Robust Portfolio Selection

VICTORIA-FESER, Maria-Pia

VICTORIA-FESER, Maria-Pia. Robust Portfolio Selection. 2000
Robust Portfolio Selection

by Maria-Pia Victoria-Feser

Université de Genève\textsuperscript{1}

August 17, 2000

\textsuperscript{1}The research was performed while the author was at the London School of Economics in 1994-1996.
1 Introduction

The Markowitz (1959) mean-variance efficient frontier is the standard theoretical model for normative investment behaviour. It is still in practice the method of choice for optimal portfolio construction although among practitioners it has nearly lost its character of ‘optimal’ tool. Indeed it is often considered not worthwhile, not because of its mathematical complexity, but because it often leads to financially irrelevant optimal portfolios.

In this paper, we discuss one of the reasons leading practitioners to the rejection of the Markowitz model and propose a new statistical method to avoid this problem. To be more precise, we discuss the problem of statistical robustness of the Markowitz optimizer and show that the latter is not robust, meaning that a few extreme assets prices or returns can lead to irrelevant ‘optimal’ portfolios. We then propose a robust Markowitz optimizer and show that it is far more stable than the classical version.

Classical mean-variance optimization assumes that the investor prefers a portfolio of securities that offers maximum expected return for some given level of risk. The latter is measured by the standard deviation of the return. Suppose there are \( N \) securities to choose from and let \( p = (p_1, \ldots, p_N) \)' be the vector of portfolio weights, such that \( \sum_{i=1}^{N} p_i = 1 \). The expected return of the portfolio can be written as

\[
R(p) = \mu p
\]

where \( \mu = (\mu_1, \ldots, \mu_N) \) is the vector of securities' expected returns. The variance of the portfolio can be written as

\[
S(p) = p' \Sigma p
\]

where \( \Sigma \) is the covariance matrix of the securities. For a given risk-aversion parameter \( \lambda \) (which varies to trace out the efficient frontier) the mean-variance procedure selects the portfolio \( p^{\text{opt}} \) which maximizes (in \( p \))

\[
R(p) - \lambda S(p)
\]

subject to

\[
p \geq 0
\]

\[
e_N' p = 1
\]

where \( e_N \) \((N \times 1)\) is a vector of ones.

---

\(^1\)Under securities we include all sorts of financial tools that the investor has access to, like assets, bonds, funds, etc.
The set of optimal portfolios for all possible levels of portfolio risk (or $\lambda$) defines the mean-variance efficient frontier. The constraint of no short-selling ($p \geq 0$) is also sometimes removed, whereas other linear constraints such as trading costs may be imposed. If there exist a borrowing interest rate, then one unique optimal portfolio can be found. It is given by the portfolio that maximizes the Sharpe ratio (excess return of the portfolio divided by its standard deviation) under the constraints given above.

The mean-variance optimizer suffers from several drawbacks. It is very often believed that optimizers produce a unique optimal portfolio for a given level of risk. The uniqueness of the solution depends on the erroneous assumption that the inputs (mean and variance) are known, whereas they must be estimated and therefore contain statistical error. The expected means, variances and covariances are usually estimated by means of the classical maximum likelihood estimator (MLE). Instead of $\mu$ we have

$$\hat{\mu} = \frac{1}{T} e_T Y$$  \hspace{1cm} (6)

where $Y$ is the matrix containing the observed returns of each securities (columnwise) for $T$ periods, and is given by

$$Y = \begin{bmatrix} y_1^T \\
\vdots \\
y_T^T \end{bmatrix} = \begin{bmatrix} y_1 \ldots y_N \end{bmatrix} = \begin{bmatrix} y_{11} & \cdots & y_{1N} \\
\vdots & \ddots & \vdots \\
y_{T1} & \cdots & y_{TN} \end{bmatrix}$$  \hspace{1cm} (7)

Moreover, instead of $\Sigma$ we have

$$\hat{\Sigma} = \frac{1}{T} (Y - e_T \hat{\mu})^T (Y - e_T \hat{\mu})$$  \hspace{1cm} (8)

Therefore, when $\hat{\mu}$ and $\hat{\Sigma}$ are inputed in the optimizer they induce statistical errors. We can distinguish two sources of errors introduced in the model. One of them has been widely discussed in the economic literature. It relies on the rather large variability of the MLE of the securities expected return. Indeed, if one constructs, for example by means of simulations, for each optimal portfolio a ‘confidence region’ of statistically equivalent portfolios, one finds that these statistically equivalent portfolios may have significantly, even radically different portfolio structures. Moreover, one can show that on average, the sharpe ratio of the selected portfolios are not only significantly below the true Sharpe ratios but also below the Sharpe ratio of an ‘ad hoc’ portfolio, like an equally weighted portfolio (see Jobson and Korkie 1981). This problem is often referred to as the “estimation error maximizers” (see Michaud 1989). To remedy to this particular problem, Jorion (1984) proposed to use an alternative estimator to the classical estimator for the expected return given
by the Bayes-Stein shrinkage estimator (Stein 1955). Simply, the observed sample means for individual assets are shrunk to some global mean (pooled mean, a Bayesian prior or the mean of the minimum variance efficient-frontier portfolio), in a way that the greater the variability in the historical data, the greater the shrinkage.

