REVSTAT – Statistical Journal
Volume 5, Number 1, March 2007, 97–114

APPLICATION OF ROBUST STATISTICS TO ASSET ALLOCATION MODELS

Authors: Roy E. Welsch
– Sloan School of Management,
Massachusetts Institute of Technology, USA
rwelsch@mit.edu

Xinfeng Zhou
– Global Index and Markets Group,
Barclays Global Investors, USA
xinfeng.zhou@barclaysglobal.com

Abstract:

- Many strategies for asset allocation involve the computation of the expected value and the covariance matrix of the returns of financial instruments. How much of each instrument to own is determined by an attempt to minimize risk — the variance of linear combinations of investments in these financial assets — subject to various constraints such as a given level of return, concentration limits, etc. The covariance matrix contains many parameters to estimate and two main problems arise. First, the data will very likely have outliers that will seriously affect the covariance matrix. Second, with so many parameters to estimate, a large number of return observations are required and the nature of markets may change substantially over such a long period. In this paper we discuss using robust covariance procedures, FAST-MCD, Iterated Bivariate Winsorization and Fast 2-D Winsorization, to address the first problem and penalization methods for the second. When back-tested on market data, these methods are shown to be effective in improving portfolio performance. Robust asset allocation methods have great potential to improve risk-adjusted portfolio returns and therefore deserve further exploration in investment management research.

Key-Words:

- *robust statistics; asset allocation; FAST-MCD; bivariate Winsorization; penalization.*

AMS Subject Classification:

- 62G35, 62P20, 91B28.

1. INTRODUCTION

Asset allocation is the process that investors use to determine the asset classes in which to invest and the weight for each asset class. Past studies have shown that asset allocation explains 75 − 90% of the return variation and is the single most important factor determining the variability of portfolio performance. Among all the asset allocation models, Harry Markowitz's mean-variance portfolio theory is by far the most well-known and well-studied model for both academic researchers and practitioners alike [17, 18]. The crux of mean-variance portfolio theory assumes that investors prefer lower standard deviations/variances for a given level of expected return. Portfolios that provide the minimum standard deviation for a given expected return are termed efficient portfolios and those that do not are termed inefficient portfolios.

For a portfolio with N risky assets to invest in, the portfolio return is the weighted average return of each asset

$$r_p \equiv w_1 r_1 + w_2 r_2 + \cdots + w_N r_N = \boldsymbol{w}'\boldsymbol{r} \tag{1.1}$$

and the expected return and the variance of the portfolio can be expressed as

$$\begin{aligned} \mu_p &= w_1\mu_1 + w_2\mu_2 + \cdots + w_N\mu_N = \boldsymbol{w}'\boldsymbol{\mu} ,\\ \mathrm{var}(r_p) &= \mathrm{var}\big(w_1 r_1 + w_2 r_2 + \cdots + w_N r_N\big) = \boldsymbol{w}'\boldsymbol{\Sigma}\,\boldsymbol{w} , \end{aligned} \tag{1.2}$$

where w_i, $\forall\, i=1,...,N$, is the weight of the i-th asset in the portfolio; r_i is the return of the i-th asset in the portfolio; μ_i is the expected return of the i-th asset in the portfolio; $\boldsymbol{w}$ is a $N\times1$ column vector of w_i's; $\boldsymbol{r}$ is a $N\times1$ column vector of r_i's; $\boldsymbol{\mu}$ is a $N\times1$ column vector of μ_i's; and $\boldsymbol{\Sigma}$ is the $N\times N$ covariance matrix of the returns of N assets.

We can formulate the following problem to assign optimal weight to each asset and identify the efficient portfolio:

$$\min_{\boldsymbol{w}} \boldsymbol{w}'\boldsymbol{\Sigma}\,\boldsymbol{w} \qquad \text{s.t.} \quad \boldsymbol{w}'\boldsymbol{\mu} = \mu_p, \quad \boldsymbol{w}'\boldsymbol{e} = 1 , \tag{1.3}$$

where μ_p is the expected portfolio return and $\boldsymbol{e}$ is $N\times1$ column vector with all elements 1. For each specified μ_p, the problem can be solved in closed form using the method of Lagrange [23]. The simple mean-variance optimization only requires two inputs–expected return vector and expected covariance matrix. The model is based on a formal quantitative objective that will always give the same solution with the same set of parameters. These all explain its popularity and its contribution to modern portfolio theory (MPT).

Nevertheless, the original form of mean-variance portfolio optimization has rarely been applied in practice because of several drawbacks. The method uses variance as the risk measure, which is often considered to be a simplistic measurement when the asset returns do not follow normal distributions. In reality, many of the financial assets' returns do have fat tails or are skewed. Besides, the one-period nature of static optimization also does not take dynamic factors into account, and some researchers argue for more complicated models based on stochastic processes and dynamic programming. However, the most serious problem of the mean-variance efficient frontier is probably the method's instability. The mean-variance frontier is very sensitive to the inputs, and these inputs are subject to random errors in the estimation of expected return and covariance. Small and statistically insignificant changes in these estimates can lead to a significant change in the composition of the efficient frontier. This may lead us to frequently and mistakenly rebalance our portfolio to stay on this elusive efficient frontier, incurring unnecessary transaction costs.

The Markowitz portfolio optimization estimates the expected return and the covariance matrix from historical return time series and treats them as true parameters for portfolio selection. The historical returns for N assets over T periods are denoted as $\boldsymbol{R}$, a $T \times N$ matrix where each column vector $\boldsymbol{r}_i$, $\forall\, i = 1, ..., N$, represents the returns of asset i over different periods and each row vector $\boldsymbol{R}_t$, $\forall\, t = 1, ..., T$, represents the returns of different assets at period t. The simple sample mean and covariance matrix are used as the parameters since they are the best unbiased estimators under the assumption of multivariate normality. Despite the simple computation involved, this approach has high complexity (large number of parameters). It suffers from the problem of high variance, which means the estimation errors can be significant and generate erroneous mean-variance efficient frontiers. This naïve "certainty equivalence" mean-variance approach often leads to extreme portfolio weights (instead of a diversified portfolio as the method anticipates) and dramatic swings in weights when there is a minor change to the expected returns or the covariance matrix [7, 10, 12]. The problem is further exacerbated if the number of observations is of the same order as the number of assets, which is often the case in financial applications to select industry sectors or individual securities.

