

# Constrained Shrinkage Estimation for Portfolio Robust Prediction

Luis P. Yapu Quispe

Universidade Federal Fluminense

**Abstract** It is possible to reformulate the portfolio optimization problem as a constrained regression. In this paper we use a shrinkage estimator combined with a constrained robust regression and apply it to portfolio robust prediction. Starting with robust estimates  $(\hat{\mu}_R, \hat{\Sigma}_R)$ , we solve the constrained optimization problem in order to obtain a robust estimation of the portfolio weights. By varying a shrinkage parameter it is possible to 'interpolate' between the robust and least-squares cases and to find an optimal value of this parameter with the best predictive power. Indeed recurrence of outliers in financial data may require some flexibility aside robustness. In particular we derive a closed formula for linear constrained regression M-estimator and present a procedure intertwining this solution with the shrinkage estimator. Monte Carlo Simulations are used to study the behavior of the optimum values of the shrinkage parameter in some distributions arising in financial data.

**Keywords** Robust Optimization, Portfolio Prediction, Shrinkage

## 1. Introduction

In the analysis of financial data we often have to implement regression analysis from historical data, the aim being to predict future values of the variables. In this paper we will work mainly with regression techniques applied to portfolio prediction. Classical applications are Markowitz portfolio optimization, Markowitz (1952), and the Capital Asset Pricing Model (CAPM), developed by many leading economists in the sixties.

CAPM is a very used method for estimating the expected return of a portfolio and evaluation of risks. It is a one-factor model:

$$r_t - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + u_t,$$

with  $t = 1, \dots, T$ , where  $r_t$  is the rate of return,  $r_{ft}$  is the risk-free rate,  $r_{mt}$  is the rate of return of the market and  $u_t$  is a random error. Typically the model is fitted by ordinary least squares (OLS).

We can put CAPM in the context of the standard linear model:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t, \quad (1)$$

with  $t = 1, \dots, T$ ,  $\boldsymbol{\beta} \in \mathbb{R}^N$  and  $u_t$  following a density  $g(\cdot)$ . It is useful to write the model in matrix notation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (2)$$

where  $\mathbf{y}$  and  $\mathbf{u}$  are  $T \times 1$  vectors and  $\mathbf{X}$  is a  $T \times N$  matrix.

In spite of well-known shortcomings, CAPM continues to be an important and widely used model. From a statistical point of view, it is known that standard OLS estimation of  $\boldsymbol{\beta}$  presents several drawbacks. In particular many authors have pointed out its high sensitivity in the presence of outliers and its loss of efficiency in the presence of small deviations from the normality assumption, see, for instance, the books by Huber (1981), Hampel et al. (1986) and Huber and Ronchetti (2009).

Robust statistics was developed to cope with the problem arising from the approximate nature of standard parametric models. Indeed robust statistics deals with deviations from the stochastic assumptions on the model and develops statistical procedures which are still reliable and reasonably efficient in a small neighborhood of the model. In particular, several well known robust regression estimators were proposed in the finance literature as alternatives to OLS to estimate  $\boldsymbol{\beta}$ . This issue was already studied by one of the creators of the CAPM, Sharpe (1971). He suggested to use least absolute deviations ( $L_1$ -estimator) instead of OLS ( $L_2$ -estimator). Chan and Lakonishok (1992) used regression quantiles, linear combinations of regression quantiles, and trimmed regression quantiles. Martin and Simin (2003) proposed to estimate  $\boldsymbol{\beta}$  using redescending M-estimators.

These robust estimators produce values of  $\boldsymbol{\beta}$  which are more reliable than those obtained by OLS in that they reflect the majority of the historical data and they are not influenced by outlying returns. In fact, robust estimators downweight abnormal observations by means of weights which are computed from the data. Following the discussion in Genton and Ronchetti (2008), robustness is important if the main goal of the analysis is to reflect the structure of the underlying process as revealed by the bulk of the data, but a familiar criticism of this approach in finance is that

\* Corresponding author:

luis.yapu@gmail.com (Luis P. Yapu Quispe)

Published online at <http://journal.sapub.org/economics>

Copyright © 2014 Scientific & Academic Publishing. All Rights Reserved

'abnormal returns are the important observations', and it has some foundation from the point of view of prediction. Indeed if abnormal returns are not errors but legitimate outlying observations, they will likely appear again in the future and downweighting them by using robust estimators will potentially result in a bias in the prediction of  $\beta$ . On the other hand, it is true that OLS will produce in this case unbiased estimators of  $\beta$  but this is achieved by paying a potentially important price of a large variability in the prediction. Therefore, we are in a typical situation of a trade-off between bias and variance and we can improve upon a simple use of either OLS or a robust estimator. This motivates the use of some form of shrinkage from the robust estimator toward OLS to achieve the minimization of the mean squared error.

That discussion on CAPM and in least-squares regression model can also be extrapolated to other models in finance based on analog statistical principles. The topic which interests us here is portfolio optimization, mainly from the point of view of prediction. The goal of portfolio optimization is to find weights  $\omega$ , which represent the percentage of capital to be invested in each asset, and to obtain an expected return with a minimum risk. Brodie et al. (2007) presented a way to express the optimization problem as a multiple regression with constraints. It is therefore possible to perform this regression using robust methods, e.g. M-estimators, least trimmed squares (LTS) or others.

Consider a portfolio with  $N$  assets and  $T$  historical returns  $\mathbf{r}_t$  forming the rows of a matrix  $\mathbf{R}$ . For an expected return  $r$  we can solve the following optimization problem:  $\hat{\omega} = \argmin \frac{1}{T} \rho(r\mathbf{1}_T - \mathbf{R}\omega)$  with constraints  $\omega' \hat{\mu} = r$  and  $\omega' \mathbf{1}_N = 1$  where  $\rho$  is a penalizing function such as squaring for the OLS estimator or the Huber's function for the robust M-estimator. We use robust estimations  $(\hat{\mu}_R, \hat{\Sigma}_R)$  and we solve the optimization problem to obtain a robust estimation for the portfolio weights  $\hat{\omega}_R^*$ . We then use a shrinkage estimator, see Eq. (24), to 'shrink' towards the OLS estimator and find an optimal value of the shrinkage parameter  $c$  for the measures of predictive power considered in Section 4 of Genton and Ronchetti (2008).

We use Monte-Carlo simulations to study the behavior of the optimum values of  $c$  for outlying returns  $\mathbf{r}_t$  generated by contamination or long-tailed skew-symmetric laws. The simulations give us empirical heuristics for actual applications in robust asset allocation. We consider specially the flexibility of skew-symmetric distributions and study these type of distributions which allow to model return distributions with significant skewness and high kurtosis as is usually the case of hedge funds (see for instance Popova et al. (2003)).

From a practical point of view, we implement the methods in the statistical software R. Some tools are already implemented (e.g. MCD estimator) but we have to program some other routines (constrained robust regression, multivariate shrinkage). Depending on the amount of data to be analyzed, execution can be expensive in time, consequently we have to take care about efficiency of the routines mainly if we want to apply Monte Carlo simulations

using resampling methods.

This paper could be considered as an application in portfolio optimization of the shrinkage estimators studied in Genton and Ronchetti (2008). They only treat the case of estimating beta in CAPM. That estimator have been generalized to multidimensional variables need in portfolio statistical analysis. Gramacy et al. (2008) use specific shrinkage estimators (LASSO and ridge regression) in finance to estimate covariances between many assets with histories of highly variable length (missing data) but they do not deal with robustness. That work have been developed and extended in Gramacy and Pantaleo (2010), where they consider a Bayesian hierarchical formulation, considering heavy-tailed errors and accounting for estimation risk.

The introduction to robust techniques to portfolio optimization is relatively recent compared with the Markowitz foundational paper. Nevertheless the subject have become very active in the last decade. We can mention the works of Vaz-de Melo and Camara (2003), Perret-Gentil and Victoria-Feser (2004), and Welsch and Zhou (2007). All three papers compute the robust portfolio policies in two steps. First, they compute a robust estimate of the covariance matrix of asset returns. Second, they solve the minimum-variance problem where the covariance matrix is replaced by its robust estimate. Recently, Demiguel and Nogales (2009) proposed solving a single nonlinear program, where portfolio optimization and robust estimation are performed in one step. They performed a theoretical study for M-estimators and S-estimators, in addition to a simulation using a mixture of a normal and a *deviation* distribution. A very recent work of Demiguel et al. (2013) have also implemented a shrinkage strategy both using shrinkage estimators of the moments of asset returns (shrinkage moments), and using shrinkage portfolios obtained by shrinking the portfolio weights directly. We have to remark that in that paper, they use shrinkage by means of a convex combination from the sample estimator (low bias), towards the target estimator (low variance). They use two calibration criteria: the expected quadratic loss minimization criterion, and the Sharpe ratio maximization criterion. We distinguish our work by the fact that we explicitly use a M-estimator as the target of the shrinkage, which enables us to use a more specific shrinkage estimator (from Genton and Ronchetti (2008)) with the calibration parameter is related to Huber's function. In fact varying that parameter allow the shrinkage model to interpolate within the family of robust estimators, the OLS estimator being a limit case for a big value of the parameter  $c$  (in fact the OLS is the limit for  $c \rightarrow \infty$ , see Section 4). This is an advantage of our shrinkage strategy compared to convex combination of estimators. Other characteristic of our work is that we use many measures of predictive errors aside the expected quadratic loss. This is specially because in our simulated study we are interested in long-tailed and asymmetric distributions. Other reference which uses skew-symmetric laws as  $St$  in portfolio optimization is Hu and Kercheval (2010) but they do not involve with shrinkage strategies.

The paper is organized as follows, in Section 2 we explain

some basic issues concerning the appearance of asymmetric and long-tailed errors and robust regression, then in Section 3 we derive the robust constrained regression model associated to portfolio allocation, in particular we have a closed formula for the shift of the estimator of the parameter vector of the linear model due to linear constraints, see. Eq. 14. In section 4, we present a shrinkage robust estimator and combine it with the constrained robust regression in a procedure for the application in portfolio optimization. Section 5 illustrates the results of the combined procedure using Monte Carlo simulation, first in an ideal standard linear model and then to simulated distributions from contaminated normal and asymmetric-long-tailed laws. Finally Section 7 presents some conclusions of our study.

