APPLICATIONS OF ROBUST STATISTICS
IN THE PORTFOLIO THEORY[*]

Bartosz Kaszuba

Abstract. The appropriate selection of portfolio components and determining their weights have a significant influence on the later performance of the investor. The classical method of calculating the weights of individual components in mean variance portfolios is based on sample mean and sample covariance matrix, which are optimal when the data come from multivariate normal distribution. In practice, the distribution of stock returns is not a normal distribution and frequently (albeit to a small extent) is contaminated by outliers; therefore, theoretically, a better approach to determine optimal weights in a portfolio would be to apply robust estimation methods. The main contribution of this paper is to present the possibilities of applying robust statistics methods in the Markowitz portfolio theory. This article contains an overview of the most important robust estimators applied in the portfolio theory. All the methods have been grouped according to the method of determining the outliers and to the accepted disorder models. Moreover, it presents the relevant achievements to date and the results of empirical research in this field. It also shows the potential problems resulting from the practical application of the robust estimation in the rolling horizon.

Keywords: robust statistics, portfolio asset allocations, robust portfolio estimation, robust risk measures.

JEL Classification: 62G35, 91G10.

1. Introduction

The portfolio theory proposed by Markowitz is based, among other elements, on an assumption that decisions are made solely on the basis of the expected return and risk (measured with variance or standard deviation), yet the major problem in practice is their estimation. In the case of assessing both parameters at the same time, classic estimators are encumbered with significant estimation error, which makes the portfolios burdened with

Marek Kośny
 Department of Operations Research, Wrocław University of Economics, Komandorska Street 118/120, 53-345 Wrocław, Poland.
 E-mail: marek.kosny@ue.wroc.pl

an even higher risk, and, as a consequence, makes them have low out of sample performance (Jagannathan, Ma, 2003, p. 1652).

Michaud (1989), Black, Litterman (1992), Chopra, Ziemba (1993) demonstrated that outside the sample, classic portfolios do not have good properties, which influence both sensitivity and portfolio estimation error. This makes portfolios – which are efficient in the case of long-term investments for multiple periods – fail to grant good investment effects. In their works, Jagannathan, Ma (2003) as well as DeMiguel, Nogales (2009) presented research concerning the comparison of alternative and classic estimation methods; nonetheless, they did not compared mean-variance portfolios with a target mean return, only portfolios with minimum variance. The reason is the high instability of the sample mean, which would impair the quality of the obtained results. The problems related to the determination of mean-variance portfolios and estimation errors while applying classic covariance methods are also described in other works (including: Jorion, 1986; Best, Grauer, 1991; Wang, 2005; Jobson, Korkie, 1980).

The aforementioned research causes an on-going analysis of alternative portfolio construction methods, which will enable a more precise estimation of portfolio weights and will contribute significantly to improving investment effects in practice. One of the concepts proposed in the literature is to apply robust statistics methods, which allows decreasing the impact of the outliers on the estimators' values.

The further part of this paper focuses on applications of the robust statistics methods in the portfolio theory and is organized as follows: Section 2 presents classic portfolio optimization issues and demonstrates differences between classic and robust estimation, which lead to two different approaches to portfolio estimation: the one-step approach and the two-step approach. It also presents the most important properties of robust estimators, which are then interpreted in the context of applications in the portfolio theory. Section 3 describes the one-step approach to portfolio estimation; Section 4 features classification and description of robust methods applied in a two-step approach to portfolio estimation. The last part contributes possible practical problems regarding the application of robust estimation in portfolio theory.

2. Robust estimation of portfolios

In this paper we consider random vector of returns of n assets $\mathbf{R} = (\mathbf{R_1}, ..., \mathbf{R_n})$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Efficient

portfolio with a target mean return μ_0 is a solution to the following optimization problem:

$$\min_{w} \mathbf{w'}\mathbf{\Sigma}\mathbf{w},$$

subject to:

$$\mathbf{w'}\mathbf{\mu} \geq \mu_0,$$
$$\mathbf{w'}\mathbf{1} = 1.$$

Disregarding the $\mathbf{w'}\mathbf{\mu} \geq \mu_0$ constraint, the resulting portfolio is a minimum-variance portfolio.

In the aforementioned case, the expression $\mathbf{w'}\mathbf{\Sigma}\mathbf{w}$ can be equally noted as

$$\sigma^2_{port} = \mathbf{w'}\mathbf{\Sigma}\mathbf{w}, \tag{1}$$

where σ^2_{port} is a portfolio variance of the sum $w_1\mathbf{R_1} + ... + w_n\mathbf{R_n}$. Analogically, one can notice a similar equality for the portfolio return: $\mu_{port} = \mathbf{w'}\mathbf{\mu}$. As, in practice, distribution of stock returns is unknown, efficient portfolios are determined using sample mean of portfolio returns $\hat{\mathbf{\mu}}$ and sample covariance matrix $\hat{\mathbf{\Sigma}}$.

The discussion in the previous section shows that robust estimators should be more appropriate to the portfolio optimization problem. Thus, portfolio variance or covariance matrix in (1) can be estimated by means of robust estimators, which decrease the influence of extreme returns. It should be emphasized that when $\hat{\sigma}^2_{port}$ is a robust estimate of portfolio variance and $\hat{\mathbf{\Sigma}}$ is a robust estimate of covariance matrix, the following inequality occurs:

$$\min_{w} \hat{\sigma}^2_{port} \neq \min_{w} \mathbf{w'}\hat{\mathbf{\Sigma}}\mathbf{w}. \tag{2}$$

Thus, we can distinguish two different approaches to portfolio optimization: the one-step approach (described in Section 3) consists in solving the optimization problem without prior estimation of location and scale parameters, and the two-step approach (described in Section 4) consists in the estimation of the covariance matrix (step one), followed by solving the optimization problem (step two).