The other major statistical error is concerned with the hypothesis underlying the Markowitz model. Indeed, the theory assumes that the security return are distributed according to a multivariate normal distribution with mean $\mu$ and variance $\Sigma$, or that the distribution of the returns is completely specified by their first two moments. In other words, it is assumed that the observed returns are realizations of a multivariate normal distribution. What happens if this hypothesis is slightly violated, like for example when one of the securities has an unexpected high or low return? To our knowledge, this question has never been raised in the context of optimal portfolio selection. In the statistical literature this problem is referred to as a problem of robustness. Simply, robustness is concerned with the stability of an estimator (here $\mu$ and $\Sigma$) of parameters (here $\mu$ and $\Sigma$) from a parametric model (here the multivariate normal distribution) when there are model mis-specifications, in particular when there are outlying observations, i.e. observations which cannot be considered as belonging to the parametric model. Although we will show numerically and analytically the non robustness of the classical optimizer, it seems natural to believe that if for example one security has an exceptional high return (outlying observation), then the expected return for this security (and also its variance) will be overestimated and in turn the optimizer will tend to allocate a higher proportion to this security. The reasoning is more complicated when several securities have extreme returns, but as we will see later, also in these cases, the optimizer will lead to biased optimal portfolios. We also propose to use robust estimators (i.e. estimators which are not influenced, or at least less influenced than the MLE, by outlying observations). These robust estimators have been developed in the statistical literature since the pioneering works of Huber (1964) and Hampel (1968).

The paper is organized as follows. In Section 2, we present an extensive simulation study. We show that only 1% of extreme returns generated randomly have the effect of leading the classical Markowitz optimizer to irrelevant portfolios. On the other hand, we show that by using a robust optimizer, the extreme returns do not influence the calculation of the optimal portfolios. In Section 3 we first show analytically how the classical Markowitz optimizer is influenced by extreme returns, and then we propose a robust version. In Section 4, we apply our results in a real situation. Finally, Section 5 concludes.
2 Simulation study

In this section we show through a simulation study that the computation of the efficient frontier can be seriously biased when the data contain extreme observations. These extreme observations are referred to in the statistical literature as outliers. In our context, they can be defined as high or low returns that are disproportional when compared to the other returns for the same security, or as observations that have a infinitesimal probability of having been generated by the multivariate normal distribution that has generated the rest of the observations. The consequence of the presence of these outliers in the data is that unfortunately they bias considerably the estimators $\hat{\mu}$ and $\hat{\Sigma}$ of the expected returns and their associated risks and correlations. This means that the estimated efficient frontier, i.e. the one based on $\hat{\mu}$ and $\hat{\Sigma}$, is actually very different from the efficient frontier based on $\mu$ and $\Sigma$. In order to show this, we simulated 40 sets of 120 observations from a multivariate normal distribution with mean

$$\mu = [ 3.25 \ 5.46 \ 2.27 \ 6.45 \ 7.02 \ 4.44 \ 6.11 ] \quad (9)$$

and variance derived from the standard deviations of the returns given by

$$\sigma = [ 0.967 \ 1.435 \ 0.133 \ 1.961 \ 2.001 \ 1.172 \ 1.731 ] \quad (10)$$

and correlation matrix given by (upper triangle only)

$$\rho = \begin{bmatrix} 1 & -0.54 & 0.07 & -0.46 & 0.73 & 0.09 \\ 1 & 0.43 & -0.35 & 0.38 & -0.44 & 0.04 \\ 1 & -0.54 & 0.25 & -0.22 & -0.31 \\ 1 & -0.09 & 0.52 & 0.07 \\ 1 & -0.23 & 0.04 \\ 1 & 0.50 \\ 1 \end{bmatrix} \quad (11)$$

We then contaminated half of the sets by, for each series of return, multiplying by 10, 1% of the returns chosen randomly.

