale, s_n.

- The initial residual estimation from step 1 and the scale estimation from step 2 are used to calculate the first iteration of weighted least square,

$$\sum_{t=1}^{T} \rho \left[\frac{R_{it}-(\hat{\alpha}_i+\hat{\beta}_i R_{mt})}{s_n}\right], \quad (13)$$

where ρ is a weight function as in table 1.

- Obtain the residual from step 3 and use to calculate the new weights.
- Reiterate step 2, 3, and 4 until convergence.
- Because of its high breakdown point and efficiency, the MM estimate become the most frequently used robust regression estimator (Begashaw et al., 2020).

2.7 τ Regression

Another robust regression with a high breakdown point proposed by (Yohan and Zaman, 1987) is now known as τ-estimation. Unlike the MM estimation, a full-scale estimation of residual is not necessary for calculating the τ-estimation. The τ-estimation also exhibits the exact fit property proposed by (Rousseeuw , 1974). If a regression parameter precisely fits more than half of the available data, then this parameter is a τ estimator. However, this estimation method is less popular than the MM-estimation method.

Mathematically, the τ-estimation for asset's beta is a pair of parameter (α and β) that minimize τ(β), where

$$\tau(\beta) = s_n^2(\beta_i)\frac{1}{nb_2}\sum_{t=1}^{T} \rho_2 \left(\frac{R_{it}-(\alpha_i+\beta_i R_{mt})}{s_n(\beta_i)}\right) \quad (14)$$

and s_n is a M-estimate for scale, that is, solution of

$$\frac{1}{b_2}\sum_{t=1}^{T} \rho_2 \left(\frac{R_{it}-(\alpha_i+\beta_i R_{mt})}{s_n(\beta_i)}\right) = b_1. \quad (15)$$

In those model, both function ρ_1 and ρ_2 must be symmetric, continuous differentiable, bounded, and strictly increasing in [0, c] and then constant for [c, ∞). Calculation for the estimator could be done by FastTau algorithm, proposed by (Salibian -Barerra et al., 2018)

3. Research Method

As stated robust statistics in portfolio optimization aim to get stable portfolio weight whenever the return data contain several outliers by (DeMiguel and Nogales, 2009). Therefore, evaluating the robust portfolio performance should be done by examining the portfolio performance and the stability of the portfolio's weight.

3.1 Simulation Procedure

The input for the Single-Index portfolio optimization model consists of the stock return data and the market return data. The market return is calculated based on the market index, whereas the index is calculated based on each asset's price. Therefore, the simulation procedure consists of two steps: simulating the stock price movement and calculating the market index based on the available stock price.

In this simulation study, we assume that each asset return follows the normal distribution with some deviation. The price of the assets at time *t* would be generated by

$$S_t = S_0\left(1+\mu\delta t+\sigma\sqrt{\delta t}Y_t\right), \quad (16)$$

where S_0 denote the price at time 0, and *D* denote the "deviation" distribution. Here Y_t is a standard normally-distributed random variable. In this study, we assumed that the stock market consists of eight assets, with specification stated in table 2.

Table 2: Stock Properties for Simulation

Stock code	S_0	μ	σ^2
1	2000	0.10	0.0021
2	2000	0.10	0.0019
3	2000	0.08	0.0021
4	2000	0.08	0.0019
5	3000	0.10	0.0021
6	3000	0.10	0.0019
7	3000	0.08	0.0021
8	3000	0.08	0.0019

To examine the effect of outliers, we replicate the experiment by change ε data randomly into normal distribution with μ = -0.15 and σ = 0.001. Three values of ε, that is, 5%, 10%, and 15%, would be used on the same original data.

We assume that the stock market provided same number of each stock, therefore the market index at time t could be calculated as

$$MI_t = \frac{\sum_{i=1}^{8} p_{it}}{\sum_{i=1}^{8} p_{i0}} \times bv, \tag{17}$$

where bv stands for base value and p_i denote the price of the i-th stock. The market index would be used to calculate market return R_{mt}.

3.2 Rolling Horizon and Turnover Analysis

Following the work of we evaluate the robust portfolio estimation using the rolling horizon method (DeMiguel and Nogales, 2009, Supandi et al., 2016; and Rosadi et al., 2020). From our preceding simulation before, for each value of ε, we can calculate the return of the market and the stocks for T periods. Based on this result, the rolling horizon method could be explained as follows.

- Choose the length of the estimation window η, where $\eta < T$.
- Calculate the beta of each stock using any estimation method, and use the estimated beta to determine asset weight using single-index model.
- Hold the portfolio obtained in step 2 for one period to get the out-of-sample return at the next time.
- Repeat step 2 and 3 for the next period, by including the newer data and discarding the oldest data. Since the length of the estimation window is η, we must do the step 2 and 3 for T-η times for each estimation method.