A number of alternative models have been developed to improve parameter estimation. For example, factor-based models try to reduce the model complexity (number of parameters) by explaining asset return variances/covariances using a limited number of common factors. Multivariate GARCH models try to address fat tails and volatility clustering by incorporating the time dependence of returns in the covariance matrix. But neither approach effectively reduces or eliminates the influences of outliers in the data. A small percentage of outliers, in some cases even a single outlier, can distort the final estimated variance and covariance. Evidence has shown that the most extreme (large positive or negative) coefficients in the estimated covariance matrix often contain the largest error and as a result, mean-variance optimization based on such a matrix routinely gives the heaviest weights — either positive or negative — to those coefficients that are most unreliable. This "error-maximization" phenomenon [24] causes the mean-variance technique to behave very badly unless such errors are corrected.

In this study, we focus on investigating robust statistical approaches to reduce the influence of outliers, to increase the stability of the portfolio and to reduce asset turnover. The remainder of the paper is organized as follows. In Section 2, we investigate and extend some robust statistical methods such as FAST-MCD, Iterated Bivariate Winsorization, and Fast 2-D Winsorization to estimate the covariance matrix. We also explore penalization methods as a direct way to reduce asset turnovers. In Section 3, we apply these methods to construct US industrial selection portfolios and show that these robust methods dramatically improve risk-adjusted portfolio performance, especially when transaction costs are taken into consideration. In Section 4, we conclude this paper by summarizing our findings and offering possible directions for future research.

2. METHODS

During the past decade, statisticians have developed a variety of robust estimation methods to estimate both the mean and the covariance matrix [4, 8, 19, 20]. However, the use of robust estimators has received relatively little attention in the finance literature overall, and in the context of estimating the expected value and the covariance matrix of asset returns in particular [13, 22]. In this study, we take the initiative to investigate the value of some robust approaches to asset allocation problems.

2.1. FAST-MCD

The general principle of robust statistical estimation is to give full weights to observations assumed to come from the main body of the data, but to reduce or completely eliminate weights for the observations from tails of the contaminated data. The minimum covariance determinant (MCD) method [3], a robust estimator introduced by Rousseeuw in 1985, eliminates perceived outliers from the estimation of the mean and the covariance matrix. It uses the mean and the covariance matrix of h data points ($T/2 \leqslant h < T$) with the smallest determinant to estimate the population mean and the covariance matrix. The method has a break-down value of $(T-h)/T$. If the data come from a multivariate normal distribution, the average of the optimal subset is an unbiased estimator of the population mean. The resulting covariance matrix is biased, but a finite sample correction factor ($c_{h,T} \geq 1$) can be used to make the covariance matrix unbiased. The multiplication factor $c_{h,T}$ can be determined through Monte Carlo simulation. For our specific purpose, the bias by itself does not affect the asset allocation since all pairs of covariances are underestimated by the same factor.

MCD has rarely been applied to high-dimensional problems because it is extremely difficult to compute. MCD estimators are solutions to highly non-convex optimization problems that have exponential complexity of the order 2^N in terms of the dimension N of the data. Therefore, these original methods are not suitable for asset allocation problems when $N > 20$. Yet, in practice, asset allocation problems often include dozens of industrial classes or hundreds of individual securities, which makes the MCD method computationally infeasible. In order to cope with computational complexity problems, a heuristic FAST-MCD algorithm developed by Rousseeuw and Van Driessen [25], provides an efficient alternative. A naïve MCD approach would compute the MCD for up to $\binom{T}{h}$ subsets, while FAST-MCD uses sampling to reduce the computation and usually offers a satisfactory heuristic estimation. Other equivariant robust covariance methods are discussed in a recent book [20] and we are experimenting with the S-estimator they recommend, SR-05.

The key step of the FAST-MCD algorithm takes advantage of the fact that, starting from any approximation to the MCD, it is possible to compute another approximation with a determinant no higher than the current one. The method is based on the following theorem related to a concentration step (C-step):

Let $H_1 \subset \{1, ..., n\}$ be any h-subset of the original cross-sectional data, put $\hat{\mu}_1 = \frac{1}{h}\sum_{t\in H_1}\boldsymbol{R}_t$ and $\hat{\boldsymbol{\Sigma}}_1 = \frac{1}{h}\sum_{t\in H_1}(\boldsymbol{R}_t - \hat{\boldsymbol{\mu}}_1)(\boldsymbol{R}_t - \hat{\boldsymbol{\mu}}_1)'$. If $\det(\hat{\boldsymbol{\Sigma}}_1) \neq 0$, define the distance $d_1(t) = \sqrt{(\boldsymbol{R}_t - \hat{\boldsymbol{\mu}}_1)\,\hat{\boldsymbol{\Sigma}}_1^{-1}(\boldsymbol{R}_t - \hat{\boldsymbol{\mu}}_1)'}$, $t = 1, ..., T$. Now take H_2 such that $\{d_1(i);\ i \in H_2\} := \{(d_1)_{1:T}, ..., (d_1)_{h:T}\}$ where $(d_1)_{1:T} \leq (d_1)_{2:T} \leq \cdots \leq (d_1)_{T:T}$ are the ordered distances, and compute $\hat{\boldsymbol{\mu}}_2$ and $\hat{\boldsymbol{\Sigma}}_2$ based on H_2. Then $\det(\hat{\boldsymbol{\Sigma}}_2) \leq \det(\hat{\boldsymbol{\Sigma}}_1)$ with equality if and only if $\hat{\boldsymbol{\mu}}_2 = \hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\Sigma}}_2 = \hat{\boldsymbol{\Sigma}}_1$.