## 2. Non-normal Errors and Robust Regression

Robust statistics is an extension of classical statistics in that it takes into account the possibility of contaminated data or more generally of model misspecification. This theory was firstly developed by Huber (1964) and Hampel (1968).

There are many ways to model errors with outlier. For instance we can consider a mixture of a normal distribution  $N$  with a large-variance distribution  $W$ . Let  $\varepsilon \in (0,1)$  be a number representing the proportion of contamination and define the neighborhood of the parametric distribution  $\mathbf{F}_\theta$  to be the set:

$$\{\mathbf{G}_\varepsilon | \mathbf{G}_\varepsilon = (1 - \varepsilon)\mathbf{F}_\theta + \varepsilon\mathbf{W}\}. \quad (3)$$

$\mathbf{G}_\varepsilon$  can be considered as a mixed distribution between  $\mathbf{F}_\theta$  and the contamination distribution  $\mathbf{W}$ . An estimator is said robust if it remains stable in a neighborhood of  $\mathbf{F}_\theta$ . Often in theoretical studies  $\mathbf{F}_\theta$  is a multivariate normal distribution in dimension  $d$ :  $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

In standard linear regression theory, least-squares estimator for the parameter  $\boldsymbol{\beta}$  is known to be non-robust. In section 3 we will use M-estimators to find robust estimates of parameters in portfolio allocation. In the context of the linear model (1), the general M-estimators minimize the *objective function*:

$$\sum_{t=1}^T \rho\left(\frac{y_t - \mathbf{x}_t' \boldsymbol{\beta}}{s}\right), \quad (4)$$

with respect to  $\boldsymbol{\beta}$  and the *loss function*  $\rho$  gives the contribution of each residual to the objective function.  $s$  is a scale parameter. Generalizing least-squares minimization, a reasonable  $\rho$  should have the following properties:

- $\rho(u) \geq 0$ ,
- $\rho(0) = 0$ ,
- $\rho(u) = \rho(-u)$ ,
- $\rho(u_i) \geq \rho(u_j)$  for  $|u_i| > |u_j|$ .

For example, for least-squares estimation we have  $\rho(u) = u^2$ .

Let  $\psi = \rho'$  denote the derivative of  $\rho$ . In this paper we will work with the *Huber* objective function and its

derivative  $\psi_c(\cdot)$  which is called *Huber function* and is defined by  $\psi_c(u) = \min(c, \max(-c, u))$ . The tuning constant  $c$  controls the level of robustness. If  $c \rightarrow \infty$  then  $\psi_\infty(u) = u$ , which corresponds to least-squares estimation.

Differentiating the objective function (4) with respect to  $\boldsymbol{\beta}$  gives the following estimating equations:

$$\sum_{t=1}^T \mathbf{x}_t \psi(y_t - \mathbf{x}_t' \boldsymbol{\beta}) = \mathbf{0}. \quad (5)$$

Define the *weight function*  $w(u) = \frac{\psi(u)}{u}$ , and denote  $w_t = w(u_t)$ . Then the estimating equation (5) can be written as:

$$\sum_{t=1}^T \mathbf{x}_t (y_t - \mathbf{x}_t' \boldsymbol{\beta}) w_t = \mathbf{0}.$$

Note that solving these estimating equations can be seen as a weighted least-squares minimization problem with objective function:

$$\sum_{t=1}^T w_t (y_t - \mathbf{x}_t' \boldsymbol{\beta})^2.$$

The weights  $w_t$ , however, depend upon the residuals, the residuals depend upon the estimated coefficients, and the estimated coefficients depend upon the weights. An iterative solution is therefore required. More details about M-estimators can be found in references, for instance Hampel et al. (1986).

At the end of the procedure we obtain the weights  $w_t$  which can be collected in a  $T \times T$  diagonal matrix  $\mathbf{W}$  and then we can calculate the M-estimator  $\hat{\boldsymbol{\beta}}_M$  in matrix notation:

$$\hat{\boldsymbol{\beta}}_M = (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W} \mathbf{y}.$$

### 2.1. Resistant Regression (LTS)

There are other robust techniques of estimation in order to reduce the influence of outliers on the fit of a model. Following the schema of Genton and Ronchetti (2008), we will use the *least trimmed squares* (LTS) regression.

LTS was proposed by Rousseeuw (1985) as another robust alternative to OLS. Let us consider a linear regression model (1). The LTS estimator  $\hat{\boldsymbol{\beta}}_{LTS}$  is defined as:

$$\hat{\boldsymbol{\beta}}_{LTS} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{t=1}^h u_{[t]}^2(\boldsymbol{\beta}),$$

where  $u_{[t]}^2(\boldsymbol{\beta})$  represents the  $t$ -th order statistics of squared residuals  $u_t^2(\boldsymbol{\beta})$  with  $u_t(\boldsymbol{\beta}) = y_t - \mathbf{x}_t' \boldsymbol{\beta}$ .

The *trimming constant*  $h$  has to satisfy  $\frac{T}{2} < h < T$ . This constant determines the robustness level of the LTS estimator, since the definition implies that  $T - h$  observations with the largest residuals do not have a direct influence on the estimator. The LTS robustness is the lowest for  $h = T$ , which corresponds to the least-squares estimator.

### 2.2. Asymmetric and Long-tailed Errors

Often returns in portfolio optimization do not follow a normal distribution and the empirical distribution presents asymmetry and thick tails. In those cases we can propose

errors following more flexible laws such as skew-symmetric distributions.

Skew-symmetric distributions were explicitly introduced in the literature by Azzalini (1985) with the aim to model departure from normality. Afterwards many generalizations have been introduced and it is nowadays a well studied topic because of its flexibility and theoretical tractability. We can mention the multivariate skew normal distribution studied by Azzalini and Dalla Valle (1996) and the multivariate skew  $t$  distribution studied in Azzalini and Capitanio (2003). Here we will only define notations.

### 2.2.1. The Multivariate Skew-normal Distribution

Given a full-rank  $d \times d$  covariance matrix  $\Omega$  define  $\omega = \text{diag}(\Omega_{11}, \dots, \Omega_{dd})^{1/2}$ , let  $\bar{\Omega} = \omega^{-1}\Omega\omega^{-1}$  be the corresponding correlation matrix and define vectors  $\xi$ ,  $\alpha \in \mathbb{R}^d$ . A  $d$ -dimensional random variable  $Z$  is said to follow a skew-normal distribution if its density function at  $\mathbf{z} \in \mathbb{R}^d$  is given by:

$$2\phi_d(\mathbf{z} - \xi; \Omega)\Phi(\alpha'\omega^{-1}(\mathbf{z} - \xi)).$$

where  $\phi_d(\mathbf{z}; \Omega)$  is the  $N_d(\mathbf{0}, \Omega)$   $d$ -dimensional normal density at  $\mathbf{z}$  with covariance matrix  $\Omega$  and  $\Phi(\cdot)$  is the  $N(0,1)$  distribution function.

We will then write  $Z:SN_d(\xi, \Omega, \alpha)$  and call  $\xi, \Omega, \alpha$  the location, dispersion and the shape or skewness parameters, respectively. If we define a new shape parameter:

$$\delta = \frac{1}{(1 + \alpha'\bar{\Omega}\alpha)}\bar{\Omega}\alpha,$$

then we can write the expressions of mean vector and covariance matrix:

$$\begin{aligned}\mu_Z &:= \mathbf{E}[Z] = \xi + \sqrt{\frac{2}{\pi}}\delta \\ \text{Var}[Z] &= \Omega - \mu_Z\mu_Z'.\end{aligned}$$

### 2.2.2. The Multivariate Skew- $t$ Distribution

In dimension 1, standard  $t$  distribution have thick tails and then it allows to model large outliers. In the multivariate case, consider random variables  $Z:SN_d(\mathbf{0}, \Omega, \alpha)$ ,  $V:\chi_v^2/v$ , independent of  $Z$ , and the constant vector  $\xi \in \mathbb{R}^d$ . We define the skew- $t$  distribution as the one corresponding to the transformation:

$$Y = \xi + V^{-1/2}Z. \quad (6)$$

We shall write  $Y:St_d(\xi, \Omega, \alpha, \nu)$ . The parameter  $\nu$  corresponds to the degrees of freedom. A small value of  $\nu$  will allow the presence of large outliers and when  $\nu \rightarrow \infty$  then  $Y$  converges to a skew-normal variable.

The density function and other formulas and properties can be found in Azzalini and Capitanio (2003). Figures 1 and 2 shows two scatterplots of a 4-dimensional skew-normal variable and skew- $t$  variable. In section 6 we will perform simulations using these distributions in the context of portfolio optimization.