The obtained out-of-sample return from the rolling horizon procedure could be used to calculate the out-of-sample mean, standard deviation, and Sharpe ratio. Suppose that r_t denote the one-period out-of-sample return for an optimal portfolio calculated at time t, where $0 < t \leq \eta$, we could calculate the out-of-sample mean and standard deviation as

$$\mu^s = \frac{1}{T-\eta} \sum_{t=\eta}^{T} \hat{r}_{t+1}^{k}, \tag{18}$$

$$\hat{\sigma}^s = \sqrt{\frac{1}{T-\eta-1} \sum_{t=\eta}^{T} (\hat{r}_{t+1}^{k} - \mu^s)^2}, \tag{19}$$

and the Sharpe Ratio become

$$\widehat{SR}^s = \frac{\hat{\mu}^s}{\hat{\sigma}^s}. \tag{20}$$

respectively.

To measure the change of each asset's weight from one period to the next period, we need to examine the weight of each asset in the portfolio along the time. Suppose $w_{i,t}^s$ denote the weight of the i-th asset in the portfolio at time t. According to, this change could be measured using portfolio turnover, i.e (DeMiguel and Nogales., 2009)

$$\mathrm{PT} = \frac{1}{T-\eta-1} \sum_{t=\eta}^{T-1} \sum_{j=1}^{N} \left| w_{i,t}^s - w_{i,t+1}^s \right|. \tag{21}$$

If the calculation is done using daily data, the above portfolio turnover measure can be understood as the average weight of asset that being traded everyday.

From the practical perspective, the daily portfolio turnover has little practical meaning since daily modification of portfolio weight leads to inefficiency. According to every transaction in the stock market must be charged, which consists of a brokerage commission fee and tax (Kellerer et al., 2001). Changing the portfolio daily could make the investor lose money since the total transaction charge will exceed the total return. On the other hand, keeping the portfolio for a week or more may cause the investor to face losses, especially when the market is bearish. Therefore, the investor should keep the asset for several days to reduce the transaction cost. Assume that in this period, no dividends were paid to the investor.

Suppose that an investor want to keep the optimal portfolio for l period of time, we introduce l-period portfolio turnover as

$$\mathrm{PT}_l = \frac{1}{T-l-\eta} \sum_{t=\eta}^{T-l} \sum_{j=1}^{N} \left| w_{i,t}^s - w_{i,t+l}^s \right|. \tag{22}$$

From equation (22), the l-period portfolio turnover is simply a generalization from the well-known portfolio turnover measure stated above. Another critical measure is the maximum turnover, defined as the maximum change of weight in the portfolio for a pre-defined period. The maximum turnover is calculated to give additional information about the most considerable change that could be happened in the portfolio weight since the portfolio turnover only represents the average change.

$$\mathrm{MT}_l = \max_{\substack{i=1,2,\ldots,n-1,n; \\ t=\eta,\eta+l,\ldots,T-l-n}} \left(\left| w_{i,t}^s - w_{i,t+l}^s \right| \right). \tag{23}$$

Rebalancing a portfolio is not accessible since the investor should pay the transaction cost. To reduce the transaction cost, an investor may desire to keep the portfolio (do not rebalance the portfolio) if the change in the asset's weight is insignificant. We developed two new formulas, one based on the change of individual asset's weight and the other based on the change of multi asset's weight. In the first formula, we assume that the investor would not rebalance the portfolio if each asset's weight changes are less than a pre-defined level C. *Hence* we name this formula 'Lower Bound Single Asset Turnover (LBSAT)'. In the second formula, or 'Lower Bound Multi Asset Turnover (LBMAT)', we assume that the investor would not rebalance the portfolio if the total change of asset's weight is less than a pre-defined level Cm. Let,

$$d_{i,t} = \left| w_{i,t}^s - w_{i,t+l}^s \right|. \tag{24}$$

the first and second formula above could be stated as

$$\mathrm{LBSAT}_l = \frac{1}{T-l-\eta} \sum_{t=\eta}^{T-l} \left[\frac{\left(\sum_{j=1}^{N} d_{i,t} \right)}{max(d_{i,t}) - C} max\left(0, max(d_{i,t}) - C\right) \right], \tag{25}$$

and

$$\mathrm{LBMAT}_l = \frac{1}{T-l-\eta} \sum_{t=\eta}^{T-l} \left[\frac{\left(\sum_{j=1}^{N} d_{i,t} \right)}{\left(\sum_{j=1}^{N} d_{i,t} \right) - C_m} max\left(0, \sum_{j=1}^{N} d_{i,t} - C_m\right) \right]. \tag{26}$$

respectively. The value of C or C_m could be determined subjectively by the investor or examined by a simulation study.

4. Simulation Result

In this simulation study, we generate eight asset paths to simulate the price movement of eight assets in the stock market, following the parameter stated in Table 2. Based on these assets in which the return contains outliers, we calculate the price and obtain data sets with a different, more significant portion of outliers. For each data set, we calculate the market index using the formula (17) by assuming that there are no other stocks in the market. These four sets of data and the obtained market index would be used to simulate the portfolio selection process using a single index model and evaluate the portfolio's performance.