If $\det(\hat{\boldsymbol{\Sigma}}_1) > 0$, the C-step yields $\hat{\boldsymbol{\Sigma}}_2$ with $\det(\hat{\boldsymbol{\Sigma}}_2) \leq \det(\hat{\boldsymbol{\Sigma}}_1)$. Basically the theorem indicates the sequence of determinants obtained through C-steps converges in a finite number of steps from any original h-subset to a subset satisfying $\det(\hat{\boldsymbol{\Sigma}}_{m+1}) = \det(\hat{\boldsymbol{\Sigma}}_m)$. Afterward, running the C-step no longer reduces the determinant. However, this process only guarantees that the resulting $\det(\hat{\boldsymbol{\Sigma}})$ is a local minimum instead of the global one. To yield the h-subset with global minimum $\det(\hat{\boldsymbol{\Sigma}})$ or at least close to optimal, many initial choices (often > 500) of H_1 are taken and C-steps are applied to each.

Simulated and empirical results showed that FAST-MCD typically gives "good" results and is orders of magnitude faster than exact MCD methods. Yet, the FAST-MCD method still requires substantial running times for large N and T, and the probability of retaining outliers in the final h-subset increases when N becomes large. We use the FAST-MCD as an affine equivariant benchmark for faster non-equivariant methods. Other examples of its use are contained in [26, 30].

2.2. Iterated bivariate Winsorization (I2D-Winsor)

The FAST-MCD estimator for the covariance matrix is positive semidefinite and affine equivariant, which means the estimator behaves properly under affine transformations of the data. If the affine equivariance requirement is dropped, much faster estimators with high breakdown points can be computed. These methods are often based on pair-wise robust correlation or covariance estimates such as coordinate-wise outlier insensitive transformations (e.g. Huber-function transformation, quadrant correlation) and bivariate outlier resistant models. All these methods have quadratic complexity in the number of variables and linear complexity in the number of observations, so they reduce the computational complexity to $O(N^2T)$.

Huber's function, defined as $H_c(x) = \min\big\{\max\{-c, x\}, c\big\},\ c > 0$, has been widely used to shrink outliers towards the median by the transformation

$$\tilde{r}_{ti} \;=\; m_i + s_i \times H_c\big((r_{ti} - m_i)/s_i\big) \;, \tag{2.1}$$

where m_i and s_i are the median and the median absolute deviation from the median of return vector $\boldsymbol{r}_i$. Essentially Huber's function brings the outliers of each variable to the boundary $m_i \pm c \times s_i$ and, as a result, reduces the impact of outliers.

The one-dimensional Winsorization approach using the Huber function has been a popular method in finance because of its intuitive appeal and easy computation. Yet for covariance analysis, the method fails to take the orientation of the bivariate data into consideration. To address the problem, bivariate Winsorization methods have also been investigated. For each pair of variables, outliers are shrunken to the border of an ellipse which includes the majority of the data by using the bivariate transformation

$$\tilde{\boldsymbol{r}}_{t,i,j} \;=\; \boldsymbol{\mu}_0 + \min\Big(\sqrt{c/D(\boldsymbol{r}_{t,i,j})},\, 1\Big)\, (\boldsymbol{r}_{t,i,j} - \boldsymbol{\mu}_0) \;, \tag{2.2}$$

where, for each pair of $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$, $\boldsymbol{r}_{t,i,j} = \begin{bmatrix} r_{ti} \\ r_{tj} \end{bmatrix}$; $\boldsymbol{\mu}_0 = \begin{bmatrix} m_i \\ m_j \end{bmatrix}$; $D(\boldsymbol{r}_{t,i,j})$ is the Mahalanobis distance based on an initial bivariate covariance matrix $\boldsymbol{\Sigma}_0$ and location $\boldsymbol{\mu}_0$: $(\boldsymbol{r}_{t,i,j} - \boldsymbol{\mu}_0)'\, \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{r}_{t,i,j} - \boldsymbol{\mu}_0)$; c is a positive constant. The transformation shrinks the outlier towards $\boldsymbol{\mu}_0$ when $D(\boldsymbol{r}_{t,i,j}) > c$.

Based on the idea of shrinking data toward the border of a two-dimensional ellipse, Chilson et al. developed an iterated bivariate Winsorization (I2D-Winsor) method to estimate covariance and applied the method to cluster correlated genes [5]. The method includes the following three steps:

Step A. For each pair of variables $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$, compute a simple robust mean and adjusted MAD for each column and construct the initial estimate of mean and covariance matrix as

$$\boldsymbol{\mu}_0 = \begin{bmatrix} m_i \\ m_j \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma}_0 = \begin{bmatrix} \dfrac{s_i}{0.6745} & 0 \\ 0 & \dfrac{s_j}{0.6745} \end{bmatrix} . \tag{2.3}$$

Step B. For each $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$, calculate the Mahalanobis distance for each return pair

$$D_{t,k} = (\boldsymbol{r}_{t,i,j} - \boldsymbol{\mu}_k)' \, \boldsymbol{\Sigma}_k^{-1} \, (\boldsymbol{r}_{t,i,j} - \boldsymbol{\mu}_k) \tag{2.4}$$

and then calculate the weight for each $\boldsymbol{r}_{t,i,j}$ as

$$z_t = \min\left(\sqrt{c/D_{t,k}}, \, 1\right) , \tag{2.5}$$

where the constant c is chosen as 5.99 (the 95% quantile of the χ_2^2 distribution).