To show the effect of the introduction of extreme returns, we first computed the efficient frontier based on the non contaminated data and on the contaminated data. Figure 1 shows the different efficient frontiers. The dotted lines represent efficient frontiers based on non contaminated samples, lines represent efficient frontiers based on contaminated samples\(^2\) and the line represented by the symbol "O" represents the true efficient frontier, i.e.

\(^2\)The efficient frontiers based on contaminated samples are actually defined for higher values of risk that do not appear in Figure 1.
the one based on the true parameters $\mu$ and $\Sigma$. We can see that the efficient frontiers based on contaminated samples are far from the true efficient frontier, whereas the efficient frontiers based on non contaminated samples are near the true efficient frontier. This means that if there are extreme returns in the data (as it is likely to happen) the Markowitz optimizer leads to irrelevant portfolios.

In a second step, we simulated similar data sets and this time computed efficient frontiers by means of a robust optimizer. Figure 2 shows the results. The dotted lines represent the efficient frontiers based on non contaminated samples, the lines represent the efficient frontiers based on contaminated samples and the line represented by the symbol ”O” represents the true efficient frontier. We can see that the different efficient frontiers are confounded, meaning that a robust optimizer is not influenced by extreme returns and thus do not lead to irrelevant portfolios.

To show the consequence of using the classical Markowitz optimizer when there are extreme returns, we fixed a level of expected return (to 6%) and computed the risk of the optimal portfolio found by a) the Markowitz optimizer on non contaminated data (MLE 0%), b) the Markowitz optimizer on contaminated data (MLE 1%), c) the robust optimizer on non contaminated data (Rob 0%), and d) the robust optimizer on contaminated data (Rob 1%). In Figure 3 are represented the distribution of the optimal portfolios risks derived from the different situations. We can clearly see that when there are no extreme returns or/and when the robust optimizer is used, we get the same results. On the other hand, when there are extreme returns and the classical Markowitz optimizer is used, the results are quite different.

Until now we have only considered the effect of extreme returns on the computation of the efficient frontiers. We now study the effect on the optimal portfolios themselves, i.e. their composition or in other words, the vector of weights $p$. For a fixed expected return (6%), we computed the vector of weights $p$ of the optimal portfolio found in the four different situations explained above. Figure 4, 5, 6, and 7 show the distribution of the weights corresponding to the 7 different assets. The symbol ”O” represents the true weight. Without extreme returns, the classical Markowitz optimizer and the robust optimizer give very similar results, i.e. weights that are near and around their true values (see Figure 4 and 6). With extreme returns, the picture is completely different. The classical Markowitz optimizer gives most of the time irrelevant portfolios (see Figure 5). Moreover, they seem to vary a lot from set to set. The robust optimizer, on the other hand, seems to be very much less influenced by these extreme returns and also more stable, thus giving optimal portfolios that are realistic (see Figure 7).

An interesting question at this point would be how would the selected optimal portfolios behave in the future? Figure 8 shows the distribution of the expected returns for the different optimal portfolios corresponding to the
fixed return of 6%, under the four different situations explained above. It seems clear that when there is no contamination and/or we use the robust optimizer, the expected returns generated by the optimal portfolios are conform to the expected return generated by the true optimal portfolio (the line in the graph). However, when there are extreme returns and when we use the classical optimizer, the expected returns generated by the optimal portfolios are significantly below the expected return generated by the true optimal portfolio.

With all these examples, it seems then that it is a safer option to use a robust optimizer in practice, since there is no guarantee that the data are free from extreme returns. In the following Section we will present the robust optimizer we have used in our simulations.

3 Robust Markowitz optimizer

In this section, we first present the statistical robustness concepts that we will then use to show that the classical Markowitz optimizer is very sensitive to extreme values. We also present the different tools available to make the Markowitz optimizer robust and will propose a robust optimizer.

3.1 Concepts of robust statistics

Robust statistics is an extension of classical statistics that takes into account the possibility of model misspecification (including outliers). Robust statistics are considered to belong to parametric statistics, because when one develops or applies the related techniques, one still makes the assumption that the underlying model is a parametric model. In our case, the parametric model is the multivariate normal model with parameters $\mu$ and $\Sigma$. However, this is just a particular case, and the theory presented below is valid for any parametric model.

The following set of distributions defines a neighbourhood of the parametric model $F_{\mu, \Sigma}$, set which includes all the possible misspecified distributions around $F_{\mu, \Sigma}$. The set is given by

$$\left\{ G_\varepsilon | G_\varepsilon = (1 - \varepsilon)F_{\mu, \Sigma} + \varepsilon W \right\}$$

where $W$ is an arbitrary distribution. $G_\varepsilon$ can be considered as a mixture distribution between $F_{\mu, \Sigma}$ and the contamination distribution $W$. One particular case is when $W = \Delta_z$, the distribution that gives a probability of one to a point $z$ chosen arbitrarily. In this case, the neighbourhood is given by

$$\left\{ F_\varepsilon | F_\varepsilon = (1 - \varepsilon)F_{\mu, \Sigma} + \varepsilon \Delta_z \right\}$$
Hence $F_\varepsilon$ generates observations from $F_{\mu, \Sigma}$ with probability $(1 - \varepsilon)$ and observations equal to any point $z$ (contamination) with probability $\varepsilon$.