The empirical distribution of the data sets for each simulated stock is presented on Figure 1.

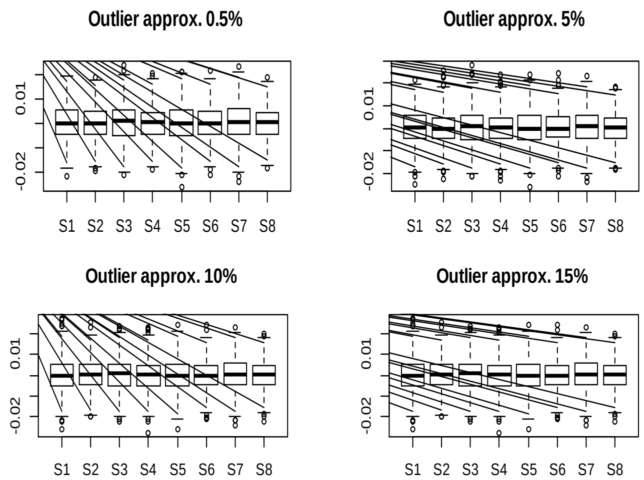

Figure 1: Boxplot of simulated stocks return data

Outliers on each stock, which would be used as the dependent variable in the single index model, became a rationale to use the robust regression method to estimate the beta for each asset. To implement the rolling horizon procedure, we choose 800 as the length of the estimation window. For each stock, 100 estimations of the beta were created based on each data set, using both classical and robust methods. The remaining data would be used as "out of sample" data to examine the portfolio's performance.

All estimation procedures in this study were done using several R packages in the R 3.5.0 software, namely (1) **stats**, for the classical least square estimation; (2) **quantreg**, for the least absolute value estimation; (3) **MASS**, for the M-estimation and the least median of square estimation; (4) **robustbase**, for the least trimmed square estimation and the MM estimation; and (5) **RobPer**, for the τ-estimation. The estimated beta obtained in this step will be used to calculate the optimal portfolio weight using Single Index Model.

4.1 Beta Estimation

The estimated beta value has an essential role in determining asset weight. To study the stability of the beta, we examine the change of estimated beta along the rolling horizon. However, to simplify the presentation, we only present the result of beta estimation for two assets from eight available assets in each data set.

Figure 2 show that for the same data asset return (that is, the second and fourth assets for each portfolio A, B, C, and D), different beta estimator yield different value of beta. Compared to the classical ordinary least square estimator, the second asset's most negligible absolute value (LAV) estimation was lower. In contrast, the same estimator was higher when used in the fourth asset. The M estimator, the MM estimator, and the τ-estimator produce a similar beta to the result of the classical ordinary least square (OLS) estimator. The most different beta resulted from the least median of square (LMS) method, which exhibits more significant variances, and represents non-stable beta along the rolling horizon. The same result was obtained from the rest of the assets.