Step C. Update $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ to $\boldsymbol{\mu}_{k+1}$ and $\boldsymbol{\Sigma}_{k+1}$ using equations

$$\begin{aligned} \boldsymbol{\mu}_{k+1} &= \sum_{i=1}^{T} z_t \, \boldsymbol{r}_{t,i,j} \Big/ \sum_{i=1}^{T} z_t \; , \\ \boldsymbol{\Sigma}_{k+1} &= \sum_{i=1}^{T} z_t^2 \, (\boldsymbol{r}_{t,i,j} - \boldsymbol{\mu}_{k+1}) \, (\boldsymbol{r}_{t,i,j} - \boldsymbol{\mu}_{k+1})' \Big/ \sum_{i=1}^{T} z_t^2 \; . \end{aligned} \tag{2.6}$$

This iteration is repeated until $\boldsymbol{\mu}_{k+1}, \boldsymbol{\Sigma}_{k+1}$ and $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$ converge as determined by the sum of absolute differences between two consecutive $\boldsymbol{\Sigma}$ being less than a predefined error. The covariance matrix of variables $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ is then set to $\boldsymbol{\Sigma}_{k+1}$. Diagonal elements of the covariance matrix are obtained using bias adjusted median absolute deviations from the median.

The I2D-Winsor method allowed parallel computation of high dimensional correlation and covariance matrices for different gene expressions and obtained good performance in heterogeneous cluster studies. But the method suffers the drawback of failing to guarantee positive semidefiniteness of the covariance matrix — a crucial requirement for mean-variance portfolio optimization. Maronna et al. [21] proposed an adjustment method to obtain a positive semidefinite covariance matrix using a pair-wise robust covariance matrix. The method is based on the observation that any positive semidefinite covariance matrix $\boldsymbol{C}$ can be expressed as $\boldsymbol{C} = \sum \hat{\lambda}_i \, \hat{\boldsymbol{a}}_i \, \hat{\boldsymbol{a}}_i'$, where $0 \leq \hat{\lambda}_1 \leq \cdots \leq \hat{\lambda}_N$ are the eigenvalues and $\hat{\boldsymbol{a}}_i$ $(i=1, ..., N)$ are the corresponding eigenvectors. If $\boldsymbol{C}$ is not positive semidefinite, then one or more of the eigenvalues are negative. To convert such a matrix to a positive semidefinite one, a natural approach is to replace these negative eigenvalues with positive ones. When $\boldsymbol{C}$ is the sample correlation, $\hat{\lambda}_i$'s are the variances of the projected data on the direction of the corresponding eigenvectors.

This indicates that in order to get rid of possibly negative eigenvalues in the quadrant covariance matrix $\hat{\boldsymbol{C}}_0$, one can replace the $\hat{\lambda}_i$'s in $\boldsymbol{C}_0 = \sum \hat{\lambda}_i \, \hat{\boldsymbol{a}}_i \, \hat{\boldsymbol{a}}_i'$ by the square of robust standard deviation estimates for the projected data. We can compute the decomposition of $\hat{\boldsymbol{C}}_0$: $\hat{\boldsymbol{C}}_0 = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}'$, where $\boldsymbol{Q}$ is the orthogonal matrix of eigenvectors and $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues. Then we can transform $\boldsymbol{R}$ to $\tilde{\boldsymbol{R}}$ using the new basis $\boldsymbol{Q}$: $\tilde{\boldsymbol{R}} = \boldsymbol{R}\boldsymbol{Q}'$ and compute the robust standard deviation estimate $(\tilde{s}_j/0.6745)$ of the columns of $\tilde{\boldsymbol{R}}$. Let $\tilde{\boldsymbol{D}}$ be the diagonal matrix whose elements are $(\tilde{s}_j/0.6745)^2$ ordered from largest to smallest. The final positive definite robust covariance matrix is $\hat{\boldsymbol{\Sigma}} = \boldsymbol{Q}\tilde{\boldsymbol{D}}\boldsymbol{Q}'$.

By transforming the I2D-Winsor robust covariance matrix using Maronna's adjustment method, we guarantee the positive semidefiniteness of the final covariance matrix and make it directly applicable to asset allocation problems.

2.3. Fast 2-D Winsorization (F2D-Winsor)

Khan et al. [11] proposed a fast two-step, two-dimensional Winsorization method (F2D-Winsor) while investigating ways to make least-angle regression (LARS) robust. Instead of repeated iteration of step B in I2D-Winsor, which is computationally expensive, Khan's method only implements step B once. In order to achieve a similar level of robustness as I2D-Winsor, F2D-Winsor constructs an informative initial covariance matrix. We again combine F2D-Winsor ideas from Khan's paper and Maronna's method to guarantee the positive semidefiniteness of the covariance matrix and design the following F2D-Winsor method:

Step A. *Initial covariance estimate.* For each pair of variables $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$, compute simple robust location (median) and scale (adjusted MAD) estimates for each variable. We then compute an initial covariance matrix using Khan's adjusted Winsorization method that is more resistant to bivariate outliers [11]. In the adjusted Winsorization method, two tuning parameters are used with c_1 for the two quadrants (separated by m_i and m_j) that contain the majority of the data and a smaller constant c_2 for the other two quadrants. For example, c_1 can be taken to be 1.96 ($\mu \pm 1.96\,\sigma$ includes 95% of the data from the normal distribution) and $c_2 = hc_1$ where $h = n_2/n_1$ with n_1 the number of observations in the major quadrants and $n_2 = T - n_1$, where T is the total number of observations. As shown in Figure 1, the data are now shrunk to the boundary of the four smaller rectangles instead of a large rectangle. As a result, the adjusted Winsorization method handles bivariate outliers better than the univariate Winsorization. However, it does raise a problem that the initial covariance matrix constructed from pairwise covariance may not be positive definite. To address the problem, Maronna's transformation is applied to convert the initial covariance matrix $\boldsymbol{\Sigma}_0$ to a positive definite one.