The next part of the article elaborates on the most important terms related to robust statistics and their interpretation in the portfolio theory. Except for the definitions listed, there are also others which enable a comparison of

various robust estimators. Some robust properties are asymptotical and sample properties. While the asymptotic properties present qualities of robust estimators in the entire sample, the sample properties are much more useful in practice, as they allow to compare estimators on a bounded sample size – just as happens in practice.

Breakdown point

A measure of the global estimator robustness is the breakdown point, proposed by Hampel (1968, 1971), which can be interpreted as the smallest fraction of "bad" data for which the estimator can take arbitrary large values. Formally, the breakdown point of an estimator T at a distribution F, denoted by $\varepsilon_{T,F}^*$, is the highest $\varepsilon^* \in (0,1)$ such that:

$$\mathop{\forall}_{\varepsilon < \varepsilon^*} \mathop{\forall}_{G \in \blacklozenge} T((1-\varepsilon)F + \varepsilon G) \text{ is bounded.}$$

For practical applications, the finite sample breakdown point,[1] introduced by Donoho and Huber (1983), is more useful; it can be interpreted as the lowest share of observations in a sample, for which the estimator can take arbitrary large values. Formally, for sample $\mathbf{x}_n = (x_1, x_2, \ldots, x_n)'$, the size of which is n, the finite sample breakdown point for estimator T_n denoted by ε_n^* is defined as follows:

$$\varepsilon_{T_n, F_n}^* = \frac{m}{n},$$

where m is the lowest amount of observations for which estimator T_n from sample $\mathbf{x}_n$, in which m observations will be replaced with arbitrary large observations $y_1, y_2, \ldots, y_m$, is bounded.

It should be noticed that classic estimators have a breakdown point of 0%, while robust estimators have a breakdown point greater than 0%. By adjusting the control constants of robust estimators, the specific breakdown point of robust estimators can be achieved.

---

[1] Researchers specify various types of sample breakdown point, depending on the method of contaminating data in the sample. If in a sample sized n, m observations are contaminated, we deal with: finite sample replacement breakdown point (Tyler, 1994) or ε-replacement breakdown point (Huber, Ronchetti, 2009). If, on the other hand, m contaminating observations are added to an n-sized sample, we deal with: finite sample addition breakdown point (Zuo, 2000) or ε-contamination breakdown point. From the perspective of application, the most frequently discussed and practical one is the finite sample replacement breakdown point.

In the context of the portfolio theory, the selection of a breakdown point can be specified on the basis of sensitivity of the selected portfolio components, in that if the portfolio contains companies which are more susceptible to the outliers (e.g. companies with low capitalization), then the breakdown point of the applied estimators should be higher than in the case of companies among which the outliers occur more seldom.

Influence function

The influence function shows the behavior of an estimator when the sample is contaminated by an infinitesimally small fraction of outliers. The influence function (IF) of an estimator T at a distribution F, in the point x is defined as (Hampel (1974):[2]

$$\text{IF}_{T,F}(x) = \lim_{\varepsilon \to 0^+} \frac{T((1-\varepsilon)F + \varepsilon\delta_x) - T(F)}{\varepsilon},$$

where δ_x is a point-mass at the point x. When an estimator has the bounded influence functions for all x, then it is called a robust estimator. After substituting empirical distribution function $\hat{F}$ to F in IF definition, one shall obtain an empirical influence function, described in detail in Section 5. If the estimator influence function is bounded, the asymptotic variance of this estimator is limited as well.

In the context of the portfolio theory, the influence function for the vector of weights in the portfolio can be analyzed. If the influence function of portfolio weights is bounded, the given portfolio is less sensitive to outliers; therefore, for those observations, robust portfolios are more stable than classic portfolios, for which the influence functions are unbounded. Additionally, it has been demonstrated (see Perret-Gentil, Victoria-Feser, 2004) that portfolios calculated by means of robust estimators of the location and scale parameters are robust portfolios, as the influence function of the efficient portfolio weights estimator depends only on the influence function of the location and scale estimator.

Based on the influence function, it is possible to analyze other properties of estimators, such as gross-error sensitivity and rejection point. The former is defined as $\sup_x | \text{IF}_{T,F}(x) |$ and it can be used to identify outliers (Perret-Gentil, Victoria-Feser, 2004). A rejection point is defined as the

---

[2] The influence function described in this part was considered by Hampel as an influence curve (IC).

smallest distance for which all observations exceeding the same have no influence on the estimator.

Mahalanobis distance

The Mahalanobis distance is used to specify how distant the given observation is from the data center. The Mahalanobis distances are defined as follows:

$$d_i = d(\mathbf{x}_i, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \sqrt{(x_i - \hat{\boldsymbol{\mu}})' \hat{\boldsymbol{\Sigma}}^{-1} (x_i - \hat{\boldsymbol{\mu}})}.$$

The outliers are identified, among other methods, by a comparison to critical values of chi-square[3] distribution, with n degrees of freedom. Robust distances are modified Mahalanobis distances in which the sample covariance matrix is replaced with robust covariance matrix.

