



Robust Portfolio Optimization Using Pseudodistances

Aida Toma, Samuela Leoni-Aubin

► To cite this version:

Aida Toma, Samuela Leoni-Aubin. Robust Portfolio Optimization Using Pseudodistances. PLoS ONE, 2015. [⟨hal-02081386⟩](#)

HAL Id: hal-02081386

<https://hal.science/hal-02081386v1>

Submitted on 27 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

RESEARCH ARTICLE

Robust Portfolio Optimization Using Pseudodistances

Aida Toma^{1,2☯*}, Samuela Leoni-Aubin^{3☯}

1 Department of Applied Mathematics, Bucharest Academy of Economic Studies, Bucharest, Romania, **2** “Gh. Mihoc—C. Iacob” Institute of Mathematical Statistics and Applied Mathematics, Romanian Academy, Bucharest, Romania, **3** INSA Lyon, ICJ, Villeurbanne Cedex, France

☯ These authors contributed equally to this work.

* Current address: Department of Applied Mathematics, Bucharest Academy of Economic Studies, Piața Romană 6, Bucharest, Romania

* aida_toma@yahoo.com



Abstract

The presence of outliers in financial asset returns is a frequently occurring phenomenon which may lead to unreliable mean-variance optimized portfolios. This fact is due to the unbounded influence that outliers can have on the mean returns and covariance estimators that are inputs in the optimization procedure. In this paper we present robust estimators of mean and covariance matrix obtained by minimizing an empirical version of a pseudodistance between the assumed model and the true model underlying the data. We prove and discuss theoretical properties of these estimators, such as affine equivariance, B-robustness, asymptotic normality and asymptotic relative efficiency. These estimators can be easily used in place of the classical estimators, thereby providing robust optimized portfolios. A Monte Carlo simulation study and applications to real data show the advantages of the proposed approach. We study both in-sample and out-of-sample performance of the proposed robust portfolios comparing them with some other portfolios known in literature.

OPEN ACCESS

Citation: Toma A, Leoni-Aubin S (2015) Robust Portfolio Optimization Using Pseudodistances. PLoS ONE 10(10): e0140546. doi:10.1371/journal.pone.0140546

Editor: Enrico Scalas, Università del Piemonte Orientale, ITALY

Received: March 7, 2015

Accepted: September 28, 2015

Published: October 15, 2015

Copyright: © 2015 Toma, Leoni-Aubin. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data were obtained from the third party MSCI (<https://www.msci.com/web/msci>). Boxplots for relevant data are included within the paper, and interested readers may obtain the full dataset by contacting Orsolya Czintos at MSCI: Orsolya.Czintos@msci.com

Funding: This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project number PN-II-RU-TE-2012-3-0007 (<http://uefiscdi.gov.ro>) (AT). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Introduction

Since Markowitz [1] formulated the idea of diversification of investments, the mean-variance approach has been widely used in practice in asset allocation and portfolio management, despite many sophisticated models proposed in literature. On the other hand, some drawbacks of the standard Markowitz approach are reported in literature (see [2]). One of the critical weaknesses of the classical mean-variance analysis is its lack of robustness. Since the classical estimators of the mean and the covariance matrix, which are inputs in the optimization procedure, are very sensitive to the presence of gross errors or atypical events in data, the weights of the resulted portfolio, which are outputs of this procedure, can be drastically affected by these atypical data. This fact was proved in [3] by using the influence function approach. In order to remove this drawback and to construct portfolios not overly affected by deviations of the data from the assumed model, many methods have been proposed in literature. For an overview on the methods used for robust portfolio optimization we refer to [4]. Among the methods which

Competing Interests: The authors have declared that no competing interests exist.

improve the stability of portfolio weights by using robust estimators of the mean and covariance, we recall those proposed by Vaz-de Melo and Camara [5] which use M-estimators, Perret-Gentil and Victoria-Feser [3] which use the translated biweight S-estimator, Welsch and Zhou [6] which use minimum covariance determinant estimator and winsorization, DeMiguel and Nogales [7] which use both M- and S-estimators, Ferrari and Paterlini [8] which use Maximum L_q -Likelihood Estimators. These contributions have the merit to consider the role of robust estimation for improving the mean-variance portfolios. On the other hand, it is known that traditional robust estimators suffer significant losses in efficiency compared with the maximum likelihood estimator. Therefore, a trade-off robustness-efficiency should be carefully analyzed.

Our contribution to robust portfolio optimization is developed within a minimum pseudodistance framework. We can say that the minimum pseudodistance methods for estimation take part to the same category with the minimum divergence methods. The minimum divergence estimators are defined by minimizing some appropriate divergence between an assumed model and the true model underlying the data. Depending on the choice of the divergence, minimum divergence estimators can afford considerable robustness at minimal loss of efficiency. However, the classical approaches based on divergence minimization require nonparametric density estimation, which can be problematic in multi-dimensional settings. Some proposals to avoid the nonparametric density estimation in minimum divergence estimation have been made in [9, 10] and [11] and robustness properties of such estimators have been studied in [12], [13].

The pseudodistance that we use was originally introduced in [10], where it was called “type-0” divergence. Corresponding minimum divergence estimators have also been studied in [10]. It was also derived (using a cross entropy argument) and extensively studied in [14] where is called γ -divergence. More details about this divergence were provided in [15]. It was also introduced in [16] in the context of decomposable pseudodistances. The minimum pseudodistance estimators for general parametric models have been presented in [16, 17] and consist in minimization of an empirical version of a pseudodistance between the assumed model and the true model underlying the data. These estimators have the advantages of not requiring any prior smoothing and conciliate robustness with high efficiency, usually requiring distinct techniques. In some papers, the name of pseudodistance was used instead of divergence since it was considered that the divergences satisfy the information processing property, which is not the case of the pseudodistance. The interest on statistical methods based on information measures and particularly on divergences has grown substantially in recent years. We refer to the monographs [15, 18] for description of research and applications in this field, as well as to some recent articles [8, 19–21] developing such methods in applicative directions.

The contribution of the present paper is as follows. First we prove and discuss theoretical properties of the minimum pseudodistance estimators of multivariate location and covariance in the case of multivariate normal distribution, such as affine equivariance, B-robustness, asymptotic normality and asymptotic relative efficiency, as well as empirical properties based on Monte Carlo simulations. The behavior of the estimators depends on a tuning positive parameter α which controls the trade-off between robustness and efficiency. When the data are consistent with normality and $\alpha \rightarrow 0$, the estimation method corresponds to the maximum likelihood method (MLE) which is known to have full asymptotic efficiency at the model. When $\alpha > 0$, the estimator gains robustness, while keeping high efficiency. The minimum pseudodistance estimators can be easily used in place of the classical mean and covariance matrix estimators, thereby providing robust and efficient mean-variance optimized portfolios. We prove asymptotic properties of portfolio weights based on minimum pseudodistance estimators, such as B-robustness and asymptotic normality. Then, on the basis of real data, we

analyze different robust portfolios based on robust minimum pseudodistance estimators, studying both their in-sample and out-of-sample behavior and comparing them with some other portfolios known in literature. Among the analyzed portfolios count: optimal mean-variance portfolios using the classical MLE or minimum pseudodistance estimators or S-estimators, minimum-variance portfolios using the classical MLE or minimum pseudodistance estimators or S-estimators, as well as the equally-weighted portfolio. We considered the cases when short selling is allowed, respectively when short selling is not allowed. The out-of-sample empirical performance of the considered portfolios is evaluated using the measures: turnover, the out-of-sample portfolio variance and the out-of-sample Sharpe ratio. Our theoretical and numerical results show that the optimal portfolios based on minimum pseudodistance estimators are much more stable to extreme events than those obtained by plugging-in the MLEs and compare well with other optimal robust portfolios known in literature.

The outline of the paper is as follows: in Section 2, we shortly describe the Markowitz's mean-variance model whose inputs are estimations of location and covariance of asset returns. The minimum pseudodistance estimators of location and covariance are introduced in Section 3. Here we prove theoretical properties of these estimators in the case of multivariate normal distribution, such as the affine equivariance and B-robustness. We also determine the asymptotic covariance matrices of the estimators and discuss the asymptotic relative efficiency. The estimators of the portfolio weights together with their asymptotic properties are presented in Section 4. In Sections 5 and 6, a Monte Carlo simulation study and then applications on real data show the advantages of the new approach. Here we illustrate both in-sample as well out-of-sample performance of the portfolios.