One way of assessing the robustness properties of an estimator $\hat{\mu}$ and $\hat{\Sigma}$ of $\mu$ and $\Sigma$ is to study their (asymptotic) stability when we assume a distribution of type $F_\varepsilon$, i.e. in a neighbourhood of the parametric model. If we consider the case when $\varepsilon$ tends to zero (i.e. an infinitesimal amount of contamination), we actually get the so-called influence function ($IF$) developed by Hampel (1968), Hampel (1974), and further developed by Hampel et al. (1986).

The $IF$ calculates the influence of an infinitesimal amount of contamination $z$ on the value of the estimator. For $\hat{\mu}$ it is given by

$$IF(z, \hat{\mu}, F_{\mu, \Sigma}) = \lim_{\varepsilon \to 0} \frac{\hat{\mu}(F_{\varepsilon}) - \hat{\mu}(F_{\theta})}{\varepsilon}$$

or when the derivative exists

$$IF(z, \hat{\mu}, F_{\mu, \Sigma}) = \left. \frac{\partial}{\partial \varepsilon} \hat{\mu}(F_{\varepsilon}) \right|_{\varepsilon=0}$$

The $IF$ is in this case vector valued, whereas as we will see, it is matrix valued for $\hat{\Sigma}$. In general however, the $IF$ can be scalar, vector or matrix valued and $z$ can be scalar or vector.

The $IF$ is a very powerful tool to assess robustness properties of estimators. Indeed, it can be shown (see Hampel et al. 1986) that the $IF$ is sufficient to describe the asymptotic bias of the estimator caused by the contamination. That is, if we find an estimator that has a bounded $IF$, then we know that its asymptotic bias is also bounded. Moreover, the $IF$ looks at the worst bias that an infinitesimal amount of contamination can cause on the estimator. Therefore, a bounded $IF$ estimator is robust in a general neighbourhood of the parametric model defined by (12).

### 3.2 Robustness properties of the Markovitz optimizer

In this section, we first derive the $IF$ of the MLE of the multivariate normal model. In a second step, we show that the robustness properties of the optimal portfolio (selected by the optimizer) depends on the $IF$ of the estimator for the parameters of the multivariate normal model. Therefore we conclude that unless the mean and covariance matrix are estimated robustly, the selected optimal portfolio can be very badly influenced by a few extreme returns.

Let $Y$ be the $(T \times N)$ matrix of $T$ independent multivariate observations from a multivariate normal distribution $N(\mu, \Sigma)$. The joint density of the
$y_t$ is given by

$$f(Y; \mu, \Sigma) = (2\pi)^{-NT} \det(\Sigma)^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left[ (Y - e_T\mu)\Sigma^{-1}(Y - e_T\mu) \right] \right\}$$

(16)

A natural estimator for $\mu$ and $\Sigma$ is the MLE. It is found by minimizing the log-likelihood function

$$\mathcal{L}(Y, \mu, \Sigma) = -\frac{NT}{2} \log(2\pi) - \frac{T}{2} \log \det(\Sigma) - \frac{1}{2} \text{tr} \left[ (Y - e_T\mu)\Sigma^{-1}(Y - e_T\mu) \right]$$

(17)

By differentiating $\mathcal{L}(Y, \mu, \Sigma)$ we find

$$d\mathcal{L}(Y, \mu, \Sigma) = \text{tr} \left\{ \left[ \left( \frac{\partial \mathcal{L}(Y, \mu, \Sigma)}{\partial \Sigma} \right) \right]' d\Sigma \right\} + \text{tr} \left\{ \left[ \left( \frac{\partial \mathcal{L}(Y, \mu, \Sigma)}{\partial \mu} \right) \right]' d\mu \right\}$$

$$= \text{tr} \left\{ \left[ -\frac{T}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1}(Y - e_T\mu)'(Y - e_T\mu)\Sigma^{-1} \right] d\Sigma \right\} + \text{tr} \left\{ \Sigma^{-1}(Y - e_T\mu)'e_T d\mu \right\}$$

(18)

So that

$$\frac{\partial \mathcal{L}(Y, \mu, \Sigma)}{\partial \mu} = e_T'(Y - e_T\mu)\Sigma^{-1} = 0$$