Affine equivariance

Affine equivariance is related to the estimation of multidimensional estimators of location and scale. Estimators $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are affine equivariant if for each non-singular matrix $\mathbf{A}$ and vector $\mathbf{b}$, the following proceeds:

$$\hat{\boldsymbol{\mu}}(\mathbf{AX} + \mathbf{b}) = \mathbf{A}\hat{\boldsymbol{\mu}}(\mathbf{X}) + \mathbf{b}, \quad \hat{\boldsymbol{\Sigma}}(\mathbf{AX} + \mathbf{b}) = \mathbf{A}\hat{\boldsymbol{\Sigma}}(\mathbf{X})\mathbf{A}'.$$

Classic estimators for a normal distribution sample have such a property. Although affine equivariance is a desirable characteristic among robust estimators, most affine equivariant estimators could be time-consuming compared to classical estimators. For most of these methods, there is no exact algorithm; therefore, the affine equivariance is often abandoned in favor of pairwise robust covariance estimators (described in Section 4), which can be calculated much faster.

In the context of the portfolio theory, for a problem $\min_w \mathbf{w}' \boldsymbol{\Sigma} \mathbf{w}$, determination of covariance matrix from data set $\mathbf{X} = (\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n})'$, for an affine equivariant estimator is equivalent to determination of covariance matrix from data set $\mathbf{AX}$, where matrix $\mathbf{A} = (\hat{a}_{ij})_{i,j=1,\ldots,p}$, and $a_{ii} = 1$, $a_{ij} = 0$, for $i \neq j$. As matrix $\mathbf{A}$ is non-singular, the aforementioned statement is false, as at least one component's weight in the portfolio amounts to 0 (at least one $a_{ii} = 0$).

---

[3] If the assumed distribution $\mathbf{F}$ is normal. This is caused by the fact that if $\mathbf{x_i} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $d^2(\mathbf{x}_i, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \sim \chi_p^2$.

The affine equivariance is also desirable when analyzing excess returns or portfolio with foreign assets. In the first case, excess return is defined as $r_t^e = r_t - r_f$, where r_t is return in period t, r_f is a risk-free rate; hence, when the estimator is not affine equivariant, then $\hat{\boldsymbol{\mu}}(\mathbf{X} + r_f) \neq \hat{\boldsymbol{\mu}}(\mathbf{X}) + r_f$, which is unacceptable in practice.

In the second case, we can assume that the Polish investor portfolio consists of two assets: US stock and Polish stock. Therefore, it is convenient to express foreign returns in Polish currency by means of the following simplified model: $r_{PLN} = r_x + r_{US} + r_x r_{US}$, where r_x is the exchange return and r_{US} is a return on US investment. Thus, in this case, we have the following matrix $\mathbf{A}$ and vector $\mathbf{b}$: $a_{11} = 1 + r_x, a_{22} = 1, a_{12} = a_{21} = 0$, $b = (r_x, 0)'$ and affine equivariance of an estimator is required by investors.

3. One-step approach

If we assume $\hat{\sigma}_{port}$, to be a sample standard deviation, we obtain the following optimization problem:

$$\min_{w \in W} \frac{1}{n} \sum_{t=1}^{n} (\mathbf{wr_t} - \hat{\mu}_{port})^2, \qquad (3)$$

where $\mathbf{w}$ is a vector of portfolio weights, W is a set of constraints (e.g. $W = \{\mathbf{w} : \mathbf{w'1} = 1\}$), $\mathbf{r_t}$ is a vector of asset returns in period t, $\hat{\mu}_{port}$ is a sample mean of portfolio returns.

Similarly to the case of linear regression or classic estimation of location and scale parameters, such a problem is sensitive to outliers, thus in order to decrease the influence of outliers, function $(\cdot)^2$ is substituted with function $\rho(\cdot)$, which allows to decrease the influence of outliers.

The next section presents estimators of robust portfolios which have been introduced in the literature to date, together with additional proposals of LTS and LMS portfolios, corresponding to LTS and LMS estimators in linear regression.

Least absolute deviations portfolio (LAD)

The LAD portfolio is a classic example of a modification to an optimization problem (3). A LAD portfolio is determined by minimizing the least absolute deviations, as follows:

$$\min_{w \in W, m \in R} \frac{1}{n} \sum_{t=1}^{n} | \mathbf{wr_t} - m |.$$

For LAD portfolios, the m parameter which minimizes the objective function is the median.

Least trimmed squares portfolio (LTS)

The principle of LTS portfolios, which employs the least trimmed squares method, consists in the determination of a set of h observations for which the portfolio variance is the lowest. LTS portfolios are determined by solving the following problem:

$$\min_{w \in W, m \in R} \frac{1}{h} \sum_{i=1}^{h} (| \mathbf{w'r}_{\pi(i)} - m |_{i:n})^2,$$

where $h \leq n$ defines the number of rejected observations ($h-n$ observations are rejected) and $\mathbf{r}_{\pi(i)}$ is such observation for which $| \mathbf{w'r}_{\pi(i)} - m |$ is i-th order statistics. The LTS method was first proposed by Rousseeuw (1984) in the context of regression. In the case of covariance matrix estimation, this method is equivalent to the MCD method described in Section 4.1. The aforementioned case is a good illustration of the differences and similarities between the robust estimation of covariance matrix and the estimation of portfolios' risk. In this case, the fundamental difference is the method of rejecting outliers: in MCD, Mahalanobis distances are used to identify the outliers, whereas in the case of LTS portfolios, distances in the Euclidean norm are applied. Therefore, the selected observation can be classified as an outlier using the MCD method, whereas it will not be rejected when applying LTS.

Least median of squares portfolio (LMS)

The LMS portfolios, which apply the least median of squares method, are determined by solving the following problem:

$$\min_{w \in W, m \in R} (\mathbf{wr_t} - m)_{h:n},$$

where $h \leq n$ defines the number of rejected observations ($h-n$ observations are rejected). The LMS method was first proposed by Rousseeuw (1984) in the context of regression. For $h = n/2$, the described portfolio is an LMS portfolio, whereas for $h > n/2$, we obtain an α-quantile portfolio

(where $\alpha = h/n$). Just like in the case of LTS portfolios, which correspond to the MCD method, an LMS portfolio has its counterpart in a covariance matrix determined using the MVE method.