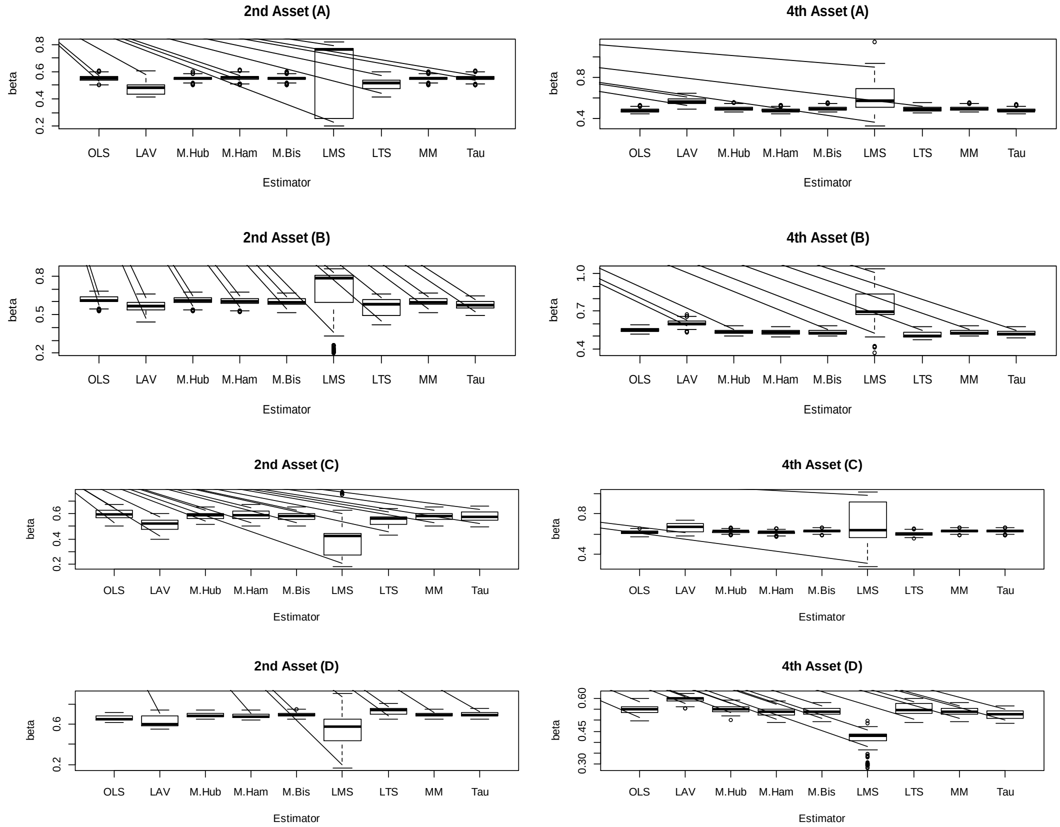

Figure 2: Boxplot of 200 value of estimated beta of the 2nd and 4th asset from data A, B, C, and D, obtained using classical method and several robust method.

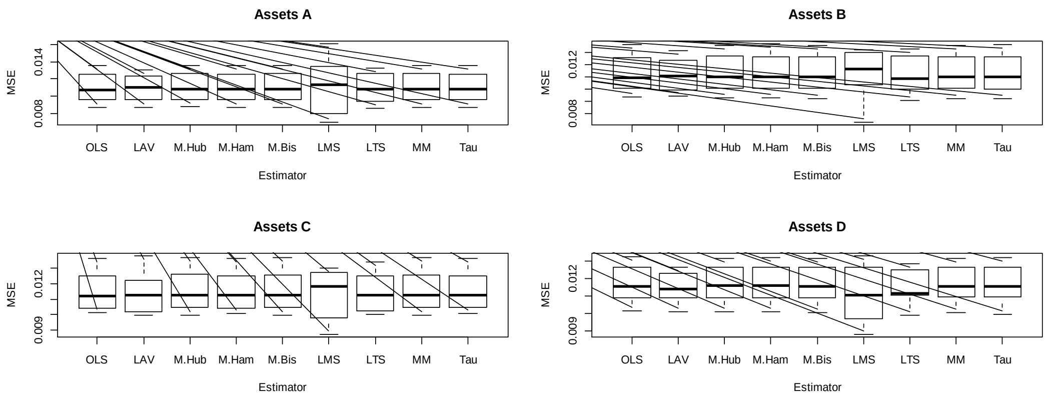

Figure 3: Boxplot of the MSE of beta estimation

Beside the stability of beta, another important issues is the accuracy of these estimators itself. Figure 3 show the mean square error (MSE) obtained for each estimation.

Figure 3 shows that in most cases, the least median of square (LMS) estimator produces a higher mean square error range (MSE) than the other estimation procedures. In several circumstances, the LMS estimator may produce a better model that exhibits a smaller MSE value. However, the MSE resulting from the classical OLS estimator seems to have little difference compared to the MSE resulting from the M-estimator, MM-estimator, and τ-estimator. In addition, the smallest range of MSE resulted from using the least trimmed square (LTS) estimation procedure.

Since the beta just became an input for the single-index portfolio optimization model, the effect of the robust estimator must be looked at not only at the beta itself but also at the characteristic of each generated portfolio. The portfolio's characteristics explanation would consist of the asset weight change and the portfolio performance (out of sample) itself.