Portfolio optimization model

We consider a portfolio formed by N financial assets. The returns of the assets are characterized by the random vector $X = (X_1, \dots, X_N)^\top$, where X_i denotes the random variable associated to the return of the asset i , $i = 1, \dots, N$. Let $p = (p_1, \dots, p_N)^\top$ be the vector of weights associated to the portfolio, where p_i represents the proportion of the investor's capital invested in the asset i . The total return of the portfolio is given by the random variable

$$p^\top X = p_1 X_1 + \dots + p_N X_N.$$

Supposing that the random vector X follows a multivariate normal distribution $\mathcal{N}_N(\mu, \Sigma)$, where μ is the vector containing the mean returns of the assets and Σ is the covariance matrix of the returns of the assets, the mean of the portfolio return can be written as $R(p) = p^\top \mu$ and the portfolio variance as $S(p) = p^\top \Sigma p$.

The Markowitz approach for optimal portfolio selection (see [1]) consists in solving the following optimization problem. For a given investor's risk aversion $\lambda > 0$, the mean-variance optimization selects the optimal portfolio p^* , solution of

$$\arg \max_p \left\{ R(p) - \frac{\lambda}{2} S(p) \right\} \quad (1)$$

with the constraint $p^\top e_N = 1$, e_N being the $N \times 1$ vector of ones. The set of optimal portfolios for all possible values of the risk aversion parameter λ defines the mean-variance efficient frontier. The solution of the above optimization problem is explicit and the optimal portfolio weights, for a fixed value of λ , are given by

$$p^* = \frac{1}{\lambda} \Sigma^{-1} (\mu - \eta e_N) \quad (2)$$

where

$$\eta = \frac{e_N^T \Sigma^{-1} \mu - \lambda}{e_N^T \Sigma^{-1} e_N}.$$

This is the case when short selling is allowed. When short selling is not allowed, we have a supplementary constraint in the optimization problem, namely all the weights p_i are positive.

When the true parameters μ and Σ and the portfolio weights are all known, then we have the true efficient frontier. An estimated efficient frontier can be obtained by using estimators of the mean and covariance matrix. Throughout this paper we denote by $\hat{\mu}$ and $\hat{\Sigma}$ the estimators of the parameters μ and Σ , and by \hat{p}^* the estimator of the optimal portfolio weights, as resulting with [Eq \(2\)](#)

$$\hat{p}^* = \frac{1}{\hat{\lambda}} \hat{\Sigma}^{-1} \left[\hat{\mu} - \frac{e_N^T \hat{\Sigma}^{-1} \hat{\mu} - \hat{\lambda}}{e_N^T \hat{\Sigma}^{-1} e_N} e_N \right]. \quad (3)$$

The mean and the covariance matrix of the returns are in practice estimated by their sample counterparts, i.e. the maximum likelihood estimators under the multivariate normal model. It is known that, under normality, the maximum likelihood estimators are the most efficient. However, in the presence of outlying observations, the asymptotic bias of these estimators can be arbitrarily large and this bias is induced to the corresponding optimal portfolio weights. For this reason, μ and Σ should be robustly estimated.

Robust estimators of the location and covariance

Minimum pseudodistance estimators

For two probability measures P and Q , admitting densities p , respectively q with respect to the Lebesgue measure, the pseudodistances that we consider are defined through

$$R_\alpha(P, Q) := \frac{1}{\alpha + 1} \ln \int p^\alpha dP + \frac{1}{\alpha(\alpha + 1)} \ln \int q^\alpha dQ - \frac{1}{\alpha} \ln \int p^\alpha dQ \quad (4)$$

for $\alpha > 0$ and satisfy the limit relation

$$R_\alpha(P, Q) \rightarrow R_0(P, Q) := \int \ln \frac{q}{p} dQ \text{ for } \alpha \downarrow 0.$$

Note that $R_0(P, Q)$ coincides with the modified Kullback-Leibler divergence. More details about [Eq \(4\)](#) and corresponding minimum divergence estimators can be found in [\[10, 14–16\]](#).

Let \mathcal{P} be a parametric model with parameter space $\Theta \subset \mathbb{R}^d$ and assume that every probability measure P_θ in \mathcal{P} has a density p_θ with respect to the Lebesgue measure. The family of minimum pseudodistance estimators of the unknown parameter θ_0 is obtained by replacing the hypothetical probability measure P_{θ_0} in the pseudodistances $R_\alpha(P_\theta, P_{\theta_0})$ by the empirical measure P_n pertaining to the sample and then minimizing $R_\alpha(P_\theta, P_n)$ with respect to θ on the parameter space. We note that the pseudodistances R_α are also used in [\[22\]](#) in order to define optimal robust M-estimators using the Hampel infinitesimal approach.

Let X^1, \dots, X^T be a sample on $X \sim \mathcal{N}_N(\mu, \Sigma)$ and denote by $\theta = (\mu, \Sigma)$ the parameter of interest. A minimum pseudodistance estimator $\hat{\theta} = (\hat{\mu}, \hat{\Sigma})$ of θ is defined by

$$\hat{\theta} := \arg \inf_{\theta} R_\alpha(P_\theta, P_n)$$

which can be written equivalently as

$$\hat{\theta} = \begin{cases} \arg \sup_{\theta} \frac{1}{TC_{\alpha}(\theta)} \sum_{i=1}^T p_{\theta}^{\alpha}(X^i) & \text{if } \alpha > 0 \\ \arg \sup_{\theta} \frac{1}{T} \sum_{i=1}^T \ln p_{\theta}(X^i) & \text{if } \alpha = 0 \end{cases} \quad (5)$$

where p_{θ} is the N -variate normal density and $C_{\alpha}(\theta) = \left(\int p_{\theta}^{\alpha+1} d\lambda \right)^{\frac{\alpha}{\alpha+1}}$. Note that, the choice $\alpha = 0$ leads to the definition of the classical MLE. Throughout the paper we will also use the notation $\|x - \mu\|_{\Sigma^{-1}} = [(x - \mu)^{\top} \Sigma^{-1} (x - \mu)]^{1/2}$. A simple calculation shows that

$$C_{\alpha}(\theta) = \frac{\left(\frac{1}{2\pi}\right)^{\frac{N\alpha^2}{2(\alpha+1)}} (\sqrt{\det \Sigma^{-1}})^{\frac{\alpha^2}{\alpha+1}}}{(\sqrt{\alpha+1})^{\frac{N\alpha}{\alpha+1}}}.$$

For $\alpha > 0$, the estimator Eq (5) can be expressed as

$$\hat{\theta} = \arg \sup_{\theta} \left(\sqrt{\det \Sigma^{-1}} \right)^{\frac{\alpha}{\alpha+1}} \sum_{i=1}^T \exp \left(-\frac{\alpha}{2} \|X^i - \mu\|_{\Sigma^{-1}}^2 \right).$$

By direct differentiation with respect to μ and Σ , we see that the estimators of these parameters are solutions of the system

$$\begin{aligned} \sum_{i=1}^T (X^i - \mu) \exp \left(-\frac{\alpha}{2} \|X^i - \mu\|_{\Sigma^{-1}}^2 \right) &= 0 \\ \sum_{i=1}^T \left[\frac{1}{\alpha+1} \Sigma - (X^i - \mu)(X^i - \mu)^{\top} \right] \exp \left(-\frac{\alpha}{2} \|X^i - \mu\|_{\Sigma^{-1}}^2 \right) &= 0 \end{aligned}$$

which can be rewritten as

$$\mu = \frac{\sum_{i=1}^T \exp \left(-\frac{\alpha}{2} \|X^i - \mu\|_{\Sigma^{-1}}^2 \right) X^i}{\sum_{i=1}^T \exp \left(-\frac{\alpha}{2} \|X^i - \mu\|_{\Sigma^{-1}}^2 \right)} \quad (6)$$

$$\Sigma = \frac{\sum_{i=1}^T (\alpha+1) \exp \left(-\frac{\alpha}{2} \|X^i - \mu\|_{\Sigma^{-1}}^2 \right) (X^i - \mu)(X^i - \mu)^{\top}}{\sum_{i=1}^T \exp \left(-\frac{\alpha}{2} \|X^i - \mu\|_{\Sigma^{-1}}^2 \right)} \quad (7)$$