(19)

$$\frac{\partial \mathcal{L}(Y, \mu, \Sigma)}{\partial \Sigma} = -\frac{T}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1}(Y - e_T\mu)'(Y - e_T\mu)\Sigma^{-1} = 0$$

(20)

define the MLE. Note that we can write (19) and (20) as

$$\frac{\partial \mathcal{L}(Y, \mu, \Sigma)}{\partial \mu} = \sum_{t=1}^{T}(y'_t - \mu)\Sigma^{-1} = 0$$

(21)

$$\frac{\partial \mathcal{L}(Y, \mu, \Sigma)}{\partial \Sigma} = \sum_{t=1}^{T} \left\{ -\frac{1}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1}(y'_t - \mu)'(y'_t - \mu)\Sigma^{-1} \right\} = 0$$

(22)

Namely, the MLE are then given by

$$\hat{\mu} = (e_T'e_T)^{-1}e_T'Y = \frac{1}{T} \sum_{t=1}^{T} y'_t$$

(23)

$$\hat{\Sigma} = \frac{1}{T}(Y - e_T\hat{\mu})'(Y - e_T\hat{\mu}) = \frac{1}{T} \sum_{t=1}^{T} (y'_t - \hat{\mu})'(y'_t - \hat{\mu})$$

(24)

The MLE of $\mu$ and $\Sigma$ can clearly be regarded as functionals $\hat{\mu}(F)$ and $\hat{\Sigma}(F)$ of some underlying distribution $F$. We can write implicity the MLE
at the empirical distribution $F(T)$ given by

$$F(T)(x) = \frac{1}{T} \sum_{t=1}^{T} \Delta_{x_t}(x)$$

(25)

where $\Delta_{x_t}(x)$ is a vector of zero and ones such that the $i$th element is 0 if the $i$th element of $x$ is strictly smaller than the $i$th element of $x_t$ and is 1 if the $i$th element of $x$ is greater than or equal to the $i$th element of $x_t$.

The MLE of $\mu$ and $\Sigma$ as functionals of the empirical distribution are defined by

$$\hat{\mu}(F(T)) = \int \cdots \int y' dF(T)(y)$$

$$\hat{\Sigma}(F(T)) = \int \cdots \int (y' - \mu(F(T)))'(y' - \mu(F(T))) dF(T)(y)$$

(26)

Asymptotically, $F(T) \rightarrow F_{\mu, \Sigma}$ so that we have

$$\hat{\mu}(F_{\mu, \Sigma}) = \int \cdots \int y' dF_{\mu, \Sigma}(y) = \mu$$

$$\hat{\Sigma}(F_{\mu, \Sigma}) = \int \cdots \int (y' - \mu)'(y' - \mu)dF_{\mu, \Sigma}(y) = \Sigma$$

(27)

We now check the robustness properties of the MLE for the multivariate normal model. The neighborhood of $F_{\mu, \Sigma}$ is given by the set of distributions $F_{\varepsilon}$ in (13). Therefore at $F_{\varepsilon}$ the MLE of $\mu$ and $\Sigma$ are defined by

$$\hat{\mu}(F_{\varepsilon}) = \int \cdots \int y' dF_{\varepsilon}(y)$$

$$= (1 - \varepsilon) \int \cdots \int y' dF_{\mu, \Sigma}(y) + \varepsilon z'$$

(30)

and

$$\hat{\Sigma}(F_{\varepsilon}) = \int \cdots \int (y' - \hat{\mu}(F_{\varepsilon}))(y' - \hat{\mu}(F_{\varepsilon})) dF_{\varepsilon}(y) =$$

$$= (1 - \varepsilon) \int \cdots \int (y' - \hat{\mu}(F_{\varepsilon}))(y' - \hat{\mu}(F_{\varepsilon})) dF_{\mu, \Sigma}(y) +$$

$$\varepsilon (z' - \hat{\mu}(F_{\varepsilon}))(z' - \hat{\mu}(F_{\varepsilon}))$$

(31)

To find the IF, we take the derivative of (30) and (31) with respect to $\varepsilon$ when $\varepsilon \rightarrow 0$. We have

$$\left. \frac{\partial}{\partial \varepsilon} \hat{\mu}(F_{\varepsilon}) \right|_{\varepsilon=0} = -\mu + z'$$

(32)
and

\[
\left. \frac{\partial}{\partial \varepsilon} \tilde{\Sigma}(F_\varepsilon) \right|_{\varepsilon=0} = -\int \cdots \left( y' - \mu \right)' (y' - \mu) \, dF_\mu \Sigma(y) - \int \cdots \left( y' - \mu \right)' \left. \frac{\partial}{\partial \varepsilon} \mu(F_\varepsilon) \right|_{\varepsilon=0} \, dF_\mu \Sigma(y) + (z' - \mu)' (z' - \mu)
\]

\[
= -\Sigma + (z' - \mu)' (z' - \mu)
\]

(33)

Hence, we can see that the IF for the MLE of \( \mu \) and \( \Sigma \) are unbounded, because for many values of \( z \), the IF become arbitrarily large.