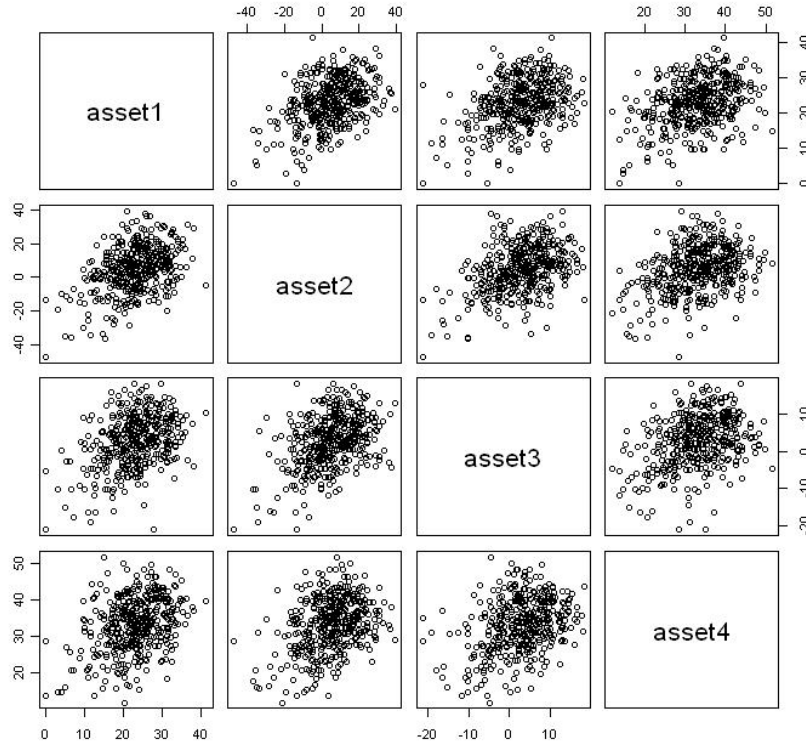
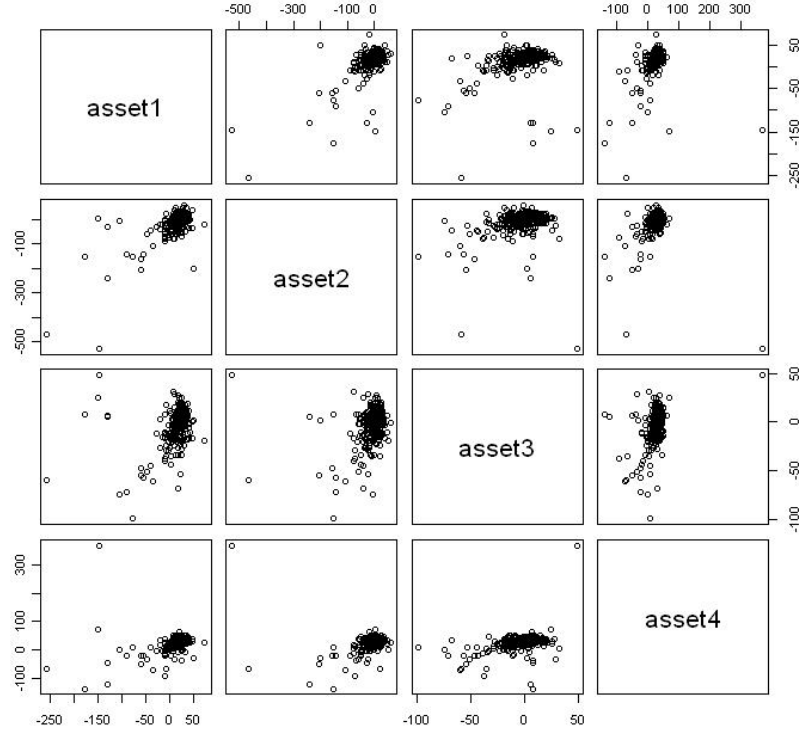


Figure 1. Scatterplot of a  $SN_4$  distribution

Figure 2. Scatterplot of a  $St_4$  distribution

### 3. Portfolio Asset Allocation

We consider  $N$  assets and denote their returns at time  $t$  by  $r_{i,t}$ ,  $i = 1, \dots, N$ ,  $t = 1, \dots, T$  and denote by  $\mathbf{r}_t = (r_{1,t}, \dots, r_{N,t})'$  the  $N \times 1$  vector of returns at time  $t$ . We assume that  $\mathbf{r}_t$  follows a multivariate distribution with  $E[\mathbf{r}_t] = \boldsymbol{\mu}$  and  $\text{Var}[\mathbf{r}_t] = \boldsymbol{\Sigma}$ .

A portfolio is defined to be a list of weights  $\omega_i$  for the assets  $i = 1, \dots, N$  that represent the amount of capital to be invested in each asset. We assume that  $\sum \omega_i = 1$  which means that capital is fully invested and denote  $\boldsymbol{\omega}$  the  $N \times 1$  vector of weights.

For a given portfolio  $\boldsymbol{\omega}$ , the expected return and variance are respectively given by:

$$E[\boldsymbol{\omega}'\mathbf{r}_t] = \boldsymbol{\omega}'\boldsymbol{\mu}, \quad (7)$$

$$\text{Var}[\boldsymbol{\omega}'\mathbf{r}_t] = \boldsymbol{\omega}'\text{Var}[\mathbf{r}_t]\boldsymbol{\omega} = \boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega}. \quad (8)$$

Following the standard Markowitz portfolio optimization procedure, we seek a portfolio  $\tilde{\boldsymbol{\omega}}$  which has minimal variance for a given expected return  $r$ . We can express the problem as:

$$\tilde{\boldsymbol{\omega}} = \underset{\boldsymbol{\omega}}{\text{argmin}} \quad \boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega},$$

with constraints

$$\boldsymbol{\omega}'\boldsymbol{\mu} = r, \quad (9)$$

$$\boldsymbol{\omega}'\mathbf{1}_N = 1, \quad (10)$$

where  $\mathbf{1}_N$  is the  $N \times 1$  vector in which every entry is equal to 1.

We can find in Brodie et al.(2007) a way to model the optimization problem using a multivariate constrained regression. Here we develop details of the derivation.

We have  $\boldsymbol{\Sigma} = E[\mathbf{r}_t\mathbf{r}_t'] - \boldsymbol{\mu}\boldsymbol{\mu}'$  and we can write:

$$\begin{aligned} \boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} &= \boldsymbol{\omega}'(E[\mathbf{r}_t\mathbf{r}_t'] - \boldsymbol{\mu}\boldsymbol{\mu}')\boldsymbol{\omega}, \\ &= E[\boldsymbol{\omega}'\mathbf{r}_t\mathbf{r}_t'\boldsymbol{\omega}] - \boldsymbol{\omega}'\boldsymbol{\mu}\boldsymbol{\mu}'\boldsymbol{\omega}. \end{aligned}$$

In fact  $\boldsymbol{\omega}'\mathbf{r}_t$  and  $\boldsymbol{\omega}'\boldsymbol{\mu}$  are scalars and using (7) we can write the last expression as:

$$\begin{aligned} \boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} &= E[(\boldsymbol{\omega}'\mathbf{r}_t)^2] - (\boldsymbol{\omega}'\boldsymbol{\mu})^2 \\ &= E[(\boldsymbol{\omega}'\mathbf{r}_t)^2] - (E[\boldsymbol{\omega}'\mathbf{r}_t])^2 \\ &= E[(\boldsymbol{\omega}'\mathbf{r}_t - E[\boldsymbol{\omega}'\mathbf{r}_t])^2]. \end{aligned}$$

Finally using (7) and the constraint (9) we have:

$$\boldsymbol{\omega}'\boldsymbol{\Sigma}\boldsymbol{\omega} = E[(\boldsymbol{\omega}'\mathbf{r}_t - r)^2]. \quad (11)$$

For the empirical implementation, we replace expectations by sample average. We set  $\hat{\boldsymbol{\mu}} = \frac{1}{T} \sum \mathbf{r}_t$  and define  $\mathbf{R}$  to be the  $T \times N$  matrix of which the  $t$ -th row is  $\mathbf{r}_t'$ .

The empirical version of expression (11) is:

$$\frac{1}{T} \sum_{t=1}^T (\boldsymbol{\omega}'\mathbf{r}_t - r)^2 = \frac{1}{T} \|\mathbf{R}\boldsymbol{\omega} - r\mathbf{1}_T\|_2^2$$

where, for a vector  $\mathbf{a}$  in  $\mathbb{R}^T$ , we use the 2-norm notation:  $\|\mathbf{a}\|_2^2 = \sum \mathbf{a}_t^2$ .

In summary, we seek to solve the new following optimization problem:

$$\hat{\boldsymbol{\omega}} = \underset{\boldsymbol{\omega}}{\text{argmin}} \quad \frac{1}{T} \|\mathbf{R}\boldsymbol{\omega} - r\mathbf{1}_T\|_2^2$$

with constraints

$$\boldsymbol{\omega}'\hat{\boldsymbol{\mu}} = r, \quad (12)$$

$$\boldsymbol{\omega}'\mathbf{1}_N = 1. \quad (13)$$

We can view this as a multiple constrained regression for the model:

$$y_t = \mathbf{r}_t \omega + \varepsilon_t,$$

$t = 1, \dots, T$ ,  $y_t = r$  for each  $t$ , and with the same constraints (12) and (13).

In the optic of robustness we replace the 2-norm by a loss function  $\rho$  which grows slower, obtaining then the problem:

$$\hat{\omega} = \underset{\omega}{\operatorname{argmin}} \frac{1}{T} \sum_{t=1}^T \rho\left(\frac{(\mathbf{R}\omega)_t - (r\mathbf{1})_t}{s}\right)$$

with constraints (12) and (13). As before  $s$  is a scale parameter which should be estimated robustly.

We have seen in the last section that the non-constrained M-estimator  $\hat{\omega}_M$  is:

$$\hat{\omega}_M = (\mathbf{R}'\mathbf{W}\mathbf{R})^{-1}\mathbf{R}'\mathbf{W}\mathbf{y}.$$

The constrained minimization is solved using Lagrange multipliers. We present the derivation in the next subsection 3.1. In the presence of  $l \leq N$  independent linear constraints  $\mathbf{C}\omega = \mathbf{v}$  we obtain the constrained M-estimator  $\hat{\omega}_{CM}$ :

$$\hat{\omega}_{CM} = \hat{\omega}_M + (\mathbf{R}'\mathbf{W}\mathbf{R})^{-1}\mathbf{C}'(\mathbf{C}(\mathbf{R}'\mathbf{W}\mathbf{R})^{-1}\mathbf{C}')^{-1}(\mathbf{v} - \mathbf{C}\hat{\omega}_M). \quad (14)$$

We observe that the constrained M-estimator  $\hat{\omega}_{CM}$  differs from the unconstrained  $\hat{\omega}_M$  by a function of the quantity  $(\mathbf{v} - \mathbf{C}\hat{\omega}_M)$ .

For our problem, the constraint matrices are:

$$\mathbf{C} = \begin{pmatrix} \hat{\mu}' \\ \mathbf{1}_N' \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} r \\ 1 \end{pmatrix}. \quad (15)$$

### 3.1. Constrained Robust Regression

Using the notation of weighted least-squares regression and in the presence of the linear constraints  $\mathbf{C}\beta = \mathbf{v}$ , we can write the Lagrangian:

$$\mathcal{L}(\beta, \lambda) = \sum_{i=1}^n w_i (y_i - \mathbf{x}_i' \beta)^2 + \lambda' (\mathbf{C}\beta - \mathbf{v}),$$

where  $\lambda$  is a  $l \times 1$  vector of lagrange multipliers,  $\mathbf{C}$  is a  $l \times n$  matrix and  $\mathbf{v}$  is a  $l \times 1$  vector.