Affine equivariance

The location and dispersion estimators defined above are affine equivariant. More precisely, if $\hat{\mu}(\mathbf{X})$ and $\hat{\Sigma}(\mathbf{X})$ are estimators corresponding to a sample $\mathbf{X} = (X^1, \dots, X^T)$, then

$$\hat{\mu}(A\mathbf{X} + b) = A\hat{\mu}(\mathbf{X}) + b \quad (8)$$

$$\hat{\Sigma}(A\mathbf{X} + b) = A\hat{\Sigma}(\mathbf{X})A^{\top} \quad (9)$$

for any $N \times N$ nonsingular matrix A and any $b \in \mathbb{R}^N$. Indeed, let A be a nonsingular matrix, $b \in \mathbb{R}^N$ and $\mathbf{Y} = (Y^1, \dots, Y^T)$, $Y^i = AX^i + b$. Estimators $\hat{\mu}(\mathbf{Y})$ and $\hat{\Sigma}(\mathbf{Y})$ are solutions of the system obtained from Eqs (6) and (7) by replacing X^i with Y^i . Then, by replacing Y^i with $AX^i + b$, we find $\hat{\mu}(\mathbf{X}) = A^{-1}(\hat{\mu}(\mathbf{Y}) - b)$ and $\hat{\Sigma}(\mathbf{X}) = A^{-1}\hat{\Sigma}(\mathbf{Y})(A^{\top})^{-1}$. Hence, Eqs (8) and (9) hold.

Influence functions

A fundamental tool used for studying statistical robustness is the influence function. Recall that, a map T defined on a set of probability measures and parameter space valued is a statistical functional corresponding to an estimator $\hat{\theta}_n$ of the parameter θ , if $\hat{\theta}_n = T(P_n)$, where P_n is the empirical measure associated to the sample. The influence function of T at P_θ is defined by

$$\text{IF}(x; T, P_\theta) := \left. \frac{\partial T(\tilde{P}_{\varepsilon x})}{\partial \varepsilon} \right|_{\varepsilon=0}$$

where $\tilde{P}_{\varepsilon x} := (1 - \varepsilon)P_\theta + \varepsilon\delta_x$, $\varepsilon > 0$, δ_x being the Dirac measure putting all mass at x . Whenever the influence function is bounded with respect to x the corresponding estimator is called robust (see [23]).

The statistical functionals associated to the minimum pseudodistance estimators of μ and Σ are $\mu(P)$ and $\Sigma(P)$ defined by the solutions of the system

$$\begin{aligned} \int (x - \mu) \exp\left(-\frac{\alpha}{2} \|x - \mu\|_{\Sigma^{-1}}^2\right) dP(x) &= 0 \\ \int \left[(x - \mu)(x - \mu)^\top \exp\left(-\frac{\alpha}{2} \|x - \mu\|_{\Sigma^{-1}}^2\right) - \right. \\ &\quad \left. - \frac{1}{\alpha + 1} \Sigma \exp\left(-\frac{\alpha}{2} \|x - \mu\|_{\Sigma^{-1}}^2\right) \right] dP(x) = 0. \end{aligned}$$

This system can be rewritten under the form

$$\int w_1(\|x - \mu\|_{\Sigma^{-1}})(x - \mu) dP(x) = 0 \quad (10)$$

$$\int \left[\frac{w_2(\|x - \mu\|_{\Sigma^{-1}})}{\|x - \mu\|_{\Sigma^{-1}}^2} (x - \mu)(x - \mu)^\top - w_3(\|x - \mu\|_{\Sigma^{-1}}) \Sigma \right] dP(x) = 0 \quad (11)$$

where

$$w_1(t) = \exp\left(-\frac{\alpha}{2} t^2\right), w_2(t) = \exp\left(-\frac{\alpha}{2} t^2\right) t^2, w_3(t) = \frac{1}{\alpha + 1} \exp\left(-\frac{\alpha}{2} t^2\right). \quad (12)$$

We note that the solutions of the system given by Eqs (10) and (11), when w_1, w_2, w_3 are arbitrary weight functions, define statistical functionals of general M-estimators of (μ, Σ) (see [24] and [25]). According to the results presented in [25], the influence functions for general M-estimators of μ and Σ are given by

$$\text{IF}(x; \mu, P_{\mu, \Sigma}) = (x - \mu) w_\mu(\|x - \mu\|_{\Sigma^{-1}}) \quad (13)$$

$$\text{IF}(x; \Sigma, P_{\mu, \Sigma}) = (x - \mu)(x - \mu)^\top w_\eta(\|x - \mu\|_{\Sigma^{-1}}) - \Sigma w_\delta(\|x - \mu\|_{\Sigma^{-1}}) \quad (14)$$

where

$$\begin{aligned} w_{\mu}(\|x - \mu\|_{\Sigma^{-1}}) &= \frac{w_1(\|x - \mu\|_{\Sigma^{-1}})}{E_{P_0}[w_1(\|y\|) + \frac{1}{N}w'_1(\|y\|)\|y\|]} \\ w_{\eta}(\|x - \mu\|_{\Sigma^{-1}}) &= \frac{N(N+2)w_2(\|x - \mu\|_{\Sigma^{-1}})}{\|x - \mu\|_{\Sigma^{-1}}^2 E_{P_0}[Nw_2(\|y\|) + w'_2(\|y\|)\|y\|]} \\ w_{\delta}(\|x - \mu\|_{\Sigma^{-1}}) &= \frac{Nw_3(\|x - \mu\|_{\Sigma^{-1}}) - 2w_2(\|x - \mu\|_{\Sigma^{-1}})}{E_{P_0}[w'_2(\|y\|)\|y\| - Nw'_3(\|y\|)\|y\|]} + \\ &\quad + \frac{(N+2)w_2(\|x - \mu\|_{\Sigma^{-1}})}{E_{P_0}[Nw_2(\|y\|) + w'_2(\|y\|)\|y\|]} \end{aligned}$$

P_0 denoting the probability measure associate to the N -variate standard normal distribution and $\|\cdot\|$ the Euclidian norm.

For the weight functions w_1, w_2, w_3 from Eq (12), corresponding to the minimum pseudo-distance estimators, we get

$$w_{\mu}(t) = (\sqrt{\alpha+1})^{N+2} \exp\left(-\frac{\alpha}{2}t^2\right) \quad (15)$$

$$w_{\eta}(t) = (\sqrt{\alpha+1})^{N+4} \exp\left(-\frac{\alpha}{2}t^2\right) \quad (16)$$

$$w_{\delta}(t) = (\sqrt{\alpha+1})^{N+2} \exp\left(-\frac{\alpha}{2}t^2\right) \quad (17)$$

and replacing in Eqs (13) and (14) we obtain

$$\text{IF}(x; \mu, P_{\mu, \Sigma}) = (\sqrt{\alpha+1})^{N+2} (x - \mu) \exp\left(-\frac{\alpha}{2}\|x - \mu\|_{\Sigma^{-1}}^2\right) \quad (18)$$

$$\text{IF}(x; \Sigma, P_{\mu, \Sigma}) = (\sqrt{\alpha+1})^{N+4} \left[(x - \mu)(x - \mu)^{\top} - \frac{1}{\alpha+1}\Sigma \right] \exp\left(-\frac{\alpha}{2}\|x - \mu\|_{\Sigma^{-1}}^2\right). \quad (19)$$

Both influence functions are bounded with respect to x . Therefore the minimum pseudodistance estimators of μ and Σ are robust.

In Fig 1 the influence function for the first component of the minimum pseudodistance estimator of the mean is represented. Here P_0 is the bivariate standard normal law and the constant corresponding to the estimator was chosed $\alpha = 0.5$. The influence function is bounded and also redescending. A major feature of the minimum pseudodistance estimators considered in this paper is that they are redescending (the influence functions tend to 0, when $\|x\|$ tends to infinity) and this represents an important advantage from the robustness point of view.