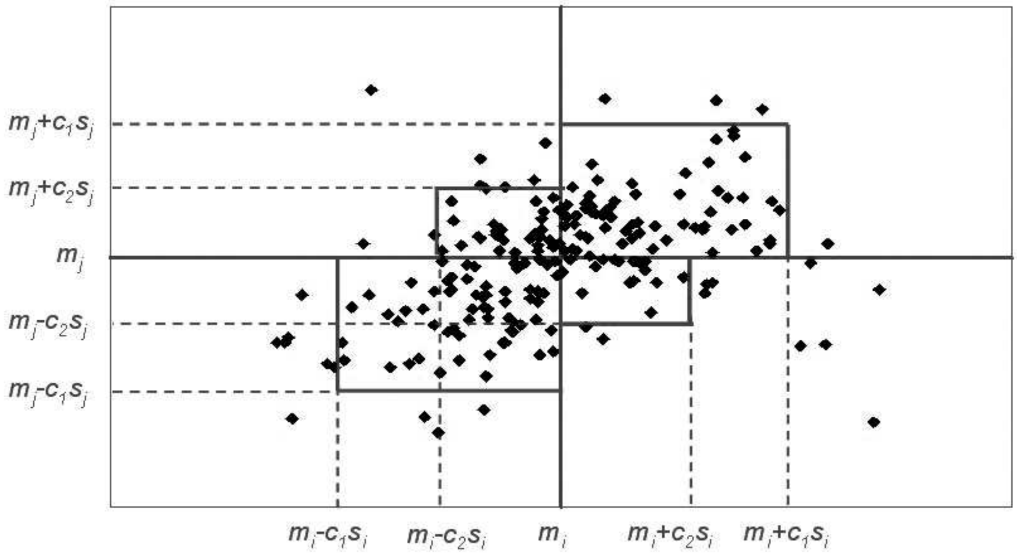

Figure 1: Adjusted Winsorization (for initial covariance) with $c_1 = 1.96$, where s_i and s_j are estimated from adjusted MAD.

Step B. *2D-Winsorization based covariance matrix*. For each pair of $(\boldsymbol{r}_i, \boldsymbol{r}_j)$, outliers are shrunk to the border of an ellipsoid by using the transformation $\tilde{\boldsymbol{r}}_{t,i,j} = \boldsymbol{\mu}_0 + \min\bigl(\sqrt{c/D_{t,0}}\,, 1\bigr)\, (\boldsymbol{r}_{t,i,j} - \boldsymbol{\mu}_0)$, with constant $c = 5.99$ (the 95% quantile of the χ^2_2 distribution). The covariance for each pair is calculated using this modified data. Maronna's transformation is again applied to guarantee the positive definiteness of the final covariance matrix.

2.4. L1-penalized mean-variance method (V1)

All these robust covariance matrix estimation methods try to increase the stability of the allocation model by increasing the stability of the mean and covariance matrix of returns over time. Since the influence of outliers is reduced, the updated return data tend to have less impact on the robust mean and covariance matrix, even if some of the new return vectors contain extreme values. In this sub-section, we also implement a different class of penalization-based robust estimators to directly increase model stability and reduce turnover.

If the expected return and covariance matrix are estimated from the historical sample $\boldsymbol{R}_1, ..., \boldsymbol{R}_T$, the original mean-variance portfolio optimization problem

$$\min_{\boldsymbol{w}} \boldsymbol{w}'\, \boldsymbol{\Sigma}\, \boldsymbol{w} \qquad \text{s.t.} \quad \boldsymbol{w}'\boldsymbol{\mu} = \mu_p, \quad \boldsymbol{w}'\boldsymbol{e} = 1 \;, \tag{2.7}$$

can be rewritten as

$$\min_{\boldsymbol{w},q} \frac{1}{T}\sum_{i=1}^{T}(\boldsymbol{w}'\boldsymbol{R}_t - q)^2 \qquad \text{s.t.} \quad \boldsymbol{w}'\boldsymbol{\mu} = \mu_p, \quad \boldsymbol{w}'\boldsymbol{e} = 1 \ . \tag{2.8}$$

Lauprete [14] and Lauprete, et al. [15] proposed penalizing deviations from the market weights ($w_{m,i} = M_i/\sum_{j=1}^{N} M_j$ with M_j being the market value of asset j) as a possible way to reduce the influences of outliers and to reduce turnover. These authors also considered using robust loss functions (M-estimators) in place of least-squares loss in (2.8). A recent paper by DeMiguel and Nogales [6] replaces M-estimators with S-estimators but omits any penalty term. If the market is efficient (or nearly efficient as many researchers believe), a penalty term serves as the prior in our optimization problem. We should penalize the final cost function if the proposed asset weights deviate from the prior. As a result, extreme deviations from the prior are unlikely. In this study, we focused on an L1 regularization method, which was the penalty function used in LASSO regression [27]. The regularized portfolio estimator can be expressed as [14]:

$$\begin{aligned} \big(\boldsymbol{w}(\lambda), q(\lambda)\big) \;&=\; \arg\min_{q\in R}\left(\frac{1}{T}\sum_{i=1}^{T}(\boldsymbol{w}'\boldsymbol{R}_t - q)^2 + \lambda|\boldsymbol{w}-\boldsymbol{w}_m|\right) \\ &\text{s.t.} \quad \boldsymbol{w}'\boldsymbol{\mu} = \mu_p, \quad \boldsymbol{w}'\boldsymbol{e} = 1 \ , \end{aligned} \tag{2.9}$$

where $\lambda > 0$ is the regularization parameter; $|\boldsymbol{w}-\boldsymbol{w}_m|$ is the L_1-norm of $\boldsymbol{w}-\boldsymbol{w}_m$: $\sum_{i=1}^{N}|w_i - w_{m,i}|$.