M-portfolio

M-portfolios (also known as Huber portfolios) were proposed by Lauprete (2001) and later investigated by DeMiguel, Nogales (2009). Just like in the case of M-estimators, they consist in substituting ρ function for the square function in problem (3). In this case, an M-portfolio is determined by solving the following problem:

$$\min_{w \in W, m \in R} \frac{1}{n} \sum_{t=1}^{n} \rho(\mathbf{wr_t} - m),$$

where ρ denotes a convex symmetric function with a unique minimum at zero. Lauprete suggested using the Huber function in which the k constant is determined in two stages using the LAD portfolio.

S-portfolios

S-portfolios were proposed by DeMiguel, Nogales (2009). Similarly to S-estimators, the S-portfolios with minimum variance are determined as follows:

$$\min_{w \in W:\, m,\, s \in R} s,$$

where s complies with:

$$\frac{1}{n} \sum_{t=1}^{n} \rho(\frac{\mathbf{wr_t} - m}{s}) = K,$$

where K is tuning constant.[4] In the aforementioned case, the ρ function should meet the requirement for the M-portfolios, and additionally it should be strictly increasing on $[0,c)$, and constant on $[c,\infty)$ for certain $c > 0$. DeMiguel, Nogales proposed using the bisquare function.

If we assume the ρ function to be $\rho(x) = \mathbf{1}_{(-1,1)}(x)$ and $K = 0.5$, the resulting S-portfolio will be an LMS portfolio.

---

[4] Selected analogically to S-estimators: $K = E_F(\rho(x))$, where F – assumed underlying distribution.

4. Two-step approach

The two-step approach consists in the estimation of the covariance matrix (step one), followed by solving the optimization problem to determine the efficient portfolio (step two). The literature presents many different robust methods, categorized into various groups based on the method of determining them or on their properties.

There are two main types of robust estimators, set apart by the method of rejecting outliers. The first model (Tukey-Huber Model) assumes that the given multi-dimensional observation comes either from distribution F or from contaminating distribution H. So if stock returns are analyzed, it is assumed that either all or no returns from the given period are outliers. The second model (FICM model) is more general; it was proposed by Alqallaf et al. (2009), and, in the case of daily rates of return, it assumes than only selected returns from a given day can be outliers. Therefore, the first model is more adequate, for example, for companies from a given sector, whereas the second one can be used to analyze companies coming from different sectors, which are less correlated.

The Tukey-Huber model, known also as ε-contaminated model or Fully Dependent Contamination Model (FDCM), is defined as follows:

$$\mathbf{X} = (1 - B)\mathbf{Y} + B\mathbf{Z},$$

where $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ are p-dimensional vectors, $\mathbf{Z} \sim F_0$ is some outlier generating distribution, whereas $\mathbf{Y} \sim H$ is some elliptical distribution, $B \sim \mathrm{B}(1, \varepsilon)$, where $\mathrm{B}(1, \varepsilon)$ is a binomial distribution with probability of success ε.

Fully independent contamination model (FICM) is defined as follows:

$$\mathbf{X} = (1 - \mathbf{B})\mathbf{Y} + B\mathbf{Z}$$

where $\mathbf{B} = diag(B_1, ..., B_p)$ is diagonal matrix, $B_i \sim \mathrm{B}(1, \varepsilon_i)$ and B_i are independent. If B_i are fully correlated, then the FICM model becomes a classic Tukey-Huber model.

Alqallaf et al. (2009) demonstrate that the former of these models is adequate for affine equivariant estimators, while the latter one is appropriate for methods based on pairwise robust correlation or covariance estimates, thus robust portfolios generated by means of robust covariance matrices can be categorized into the following groups:

1. Efficient portfolios assuming FDCM (FDCM portfolios): it is assumed that the occurrence of outliers can be usually observed on the same day for all companies included in the portfolio. Such portfolios should be the most effective for the analysis of companies from the same sector or trade.

2. Efficient portfolios assuming FICM (FICM portfolios): it is assumed that outliers occur independently for each company in the portfolio, whereas the occurrence of an outlier for each company from the portfolio on the same day is very rare. Such portfolios should be the most effective for analyzing companies from various sectors with different characteristics.

4.1. FDCM Portfolios

The FDCM portfolio group encompasses all portfolios based on affine equivariant estimators of covariance matrix. In this group, two main methods, depending on the method of determining the matrix, can be distinguished:

- Methods based on covariance matrix estimation for elliptic distributions.
- Methods based on projection pursuit.

The first group includes for example: M-estimators, S-estimators, MVE or MCD, whereas the most popular estimator belonging to the second group is the Stahel-Donoho estimator.

M-estimators, introduced by Maronna (1976), the main disadvantage of which is that for a great number of dimensions they have a very low breakdown point.

Constrained M-estimators (CM − estimators), proposed by Kent and Tayler (1996), combine properties of good local robustness of M-estimators, and good global robustness of S-estimators. Thanks to tuning constants, CM-estimators enable the appropriate selection of the influence function and the estimator efficiency. Moreover, modification of estimator efficiency has no influence on the breakdown point (Kent, Tyler, 2001).

S-estimators were first introduced (in the context of regression) by Rousseeuw, Yohai (1984), whereas in the context of estimating covariance matrix, they were introduced by Davies (1987) as well as described and compared to M-estimators by Lopuhaa (1989). Rocke (1996) demonstrated that when the number of dimensions is large, even with a breakdown point close to 50%, the M-estimators are sensitive to outliers; therefore, he proposed to apply translated biweight function, or biflat function, depending on the tuning constants which allow the specified point to be reached. Portfo-

lios based on S-estimators with abiweight function were examined by Perret-Gentil, Victoria-Feser (2004).