Asymptotic normality

For general parametric models, under some regularity conditions, the minimum pseudodistance estimators are asymptotically normal distributed (see [16]). In this section, we derive the asymptotic covariance matrices of the mean and the covariance matrix minimum pseudodistance estimators. We adopt the influence function approach and make use of the general results for affine equivariant location and dispersion M-estimators as presented in [26] and [23].

When the observations correspond to the standard N -variate normal law P_0 , under appropriate conditions, $\hat{\mu}$ is asymptotically normal distributed with the asymptotic covariance

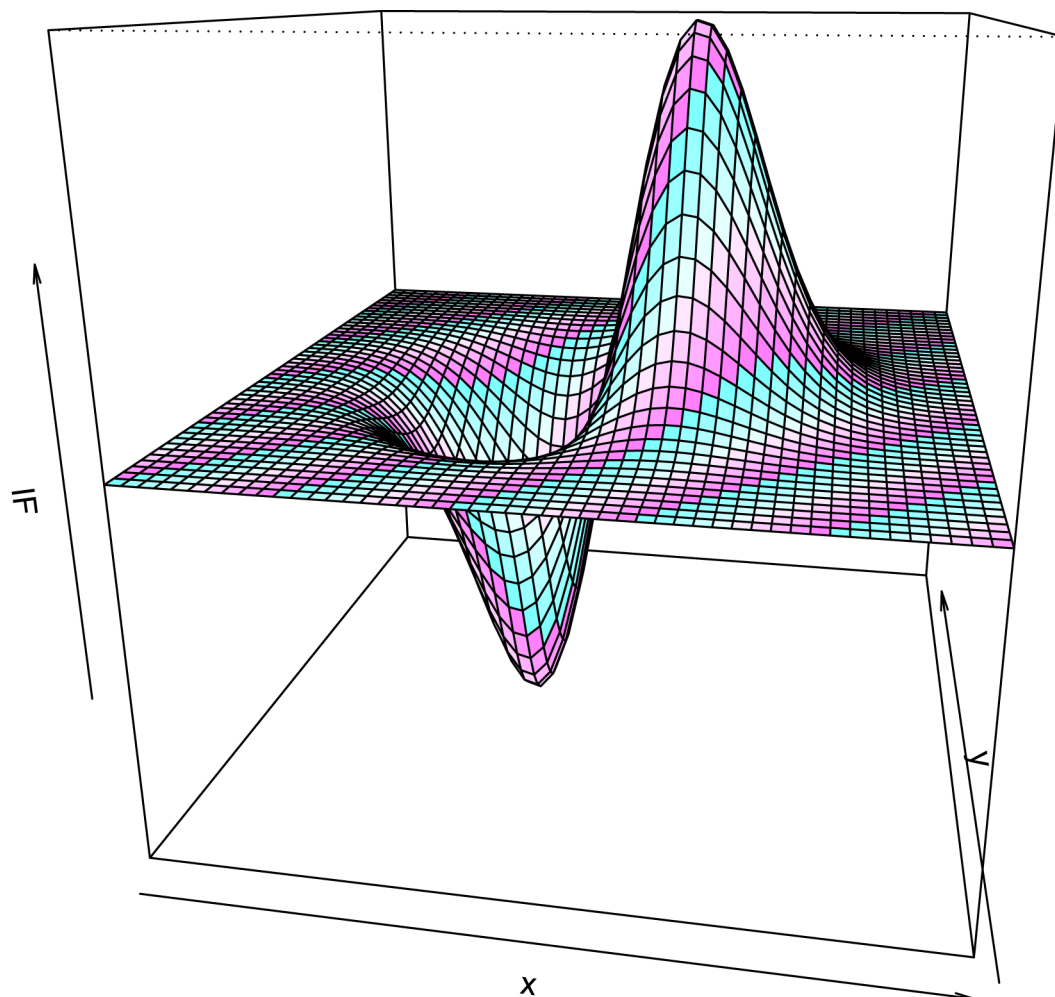


Fig 1. The influence function for the first component of the minimum pseudodistance estimator of the mean.

doi:10.1371/journal.pone.0140546.g001

matrix

$$V(\mu, P_0) = E_{P_0} \{ \text{IF}(Z; \mu, P_0) \text{IF}(Z; \mu, P_0)^T \} = d_\mu I \quad (20)$$

where $d_\mu := E_{P_0} \{ \|Z\|^2 w_\mu^2(\|Z\|) \} / N$ and I is the identity matrix. Formula (20) has been established in [26] for general affine equivariant location M-estimators. The estimator $\hat{\mu}$ belongs to this class. For the weight w_μ from Eq (15) we get $d_\mu = (\alpha + 1)^{N+2} / (\sqrt{2\alpha + 1})^{N+2}$, hence the asymptotic covariance matrix of the minimum pseudodistance estimator $\hat{\mu}$ is

$$V(\mu, P_0) = \left(\frac{\alpha + 1}{\sqrt{2\alpha + 1}} \right)^{N+2} I.$$

When the observations correspond to the normal law $P_{\mu, \Sigma}$, the asymptotic covariance matrix of

$\hat{\mu}$ is given by

$$\begin{aligned} V(\mu, P_{\mu, \Sigma}) &= E_{P_{\mu, \Sigma}} \{ \text{IF}(X; \mu, P_{\mu, \Sigma}) \text{IF}(X; \mu, P_{\mu, \Sigma})^\top \} \\ &= d_\mu \Sigma = \left(\frac{\alpha + 1}{\sqrt{2\alpha + 1}} \right)^{N+2} \Sigma. \end{aligned} \quad (21)$$

Similar results hold for $\text{vecs}(\hat{\Sigma})$, where vecs is the operation that stacks the $N + N(N - 1)/2$ non-redundant elements of Σ into a vector, as follows:

$$\text{vecs}(\Sigma) := (\sigma_{11}/\sqrt{2}, \dots, \sigma_{NN}/\sqrt{2}, \sigma_{21}, \sigma_{31}, \dots, \sigma_{N, N-1})^\top.$$

According to the results presented in [26], when the observations come from the N -variate standard normal law P_0 , the asymptotic covariance matrix corresponding to an affine equivariant M-estimator of the covariance matrix is given by

$$\begin{aligned} V(\Sigma, P_0) &= E_{P_0} \{ \text{vecsIF}(Z; \Sigma, P_0) \text{vecsIF}(Z; \Sigma, P_0)^\top \} \\ &= d_\eta \left(I - \frac{1}{N} w w^\top \right) + d_\tau \cdot \frac{1}{N} w w^\top \end{aligned}$$

where $w^\top := (e_N^\top, \mathbf{0}_{N(N-1)/2}^\top)$, $d_\eta := E_{P_0} \{ \|Z\|^4 w_\eta^2(\|Z\|) \} / (N(N+2))$ and $d_\tau := E_{P_0} \{ w_\tau^2(\|Z\|) \} / (2N)$ with $w_\tau := t^2 w_\eta(t) - N w_\delta(t)$, w_η , w_δ and w_τ being specific to the M-estimator in question.