In matrix notation:

$$\mathcal{L}(\beta, \lambda) = (\mathbf{y} - \mathbf{X}\beta)' \mathbf{W} (\mathbf{y} - \mathbf{X}\beta) + \lambda' (\mathbf{C}\beta - \mathbf{v}).$$

Differentiation with respect to  $\beta$  and  $\lambda$  gives the equations:

$$\frac{\partial \mathcal{L}}{\partial \beta} = -2\mathbf{X}'\mathbf{W}(\mathbf{y} - \mathbf{X}\beta) + \mathbf{C}'\lambda = \mathbf{0},$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \mathbf{C}\beta - \mathbf{v} = \mathbf{0}.$$

From the first equation we find:

$$\beta = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}[\mathbf{X}'\mathbf{W}\mathbf{y} - \frac{1}{2}\mathbf{C}'\lambda], \quad (16)$$

and replacing this into the second equation we get:

$$\mathbf{v} = \mathbf{C}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} - \frac{1}{2}\mathbf{C}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{C}'\lambda. \quad (17)$$

From this, we obtain the value of  $\lambda$ :

$$\lambda = 2(\mathbf{C}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} - \mathbf{v}), \quad (18)$$

and replacing this into expression (16) we find the expression of the constrained estimator:

$$\hat{\beta} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}[\mathbf{X}'\mathbf{W}\mathbf{y} - \mathbf{C}'(\mathbf{C}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\hat{\beta}_M - \mathbf{v})]$$

Recall the formula of the non-constrained weighted estimator:

$$\hat{\beta}_M = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}.$$

Using this the final expression of the constrained M-estimator can be written:

$$\hat{\beta} = \hat{\beta}_M - (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{C}'(\mathbf{C}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\hat{\beta}_M - \mathbf{v}). \quad (19)$$

## 4. Shrinkage Robust Estimator

Genton and Ronchetti (2008) have defined a robust estimator with shrinkage for the linear model:

$$\tilde{\beta}_c = \hat{\beta}_R + (\sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t')^{-1} \sum_{t=1}^T \hat{\sigma} \psi_c\left(\frac{y_t - \mathbf{x}_t' \hat{\beta}_R}{\hat{\sigma}}\right) \mathbf{x}_t', \quad (20)$$

where  $\hat{\beta}_R$  is a robust estimator of  $\beta$ ,  $\hat{\sigma}$  is a robust estimator of scale such as the median absolute deviation (MAD), and  $\psi_c(\cdot)$  is the Huber's function. As we have seen in Section 2, there are many proposals for the robust estimator  $\hat{\beta}_R$ , see for instance Hampel et al. (1986).

The tuning constant  $c$  allows us to control the level of shrinkage. If  $c = 0$  we find the robust estimator  $\hat{\beta}_R$  and if  $c \rightarrow \infty$  we find the least-squares estimator (OLS).

Indeed, in  $c = 0$  then  $\psi_0(u) = 0$  for all values of  $u$  then the rightmost expression in (20) is zero and we have  $\tilde{\beta}_c = \hat{\beta}_R$ , the robust estimator. On the other side, if  $c \rightarrow \infty$  then  $\psi_\infty(u) = u$  and equation (20) simplifies to:

$$\begin{aligned} \tilde{\beta}_\infty &= \hat{\beta}_R + \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \sum_{i=1}^n (y_i - \mathbf{x}_i' \hat{\beta}_R) \mathbf{x}_i \\ &= \hat{\beta}_R + \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \left(\sum_{i=1}^n \mathbf{x}_i y_i - \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' \hat{\beta}_R\right) \end{aligned}$$

Distributing the expression  $(\cdot)^{-1}$  and recognizing the expression of ordinary least-squares estimator  $\hat{\beta}_{OLS}$  we obtain:

$$\tilde{\beta}_\infty = \hat{\beta}_R + \hat{\beta}_{OLS} - \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i'\right) \hat{\beta}_R$$

The rightmost expression simplifies and we obtain the limit expression:

$$\tilde{\beta}_\infty = \hat{\beta}_R + \hat{\beta}_{OLS} - \hat{\beta}_R = \hat{\beta}_{OLS} \quad (21)$$

As was discussed in the Section 1, if the goal of modeling is to find a model which reflects the bulk of the data then the robust estimation is the most adequate method because outliers are under-weighted. Nevertheless from a predictive point of view, outliers in finance could be considered as interesting data. So it is important to study if, by varying the constant  $c$ , we can improve the predictive power with respect to least-squares or robust estimators.

There are many choices for measuring the quality of the prediction. For normal-distributed errors, the most used

criterion is the mean squared error (MSE). However for asymmetric or long-tailed distributions there is not a standard choice.

Following Genton and Ronchetti (2008) we consider a family of measures:

$$Q_c(p) = \left( \frac{1}{m} \sum_{i=1}^m |r_i - \hat{r}_i|^p \right)^{1/p}, \quad (22)$$

where  $c$  is the shrinkage constant used to estimate the expected returns  $\hat{r}_i$ . The case  $p = 2$  give us the MSE measure. We will be interested in  $p = 1/2$ ,  $p = 1$  and  $p = 2$ .

An important tool to compare the shrinkage estimators is the relative gain:

$$RG_{c,d}(p) = \frac{Q_d(p) - Q_c(p)}{Q_d(p)}. \quad (23)$$

It will be useful to analyze if shrinkage with a level  $c$  offers more predictive power than the robust estimator ( $d = 0$ ) or the OLS estimator ( $d \rightarrow \infty$ ).

#### 4.1. Application to Portfolio Optimization

In classical portfolio optimization the first stage in general is to estimate the mean  $\mu = E[r]$  and the covariances  $\Sigma = \text{Var}[r]$ . We can use the historical data and use robust methods to obtain (robust) estimators  $\hat{\mu}_R$  and  $\hat{\Sigma}_R$ , as explained by Welsch and Zhou (2007). Some methods such as minimum covariance determinant MCD or FAST-MCD are already implemented in statistical software such as R.

We have seen in section 3 how to write the classical portfolio optimization problem as a multivariate constrained regression problem and then we considered the robust setting of the problem. In this formulation, we need only a robust estimator of  $\mu$  which enters into de constrains matrix  $C$  (see formula (15)). We will use the MCD estimator which is already implemented in R.

In subsection 3.1 we have found the formula of the robust constrained estimator  $\hat{\omega}_{CM}$  of the portfolio weights. We can now try to apply the shrinkage to  $\hat{\omega}_{CM}$  but then the constraints are no more satisfied. We need to use equation (19) one more time but for this we have to recalculate the matrix  $W$ .

To be precise we present next a detailed description of the procedure step by step.

1. Calculate robust estimator  $\hat{\omega}_M$  and the regression matrix of weights  $W$ . This can be do with standard routines of statistical software such as R.

2. Use  $W$  to obtain the constrained estimator  $\hat{\omega}_{CM}$  given by formula

$$\hat{\omega}_{CM} = \omega_M + (R'WR)^{-1}C'(C(R'WR)^{-1}C')^{-1}(v - C\omega_M).$$

3. Let  $c$  be de tuning constant of shrinkage then calculate the shrinkage estimator:

$$\tilde{\omega}_c = \hat{\omega}_{CM} + (\sum_{t=1}^T r_t r_t')^{-1} \sum_{t=1}^T \hat{\sigma} \psi_c \left( \frac{y_t - r_t' \hat{\omega}_{CM}}{\hat{\sigma}} \right) r_t. \quad (24)$$

4. Using  $\tilde{\omega}_c$ , calculate the new weight matrix  $W_c$  associated with this estimator. We need to calculate standardized residuals  $r_t = (y_t - R\tilde{\omega}_c)/\hat{\sigma}_c$ , where  $\hat{\sigma}_c$  is

a robust estimate of scale of the residuals. The matrix  $W_c$  is diagonal with  $t$ -th component:

$$w_t = \frac{\psi_c(r_t)}{r_t}.$$

5. Use  $W_c$  to obtain the shrinkage constrained estimator  $\hat{\omega}_{Cc}$  given by formula

$$\hat{\omega}_{Cc} = \tilde{\omega}_c + (R'W_cR)^{-1}C'(C(R'W_cR)^{-1}C')^{-1}(v - C\tilde{\omega}_c).$$

We will use Monte-Carlo simulations to study the behavior of the optimum values of  $c$  with respect to the different prediction error measures (values of  $p$  in (22)) for normal and outlying returns  $r_t$  generated by contamination (mixture) and by skew-symmetric laws.

A suitable shrinkage could vary in time depending on new available historical data. Anyway at any moment the Monte Carlo simulation can be performed to assess an optimum value of this shrinkage parameter for future estimations of the portfolio weights.

We remark that the computation complexity of this procedure depends of the actual implementation of the robust estimation and the matrix operations involved. Supposing that Iteratively Reweighed Least Squares is used in step 1. and only a few iterations are sufficient for convergence, the number of operations involved in all the steps is of the order  $O(N^3 + N^2T + T^2N)$ . Taking  $N$  (the number of the assets) fixed, the computational complexity becomes  $O(T^2)$ . In consequence, even if it is possible to assess efficiency of Monte Carlo simulation by changing  $T$ , the computational work increases too, as well as the collateral computational errors. A better study of this remains to be done.

## 5. Monte Carlo Simulations

In this section we perform simulations based on the paper of Genton and Ronchetti (2008) but in the multivariate case. We will apply the general strategy to portfolio optimization in section 6. We consider the linear model:

$$y_t = x_t' \beta + u_t,$$

with  $t = 1, \dots, T$ ,  $\beta$  a  $N \times 1$  vector,  $u_t$  i.i.d.  $N(0,1)$  errors and  $x_t$  a multivariate normal. As the first example we consider two covariates:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + u_t,$$

and we take the values  $\beta_0 = 1$ ,  $\beta_1 = 0.5$ ,  $\beta_2 = 1.5$ , with  $x_{t1}$ ,  $x_{t2}$  independent  $N(0,2^2)$  variables.