4.2 Assets Weight

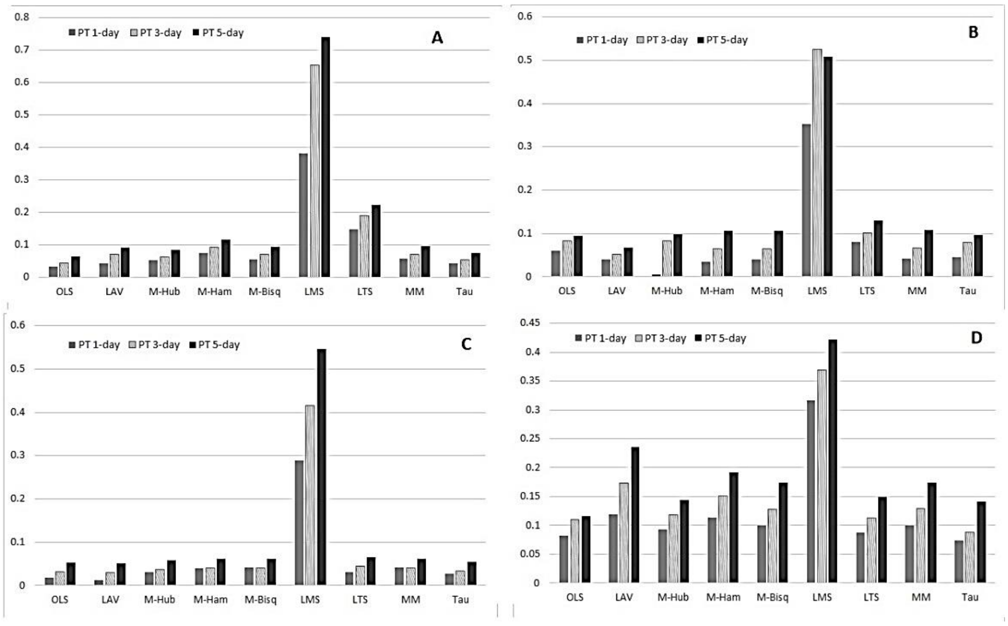

Figure 4: Portfolio Turnover (1-day, 3-day, and 5-day)

In this sub-section we examine the change of asset's weight in the optimal portfolio calculated using Single Index Model with various beta estimators.

As shown in Figure 4, the least median of square (LMS) generates a portfolio with the highest turnover. This condition means that the portfolio weight needed to be more stable and undergo a change in a short time. Generally, for one-day and three days turnovers, the τ-estimator can produce a portfolio with a smaller turnover than the classical ordinary least square estimator. According to figure 4B, the most negligible absolute value (LAV) estimator, the M-estimator, and the MM-estimator also could produce a portfolio with a smaller turnover than the classical one. However, this condition is not held in Figures 4A and 4D.

4.3 Portfolio Performance

For each portfolio calculated at each date using any classical or robust method, we obtain the total return during 100-days, calculate the standard deviation, and obtain the Sharpe Ratio (the ratio between excess return to standard deviation). A higher Sharpe ratio value corresponds to a better portfolio since more return could be obtained per level increase of investment risk.

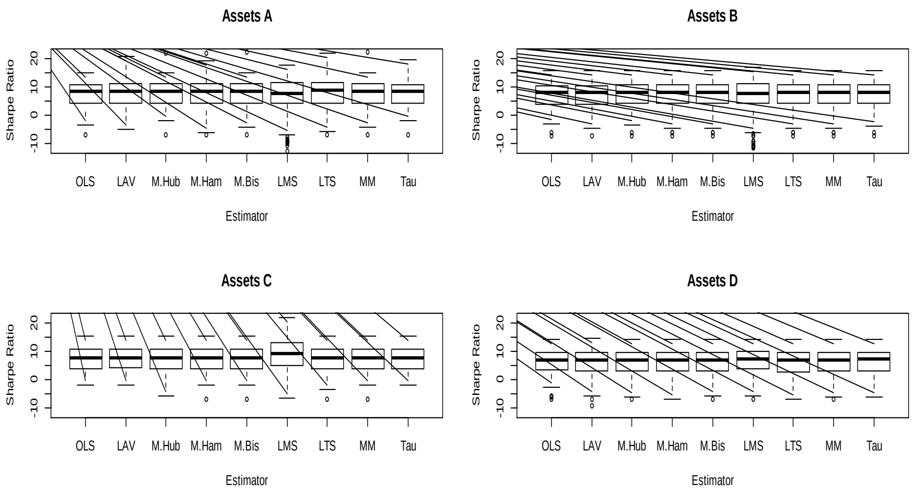

Figure 5: Sharpe Ratio for Portfolio

From Figure 5, it could be deduced that there were minor differences in the Sharpe Ratio of the portfolio resulting from each estimator. Compared to the classical OLS estimator, several robust estimators such as the LAV, LTS, M-estimator, and MM-estimator produce a higher Sharpe Ratio in assets A but lower in assets C and D. In contrast, LMS produces many "outliers" with lower Sharpe Ratios in assets A and B but higher Sharpe Ratios without outliers in assets C. Hence, we could not recommend the usage of this estimator in portfolio optimization based on the Single-Index Model due to the unpredictable result compared to the classical ones.

Among these estimators, only the consistent τ-estimator produces a higher Sharpe Ratio than the classical one in all conditions. In addition, the previous section shows that generally, τ-estimator could be used to calculate optimal portfolio with smaller turnover.

5. Empirical Study

For the empirical study, we chose 36 blue-chip stocks traded in the Indonesia Stock Exchange and obtained the return data from October 2015 to July 2017. Following the same procedure as the simulation study, we estimate the beta for each stock for 20 days and calculate the optimal portfolio weight using Single-Index Model.

In this empirical study, we found that the beta estimator exhibit the same characteristics as presented in the simulation study. Our optimal portfolio consists of 5 to 8 assets, where the least median of square (LMS) estimator corresponds to the highest number of assets (7-8 assets) and the least absolute value (LAV) estimator yield the portfolio with the smallest number of assets (5 or 6 assets). A significant result related to the usage of lower bounded asset turnover, which means that the investor ignores the weight change less than the lower bound, is presented in Tables 3 and 4 below.