In our case, w_η and w_δ are given by Eqs (16) and (17), hence

$$w_\tau(t) = (\sqrt{\alpha + 1})^{N+4} \left[t^2 - \frac{N}{\alpha + 1} \right] \exp \left(-\frac{\alpha}{2} t^2 \right).$$

After some calculation, we obtain

$$d_\eta = \left(\frac{\alpha + 1}{\sqrt{2\alpha + 1}} \right)^{N+4} \text{ and } d_\tau = \frac{N\alpha^2(\alpha + 1)^{N+2}}{2(\sqrt{2\alpha + 1})^{N+4}} + \left(\frac{\alpha + 1}{\sqrt{2\alpha + 1}} \right)^{N+4},$$

therefore,

$$V(\Sigma, P_0) = \left(\frac{\alpha + 1}{\sqrt{2\alpha + 1}} \right)^{N+4} I + \frac{\alpha^2(\alpha + 1)^{N+2}}{2(\sqrt{2\alpha + 1})^{N+4}} w w^\top. \quad (22)$$

When the observations correspond to the law $P_{\mu, \Sigma}$, the asymptotic covariance matrix of $\text{vecs}(\hat{\Sigma})$ can be established by using the formula from [23] p.282, which in our notations writes as follows

$$V(\Sigma, P_{\mu, \Sigma}) = \left[\frac{\partial \text{vecs} \left[\Sigma^{\frac{1}{2}} S \Sigma^{\frac{1}{2}} \right]}{\partial \text{vecs} S} \right] V(\Sigma, P_0) \left[\frac{\partial \text{vecs} \left[\Sigma^{\frac{1}{2}} S \Sigma^{\frac{1}{2}} \right]}{\partial \text{vecs} \Sigma} \right]^\top. \quad (23)$$

According to [23] p.272, for a given $N \times N$ matrix Σ_* , it holds

$$\left[\frac{\partial \text{vecs} \left[\Sigma^{\frac{1}{2}} S \Sigma^{\frac{1}{2}} \right]}{\partial \text{vecs} S} \right] \text{vecs} \Sigma_* = \text{vecs} \left(\Sigma^{\frac{1}{2}} \Sigma_* \Sigma^{\frac{1}{2}} \right).$$

Particularly,

$$\left[\frac{\partial \text{vecs} \left[\begin{smallmatrix} \Sigma_2^1 & S \Sigma_2^1 \\ \Sigma_2^1 & S \Sigma_2^1 \end{smallmatrix} \right]}{\partial \text{vecs} S} \right] \text{vecs} I = \text{vecs} \Sigma. \quad (24)$$

Note that $w = \sqrt{2} \text{vecs} I$ and combining Eqs (23), (22) and (24), we get

$$\begin{aligned} V(\Sigma, P_{\mu, \Sigma}) &= \left(\frac{\alpha + 1}{\sqrt{2\alpha + 1}} \right)^{N+4} \left[\frac{\partial \text{vecs} \left[\begin{smallmatrix} \Sigma_2^1 & S \Sigma_2^1 \\ \Sigma_2^1 & S \Sigma_2^1 \end{smallmatrix} \right]}{\partial \text{vecs} S} \right] \left[\frac{\partial \text{vecs} \left[\begin{smallmatrix} \Sigma_2^1 & S \Sigma_2^1 \\ \Sigma_2^1 & S \Sigma_2^1 \end{smallmatrix} \right]}{\partial \text{vecs} S} \right]^\top + \\ &+ \frac{\alpha^2 (\alpha + 1)^{N+2}}{(\sqrt{2\alpha + 1})^{N+4}} \text{vecs} \Sigma (\text{vecs} \Sigma)^\top. \end{aligned}$$

For symmetry reasons, the minimum pseudodistance location and covariance estimators are asymptotically uncorrelated and hence asymptotically independent. This is valid for location and covariance M-estimators in general, as it is underlined in various articles, for example in [24].

Asymptotic relative efficiency

In order to assess the efficiency of the proposed estimators with respect to that of the MLE, we adopt as measure the asymptotic relative efficiency (ARE). For a parameter θ taking values in \mathbb{R}^d and an estimator $\hat{\theta}$ which is asymptotically d -variate normal with mean θ and nonsingular covariance matrix $V(\theta, P)$, the asymptotic relative efficiency with respect to that of the MLE is defined as

$$\text{ARE}(\hat{\theta}, P) = \left(\frac{\det V_0(\theta, P)}{\det V(\theta, P)} \right)^{1/d},$$

$V_0(\theta, P)$ being the asymptotic covariance matrix of the MLE of θ when the observations follow the law P (see [27]). Although the asymptotically most efficient estimator is given by the MLE, the particular MLE can be drastically inefficient when the underlying distribution departs even a little bit from the assumed nominal distribution. Therefore the trade-off between robustness and efficiency should be carefully analyzed.

Due to the asymptotic independence of the mean and the covariance matrix minimum pseudodistance estimators, the asymptotic relative efficiency of $\hat{\theta} = (\hat{\mu}^\top, \text{vecs}(\hat{\Sigma})^\top)^\top$ can be expressed as

$$\begin{aligned} \text{ARE}(\hat{\theta}, P_{\mu, \Sigma}) &= \left(\frac{\det V_0(\theta, P_{\mu, \Sigma})}{\det V(\theta, P_{\mu, \Sigma})} \right)^{\frac{2}{N(N+3)}} \\ &= \left(\frac{\det V_0(\mu, P_{\mu, \Sigma}) \det V_0(\Sigma, P_{\mu, \Sigma})}{\det V(\mu, P_{\mu, \Sigma}) \det V(\Sigma, P_{\mu, \Sigma})} \right)^{\frac{2}{N(N+3)}}. \end{aligned} \quad (25)$$

Using Eqs (21) and (23), formula (25) can be written as

$$\text{ARE}(\hat{\theta}, P_{\mu, \Sigma}) = \left(\frac{\det V_0(\mu, P_0) \det V_0(\Sigma, P_0)}{\det V(\mu, P_0) \det V(\Sigma, P_0)} \right)^{\frac{2}{N(N+3)}}.$$

Table 1. Asymptotic relative efficiency of the minimum pseudodistance estimators.

N	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
1	1	0.98151	0.93871	0.76904	0.63774	0.53033
2	1	0.97704	0.92429	0.72086	0.57042	0.45266
3	1	0.97273	0.91051	0.67698	0.51187	0.38814
4	1	0.96851	0.89718	0.63647	0.46018	0.33370
5	1	0.96435	0.88419	0.59879	0.41420	0.28738
6	1	0.96025	0.87148	0.56360	0.37311	0.24778
7	1	0.95619	0.85902	0.53065	0.33629	0.21380
8	1	0.95215	0.84679	0.49975	0.30322	0.18460
9	1	0.94815	0.83477	0.47073	0.27350	0.15946
10	1	0.94418	0.82294	0.44345	0.24674	0.13779

doi:10.1371/journal.pone.0140546.t001

A direct calculation shows that

$$\begin{aligned}\det V(\mu, P_0) &= \left(\frac{\alpha + 1}{\sqrt{2\alpha + 1}} \right)^{N(N+2)} \\ \det V(\Sigma, P_0) &= \left(\frac{\alpha + 1}{\sqrt{2\alpha + 1}} \right)^{\frac{N(N+1)(N+4)}{2}} \left(1 + \frac{N\alpha^2}{2(\alpha + 1)^2} \right).\end{aligned}$$

Particularly, for $\alpha = 0$, we find the similar quantities for the MLE, namely $\det V_0(\mu, P_0) = 1$ and $\det V_0(\Sigma, P_0) = 1$. Hence

$$\text{ARE}(\hat{\theta}, P_{\mu, \Sigma}) = \frac{1}{\left(\frac{\alpha + 1}{\sqrt{2\alpha + 1}} \right)^{\frac{N^2 + 7N + 8}{N + 3}} \left(1 + \frac{N\alpha^2}{2(\alpha + 1)^2} \right)^{\frac{2}{N(N+3)}}}. \quad (26)$$

Note that, for fixed N and α , $\text{ARE}(\hat{\theta}, P_{\mu, \Sigma})$ is the same, whatever μ or Σ .

In Table 1 values of the asymptotic relative efficiency Eq (26) are given. As it can be seen, when N or α increases, the asymptotic relative efficiency $\text{ARE}(\hat{\theta}, P_{\mu, \Sigma})$ decreases. Therefore, values of α close to zero will provide high efficiency and in the meantime the robustness of the estimation procedure.

The estimator of the optimal portfolio weights

We consider the estimator \hat{p}^* of the optimal portfolio weights, as given by Eq (3), with $\hat{\mu}$ and $\hat{\Sigma}$ minimum pseudodistance estimators.