We take  $T = 400$  and include 10% of outliers (contamination) for  $x_{t1}$ ,  $x_{t2}$  and  $y_t$  from  $N(0,5^2)$ .

We use M and LTS estimators,  $\hat{\beta}_M$  and  $\hat{\beta}_{LTS}$ , to obtain robust estimates of  $\beta$ . The estimated values are:

Variable	M-estimator	LTS	OLS
$\hat{\beta}_0$	1.016	1.073	0.886
$\hat{\beta}_1$	0.466	0.611	0.234
$\hat{\beta}_2$	1.321	1.459	0.790

We observe that the OLS estimates are biased and will not be useful for future predictions. We can now use the shrinkage robust estimator (called  $SR_c$  in the sequel),  $\hat{\beta}_c$ , of  $\beta$  with shrinkage constant  $c$ . In order to analyze the effect of outliers, we simulate 1000 training data sets of size  $n = 100$  each and containing outliers as indicated. For each sample we estimate  $\beta$  by LTS, OLS and  $SR_c$  with  $c = 1, \dots, 10$ . Figures 4 show boxplots of these estimates over the 1000 simulated training data.

The LTS estimators of  $\beta_1$  and  $\beta_2$  have smaller bias than OLS estimators. We observe that for some values of  $c$ , the variance of  $SR_c$  is reduced at the cost of a small increase in bias.

Next we investigate the effect of outliers and shrinkage on the prediction of future observations. More precisely, for each of the 1000 estimates  $\hat{\beta}$ , we compute the predicted values  $\hat{y}_t = \mathbf{x}_t' \hat{\beta}$ ,  $t = 1, \dots, 400$ .

We consider the measures of quality of prediction with the shrinkage robust estimator  $SR_c$ ,  $Q_c(p)$ , defined in (22) with three choices for  $p$ . For  $p = 2$  we have the root mean square error (RMSE), for  $p = 1$  the mean absolute error (MAE) and for  $p = 1/2$  the square root absolute error (STAE).

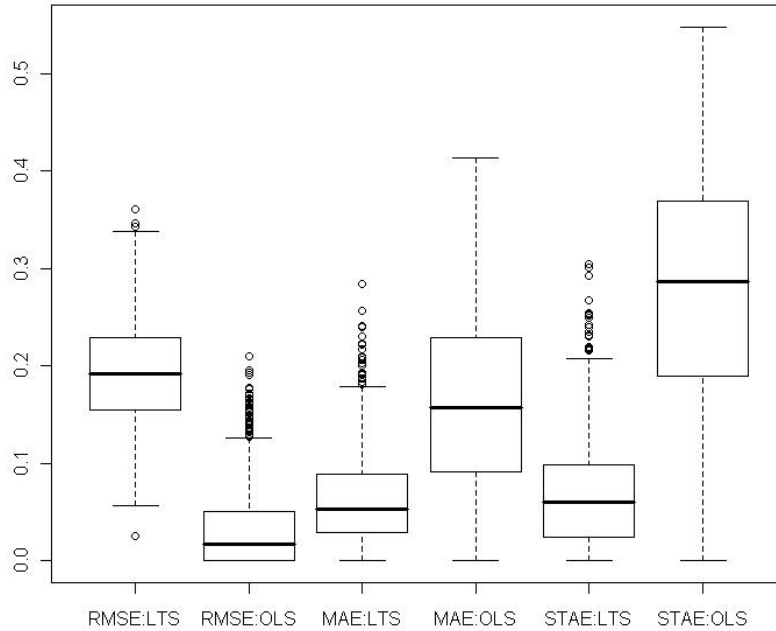
In Table 1 we report the frequencies of selection of a minimum measure of prediction for a range of values of  $c = 0, 1, \dots, 10, \infty$  over the 1000 replicates. As can be seen, the optimal  $c$  which minimizes a certain measure of quality of prediction is not exclusively concentrated at the 'limit' estimators LTS and OLS. The RMSE measure is related with least squares estimation, consequently OLS is selected most of times. We observe that a shrinkage constant of  $c = 3$  is optimal for MAE and  $c = 2$  is optimal for STAE. If we are interested in a more precise value of the constant  $c$ , it is possible to refine the search around the values 2 or 3 of the parameter  $c$  and use a smaller step size.

We have defined the relative gain by  $RG_{c,d}(p)$  in (23). Denote by  $c^*$  the value of  $c$  minimizing  $Q_c(p)$  for a fixed  $p$ . Figure 3 depicts boxplots over the 1000 replicates of  $RG_{c^*,0}(p)$  and  $RG_{c^*,\infty}(p)$  for  $p=1/2, 1$  and  $2$ , that is, the relative gain compared to LTS and OLS estimators using STAE, MAE and RMSE respectively.

We remark that in terms RMSE, the gains of  $SR_c$  compared to OLS are small. In terms of MAE and STAE, the gains can reach 30% – 40%. The gains compared to LTS go rather in the other direction.

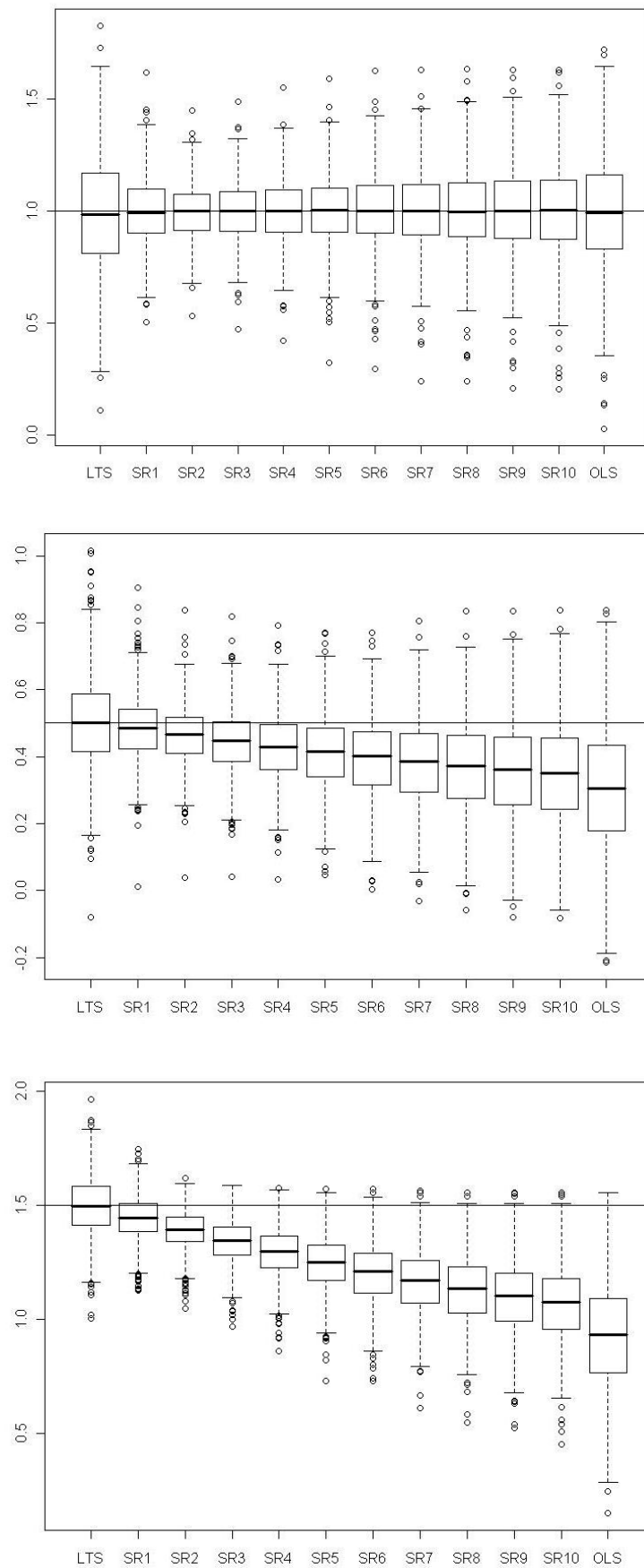
**Table 1.** Normal contamination: frequencies of selection of a minimum measure of prediction

Value of $c$ :	LTS	1	2	3	4	5	6	7	8	9	10	OLS
MAE	28	122	211	224	173	83	53	38	23	14	22	9
RMSE	0	0	0	1	6	22	54	81	129	154	210	343
STAE	67	215	288	183	114	50	30	19	13	6	7	8



**Figure 3.** Normal contamination: relative gains obtained with shrinkage robust estimators compared to LTS and OLS on various measures of prediction (RMSE, MAE, STAE)





**Figure 4.** Normal contamination: boxplots of  $\beta_0$ ,  $\beta_1$  and  $\beta_2$  for several values of the shrinkage constant  $c$

## 6. Application to Portfolio Optimization

Monte Carlo simulations can give us empirical heuristics for actual applications of the shrinkage robust asset allocation. We have already mentioned the flexibility of skew-symmetric distributions and we will especially study these types of distributions which allow to model return distributions which have significant skewness and high kurtosis such as hedge funds (see for instance Popova et al. (2003)). In this paper we will perform only a empirical study.

### 6.1. Normal Data with Normal Contamination

We consider a portfolio of  $N = 4$  assets. In this example, we suppose that returns are generated from a 4-dimensional normal distribution

$$\mathbf{r}_t: N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (25)$$

with parameters:

$$\boldsymbol{\mu} = \begin{pmatrix} 30 \\ 20 \\ 10 \\ 40 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 100 & 140 & 70 & 70 \\ 140 & 400 & 140 & 140 \\ 70 & 140 & 100 & 70 \\ 70 & 140 & 70 & 100 \end{pmatrix}. \quad (26)$$

We will include 10% of contamination from  $N_4(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$  where:

$$\boldsymbol{\mu}_1 = \begin{pmatrix} -20 \\ -20 \\ -20 \\ -20 \end{pmatrix}, \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 2500 & 0 & 0 & 0 \\ 0 & 2500 & 0 & 0 \\ 0 & 0 & 2500 & 0 \\ 0 & 0 & 0 & 2500 \end{pmatrix}. \quad (27)$$

We interpret this as independent returns with large variance. We used a general scaling constant of 10 and this explains the order of magnitude of mean vectors and variance matrices.