How will this result influence the optimizer? For a given \( \lambda \) (risk-aversion parameter), the optimal portfolio is obtained by solving simultaneously

\[
\mu' - 2\lambda \Sigma p = 0
\]

(34)

\[
p \geq 0
\]

(35)

\[
e'_N p = 1
\]

(36)

\( p \) depends on \( \mu \) and \( \Sigma \) and should be written as \( p = p(\mu, \Sigma) \). However, generally \( \mu \) and \( \Sigma \) are unknown and are replaced by estimators \( \tilde{\mu} \) and \( \tilde{\Sigma} \), so that we have \( \tilde{p} = p(\tilde{\mu}, \tilde{\Sigma}) \). The real problem is then to solve

\[
\tilde{\mu}' - 2\lambda \tilde{\Sigma} \tilde{p} = 0
\]

(37)

\[
\tilde{p} \geq 0
\]

(38)

\[
e'_N \tilde{p} = 1
\]

(39)

The estimators can be written as functional of a given distribution. To study the effect of a small amount of contamination on the efficient frontier, we suppose the distribution \( F_\varepsilon \). We then have

\[
\tilde{\mu}(F_\varepsilon)' - 2\lambda \tilde{\Sigma}(F_\varepsilon)p(\tilde{\mu}(F_\varepsilon), \tilde{\Sigma}(F_\varepsilon)) = 0
\]

(40)

\[
p(\tilde{\mu}(F_\varepsilon), \tilde{\Sigma}(F_\varepsilon)) \geq 0
\]

(41)

\[
e'_N p(\tilde{\mu}(F_\varepsilon), \tilde{\Sigma}(F_\varepsilon)) = 1
\]

(42)

We now take the derivatives of (40), (41) and (42) with respect to \( \varepsilon \) at \( \varepsilon = 0 \). For (40), we get

\[
\left. \frac{\partial}{\partial \varepsilon} \tilde{\mu}(F_\varepsilon)' \right|_{\varepsilon=0} - 2\lambda \left. \frac{\partial}{\partial \varepsilon} \tilde{\Sigma}(F_\varepsilon) \right|_{\varepsilon=0} p - 2\lambda \Sigma \left. \frac{\partial}{\partial \varepsilon} \tilde{p} \right|_{\varepsilon=0} = 0
\]

(43)
Equation (41) doesn’t give information, and for (42), we have
\[ e_N' \frac{\partial}{\partial \epsilon} \tilde{p} \bigg|_{\epsilon=0} = 0 \] (44)

Equation (43) that can be written as
\[ \frac{\partial}{\partial \epsilon} \tilde{p} \bigg|_{\epsilon=0} = \frac{1}{2\lambda} \Sigma^{-1} \frac{\partial}{\partial \epsilon} \tilde{\mu}(F_{\epsilon})' - \Sigma^{-1} \frac{\partial}{\partial \epsilon} \tilde{\Sigma}(F_{\epsilon}) \bigg|_{\epsilon=0} \] (45)
tells us that the effect of the contamination on the optimal portfolio depends on the effect of the contamination on the estimator of the mean vector and covariance matrix. (44) says that the sum of the effects on the weights defining the optimal portfolio cancel each other. This is an expected result because from the structure of the problem, the weights are constrained to sum to 1.

Therefore, we conclude that unless the estimators of the mean vector and of the covariance matrix are estimated robustly, the Markowitz optimizer can lead to irrelevant portfolios, i.e. it would allocate a very high weight to a security with an extreme high return, that would have received a low weight without the presence of this extreme return. As we have seen above, the classical estimators for \( \mu \) and \( \Sigma \), i.e. the MLE, are not robust to infinitesimal amounts of contamination, therefore, the classical Markowitz optimizer is not robust.

In the next subsection, we review different estimators for \( \mu \) and \( \Sigma \) that are robust and propose a robust optimizer.

### 3.3 Robust portfolio selection

In this subsection, we shortly review some of the robust estimators for \( \mu \) and \( \Sigma \) proposed in the statistical literature. They will enable us to propose a robust optimizer.