The term $\lambda|\boldsymbol{w}-\boldsymbol{w}_m|$ reflects the investor's a priori confidence in the market portfolio $\boldsymbol{w}_m$. A large λ means large penalty for any deviation and strong confidence in $\boldsymbol{w}_m$; a small λ reflects weak confidence in $\boldsymbol{w}_m$. We choose the parameter λ using 5-fold cross validation. For any given λ, we implement the following steps:

Step A. Divide the T observations randomly into 5 subsets of $T/5$ observations. Call these subsets $T(i)$ for $i = 1, ..., 5$. For every i, run the optimization to yield the optimal $\big(\hat{\boldsymbol{w}}(\lambda), \hat{q}(\lambda)\big)$ for the in-sample data:

$$\begin{aligned} \big(\hat{\boldsymbol{w}}(\lambda), \hat{q}(\lambda)\big) \;&=\; \arg\min_{q\in R}\left(\frac{1}{0.8\,T}\sum_{t\in T\setminus T(i)}(\boldsymbol{w}'\boldsymbol{R}_t - q)^2 + \lambda|\boldsymbol{w}-\boldsymbol{w}_m|\right) \\ &\text{s.t.} \quad \boldsymbol{w}'\boldsymbol{\mu} = \mu_p, \quad \boldsymbol{w}'\boldsymbol{e} = 1 \ . \end{aligned} \tag{2.10}$$

Step B. For every $i = 1, ..., 5$ apply $\big(\hat{w}(\lambda), \hat{q}(\lambda)\big)$ to the out-of-sample data to calculate a sum of squared errors, $PE_\lambda(i) = \sum_{t\in T(i)}\big[\big(\hat{\boldsymbol{w}}_\lambda(i)'\boldsymbol{R}_t - \hat{q}_\lambda(i)\big)^2\big]$.

Step C. Calculate the total sum of squared errors $PE_\lambda = \sum_{i=1}^{5} PE_\lambda(i)$.

A series of candidate values of λ from 0.01 to 2 are tested to yield a value of λ with minimum total sum of squared errors PE_λ. Once λ is selected, $\boldsymbol{w}(\lambda)$ and $q(\lambda)$ can be solved as the "optimal" solution to the corresponding quadratic optimization problem. The lower bound of 0.01 was found by experimentation and may be different for other data sets.

3. APPLICATION RESULTS

In this section, we show a real asset allocation application using daily returns on 51 MSCI US industry sector indexes, from 01/03/1995 to 02/07/2005 (2600 trading days of data). Combining the stocks in these industry indexes ($\sim$700 stocks included) forms a general index for US equity markets broader than the S&P 500. The robust methods discussed in Section 2 are applied to find the "optimal" weights for each industry.

For every estimator, we use the following portfolio rebalancing strategy: estimate the industry sector weights using the most recent 100 daily returns and rebalance the portfolio weights every five trading days (a week). Since there are 2600 trading days in the data, there are 500 rebalances in total. In practice, there are transaction costs when we change the weights of each asset using updated information. So we will compare the results both without considering transaction costs and with 5 cents for each $100 bought or sold. We apply a target return constraint and convexity constraint to all estimates:

$$\boldsymbol{w}'\boldsymbol{\mu} = \mu_p\ , \qquad \boldsymbol{w}'\boldsymbol{e} = 1\ . \tag{3.1}$$

The resulting stream of ex-post portfolio returns is collected for each estimator/target return combination. We calculate the following statistics of the ex-post returns of each estimator/target return combination:

Mean: the sample mean of weekly ex-post returns;

STD: the sample standard deviation of weekly ex-post returns;

Information Ratio: $IR = mean/STD \times \sqrt{52}$, where the standardization by $\sqrt{52}$ makes the information ratio an annual estimate assuming 260 trading days per year;

α-VaR for $\alpha = 5\%$ and 1%: the loss at the α-quantile of the weekly ex-post return;

MaxDD: the maximum drawdown, which is the maximum loss in a week;

CRet: cumulative return;

Turnover: weekly asset turnover, defined as the mean of the absolute weight changes $\left(\sum_{i=1}^{51} |w_{t,i} - w_{t-1,i}|\right)$ for 500 updates;

Cret_cost: cumulative return with transaction costs;

IRcost: Information ratio with transaction costs.

Except for the market model, which uses market weights and the corresponding market returns, a range of target expected annual portfolio returns from 10% to 20% are used for portfolio construction. Table 1 shows the summarized results for annual expected return $\mu_p = 15\%$ for V (mean-variance optimization with simple mean and covariance matrix), FAST-MCD, I2D-Winsor, F2D-Winsor, V1 models and market index. More extensive tables are in Zhou [31].

Table 1: Performance of V, FAST-MCD, I2D-Winsor, F2D-Winsor, V1 models and Market index for $\mu_p = 15\%$. For FAST-MCD, I2D-Winsor and F2D-Winsor, the median instead of the mean of the returns is used as the expected return of each asset.

$\mu_p = 15\%$	V	FAST-MCD	I2D-Winsor	F2D-Winsor	V1	Market
mean	0.065%	0.096%	0.156%	0.155%	0.198%	0.160%
STD	1.962%	2.025%	1.948%	2.007%	2.431%	2.343%
IR	0.239	0.341	0.578	0.558	0.589	0.491
VaR(0.05)	3.06%	3.10%	3.10%	3.23%	3.71%	4.06%
VaR(0.01)	5.78%	6.33%	5.80%	5.52%	6.65%	5.28%
MaxDD	−7.48%	−8.57%	−9.39%	−9.40%	−8.35%	−10.01%
Cret	1.256	1.457	1.983	1.965	2.328	1.935
Cret_cost	0.845	0.888	1.801	1.803	2.252	1.923
IRcost	−0.054	−0.013	0.507	0.497	0.569	0.487
Turnover	1.59	1.99	0.39	0.35	0.13	0.02

Both the pair-wise Winsorization methods and the penalization method yield significantly better results than mean-variance optimization with the simple mean and covariance matrix as inputs. The V method has significant asset turnover (159%) and as a result the IRcost — the most popular performance measure — is negative after the transaction costs are taken into consideration. In contrast, I2D-Winsor, F2D-Winsor and V1 methods have much lower turnovers (0.39, 0.35 and 0.13 respectively) and yield an IRcost of 0.507, 0.497 and 0.569 respectively, which are much higher than the V method. All these methods also beat the market in VaR (5%), MaxDD and IRcost, which clearly shows their value in active portfolio management.