MVE and MCD estimators (Minimum Volume Ellipsoid, Minimum Covariance Determinant) were introduced by Rousseeuw (1984, p. 877), and described in detail in another publication by the same author (see Rousseeuw, 1985). The MVE estimator is a generalization of the least median of squares (LMS) estimator. The MVE estimator is used as an initial estimator to calculate S-estimators, which is influenced by its low maximum bias. The MCD estimator is a generalization of the Least Trimmed of Squares (LTS) estimator. The breakdown point of the MCD estimator is the same as the breakdown point of the MVE estimator, yet the MCD has more advantages than the MVE (Butler, Davies, Jhun, 1993; Davies 1992). Portfolios based on MCD estimators were investigated by Zhou (2006) and Welsch, Zhou, (2007), and, in a modified version, by Mendes, Leal (2005).

Stahel-Donoho Estimator (SDE) was defined independently by Stahel (1981) and Donoho (1982); it was the first equivariant estimator of the location and scale parameter for multidimensional observations to be characterized by a high breakdown point, regardless of the number of dimensions. The SDE estimator employs projection pursuit methodology; the method of determining this estimator is described, for example, by Maronna, Martin, Yohai (2006). Maronna, Yohai (1995) demonstrated the high efficiency of the SDE estimator, both for multidimensional normal distribution, and for Cauchy distribution. Maronna, Zamar (2002) showed good properties of the SDE estimator for simulation data, yet in the case of real data, the SDE estimator required large amounts of data to maintain a high breakdown point. Maronna, Yohai (1995) also demonstrated that the SDE estimator has better qualities than comparable S-estimators and M-estimators.

Other robust estimators of covariance matrix

There also exist research studies done on other robust estimators, less frequently applied in practice, which include: MM-estimators, described initially by Yohai (1987) in the context of regression and examined further by Lopuhaa (1992) as well as by Tatsuoka, Tyler (2000). A detailed description can be found in Salibian-Barrera, Van Aelst, Willems (2006). Minimum weighted covariance determinant, described by Roelant, Van Aelst, Willems (2009), has the same breakdown point as the MCD estimator, whereas its efficiency in multidimensional distributions of t-Student is higher (yet still

remains at a rather low level). Other estimators, for instance: Nearest-Neighbor Variance Estimator (Wang, Raftery, 2002), τ-estimator of location and scale (Lopuhaa, 1991); estimators based on projection pursuit method, e.g. affin-equivariant location estimator of Donoho-Gasko (Donoho, Gasko, 1992), P-estimator of covariance matrix (Maronna, Stahel, Yohai, 1992).

4.2. FICM Portfolios

The FICM portfolios group includes portfolios which apply methods based on pairwise robust correlation or covariance estimates. In terms of the estimation method, the following three method groups can be distinguished (e.g. Chilson et al., 2004 or Alqallaf et al., 2002):

• Methods based on classical rank estimators – these methods apply classical rank estimators, such as Spearman's ρ or Kendall's τ.

• Methods consisting in the rejection of outliers for each random variable, followed by a calculation of covariance for two variables – one example of such an estimator can be the QC estimator (Huber, 1981, pp. 203-204).

• Two-dimensional methods of rejecting outliers, such as Gnanadesikan-Kettenring Estimator (Gnanadesikan, Kettenring, 1972), 2D-Winsorization method (proposed by Khan, Van Aelst, Zamar, 2007), or 2D-Huber (proposed by Chilson et al., 2004). The last two were used by Welsch, Zhou (2007) in order to construct robust portfolios.

For the aforementioned methods, the obtained matrix is neither affine equivariant nor positive-definite. For this purpose, the method of orthogonalization is used, proposed by Maronna, Zamar (2002), which allows to obtain a positive-definite matrix and an "approximately" affine equivariant matrix.

Algorithms of estimating robust covariance matrices were described by Maronna, Martin, Yohai (2006) and implemented in the rrcov package of the R program (Todorov, Filzmoser, 2009) for the following estimators of covariance matrices: CM-estimators, S-estimators, orthogonalized Gnanadesikan-Kettenring estimator, MVE, MCD, MM and Stahel-Donoho estimator.

5. Rolling portfolios and stability of weights

This section presents the undesirable effects of applying robust estima-tors, such as an increase in transaction costs. From the point of view of an

investor, the stability of weights in a portfolio constructed by them through-out the entire duration of the investment is a significant element.

In most cases, the researchers analyze out of sample portfolios beha-vior. For this purpose, rolling portfolios were compared, which are deter-mined in the following manner: at period t, weights of optimal portfolio were determined on the basis of $T < N$ last observations (estimation win-dow), where N is the total number of observations. Next, the time series of the obtained portfolio returns were analyzed at period $t+1$, and weights determined at period $t+1$ with weights determined at period t.

Assuming that the investor constructs rolling portfolios in accordance with the above described methodology, it is important for the difference between weights determined at period t (on the basis of the last n observa-tions) and weights determined at period $t+1$ (also on the basis of the last n observations) to be as low as possible throughout the entire duration of the investment; therefore, it is important that a change in one observation does not significantly influence the weights on the portfolios. To simplify, one can investigate the difference between weights determined at period t (on the basis of the last n observations), and weights determined at period $t+1$ (on the basis of the last $n+1$ observations), which results in an empirical influence function (also known as sensitivity curve), defined as follows (Croux, 1998):

$$\text{EIF}_{T,F_n}(x_{t+1}) = (n+1)(T_{t+1}(x_1,...,x_t,x_{t+1}) - T_t(x_1,...,x_t)), \qquad (4)$$

where $T_t = T(F_t)$ is an estimator of portfolio weights and F_t is an empirical distribution function from a sample of size t.