Robust estimators for location an scale with multivariate data have first been proposed by Gnanadesikan and Kettenring (1972). Following their work, a number of other propositions have appeared since in the statistical literature. In order to review some of the propositions, we have to present some desirable properties that robust estimators should fulfil. One of them is the property of \textit{affine equivariance} and is fulfilled by estimators \( \tilde{\mu}(Y) \) of location and estimators \( \tilde{\Sigma}(Y) \) of scale that satisfy

\[ \tilde{\mu}(YA + e_T b) = \tilde{\mu}(Y)A + b \] (46)
\[ \tilde{\Sigma}(YA + e_T b) = A' \tilde{\Sigma}(Y)A \] (47)
This property is considered important because it is "natural". Location estimators that are not affine equivariant do not necessarily lie in the convex hull of the sample, i.e. the estimates are not necessarily in the center of the data. Scale estimators that are not affine equivariant lead to matrices that are not necessarily semidefinite positive, a property of covariance matrices. However, non affine equivariant estimators are very often intuitive and are simple to compute. Moreover, the problem of non semidefinite covariance matrices can be overcome as we will see later.

Another important property is the statistical property of global robustness. That is, it is often not sufficient to know that an estimator is robust to infinitesimal amounts of contamination in the data, one also requires an estimator that is robust to a relatively large amount of contamination. This concept is measured by means of the breakdown point. It gives the maximum proportion of contamination introduced in the data that the estimator can bear before it breaks down, or before it can take arbitrary values. For example, with univariate data, the sample mean has a (sample) breakdown point of one over the sample size, since only one observation can make the sample mean have an arbitrary value. On the other hand, the sample median has a breakdown point of approximately \( \frac{1}{2} \) since one needs one half of the data on one side of the median to be outliers before the sample median takes arbitrary values. The breakdown point varies between 0 and \( \frac{1}{2} \). However, very robust estimators, i.e. estimators with high breakdown points, usually are less efficient than less robust estimators or than classical estimators. Moreover, they tend to be more computer intensive. Therefore, it is not always obvious to choose among the different robust estimators proposed in the literature, one has to make compromises between a) ease of computation, b) robustness properties and c) efficiency considerations.

Affine equivariant robust estimators of location and scale have been first formally introduced by Maronna (1976) and then generalized by Huber (1977). These are affine equivariant M-estimators and are implicitly defined as the solution in \( \mu \) and \( \Sigma \) of

\[
\mu = \frac{\frac{1}{T} \sum_{t=1}^{T} u_1(s_t) y'_t}{\frac{1}{T} \sum_{t=1}^{T} u_1(s_t)}
\]

\[
\Sigma = \frac{\frac{1}{T} \sum_{t=1}^{T} u_2(s_t)(y'_t - \mu')(y'_t - \mu)}{\frac{1}{T} \sum_{t=1}^{T} u_3(s_t)}
\]

where \( s_t = (y'_t - \mu)\Sigma^{-1}(y'_t - \mu)' \) and with \( u_1, u_2 \) and \( u_3 \) being real-valued weight functions. The intuitive interpretation behind these M-estimators is that they are adaptively weighted sample means and sample covariance matrices with the weights determined by the data through the adaptive Mahalanobis distances \( s_t, 1 \leq t \leq T \). Their properties have been well
studied, see e.g. (Huber 1981) and Hampel et al. (1986). A disappointing feature of M-estimators is their relatively low breakdown point which is not more than $\frac{1}{2}$ (see Stahel 1981).

Among the affine equivariant robust estimators with high breakdown point, the most well known is Rousseeuw (1983) and Rousseeuw (1985) \textit{minimum volume ellipsoid} (MVE) estimator. The location estimator is the center of the minimal volume ellipsoid covering (at least) $h$ data points, where $h$ can be as small as one-half of the data. The scale estimator is given by the ellipsoid itself, multiplied by a suitable factor to obtain consistency. The MVE has a breakdown point of $\frac{1}{2}$. To compute the MVE, several algorithm have been proposed in the literature. Since it is not feasible to consider all the combinations of $h$ data, one resorts to approximate algorithms. Rousseeuw and Leroy (1987) and Rousseeuw and Van Zomeren (1990) proposed the first algorithm which was later criticized by Woodruff and Rocke (1994) and Atkinson (1994) who proposed other algorithm that were, among other advantages, faster and more reliable with small sample sizes. The MVE suffers from an important drawback, it is not very efficient. Therefore, it is never used in itself but as an initial solution on which to base a one-step improvement. Rousseeuw and Leroy (1987) proposed to use the MVE to detect the outlying observations, then remove them from the sample, and compute the MLE for location and scale on the clean sample. To be more precise, let $\hat{\mu}_{\text{MVE}}$ and $\hat{\Sigma}_{\text{MVE}}$ be the MVE of location and scale, the weights $w_t$, $t = 1, \ldots, T$, are defined as

$$w_t = \begin{cases} 1 & \text{if } (y_t' - \hat{\mu}_{\text{MVE}})\hat{\Sigma}_{\text{MVE}}^{-1}(y_t' - \hat{\mu}_{\text{MVE}})' \leq \chi^2_{N,0.975} \\ 0 & \text{otherwise} \end{cases} \quad (50)$$