The benefit of FAST-MCD is modest compared with the V method and it is inferior to the market. The reason most likely lies in the strict assumptions of the MCD approaches. Although both MCD methods and pair-wise robust estimators are designed to eliminate the effects of outliers, MCD models use a restrictive contamination model assuming complete dependence of outliers for different assets. Basically MCD models assume that each row of returns, $\boldsymbol{R}_t$, is either from the core distribution F_0 or outlier generating distribution H. The data are from the following mixed model:

$$F = (1-\varepsilon) F_0 + \varepsilon H , \qquad 0 < \varepsilon < \frac{1}{2} , \tag{3.2}$$

where F is the mixed model; F_0 is a multivariate normal distribution; H is an arbitrary multivariate distribution that generates outliers.

Such a contamination model is rather restrictive for our application. By looking at N-dimensional outliers, the models assume that all asset returns for any given day are either from a core distribution F_0 or outlier generating distribution H. This assumption is only true if the market is the only factor that determines asset returns or there are high correlations between different assets' returns. In practice, the market return by itself only explains a small percentage of the variance of asset returns. Industrial factors and idiosyncratic risk have been shown to explain the majority of the return variances. The pair-wise models [1, 2] use a much more flexible mixed model for data:

$$\boldsymbol{R}_t = (\boldsymbol{I}-\boldsymbol{B}) \boldsymbol{Y}_t + \boldsymbol{B} \boldsymbol{Z}_t , \tag{3.3}$$

with $\boldsymbol{B} = \operatorname{diag}\big([B_1 \; B_2 \; \cdots \; B_N]\big)$, $\boldsymbol{Y}_t$ multivariate normal, $\boldsymbol{Z}_t$ an arbitrary random vector, and the B_i, Bernoulli random variables with success probability ε_i. We can assume any format for the correlation matrix matrix of $(B_1, B_2, \ldots, B_N)$. MCD models assume complete dependence $B_1 = B_2 = \cdots = B_N$, while pair-wise models often assume independent B_i and B_j, $i \neq j$, or independently evaluate the correlation for each pair of B_i and B_j. As a result, pair-wise robust estimators offer more flexibility to calculate the covariances. Once the positive semidefiniteness property of the covariance matrix is guaranteed through transformation, they provide far better results than FAST-MCD.

As shown in Table 2, pair-wise Winsorization methods are also faster than the FAST-MCD method (10 hours) for the same data set. The sampling process of FAST-MCD is much faster than the original MCD method, but the C-steps still require extensive computation. Between the two pair-wise Winsorization methods, F2D-Huber (35 minutes) is faster because it eliminates the repeated iteration step in I2D-Winsor (3 hours), while I2D-Winsor is likely to yield a more robust estimation of the covariance and indeed gives slightly better results than F2D-Huber in our study. It is also worth noting that the estimated covariance matrix often slightly underestimates the real covariance, so the estimation is biased. Yet it is believed that for the constant $c = 5.99$ (the 95% quantile of the χ^2_2 distribution) that we chose, the bias would be small. Furthermore, the asset weights depend on the relative size of the covariance, so the impact of bias on our problem is even smaller.

Table 2: Run Time for 500 Rebalancings. All programs were run on a computer with 3 GHz CPU and 3 GB of RAM.

Time	V	FAST-MCD	I2D-Winsor	F2D-winsor	V1	Market
500 Rebalances	40 sec	10 hr	3 hr	35 min	4 hr	< 4 sec

Penalization methods are more computationally intensive than pair-wise Winsorization methods. The addition of the penalty term extends the dimension of the optimization problems and increases the number of constraints. The cross-validation of each penalty coefficient λ increases the computation further by ~ 25 fold. Unlike robust estimation of the mean and covariance matrix, which only need to calculate the parameters once for all μ_p, the optimization problem needs to be performed for every μ_p. As a result, the run times of penalization methods are often longer.

Though computationally intensive, the V1 method using the market index as the prior carries great advantages. It yields the best information ratio with or without transaction costs. Because of the L1 penalty term, most asset weights are mainly restricted to the market weight, which dramatically reduces the asset turnover compared with pair-wise Winsorization methods. Penalization methods are especially valuable when the number of assets is of the same order of magnitude as the number of observations (in our study, $T = 2N$), since the covariance matrix is often ill-conditioned.

We also compared our methods with some of the factor-based models, e.g., CAPM model, Principal Component Analysis model, Shrinkage model ([16]) and multivariate GARCH models (e.g., Constant Conditional Correlation GARCH and Dynamic Conditional Correlation GARCH [28, 29]). The results [31] show that both pair-wise Winsorization methods and penalization methods perform better than these traditional approaches.

4. CONCLUSION

The implementation of the mean-variance portfolio optimization is limited in practice by difficulties in estimating model inputs, expected returns and the covariance matrices of different assets, and the sensitivity of asset weights assigned to these inputs. Traditionally, sample means and covariance matrices from historical data were used, which are subject to large estimation errors.