The influence function of the estimator \hat{p}^* is proportional to the influence functions of the estimators $\hat{\mu}$ and $\hat{\Sigma}$. More precisely,

$$\begin{aligned}\text{IF}(x; p^*, P_{\mu, \Sigma}) &= -\Sigma^{-1} \text{IF}(x; \Sigma, P_{\mu, \Sigma}) p^* + \frac{1}{\lambda} \Sigma^{-1} \{ \text{IF}(x; \mu, P_{\mu, \Sigma}) + \\ &+ \frac{e_N^\top \Sigma^{-1} [\text{IF}(x; \Sigma, P_{\mu, \Sigma}) \Sigma^{-1} \mu - \text{IF}(x; \mu, P_{\mu, \Sigma})] e_N}{e_N^\top \Sigma^{-1} e_N} + \\ &+ \frac{(e_N^\top \Sigma^{-1} \text{IF}(x; \Sigma, P_{\mu, \Sigma}) \Sigma^{-1} e_N) (e_N^\top \Sigma^{-1} \mu - \lambda) e_N}{(e_N^\top \Sigma^{-1} e_N)^2} \} \end{aligned} \quad (27)$$

where $IF(x; \mu, P_{\mu, \Sigma})$ and $IF(x; \Sigma, P_{\mu, \Sigma})$ are those from Eqs (18) and (19). This formula is obtained by considering the statistical functional associated to the optimal portfolio weights,

$$p^*(P) = \frac{1}{\lambda} \Sigma^{-1}(P) \left[\mu(P) - \frac{e_N^\top \Sigma^{-1}(P) \mu(P) - \lambda}{e_N^\top \Sigma^{-1}(P) e_N} e_N \right]$$

where $\Sigma^{-1}(P)$ denotes the statistical functional corresponding to $\hat{\Sigma}^{-1}$, and then deriving the influence function, taking also into account that

$$IF(x; \Sigma^{-1}, P_{\mu, \Sigma}) = -\Sigma^{-1} IF(x; \Sigma, P_{\mu, \Sigma}) \Sigma^{-1}.$$

On the basis of the direct proportionality between the influence function $IF(x; p^*, P_{\mu, \Sigma})$ and the influence functions $IF(x; \mu, P_{\mu, \Sigma})$ and $IF(x; \Sigma, P_{\mu, \Sigma})$, we deduce that the global robustness of $\hat{\mu}$ and $\hat{\Sigma}$ is transferred to the plug-in estimator \hat{p}^* .

The consistency of the estimator of the optimal portfolio weights can be obtained using continuity arguments (see also [28]).

Note that, for the market parameters (μ, Σ) , Eq (1) is a convex optimization problem and possesses at least one optimal solution since the space of the feasible portfolios on which we optimize is compact. In addition, the solution is unique and is given by Eq (2). Since the problem (1) has a unique solution for each (μ, Σ) in a neighborhood of a given (μ_0, Σ_0) , for a fixed λ , the mapping $(\mu, \Sigma) \rightarrow p^* = p^*(\lambda, \mu, \Sigma)$ is continuous in a neighborhood of (μ_0, Σ_0) . Indeed, according to Theorem 4.2.1 from [29], the optimal set mapping is Hausdorff upper semicontinuous due to the compactness of the feasibility set and due to the continuity of the objective function. As the solution is unique, upper semicontinuity of the set-valued map yields continuity in the usual sense.

Then, for a given λ , $\hat{p}^* = \hat{p}^*(\lambda, \hat{\mu}, \hat{\Sigma})$ is a consistent estimator of the true optimal portfolio weights $p^* = p^*(\lambda, \mu, \Sigma)$, if $\hat{\mu}$, $\hat{\Sigma}$ are consistent estimators of μ and Σ . Indeed, the almost sure convergence and the convergence in probability remain valid after continuous transformations. The consistency of the estimators $\hat{\mu}$, $\hat{\Sigma}$ is implied by arguments from the theory of M-estimators, as developed by Huber (see [24] p.176). Using the uniqueness of the optimal solution, we have the continuity of the function $(\mu, \Sigma) \rightarrow p^*(\lambda, \mu, \Sigma)$, as noted above. Therefore, the consistency of the estimator of the optimal portfolio weights holds.

On the other hand, by using the multivariate Delta method, the asymptotic normality of \hat{p}^* is kept, as well. Given the i.i.d. observations X^1, \dots, X^T from $P_{\mu, \Sigma}$, since $\hat{\mu}$ and $\text{vecs}(\hat{\Sigma})$ are asymptotically normal and the function $h(\theta) = \frac{1}{\lambda} \Sigma^{-1}(\mu - \eta e_N)$ with $\theta = (\mu^\top, (\text{vecs} \Sigma)^\top)^\top$ is differentiable, by applying the multivariate Delta method, it holds $\sqrt{n}(\hat{p}^* - p^*) \rightarrow \mathcal{N}(0, V(p^*, P_{\mu, \Sigma}))$ where $V(p^*, P_{\mu, \Sigma}) = Dh(\theta) V(\theta, P_{\mu, \Sigma}) Dh(\theta)^\top$, $Dh(\theta)$ is the differential of h in θ and

$$V(\theta, P_{\mu, \Sigma}) = \begin{pmatrix} V(\mu, P_{\mu, \Sigma}) & 0 \\ 0 & V(\Sigma, P_{\mu, \Sigma}) \end{pmatrix}.$$

Monte Carlo simulations

We performed Monte Carlo simulations in order to assess the performance of the minimum pseudodistance estimators of the mean and covariance matrix, for both contaminated and non-contaminated data.

In a first study, we considered the multivariate normal distribution $\mathcal{N}_N(\mu_0, \Sigma_0)$, with $\mu_0 = \mathbf{0}$ and Σ_0 a $N \times N$ matrix with variances equal to 1 and covariances all equal to 0.2. We generated

samples of size T in which about $(1 - \varepsilon)T$ observations are from $\mathcal{N}_N(\mu_0, \Sigma_0)$, while a smaller portion εT is from the contaminating distribution $\mathcal{N}_N(\mu_c, \Sigma_c)$ with $\mu_c = -4$ and $\Sigma_c = 4\Sigma_0$. We considered $N \in \{2, 5, 10, 20, 50\}$ and $\varepsilon \in \{0, 0.05, 0.1, 0.2\}$. For each setting, we generated 1000 samples and for each sample we computed minimum pseudodistance estimates $\hat{\mu}$ and $\hat{\Sigma}$ corresponding to $\alpha \in \{0, 0.1, 0.2, 0.5, 0.75, 1\}$.

The estimates $\hat{\mu}$ and $\hat{\Sigma}$, which are solutions of the system of Eqs (6) and (7), were obtained using the following reweighting algorithm.

Let $s \in \{0, 1, \dots, s^*\}$ denotes the iteration step.

1. If $s = 0$

$\mu^{(s)}$ and $\Sigma^{(s)}$ are set to be initial estimates of location and scale;

2. For $0 < s < s^*$,

$$\begin{aligned}\mu^{(s)} &= \sum_{i=1}^T w_i^{(s-1)} X^i \\ \Sigma^{(s)} &= \sum_{i=1}^T (\alpha + 1) w_i^{(s-1)} (X^i - \mu^{(s)})(X^i - \mu^{(s)})^\top\end{aligned}$$

where

$$w_i^{(s)} = \frac{\exp\left(-\frac{\alpha}{2}(X^i - \mu^{(s)})^\top (\Sigma^{(s)})^{-1} (X^i - \mu^{(s)})\right)}{\sum_{i=1}^T \exp\left(-\frac{\alpha}{2}(X^i - \mu^{(s)})^\top (\Sigma^{(s)})^{-1} (X^i - \mu^{(s)})\right)}.$$

At step 1, we used maximum likelihood estimates as initial estimates of location and covariance. For details on general convergence behavior of reweighting algorithms we refer to [30]. If $\alpha > 0$, the above procedure associates low weights to the observations that disagree sensibly with the model. If $\alpha = 0$, all the observations receive the same weight and the estimators are the maximum likelihood ones

$$\begin{aligned}\hat{\mu}_{ML} &= \frac{1}{T} \sum_{i=1}^T X^i \\ \hat{\Sigma}_{ML} &= \frac{1}{T} \sum_{i=1}^T (X^i - \hat{\mu}_{ML})(X^i - \hat{\mu}_{ML})^\top.\end{aligned}$$

We present simulation based estimates of the mean square error given by

$$\widehat{\text{MSE}} = \frac{1}{n_s} \sum_{i=1}^{n_s} \|\hat{\theta}_i - \theta_0\|^2$$

where n_s is the number of samples (in our case $n_s = 1000$), $\theta_0 = (\mu_0^\top, \text{vech}(\Sigma_0)^\top)^\top$ and $\hat{\theta}_i = (\hat{\mu}_i^\top, \text{vech}(\hat{\Sigma}_i)^\top)^\top$ is an estimation corresponding to the sample i . Here $\text{vech}(\Sigma)$ is “the vector half”, namely the $n(n+1)/2$ -dimensional column vector obtained by stacking the columns of the lower triangle of Σ , including the diagonal, one below the other.