Using the Taylor expansion, it is possible to approximate the empirical influence function (Rousseeuw, Leroy, 1987, p. 186):

$$\text{EIF}_{T,F_t}(x) \approx \text{IF}_{T,F}(x).$$

Thus, the influence function in a suitably large sample approximates the empirical influence function well. Knowing the shape of the influence func-tion of robust portfolio weights, one can calculate the approximate maxi-mum change for the given observation.

In the case of M-portfolios, the influence function of the M-portfolio and S-portfolio weights (DeMiguel, Nogales, 2009, pp. 567-568) is propor-tional to the ψ function, thus portfolios with the Huber function or LAD[5]

---

[5] LAD portfolios are M-portfolios with function $\rho(x) = |x|$.

portfolios have a bounded influence function, which for the classic portfo-lios[6] is unbounded.

Similarly, for portfolios constructed by means the two-step approach, it is possible to demonstrate that their influence function is bounded, provided that the influence functions of the location and scale estimators applied in determination of optimal weights are bounded as well (Perret-Gentil, Victoria-Feser, 2004).

Unfortunately, decreasing sensitivity to outliers increases sensitivity to lesser observations (especially in a small sample). In practice, the share of outliers is minor, thus for a rolling portfolio, most observations cause grea-ter changes in weights for robust portfolios than for classic ones, whereas only for a small number of observations (outlying ones), robust portfolios are less sensitive than classic portfolios. Hence, as the breakdown point for the given estimator increases, its sensitivity for lesser observations grows, while its sensitivity to observations which are more distant from the bulk of data decreases. Such an effect can be observed, for instance, in the results of research conducted by DeMiguel, Nogales (2009), where for portfolios with minimum variance and short-selling constraints, S-portfolios had almost 8% higher transaction costs. Yet, in the same research, M-portfolios had much lower transaction costs than classic portfolios, but their risk was statistically significantly higher. The same occurred in research by Welsch, Zhou (2007), where some robust estimators (MCD) achieved higher transaction cost, while others (I2D-Winsor, F2D-Winsor) – lower, in comparison to classic portfolios.

The following empirical example confirms the aforementioned discus-sion. In this example we constructed the rolling minimum variance portfo-lios with no short-selling constraints. We used an empirical data set with 5 assets from the DAX index: Adidas, Allianz, Bayer, Beiersdorf, BMW. We used daily logarithmic returns from the period between 3.01.2003 and 22.02.2012 and an estimation window length of 120 days. To determine the robust portfolio we used the Minimum Covariance Determinant estimator with 5% breakdown point, while to determine the classic portfolio we used sample covariance matrix. For both estimators we determined the $\text{EIF}(x_{t+1})$ as in (4), so we calculated the differences between weights deter-mined at period t (on the basis of the last 120 observations), and weights determined at period $t+1$ (on the basis of the last 121 observations). For

---

[6] Classic portfolios are M-portfolios with function $\rho(x) = (x)^2$.

each $\text{EIF}(\mathbf{x}_{t+1})$ we calculated the robust distance of observation $\mathbf{x}_{t+1}$ from the sample $\mathbf{X}_t = (\mathbf{x}_{t-119}, \ldots, \mathbf{x}_t)'$, so we determined $d(\mathbf{x}_{t+1}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$, where $\hat{\boldsymbol{\mu}}(\mathbf{X}_t)$ and $\hat{\boldsymbol{\Sigma}}(\mathbf{X}_t)$ are robust MCD estimators. Next, we divided $\text{EIF}(\mathbf{x}_{t+1})$ into two groups: extreme returns – this group contains those $\text{EIF}(\mathbf{x}_{t+1})$ for which robust distance $d(\mathbf{x}_{t+1}, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ is in a set of 5% greatest distances, the bulk of data – contains the remaining EIF's.

Fig. 1 presents box-plots of EIF's of MCD portfolios and classic portfolios within two subsets: extreme returns and the bulk of data. It can be seen that classic portfolios are more sensitive to extreme returns than robust portfolios and, moreover, for robust portfolios we observe no influence of the majority of extreme returns. Unfortunately, in the bulk of data we observe the opposite behavior: robust portfolios are more sensitive to non-extreme returns than classic portfolios.
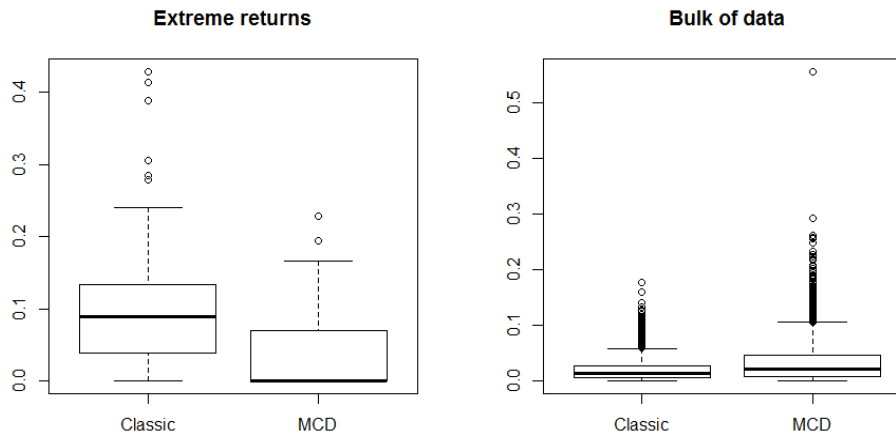

Fig. 1. Boxplots of EIF's of robust portfolios and classic portfolios

Source: own elaboration.