Recall that in portfolio optimization, the regression equation is:

$$y_t = \mathbf{r}_t \boldsymbol{\omega} + \varepsilon_t \quad (28)$$

$t = 1, \dots, T$ ,  $y_t = r$  the expected return of the portfolio and with constraints (12) and (13).

We simulate a contaminated test sample of size  $T = 400$  and take  $r = 25$ . In practice we don't know the theoretical mean vector  $\boldsymbol{\mu}$  and we only have the matrix  $\mathbf{R}$  of all returns. As long as we have outliers we need a robust estimate of  $\boldsymbol{\mu}$  denoted  $\hat{\boldsymbol{\mu}}$ . This can be performed using the method called "fast MCD" developed by Rousseeuw and Van Driessen (1999) which is more general and compute a robust covariance matrix estimator too. The robust estimate of  $\boldsymbol{\mu}$  is:

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} 30.09 \\ 19.82 \\ 10.04 \\ 40.02 \end{pmatrix}. \quad (29)$$

In Figures 5-8 we simulate 500 data sets of size 400 each and compare the M-estimates and OLS estimates of contaminated data with the OLS estimates of non-contaminated data. The bias is much smaller for the M-estimates but for some variables it is not null. Anyway in this paper we worked with M-estimators because we could find analytical formulas for constrained regression (Section 3.1). These formulas allowed us to implement shrinkage and to verify the constraints. There is also the possibility to use more resistant estimators but in that case it is necessary to use other methods to project the shrinked weights onto the constrained subspace.

Following the algorithm described in subsection 4.1 we will use M-estimator as initial robust estimate. The estimates are:

Variable	Constrained M	Constrained OLS
$\hat{\omega}_0$	0.4808	0.3776
$\hat{\omega}_1$	-0.3083	-0.2973
$\hat{\omega}_2$	0.5598	0.5853
$\hat{\omega}_3$	0.2677	0.3344

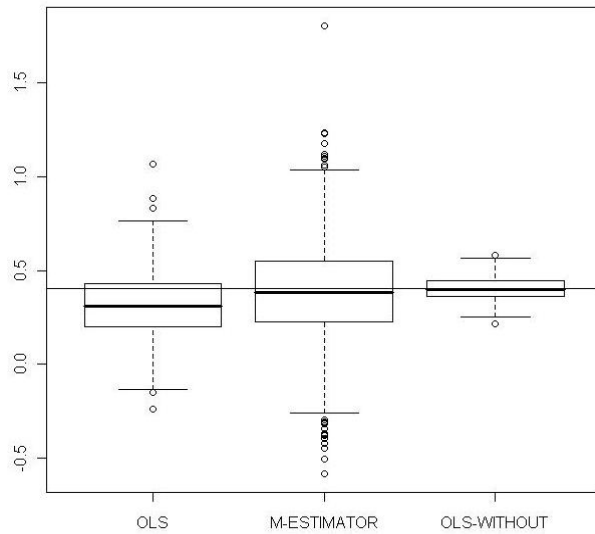
As these values are computed using constraints, it is difficult to assess the standard errors and intervals of confidence analytically. The Monte Carlo simulations will show that the interquartile ranges (IQR) are large and this reflects the instability of classical portfolio optimization.

We simulate 1000 training data sets of size  $n = 100$  each and containing the same kind of contamination. For each sample we estimate  $\boldsymbol{\omega}$  by M-estimator, OLS and  $SR_c$  with  $c = 1, \dots, 10$ . Figures 10-13 show boxplots of these estimates over the 1000 simulated training data. We observe that the shrinkage estimators of  $\omega_2$  and  $\omega_3$  show the biasing effects but those are much less important than in non-constrained regression as presented in section 5. The other effect we can see is that variabilities are large but we observe reduction of variability with some values of the shrinking constant  $c$ .

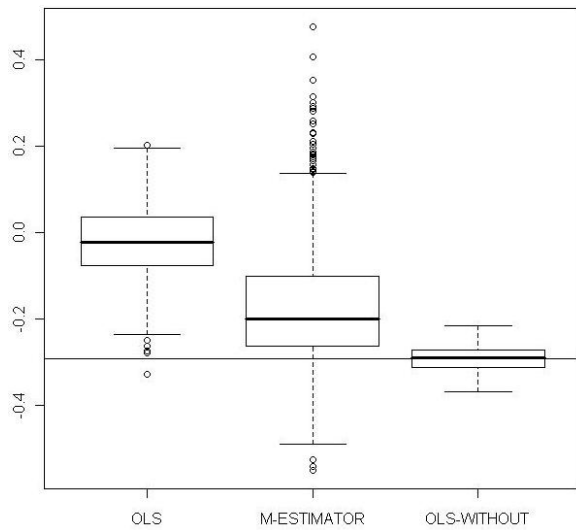
Now in Table 2 we report the frequencies of selection of a minimum measure of prediction for a range of values of  $c = 0, 1, \dots, 10, \infty$  over the 1000 replicates. The optimal  $c$  which minimizes the quality of prediction is around 1 for the three measures MAE, RMSE and STAE. At this stage, our simulations showed that with weaker contamination M-estimation is optimal and with less percentage of contamination the optima are very instable.

**Table 2.** Portfolio with normal contamination: frequencies of selection of a minimum measure of prediction

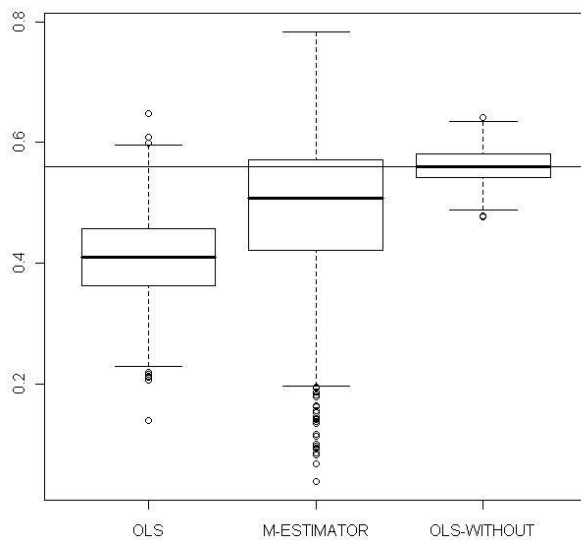
Value of c:	LTS	1	2	3	4	5	6	7	8	9	10	OLS
MAE	107	129	107	93	102	93	71	66	57	68	53	54
RMSE	84	120	108	94	102	89	78	73	63	78	51	60
STAE	111	133	99	94	109	94	66	68	60	58	56	52



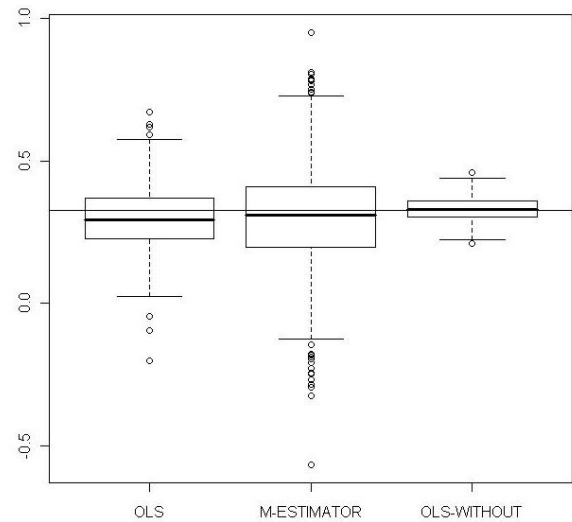
**Figure 5.** OLS-estimates, M-estimates and non-contaminated OLS estimates of  $\omega_1$



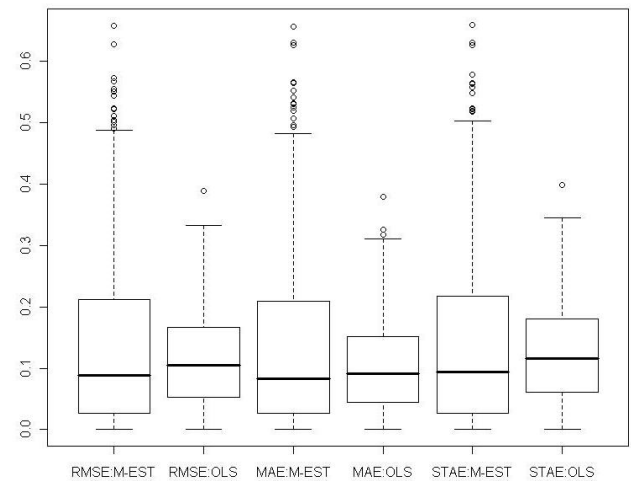
**Figure 6.** OLS-estimates, M-estimates and non-contaminated OLS estimates of  $\omega_2$



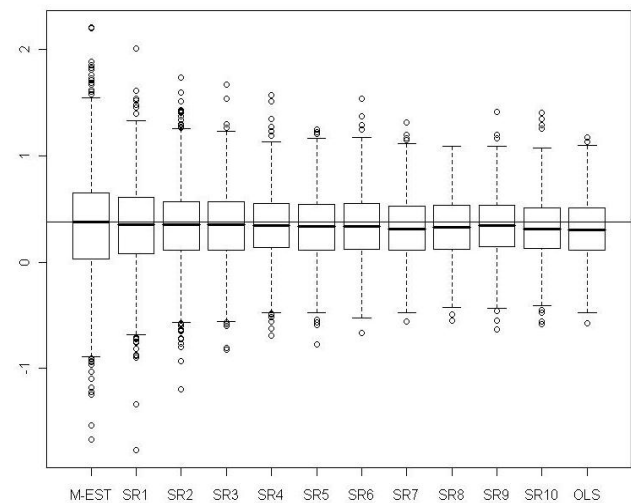
**Figure 7.** OLS-estimates, M-estimates and non-contaminated OLS estimates of  $\omega_3$