This paper investigates some of the recently developed robust statistical methods such as FAST-MCD, Iterative 2-D Winsorization, Fast 2-D Winsorization and penalization methods. These methods prove to be valuable tools in improving risk-adjusted portfolio performance and reducing asset turnover. Results also show that the V1 penalization method outperforms the 2-D Winsorization methods. However, they achieve this at the cost of significantly higher computational complexity. The computational problem may be overcome by the recently developed LARS algorithm [9]. LARS greatly speeds up computations for LASSO since all solutions for all λ can be found in about the same time as one-least-squares regression, which removes the need for a grid search on λ. If the LARS algorithm can be successfully applied to penalized portfolio optimization, then penalization methods can be used to allocate weights for 700 individual stocks directly instead of 51 sector index funds. This is work in progress.

ACKNOWLEDGMENTS

This research was supported, in part, by the Singapore-MIT Alliance and the MIT Center for Computational Research in Economics and Management Science.

REFERENCES

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

[2] Alqallaf, F.A.; Van Aelst, S.; Yohai, V.J. and Zamar, R.H. (2005). *A model for contamination of multivariate data*, Working Paper, University of British Columbia.

[3] Balvers, R.J.; Cosimano, T.F. and McDonald, B. (1990). Predicting stock returns in an efficient market, *Journal of Finance*, **45**(4), 1109–1128.

[4] Brunelli, R. and Messelodi, S. (1995). Robust estimation of correlation with applications to computer vision, *Pattern Recognition*, **28**(6), 833–841.

[5] Chilson, J.; Ng, R.; Wagner, A. and Zamar, R. (2004). Parallel computation of high dimensional robust correlation and covariance matrices, *Proceedings of the ACM SIGKDD*, 533–538.

[6] DeMiguel, V. and Nogales, F.J. (2006). *Portfolio selection with robust estimation of risk*, Working Paper, London Business School.

[7] Dickenson, J. (1974). Some statistical aspects of portfolio analysis, *Statistician*, **23**(1), 5–16.

[8] Dueck, A. and Lohr, S. (2005). Robust estimation of multivariate covariance components, *Pattern Recognition*, **61**, 162–169.

[9] Efron, B.; Hastie, T. and Johnstone, I. (2004). Least angle regression, *Annals of Statistics*, **32**(2), 407–499.

[10] Jorion, P. (1976). International portfolio diversification with estimation risk, *Journal of Business*, **58**(3), 259–278.

[11] Khan, J.A.; Van Aelst, S. and Zamar, R.H. (2002). *Robust linear model selection based on least angle regression*, Working Paper, Statistics Department, University of British Columbia.

[12] Klein, R.W. and Bawa, V.S. (1976). Effect of estimation risk on optimal portfolio choice, *Journal of Financial Economics*, **3**, 215–231.

[13] Knez, P.J. and Ready, M.J. (1997). On the robustness of size and book-to-market in cross-sectional regressions, *Journal of Finance*, **52**(4), 1355–1382.

[14] Lauprete, G.J. (2001). *Portfolio risk minimization under departures from normality*, MIT PhD Thesis.

[15] Lauprete, G.; Samarov, A. and Welsch, R. (2002). Robust portfolio optimization, *Metrika*, **55**, 139–149. Also appeared in "Developments in Robust Statistics" (R. Dutter, P. Filzmoser, U. Gather and P. Rouseeuw, Eds.), Physica-Verlag, Heidelberg, 235–245.

[16] Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices, *Journal of Multivariate Analysis*, **88**(2), 365–411.

[17] Markowitz, H. (1952). Portfolio selection, *Journal of Finance*, **7**, 77–97.

[18] Markowitz, H. (1959). *Portfolio Selection*, Blackwell, Cambridge.

[19] Maronna, R.A. (1976). Robust M-estimators of multivariate location and scatter, *The Annals of Statistics*, **4**, 51–67.

[20] Maronna, R.A.; Martin, R.D. and Yohai, V.J. (2006). *Robust Statistics: Theory and Methods*, Wiley, New Jersey.

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

[22] Martin, R.D. and Simin, T.T. (2003). Outlier-resistant estimates of beta, *Financial Analysts Journal*, **59**(5), 56–69.

[23] Merton, R. (1972). An analytical derivation of the efficient portfolio frontier, *Journal of Financial and Quantitative Analysis*, **7**, 1851–1872.

[24] Michaud, R. (1989). The Markowitz optimization enigma: Is optimized optimal? *Financial Analysts Journal*, **45**, 31–42.

[25] Rousseeuw, P.J. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator, *Technometrics*, **41**(3), 212–223.

[26] Scherer, B. and Martin, R.D. (2005). *Introduction to Modern Portfolio Optimization with NUOPT and S-Plus*, Springer, New York.

[27] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B*, **58**(1), 267–288.

[28] Tse, Y.K. (2000). A test for constant correlations in a multivariate GARCH model, *Journal of Econometrics*, **98**(1), 107–127.

[29] Tse, Y.K. and Tsui, A.K.C. (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations, *Journal of Business*, **20**(3), 351–362.

[30] Vaz de Melo, B. and Câmara, R.P. (2003). Robust modeling of multivariate financial data, *COPPEAD Working Paper Series*, No. 355, Graduate School of Business, Federal University at Rio de Janeiro, Brazil.

[31] Zhou, X. (2006). *Application of robust statistics to asset allocation models*, MIT MS Thesis.

Robust asset allocation strategies: relaxed versus classical robustness. IMA Journal of Management Mathematics. doi: 10.1093/imaman/dps023. Google Scholar. Rockafellar, R. T. (1970). Convex analysis. Princeton: Princeton University Press. Google Scholar.Â Application of robust statistics to asset allocation models. REVSTAT Statistical Journal, 5(1), 97â€“114. Google Scholar. Yaari, M. E. (1987). The dual theory of choice under risk. Econometrica, 55, 95â€“115. CrossRefGoogle Scholar. Zhu, S. S., & Fukushima, M. (2009). Worst-case conditional value at risk with application to robust portfolio management.