This analysis illustrates the most desirable property of robust portfolios: extreme returns have a significantly less influence on robust portfolio weights than on classic portfolios. It also illustrates the most undesirable property of robust portfolios – non-extreme returns cause greater changes in weights for robust portfolios than for classic ones.

6. Summary

This paper presents a review of robust statistics methods applied in the portfolio theory and the results of research in this field obtained to date. It also covers the most important definitions specifying the properties of robust estimators together with their interpretation in the portfolio theory. The article presents possible approaches to the construction of robust portfolios by applying the current research results in the field of robust statistics. It also describes the method of constructing robust portfolios, where the first stage is the selection of the optimization method (one-step or two-step approach), which also influences the method of rejecting (decreasing the influence) of outliers. The next stage is to choose the method, whereas for two-step methods, one can distinguish two groups of methods, differing in the properties of selected companies (contamination model), while in each group one can find various methods of robust estimation, which also influence the method of rejecting (decreasing influence of) outliers. Thus, the selection of robust estimators results in a different way of identifying outlying observations as well as different course of the optimization process. This article also presents the undesirable effects of applying robust estimators, such as the increase in transaction costs which can arise from the improper application of estimators to the examined sample, for example by choosing an excessively robust estimator with a high breakdown point.

Literature

Alqallaf F.A., Konis K.P., Martin R.D., Zamar R.H. (2002). *Scalable robust covariance and correlation estimates for data mining*. In: *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Edmonton*. Alberta. Pp. 14-23.

Alqallaf F., Van Aelst S., Yohai V.J., Zamar R.H. (2009). *Propagation of outliers in multivariate data*. The Annals of Statistics. Vol. 37(1). Pp. 311-331.

Best M.J., Grauer R.R. (1991). *On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results*. The Review of Financial Studies. Vol. 4. Pp. 315-342.

Black F., Litterman R. (1992). *Global portfolio optimization*. Financial Analysts Journal. Vol. 48(5). Pp. 28-43.

Butler R.W., Davies P.L., Jhun M. (1993). *Asymptotics for the minimum covariance determinant estimator*. The Annals of Statistics. Vol. 21(3). Pp. 1385-1400.

Chilson J., Ng R., Wagner A., Zamar R. (2004). *Parallel computation of high dimensional robust correlation and covariance matrix*. In: *Proceedings of the ACM SIGKDD*. Pp. 533-538.

Chopra V.K., Ziemba W.T. (1993). *The effect of errors in means, variances and covariances on optimal portfolio choice*. The Journal of Portfolio Management. Vol. 19. Pp. 6-11.

Croux C. (1998). *Limit behaviour of the empirical influence function of the median*. Statistics and Probability Letters. Vol. 37. Pp. 331-340.

Davies P.L. (1987). *Asymptotic behaviour of s-estimates of multivariate location parameters and dispersion matrices*. The Annals of Statistics. Vol. 15. Pp. 1269-1292.

Davies P.L. (1992). *The asymptotics of Rousseeuw's minimum volume ellipsoid estimator*. The Annals of Statistics. Vol. 20. Pp. 1828-1843.

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

DeMiguel V., Garlappi L., Uppal R. (2005). *Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?* The Review of Financial Studies. Vol. 22(5). Pp. 1915-1953.

Donoho D.L. (1982). *Breakdown Properties of Multivariate Location Estimators*. Ph.D. qualifying paper. Harvard University.

Donoho D.L., Huber P.J. (1983). *The notion of breakdown point*. In: P.J. Bickel, K. Doksumand, J.L. Hodges (Eds.). *A Festschrift for Erich Lehmann*. Wadsworth. Belmont. CA. Pp. 157-184.

Donoho D.L., Gasko M. (1992). *Breakdown properties of location estimates based on halfspace depth and projected outlyingness*. The Annals of Statistics. Vol. 20(4). Pp. 1803-1827.

Gnanadesikan R., Kettenring J.R. (1972). *Robust estimates, residuals, and outlier detection with multiresponse data*. Biometrics. Vol. 28. Pp. 81-124.

Hampel F.R. (1968). *Contributions to the Theory of Robust Estimation*. Ph.D. thesis. University of California. Berkeley.

Hampel F.R. (1971). *A general qualitative definition of robustness*. The Annals of Mathematical Statistics. Vol. 42. Pp. 1887-1896.

Hampel F.R. (1974). *The influence curve and its role in robust estimation*. Journal of the American Statistical Association. Vol. 69. Pp. 383-393

Huber P.J. (1981). *Robust Statistics*. John Wiley and Sons. New York.

Huber P.J., Ronchetti E.M. (2009). *Robust Statistics*. Wiley Series in Probability and Statistics. 2nd edition. John Wiley and Sons. New York.

Jagannathan R., Ma T. (2003). *Risk reduction in large portfolios: Why imposing the wrong constraints helps*. The Journal of Finance. Vol. 58(4). Pp. 1651-1684.

Jobson J.D., Korkie B.M. (1980). *Estimation for Markowitz efficient portfolios*. Journal of the American Statistical Association. Vol. 75. Pp. 544-554.

Jorion P. (1986). *Bayes-Stein estimation for portfolio analysis*. The Journal of Financial and Quantitative Analysis. Vol. 21. Pp. 279-292.

Kent J.T., Tyler D.E. (1996). *Constrained M-estimation for multivariate location and scatter*. The Annals of Statistics. Vol. 24(3). Pp. 1346-1370.