Tables 2 and 3 present simulation based estimates of the mean square error, when the sample size is $T = 10 \cdot N$, respectively when $T = 100 \cdot N$.

When there is no contamination, the MLE ($\alpha = 0$) performs the best, whatever the dimension N . On the other hand, the estimations obtained with the minimum pseudodistance estimators in this case are very close to those provided by the MLE, when α is not far from zero (for example $\alpha = 0.1$ and $\alpha = 0.2$). In the presence of contamination, the minimum

Table 2. Simulation based estimates of the mean square error, when $T = 10 \cdot N$.

$\varepsilon = 0\%$						
N	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
2	0.343	0.358	0.384	0.559	0.804	1.177
5	0.513	0.530	0.593	1.340	4.806	5.324
10	0.760	0.817	0.945	10.471	11.389	12.197
20	1.290	1.429	2.069	29.979	38.830	47.064
$\varepsilon = 5\%$						
N	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
2	4.425	0.888	0.533	0.593	0.849	1.202
5	18.816	1.077	0.662	1.565	4.985	5.437
10	41.312	0.951	1.022	10.646	11.470	12.600
20	145.172	1.517	2.273	30.339	39.561	47.072
$\varepsilon = 10\%$						
N	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
2	11.554	4.605	0.945	0.694	0.923	1.294
5	43.446	4.075	0.749	1.758	5.052	5.449
10	143.395	1.325	1.091	10.720	11.454	12.648
20	503.319	1.648	2.422	30.776	39.758	47.693
$\varepsilon = 20\%$						
N	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
2	32.696	24.612	9.955	1.118	1.190	1.475
5	132.542	53.841	1.869	2.171	5.362	5.625
10	441.209	19.233	1.241	10.751	11.751	12.745
20	1613.373	1.930	3.742	31.361	40.292	49.644

doi:10.1371/journal.pone.0140546.t002

pseudodistance estimators give much better results than the MLE, in all considered cases. In most cases, the choices $\alpha = 0.1$, $\alpha = 0.2$ provide the best results in terms of robustness. In the meantime, these choices correspond to estimation procedures with high asymptotic relative efficiency, according to the results from Table 1. When the contamination is more pronounced, i.e. $\varepsilon = 10\%$ or $\varepsilon = 20\%$, and the dimension N is low, i.e. $N = 2$, the choices $\alpha = 0.5$ and $\alpha = 0.75$ provide better robust estimates, but the asymptotic relative efficiencies of the corresponding estimation procedures are too low. Thus, values of α close to zero, such as $\alpha = 0.1$, $\alpha = 0.2$, represent choices that offer an equilibrium between robustness and efficiency. We notice that, even for highly multivariate data ($N = 50$ in our examples), when the number of observations is sufficiently high, the minimum pseudodistance estimators with $\alpha = 0.1$, $\alpha = 0.2$ give excellent results both in the case of clean data, as well as in case of contamination. The simulation results presented in Tables 2 and 3 show that increasing sample size leads to improved estimations.

Moreover, we present an example for highly correlated data. Here we considered the normal distribution $\mathcal{N}_N(\mu, \Sigma)$ with $N = 5$, $\mu = \mathbf{0}$, Σ the matrix with the diagonal terms 2 and the other entries 1. The contaminating distribution is $\mathcal{N}_N(\mu_c, \Sigma_c)$, where $\mu_c = \mathbf{4}$ and $\Sigma_c = \Sigma$, the contamination level being $\varepsilon \in \{0, 0.1, 0.2\}$. In this case, the mean square errors of the estimations are given in Table 4, the number of considered samples being again $n_s = 1000$. These results show that even in the case when the correlation level of the data is high and the data are contaminated, the minimum pseudodistance estimators give significantly better results than the MLE, the best results being obtained for $\alpha = 0.75$ and $\alpha = 1$.

Table 3. Simulation based estimates of the mean square error, when $T = 100 \cdot N$.

$\varepsilon = 0\%$						
N	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
2	0.035	0.035	0.039	0.051	0.067	0.084
5	0.050	0.052	0.057	0.087	0.135	0.204
10	0.075	0.081	0.093	0.185	0.395	8.703
20	0.129	0.142	0.181	0.910	28.127	31.790
50	0.287	0.359	0.685	171.329	219.263	262.867
$\varepsilon = 5\%$						
N	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
2	2.504	0.304	0.076	0.060	0.068	0.092
5	10.863	0.207	0.066	0.092	0.136	0.217
10	37.549	0.129	0.100	0.191	0.409	9.329
20	136.364	0.157	0.194	0.891	28.354	33.200
50	809.024	0.377	0.724	186.600	268.848	321.844
$\varepsilon = 10\%$						
N	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
2	8.910	2.493	0.285	0.066	0.073	0.096
5	39.015	2.203	0.089	0.098	0.142	0.240
10	133.655	0.404	0.107	0.207	0.474	9.746
20	493.386	0.182	0.203	1.249	28.467	33.925
50	2922.483	0.400	0.772	189.327	273.664	327.170
$\varepsilon = 20\%$						
N	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
2	28.470	18.606	7.524	0.106	0.102	0.115
5	124.702	44.374	0.327	0.113	0.168	0.272
10	429.087	17.016	0.128	0.232	0.592	10.423
20	1576.600	0.335	0.229	3.460	29.091	34.815
50	9334.156	0.453	0.887	190.613	282.404	348.252

doi:10.1371/journal.pone.0140546.t003

Table 4. Simulation based estimates of the mean square error, for highly correlated data, $T = 100 \cdot N$.

	$\alpha = 0$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1$
$\varepsilon = 0\%$	0.1999	0.2033	0.2281	0.3311	0.5232	0.8612
$\varepsilon = 10\%$	32.3561	23.0288	13.9551	1.7271	0.9666	1.1272
$\varepsilon = 20\%$	101.7404	97.5348	88.9003	34.6433	5.3518	2.0940

doi:10.1371/journal.pone.0140546.t004

Application for financial data

In-sample analysis

We analyze 172 monthly log-returns of 8 MSCI Indexes (France, Germany, Italy, Japan, Pacific Ex JP, Spain, United Kingdom and USA) from January 1998 to April 2012 with the aim to construct robust and efficient portfolios. The data are provided by MSCI (see [31]). Boxplots for these data are presented in Fig 2.

For these indexes, estimates of the expected return and of the variance are represented in Fig 3. Note that the estimates of the expected returns obtained with the minimum pseudodistance estimators are larger than the maximum likelihood ones. In the meantime, the minimum pseudodistance estimates of the variances are smaller than those provided by the MLE.