Then the reweighted estimators are given by

$$\tilde{\mu} = \frac{\sum_{t=1}^{T} w_t y_t'}{\sum_{t=1}^{T} w_t} \quad (51)$$

and

$$\tilde{\Sigma} = \frac{\sum_{t=1}^{T} w_t (y_t' - \tilde{\mu})(y_t' - \tilde{\mu})'}{\sum_{t=1}^{T} w_t - 1} \quad (52)$$

This robust estimator is proposed in the statistical software Splus which runs on workstations as well as on Windows for PCs.

MVE are used as starting point to other iteratively defined robust estimators of location and scale, such as \textit{S-estimators} (see Rousseeuw and Yohai 1984 and Davies 1987). For other such estimators, the reader is referred to the survey by Tyler (1991). Recently, Cheng and Victoria-Feser (1998) have also reviewed high breakdown point estimators of location and scale and proposed one for the situation of missing data.
A robust optimizer is then obtained by estimating the location and scale by means of a robust estimator, and then use these estimators in the mathematical problem given by (3), (4) and (5). In particular, we used the routine given in Splus.

4 Application to CTA

In this section, we analyse a real example. The data come from the TASS Management database TASS 2.3 and represent the monthly returns of about 300 different Commodity Trading Advisors (CTA) around the world. We selected 7 CTA and considered their (published) rate of return from January 1985 to October 1994. These rates of returns are presented in the barplots given in Figure 9 and Figure 10. We can first notice that some of the CTA have a few extreme returns like CTA nb 3, 4 and 5, especially at the beginning of the considered period. We then would expect the classical optimizer to give a greater weight to these CTAs, whereas the robust optimizer would not be influence by these extreme returns.

We computed the efficient frontier based on the classical and robust optimizer and drawn them in Figure 11. Unfortunately, the efficient frontiers don’t tell us very much in this case as to what to expect from the optimal portfolios selected by the two methods. Because the classical efficient frontier is above the robust efficient frontier doesn’t tell us that is it better, it just says that the estimated returns and risks are larger than with the robust optimizer. An interesting information is given by the composition of the optimal portfolios. If we look at the weights, we remark the the classical optimizer tends to give a greater weight to CTA nb 2, 4 and 5 and gives a lower weight to CTA nb 7. This can be partially explained by the fact that both CTA nb 4 and 5 appear to have some extreme high returns, whereas CTA nb 7 has one extreme low return.

A further step in the analysis would be to check which observations are considered as outliers. Figure 12 is a barplot of the weights given to the different observations. It seems that the MVE detects outliers at the beginning of the series, as suggested by the barplots of the rates of return.

We therefore believe that it is really a safer option to at least consider the use of a robust optimizer that will enable the user to not only be safe againts etreme returns, but also to find the extreme returns and thus assets that are over or under estimated.
5 Conclusion

In this paper, we have shown that the classical Markowitz optimizer suffers from an important drawback, which is its non-robustness to extreme returns. We showed this phenomenon by means of simulations and then analytically by using the concept of the $IF$. We then reviewed some of the robust estimators of location and scale proposed in the statistical literature and proposed a robust optimizer. Finally, by means of a real example, we showed the usefulness of at least considering the use of a robust optimizer when selecting optimal portfolios.
References


Figure 1: Classically estimated efficient frontiers

Figure 2: Robustly estimated efficient frontiers
Boxplots of risks for fixed return (6%)

Figure 3

Classical optimal portfolios without contamination

Figure 4
Classical optimal portfolios with contamination

Figure 5

Robust optimal portfolios without contamination

Figure 6
Robust optimal portfolios with contamination

Figure 7

Efficiency of the optimal portfolios

Figure 8
Published returns for CTA1

Published returns for CTA2

Published returns for CTA3

Published returns for CTA4

Figure 9: Published returns for some CTAs
Published returns for CTA5

Published returns for CTA6

Published returns for CTA7

Figure 10: Published returns for some CTAs
Efficient frontiers on real dataset

![Graph showing classical and robust efficient frontiers on a risk-return plane.]

Figure 11

Outliers detection plot

![Plot showing outliers detection.]

Figure 12