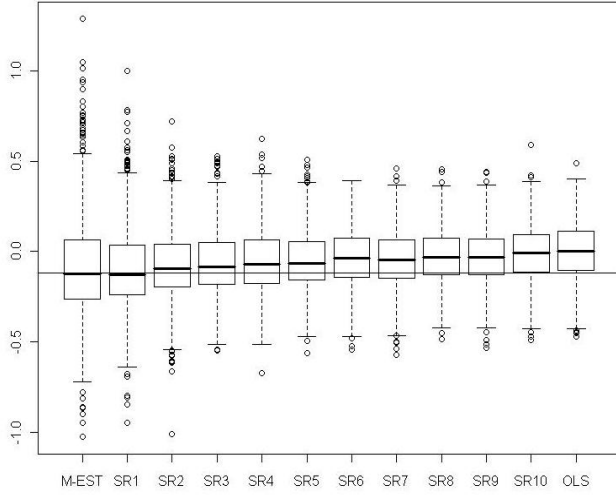
**Figure 8.** OLS-estimates, M-estimates and non-contaminated OLS estimates of  $\omega_4$



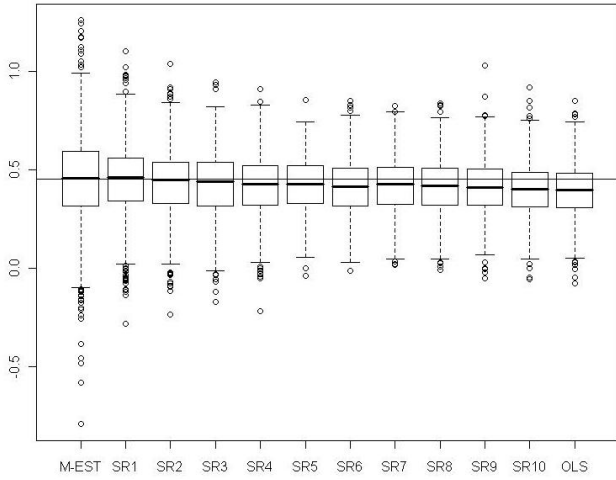
**Figure 9.** Portfolio with normal contamination: relative gains obtained with shrinkage robust estimators compared to M-estimator and OLS on various measures of prediction (RMSE, MAE, STAE)



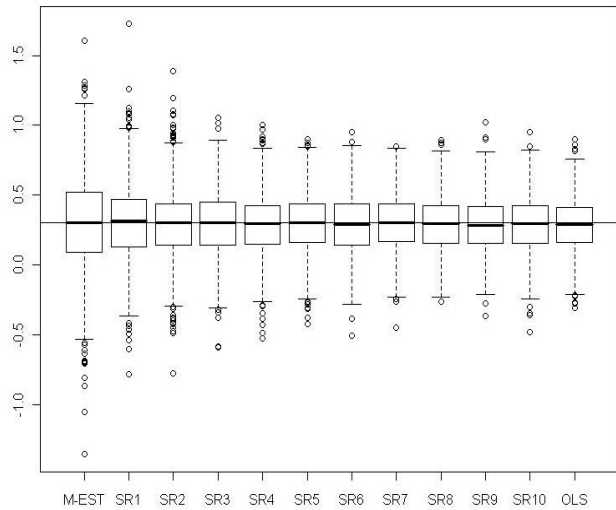
**Figure 10.** Portfolio with normal contamination: Boxplots of  $\omega_1$  for several values of shrinkage



**Figure 11.** Portfolio with normal contamination: Boxplots of  $\omega_2$  for several values of shrinkage



**Figure 12.** Portfolio with normal contamination: Boxplots of  $\omega_3$  for several values of shrinkage



**Figure 13.** Portfolio with normal contamination: Boxplots of  $\omega_4$  for several values of shrinkage

## 6.2. Skew-Normal and Skew- $t$ Data

In this example, we suppose that returns are generated from a 4-dimensional skew-normal distribution:

$$\mathbf{r}_t : SN_4(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}) \quad (30)$$

with parameters:

$$\boldsymbol{\xi} = \begin{pmatrix} 30 \\ 20 \\ 10 \\ 40 \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} 100 & 140 & 70 & 70 \\ 140 & 400 & 140 & 140 \\ 70 & 140 & 100 & 70 \\ 70 & 140 & 70 & 100 \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} -30 \\ -30 \\ -30 \\ -30 \end{pmatrix}. \quad (31)$$

We simulate a skew-normal test sample of size  $T = 400$  and take  $r = 25$  to be the expended return of the portfolio as in the last subsection. The robust estimate of  $\boldsymbol{\mu}$  is:

$$\hat{\boldsymbol{\mu}} = \begin{pmatrix} 23.56 \\ 6.38 \\ 3.58 \\ 33.53 \end{pmatrix}. \quad (32)$$

The estimated weights are the following:

Variable	Constrained M	Constrained OLS
$\hat{\omega}_0$	0.4825	0.4597
$\hat{\omega}_1$	-0.0611	-0.0887
$\hat{\omega}_2$	0.1795	0.2123
$\hat{\omega}_3$	0.3989	0.4168

As before we simulate 1000 training data sets of size  $n = 100$  each and containing the same kind of contamination. For each sample we estimate  $\boldsymbol{\omega}$  by M-estimator, OLS and  $SR_c$  with  $c = 1, \dots, 10$ . Figures 15-18 show boxplots of these estimates over the 1000 simulated training data. We observe that the shrinkage estimators of  $\omega_2$  and  $\omega_3$  show the biasing effect when  $SR_c$  tends to OLS. Now the IRQ are smaller than in the normal contamination case. The IRQ are in general less than 0.1 excepting the IRQ for  $\omega_1$  which is around 0.2 for the M-estimator. As before we observe the reduction of variability with some values of the shrinking constant  $c$ .

Now in Table 3 we report the frequencies of selection of a minimum measure of prediction for a range of values of  $c = 0, 1, \dots, 10, \infty$  over the 1000 replicates. The optimal  $c$  which minimizes the quality of prediction for MAE and STAE is 1, for RMSE it is around 5. Others simulations showed that these optimum values are more or less instable around 2.

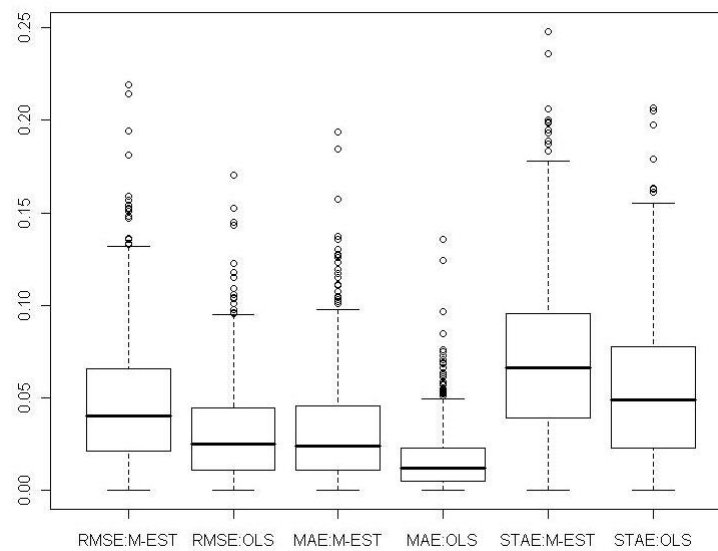
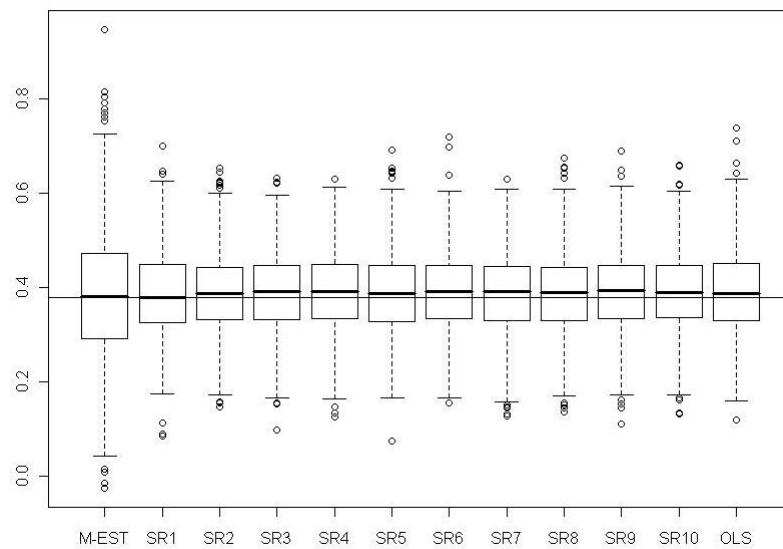
Finally, Table 4 summarizes the computations for the Skew- $t$  model, using the same parameters as the skew-normal and using 3 degrees of freedom. Small degree of freedom value allows for more outliers. Following Huisman thesis (1999), 3 to 6 degrees of freedom are usual in financial data. The optimum value for  $c$  is about 5.