Table 3: Value of Lower Bounded Single Asset Turnover with various C

Estimator	3-day turnover				5-day turnover			
	0.00	**0.03**	**0.05**	**0.08**	**0.00**	**0.03**	**0.05**	**0.08**
OLS	0.0284	0.0284	0.0204	0.0099	0.0355	0.0355	0.0340	0.0207
LAV	0.0289	0.0282	0.0263	0.0169	0.0358	0.0358	0.0350	0.0299
M-Huber	0.0392	0.0392	0.0299	0.0233	0.0429	0.0429	0.0413	0.0260
M-Tukey	0.0258	0.0258	0.0169	0.0046	0.0335	0.0335	0.0319	0.0131
M-Hampel	0.0296	0.0296	0.0202	0.0094	0.0358	0.0358	0.0343	0.0155
LMS	0.0275	0.0237	0.0192	0.0068	0.0362	0.0372	0.0372	0.0258
LTS	0.0366	0.0359	0.0286	0.0259	0.0508	0.0508	0.0462	0.0413
MM	0.0323	0.0323	0.0255	0.0145	0.0358	0.0358	0.0342	0.0190
Tau	0.0251	0.0251	0.0151	0.0045	0.0322	0.0322	0.0291	0.0156

Table 4: Value of Lower Bounded Multiple Asset Turnover with various C_m

Estimator	3-day turnover				5-day turnover			
	0.00	**0.15**	**0.20**	**0.25**	**0.00**	**0.15**	**0.20**	**0.25**
OLS	0.0284	0.0244	0.0176	0.0160	0.0355	0.0344	0.0328	0.0235
LAV	0.0289	0.0241	0.0202	0.0151	0.0358	0.0339	0.0325	0.0285
M-Huber	0.0392	0.0354	0.0287	0.0272	0.0429	0.0418	0.0386	0.0314
M-Tukey	0.0258	0.0209	0.0141	0.0108	0.0335	0.0324	0.0308	0.0217
M-Hampel	0.0296	0.0257	0.0190	0.0174	0.0358	0.0348	0.0333	0.0259
LMS	0.0275	0.0217	0.0193	0.0141	0.0362	0.0407	0.0390	0.0354
LTS	0.0366	0.0322	0.0259	0.0241	0.0508	0.0498	0.0436	0.0416
MM	0.0323	0.0295	0.0203	0.0184	0.0358	0.0347	0.0315	0.0223
Tau	0.0251	0.0204	0.0125	0.0106	0.0322	0.0311	0.0264	0.0209

As seen in Table 3 and 4, placing the lower bond would affect the value of portfolio turnover. Since the assets turnover d_i are rarely constant, the rank of estimator based on the portfolio turnover could be changed. In general, the value of 3-day based turnover measure is more affected by the lower bound, compared to the 5-day based turnover measure. However, based on our study, the usage of lower bound either single or multiple still placing the τ-estimator as the smallest turnover measure, and are not affected by the value of the bound. On the contrary, the least trimmed square (LTS) estimator frequently produce portfolio with largest turnover measure. Among the portfolio calculated based on M-estimator, the choice of weight function (e.g. Hampel, Bisquare, or Huber function) significantly affect the portfolio turnover.

Table 5: Performance of Single Index Portfolio obtained by classical Ordinary Least Square (OLS) estimator and several robust estimators.

Statis-tics	Beta Estimator								
	OLS	**LAV**	**M-Huber**	**M-Tukey**	**M- Hampel**	**LMS**	**LTS**	**MM**	**Tau**
Total Return									
Mean	0.0305	0.0307	0.0305	0.0305	0.0305	-0.0033	0.0306	0.0305	0.0305
Max	0.075	0.0844	0.0743	0.0752	0.0745	0.0524	0.0755	0.0743	0.0743
Min	-0.034	-0.0327	-0.0345	-0.0347	-0.0346	-0.0558	-0.0368	-0.0346	-0.0337
Standard deviation of Return									
Mean	0.0108	0.0115	0.0108	0.0108	0.0108	0.0101	0.0107	0.0108	0.0107
Max	0.0171	0.0183	0.0167	0.0170	0.0168	0.0125	0.0163	0.0167	0.0164
Min	0.0058	0.0060	0.0058	0.0058	0.0058	0.0077	0.0059	0.0058	0.0058
Sharpe Ratio									
Mean	2.341	2.046	2.374	2.352	2.367	-0.753	2.372	2.373	2.393
Min	-4.736	-3.919	-4.876	-4.826	-4.872	-7.799	-5.33	-4.873	-4.843
Max	8.106	6.614	7.647	7.915	7.768	4.261	7.167	7.646	7.408

From Table 5, we can analyze the average, minimum and maximum value of the total return, the standard deviation of the return, and the Sharpe Ratio of each portfolio in this empirical study. Compared to these estimators, on average, the least absolute value (LAV) produces the highest return and highest risk. In contrast, the least median of square (LMS) makes the lowest return and, of course, the lowest standard deviation. This result may correspond to the principle of “higher risk yield higher return” on the financial investment. The Sharpe Ratio shows that the τ-estimator produces the best portfolio with the highest average value and minimum value of the Sharpe Ratio. This result was consistent with the simulation study, which placed the τ-estimator as a better candidate to replace the classical estimator.