Kent J.T., Tyler D.E. (2001). *Regularity and uniqueness for constrained m-estimates and redescending m-estimates*. Annals of Statistics. Vol. 29(1). Pp. 252-265.

Khan J.A., Van Aelst S., Zamar R.H. (2007). *Robust linear model selection based on Least Angle Regression*. Journal of the American Statistical Association. Vol. 102. Pp. 1289-1299.

Lauprete G.J. (2001). *Portfolio Risk Minimization under Departures from Normality*. Ph.D. thesis. Sloan School of Management, MIT. Cambridge.

Lopuhaa H.P. (1989). *On the relation between S-estimators and M-estimators of multivariate location and covariance*. The Annals of Statistics. Vol. 17. Pp. 1662-1683.

Lopuhaa H.P. (1991). *Multivariate τ-estimators for location and scatter*. The Canadian Journal of Statistics. Vol. 19. Pp. 307-321.

Lopuhaa H.P. (1992). *Highly efficient estimators of multivariate location with high breakdown point*. The Annals of Statistics. Vol. 20. Pp. 398-413.

Maronna R.A. (1976). *Robust M-estimators of multivariate location and scatter*. Annals of Statistics. Vol. 4. Pp. 51-67.

Maronna R.A., Martin R., Yohai V.J. (2006). *Robust Statistics: Theory and Methods*. John Wiley. New York.

Maronna R.A., Stahel W.A., Yohai V.J. (1992). *Bias-robust estimators of multivariate scatter based on projections*. Journal of Multivariate Analysis. Vol. 42. Pp. 141-161.

Maronna R., Yohai V. (1995). *The behavior of the Stahel-Donoho robust multivariate estimator*. Journal of the American Statistical Association. Vol. 90(429). Pp. 330-341.

Maronna R., Zamar R. (2002). *Robust estimates of location and dispersion for high-dimensional data sets*. Technometrics. Vol. 44(4). Pp. 307-317.

Mendes B.V.M., Leal C.R.P. (2005). *Robust multivariate modeling in finance*. International Journal of Managerial Finance. Vol. 1(2). Pp. 95-106.

Michaud R. (1989). *The Markowitz optimization enigma: Is "optimized" optimal?* Financial Analysts Journal. Vol. 45. Pp. 31-42.

Papahristodoulou C., Dotzauer E. (2004). *Optimal portfolios using linear programming models*. The Journal of the Operational Research Society. Vol. 55. Pp. 1169-1177.

Perret-Gentil C., Victoria-Feser M.P. (2004). *Robust Mean-Variance Portfolio Selection*. FAME Research Paper 140. International Center for Financial Asset Management and Engineering. Geneva.

Rocke D.M. (1996). *Robustness properties of S-estimators of multivariate location and shape in high dimension*. The Annals of Statistics. Vol. 24. Pp. 1327-1345.

Roelant E., Van Aelst S., Willems G. (2009). *The minimum weighted covariance determinant estimator*. Metrika. Vol. 70(2). Pp. 177-204.

Rousseeuw P.J. (1984). *Least median of squares regression*. Journal of the American Statistical Association. Vol. 79. Pp. 871-880.

Rousseeuw P.J. (1985). *Multivariate estimation with high breakdown point*. In: W. Grossmann, G. Pflug, I. Vincze, W. Wertz (Eds.). *Mathematical Statistics and Applications*. Vol. B. Pp. 283-297.

Rousseeuw P.J., Leroy A.M. (1987). *Robust Regression and Outlier Detection*. John Wiley & Sons. New York.

Rousseeuw P.J., Yohai V.J. (1984). *Robust regression by means of S-estimators*. In: J. Franke, W. Hardle, R.D. Martin (Eds.). *Robust and Nonlinear Time Series Analysis*. Springer-Verlag. New York. Pp. 256-272.

Salibian-Barrera M., Van Aelst S., Willems G. (2006). *PCA based on multivariate MM-estimators with fast and robust bootstrap*. Journal of the American Statistical Association. Vol. 101. Pp. 1198-1211.

Stahel W.A. (1981). *Breakdown of Covariance Estimators*. Res Rep 31. Fachgruppe für Statistik. E.T.H. Zurich.

Tatsuoka K.S., Tyler D.E. (2000). *The uniqueness of S- and M-Functionals under non--elliptical distributions*. The Annals of Statistics. Vol. 28. Pp. 1219-1243.

Todorov V., Filzmoser P. (2009). *An object oriented framework for robust multivariate analysis*. Journal of Statistical Software. Vol. 32(3). Pp. 1-47.

Tyler D.E. (1994). *Finite sample breakdown points of projection based multivariate location and scatter statistics*. The Annals of Statistics. Vol. 22(2). Pp. 1024-1044.

Wang Z. (2005). *A shrinkage approach to model uncertainty and asset allocation*. The Review of Financial Studies. Vol. 18(2). Pp. 673-705.

Wang N., Raftery A.E. (2002). *Nearest-neighbor variance estimation (NNVE): robust covariance estimation via nearest-neighbor cleaning*. Journal of the American Statistical Association. Vol. 97(460). Pp. 994-1006.

Welsch R.E., Zhou X. (2007). *Application of robust statistics to asset allocation models*. REVSTAT – Statistical Journal. Vol. 5(1). Pp. 97-114.

Yohai V.J. (1987). *High breakdown-point and high efficiency robust estimates for regression*. The Annals of Statistics Vol. 15. Pp. 642-656.

Zhou X. (2006). *Application of Robust Statistics to Asset Allocation Models*. MSc. thesis. MIT.

Zuo Y. (2000). *Multivariate monotone location estimators*. Sankhyā: The Indian Journal of Statistics, Series A. Vol. 62(2). Pp. 161-177.