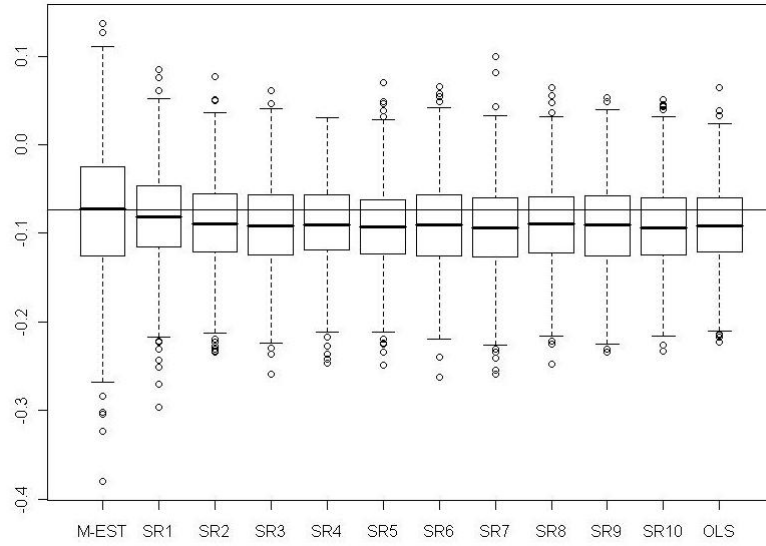
**Table 3.** Portfolio skew-normal contamination: frequencies of selection of a minimum measure of prediction

Value of c:	LTS	1	2	3	4	5	6	7	8	9	10	OLS
MAE	40	103	94	90	88	77	84	77	93	82	91	81
RMSE	39	62	90	92	84	110	85	79	81	96	87	95
STAE	55	103	83	99	78	88	77	82	88	75	84	88

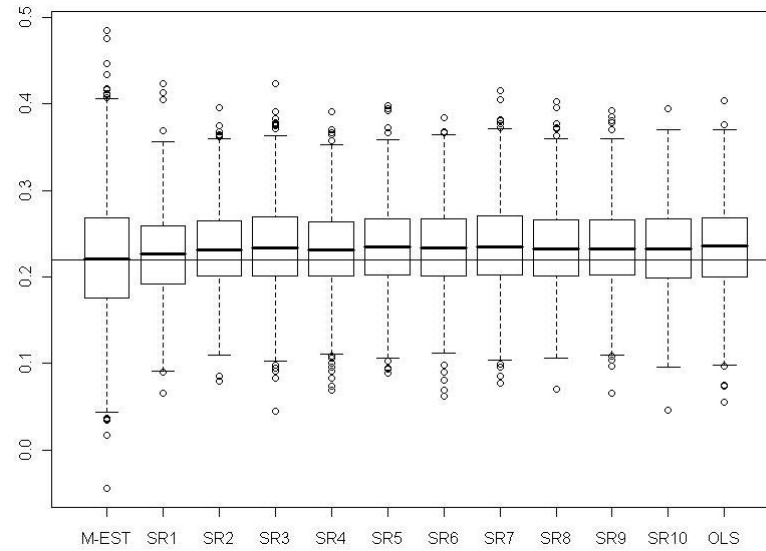
**Table 4.** Portfolio skew-*t* contamination: frequencies of selection of a minimum measure of prediction

Value of c:	LTS	1	2	3	4	5	6	7	8	9	10	OLS
MAE	71	82	84	87	90	92	84	105	76	85	78	66
RMSE	92	104	94	98	92	78	77	71	83	64	94	53
STAE	61	83	80	83	85	107	88	89	86	80	86	72

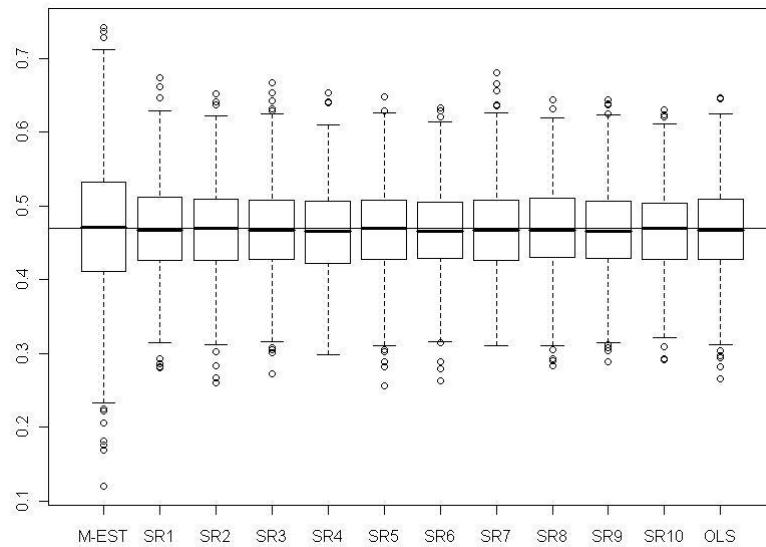
**Figure 14.** Portfolio with skew-normal returns: relative gains obtained with shrinkage robust estimators compared to M-estimator and OLS on various measures of prediction (RMSE, MAE, STAE)**Figure 15.** Portfolio with skew-normal returns: Boxplots of  $\omega_1$  for several values of shrinkage



**Figure 16.** Portfolio with skew-normal returns: Boxplots of  $\omega_2$  for several values of shrinkage



**Figure 17.** Portfolio with skew-normal returns: Boxplots of  $\omega_3$  for several values of shrinkage



**Figure 18.** Portfolio with skew-normal returns: Boxplots of  $\omega_4$  for several values of shrinkage

## 7. Conclusions

In this paper, we have implemented a multivariate version of the shrinkage robust estimators described in Genton and Ronchetti (2008). The aim was to apply the method to the estimation of weights for portfolio optimization. The greatest difficulty was to combine the general method with the constraints which are present in the definition of portfolio optimization. We have seen in Section 6 that the shrinkage constant  $c$  is more instable than in the non-constrained case (Section 5). The origin of the effect is very probably the high instability of the estimation of portfolio weights even with M-estimators. Anyway, the simulations show a optimal shrinkage constant of about 1 for our skew-normal returns and about 5 for our skew- $t$  returns. Location, scale and shape parameters were the same for both laws. We used a skew- $t$  distribution with 3 degrees of freedom, and consequently large outliers were allowed.

The Monte Carlo simulations give us only empirical heuristics for actual applications of the robust portfolio allocation. In the future this can be followed by a theoretical study to find more general properties relating asymmetry and shrinkage.

## ACKNOWLEDGEMENTS

The core of this work was done while the author was a master student in statistics in the University of Geneva in 2008. The author is very grateful to Prof. Marc Genton and Prof. Elvezio Ronchetti for helpful advises and remarks.

## REFERENCES

- [1] Azzalini, A. (1985) *A class of distributions which includes the normal ones*, *Scand. J.Statist.* 12, pp. 171-178.
- [2] Azzalini, A., Capitanio, A. (2003) *Distributions generated by perturbation of symmetry with emphasis on a multivariate skew  $t$  distribution*, *J. Roy.Statist.Soc., series B* vol 65(2003), pp. 367-389.
- [3] Azzalini, A., Dalla Valle, A. (1996) *The multivariate skew normal distribution*, *Biometrika* 83, pp. 715-726.
- [4] Brodie, J., Daubechies, I., De Mol, C., Giannone, D. (2007) *Sparse and stable Markowitz portfolios*, No 6474, *CEPR Discussion Papers from C.E.P.R. Discussion Papers*.
- [5] Chan, L.K.C. and Lakonishok, J. (1992), *Robust measurement of beta risk*, *Journal of Financial and Quantitative Analysis* 27, 265-282.
- [6] DeMiguel, V., Nogales, F.J., (2009). *Portfolio selection with robust estimation*, *Operations Research* 57, 560-577.
- [7] DeMiguel, V., Martin-Utrera, A., Nogales, F.J., (2013), *Size matters: Optimal calibration of shrinkage estimators for portfolio selection*, *Journal of Banking & Finance* 37 (2013) 3018-3034.
- [8] Genton, M., Ronchetti, E. (2008) *Robust Prediction of Beta*, in *Kontoghiorghes, E. J., Rustem, B. and Winker, P. (eds.), Computational Methods in Financial Engineering, Essays in Honour of Manfred Gilli*, Springer, 147-161.
- [9] Gramacy, R. B., Lee, J. H., and Silva, R. (2008). *On estimating covariances between many assets with histories of highly variable length.* *Tech. Rep. 0710.5837*, arXiv. Url: <http://arxiv.org/abs/0710.5837>.
- [10] Gramacy R. and Pantaleo E., (2010) *Shrinkage Regression for Multivariate Inference with Missing Data, and an Application to Portfolio Balancing*, *Bayesian Analysis* 5, Number 2, pp. 237-262.
- [11] Hampel, F.R., Ronchetti, E., Rousseeuw, P.J., et Stahel (1986) *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.
- [12] Hampel, F.R., (1968) *Contribution to the theory of Robust Estimation*, *Ph. D. thesis*, University of California, Berkeley.
- [13] Hu W., Kercheval A. (2010), *Portfolio optimization for student  $t$  and skewed  $t$  returns*, *Quantitative Finance*, Volume 10, Issue 1 Jan. 2010, p. 91-105.
- [14] Huber, P.J. (1964) *Robust estimation of a location parameter*, *Annals of mathematical Statistics* 35, 73-101.
- [15] Huber P.J., Ronchetti E.M. (2009), *Robust Statistics*, Wiley, New York, 2nd edition.
- [16] Huisman R. (1999) *Adventures in international financial markets*, *PhD. Thesis*, Maastricht University.
- [17] Markowitz H. (1952) *Portfolio Selection*, *Journal of Finance*, 7:1, pp.77-91.
- [18] Martin, R.D. and Simin, T. (2003), *Outlier resistant estimates of beta*, *Financial Analysts Journal* 59, 56-69.
- [19] Perret-Gentil, C., M.-P. Victoria-Feser. (2004). *Robust mean-variance portfolio selection*, *FAME Research Paper 140*, International Center for Financial Asset Management and Engineering, Geneva.
- [20] Popova, I., Morton, D., Popova, E., Yau, J. (2003) *Optimal hedge fund allocation with asymmetric preferences and distributions*, *Technical Report*, University of Texas.
- [21] Rousseeuw, P.J. (1985) *Multivariate estimation with high breakdown point*, in W.Grossman, G. Pflug, I. Vincze, and W. Wertz eds., *Mathematical statistics and Applications*, Vol. B, Reidel, Dordrecht, The Netherlands, 283-197.
- [22] Rousseeuw, P.J. and Van Driessen, K. (1999) *A Fast Algorithm for the Minimum Covariance Determinant Estimator*, *Technometrics*, 41, 212-223.
- [23] Sharpe, W.F. (1971), *Mean-absolute-deviation characteristic lines for securities and portfolios*, *Management Science* 18, B1-B13.
- [24] Vaz-de Melo, B., R. P. Camara. (2003). *Robust modeling of multivariate financial data*, *Coppead Working Paper Series 355*, Federal University at Rio de Janeiro, Rio de Janeiro, Brazil.
- [25] Welsch, R., Zhou, X. (2007) *Application of robust statistics to asset allocation models*, *Statistical Journal* volume 5, number 1, March 2007. pp. 97-114.