6. Discussion

In this study, we implement several robust estimation methods to estimate the asset's beta (β) value using historical data. Following beta can be interpreted as an asset's risk measure (Elton et al., 2012). Assets or stocks with a beta ≈ 1 indicates that its price moves in accordance with stock market movements, while beta ≈ of 0 indicates that the asset's price is little affected by changes in the stock market. Applying a robust estimator in beta measurement will produce a better estimate that is not affected by any outlier data or extreme price movement.

This study provides an alternative approach for robust portfolio optimization, in addition to using a robust covariance matrix estimator (eg. Rosadi, 2020; Supandi et al., 2017; Kaszuba, 2012; DeMiguel and Nogales, 2009) as well as using robust risk measures (Kaszuba, 2013b). As well as other studies that implement the single-index portfolio optimization model e.g., the essential benefit of this approach is the opportunity for an investor to choose the best assets in the market (Nandan and Sritavasta, 2017; Yuwono and Ramdhani, 2017). One can calculate the beta of all available assets, but the optimization will pick several assets that perform better than the others.

We affirm that the rolling horizon approach proposed is an important procedures to evaluate portfolio optimization methods by (DeMiguel and Nogales, 2009). They also propose a measure known as portfolio turnover, which represents the change of assets' weight in the portfolio while oldest data are removed and the recent data are added. In this study, we develop new turnover measure namely the *l*-period portfolio turnover, maximum turnover, lower bound single-asset turnover, and lower bound multiple-asset turnover. These turnover measures are defined to represent the investor behavior, by assuming (1) an investor might not change the portfolio before a specific time period, (2) an investor will ignore a slight change in the weight of the optimal portfolio. These turnover measure can be used on any portfolio optimization procedure that studied through rolling horizon approach.

7. Conclusion

In this research, we compare the turnover and the performance of single-index portfolios estimated by classical ordinary least squares and several robust estimators. Compared to the other robust estimators, the τ-estimator is a better candidate to replace the classical standard least square (OLS) to estimate the asset's beta and use the result in portfolio optimization based on Single-Index Model.

Future research is needed to examine the usage of the τ-estimator in the single index or multi-index portfolio, with and without the short sell. Related to the portfolio turnover measure, minimizing turnover may become an interesting topic

References

Ahuja, R., 2017. Sharpe Single Index Model: Evidence from Bombay stock Exchange (BSE) in India. Research Bulletin, 43(1), Pp. 166-179.

Andersen, R., 2008. Modern Methods for Robust Regeression. No. 152, Sage.

Andrews, D, F., 1974. A Robust Method for Multiple Linear Regression. Technometrics. 16, Pp. 523-531.

Begashaw, G. B., and Yohannes, Y. B., 2020. Review of outlier detection and identifying using robust regression model. International Journal of Systems Science and Applied Mathematics, 5(1), Pp. 4-11.

Benati, S., 2015. Using medians in portfolio optimization. Journal of the Operational Research Society, 66(5),Pp. 720-731.

DeMiguel, V., and Nogales, F. J., 2009. Portfolio selection with robust estimation. Operations Research, 57(3), Pp. 560-577.

Dielman, T.E., 2005. Least absolute value regression: recent contributions. Journal of Statistical Computation and Simulation, 75(4), Pp. 263-286.

Drapper, N. and Smith, H., 1999. Applied Regression Analysis, Third edition, John Wiley and Sons, Canada.

Elton, E.J., Gruber, M.J., and Padberg, M.W., 1976. Simple criteria for optimal portfolio selection. The Journal of Finance, 31(5), Pp. 1341-1357.

Elton, E.J., Gruber, M.J., Goetzmann, W.N., Brown, S.J., 2014. Modern Portfolio Theory and Investment Analysis, 12th edition. New York: John Wiley and Sons.

Fabozzi, F.J., Huang, D., and Zhou, G., 2010. Robust portfolios: contributions from operations research and finance. Annals of Operations Research, 176(1), Pp. 191-220.

Goldfarb, D., and Iyengar, G., 2003. Robust portfolio selection problems. Mathematics of operations research, 28(1), Pp.1-38.

Grossi, L., and Laurini, F., 2011. Robust estimation of efficient mean–variance frontiers. Advances in Data Analysis and Classification, 5(1), Pp. 3-22.

Huo, L., Kim, T.H., and Kim, Y., 2012. Robust estimation of covariance and its application to portfolio optimization. Finance Research Letters, 9(3), Pp. 121-134.

Kaszuba, B., 2012. Applications of robust statistics in the portfolio theory. Mathematical Economics, 8(15), Pp. 63-82.

Kaszuba, B., 2013. Empirical Comparison of Robust Portfolios' Investment Effects. The Review of Finance and Banking. 85, Pp. 47-61.

Kellerer, H., Mansini, R., and Speranza, M. G., 2000. Selecting portfolios with fixed costs and minimum transaction lots. Annals of Operations Research, 99(1-4), Pp. 287-304.

Lauprete, G.J., Samarov A.M., and Welsch, R.E., 2002. Robust Portfolio Optimization. Metrika, 25, Pp. 139-149.

Lauprete, G. J., Samarov, A. M., and Welsch, R. E., 2003. Robust portfolio optimization. In Developments in Robust Statistics, Pp. 235-245. Physica, Heidelberg.

Markowitz, H., 1952. Portfolio selection. The journal of finance, 7(1), Pp. 77-91.

Maronna, R. A., Martin, R. D., and Yohai, V., 2006. Robust statistics Theory and Methods. John Wiley and Sons, Chichester.

Maronna, R. A., and Zamar, R. H., 2002. Robust estimates of location and dispersion for high-dimensional datasets. Technometrics, 44(4), Pp. 307-317.

Massart, D. L., Kaufman, L., Rousseeuw, P. J., and Leroy, A., 1986. Least median of squares: a robust method for outlier and model error detection in regression and calibration. Analytica Chimica Acta, 187, Pp. 171-179.

Murthy, J., 2018. The Construction of Optimal Portfolio Using Sharpe's Single Index Model-An Empirical Study on Nifty Metal Index. Sumedha Journal of Management, 7(1), Pp. 126-134.

Nandan, T., and Srivastava, N., 2017. Construction of Optimal Portfolio Using Sharpe's Single Index Model: An Empirical Study on Nifty 50 Stocks. Journal of Management Research and Analysis, 4(2), Pp.74-83.

Partono, T., Yulianto, A., and Vidayanto, H., 2017. The Analysis of Optimal Portfolio Forming with Single Index Model on Indonesian Most Trusted Companies. International Research Journal of Finance and Economics, (163),Pp. 50-59.

Perret-Gentil, C. and Victoria-Feser, M-P., 2003. Robust Mean-variance Portfolio Selection. Cahiers du departement d'Econometrie, Faculte des Sciences economiques et sociales, Universite de Geneve.

Rosadi, D., Setiawan, E. P., Templ, M., and Filzmoser, P., 2020. Robust covariance estimators for mean-variance portfolio optimization with transaction lots. Operations Research Perspectives, 7, 100154.

Rousseeuw, P. J., and Leroy, A. M., 2005. Robust regression and outlier detection (Vol. 589). John wiley and sons.

Salibian-Barrera, M., Willems, G., and Zamar, R., 2008. The fast-τ estimator for regression. Journal of Computational and Graphical Statistics, 17(3), Pp. 659-682.

Sharpe, W. F., 1963. A simplified model for portfolio analysis. Management science, 9(2), Pp.277-293.

Setiawan, E. P., and Rosadi, D., 2017. Robust Single-Index Model with Adjusted Beta: A Case-Study in Indonesia Stock Exchange. The 2nd International Statistical Institute-Regional Statistics Conference, Bali, Indonesia.

Supandi, E. D., and Rosadi, D., 2015, August). An application of Constrained M-Estimator in construction of robust portfolio. In Research and Education in Mathematics (ICREM7), 2015 International Conference on. Pp. 268-273. IEEE.

Supandi, E. D., and Rosadi, D., 2017. An Empirical Comparison between Robust Estimation and Robust Optimization to Mean-Variance Portfolio. Journal of Modern Applied Statistical Methods, 16(1), Pp. 32.

Treynor, J. L., and Black, F., 1973. How to use security analysis to improve portfolio selection. The Journal of Business, 46(1), Pp. 66-86.

Tütüncü, R. H., and Koenig, M., 2004. Robust asset allocation. Annals of Operations Research, 132(1-4), Pp. 157-187.

Vaz de Melo Mendes, B., and Pereira Câmara Leal, R., 2005. Robust multivariate modeling in finance. International Journal of Managerial Finance, 1(2), Pp. 95-106.

Welsch, R. E., and Zhou, X., 2007. Application of robust statistics to asset allocation models. Revstat, 5(1),Pp. 97-114.

Wilcox, R. R., 2011. Introduction to robust estimation and hypothesis testing. Academic Press, London.

Yuwono, T., and Ramdhani, D., 2017. Comparison Analysis of Portfolio Using Markowitz Model and Single Index Model: Case in Jakarta Islamic Index. Journal of Multidisciplinary Academic, 1(1), Pp. 25-31.

Yohai, V. J., 1987. High breakdown-point and high efficiency robust estimates for regression. The Annals of Statistics, Pp. 642-656.