Boxplots

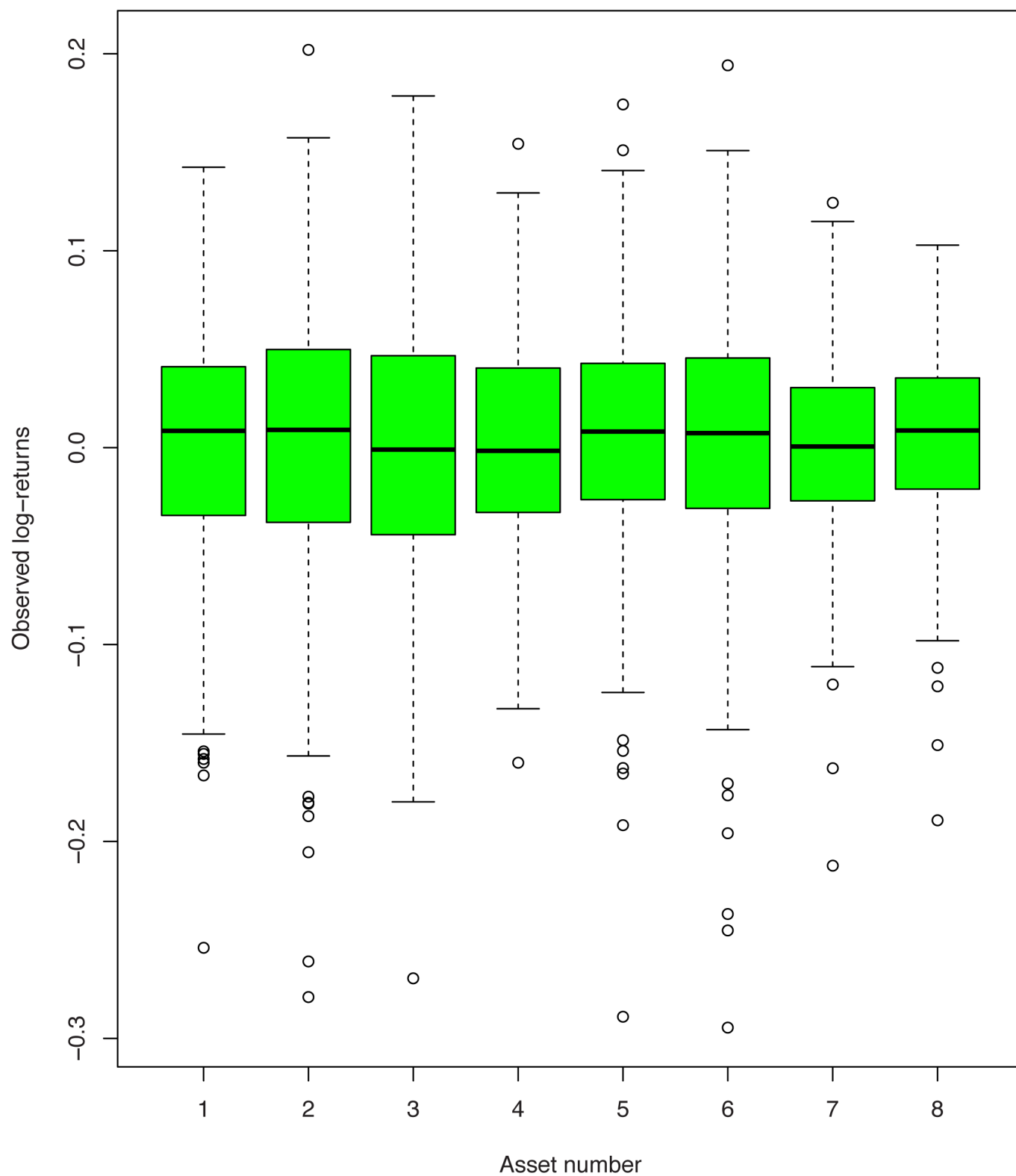


Fig 2. Boxplots for monthly log-returns corresponding to the MSCI Indexes. (1: France, 2: Germany, 3: Italy, 4: Japan, 5: Pacific ex JP, 6: Spain, 7: United Kingdom, 8: USA.)

doi:10.1371/journal.pone.0140546.g002

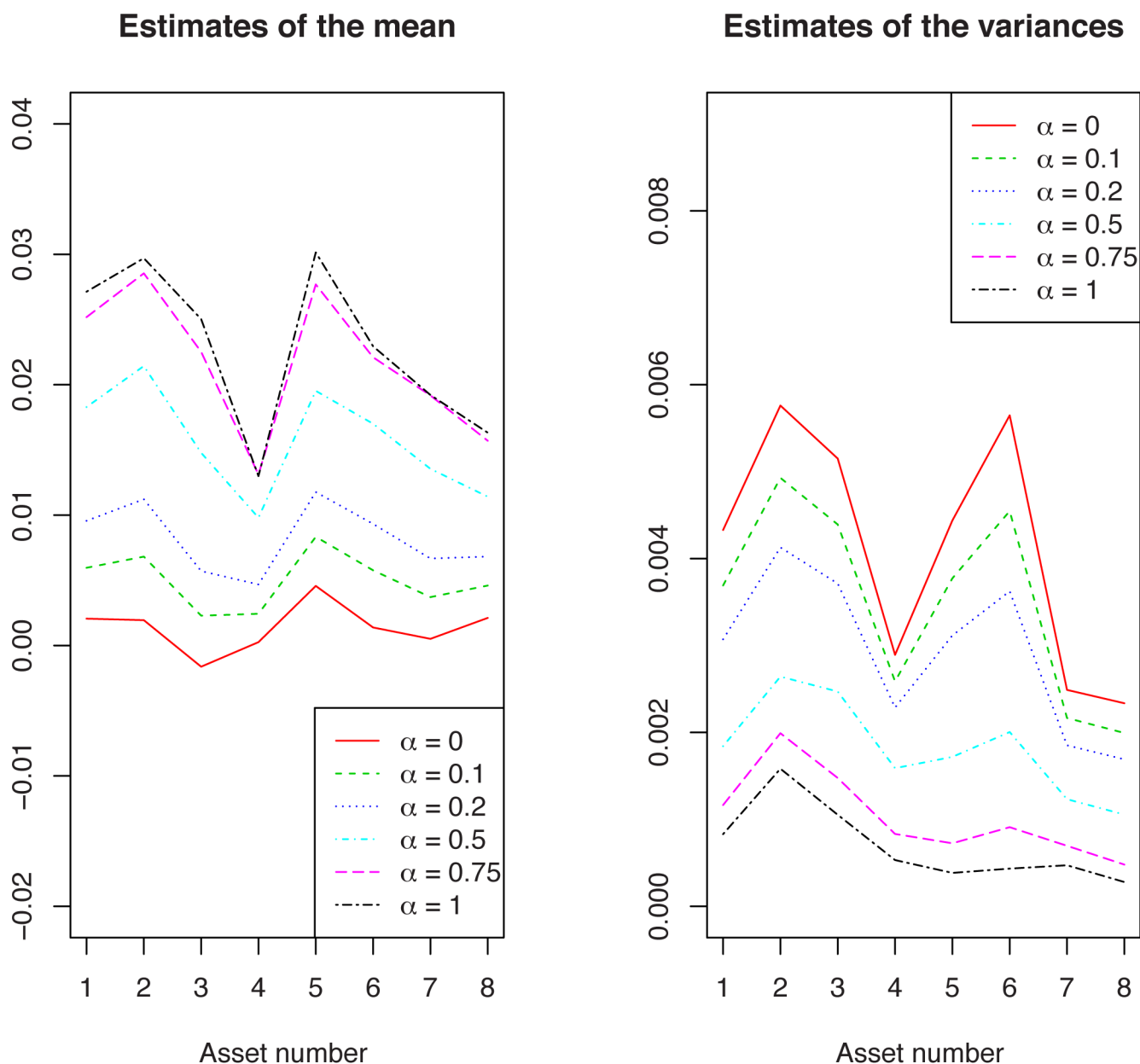


Fig 3. Expected returns estimates (left) and variance estimates (right) for the 8 MSCI Indexes. (1: France, 2: Germany, 3: Italy, 4: Japan, 5: Pacific ex JP, 6: Spain, 7: United Kingdom, 8: USA.)

doi:10.1371/journal.pone.0140546.g003

Estimates of the mean vector and of the covariance matrix computed with different minimum pseudodistance estimators are used to determine efficient frontiers. In Fig 4 we plot efficient frontiers for the case “short selling not allowed”. Similar results can be obtained in the case “short selling allowed”. In both cases, the frontiers based on the minimum pseudodistance estimations dominate those based on the classical maximum likelihood estimations, yielding portfolios with larger expected returns and smaller risks. Thus, the robust estimates reduce the volatility effects which typically affects the results of the traditional approaches.

The next step in our analysis is to identify the influential observations which are responsible for the shift of the efficient frontier. We perform this study in the case “short selling”. In this sense, we use the data influence measure (DIM) as diagnostic tool (see [3]). This is defined as the Euclidian norm of the influence function of the estimator of weights based on maximum likelihood estimators of μ and Σ . More precisely,

$$\text{DIM}(x, p^*) = [\text{IF}(x; p^*, P_{\mu, \Sigma})^\top \text{IF}(x; p^*, P_{\mu, \Sigma})]^{1/2}$$

where $\text{IF}(x; p^*, P_{\mu, \Sigma})$ is given by Eq (27) with $\text{IF}(x; \mu, P_{\mu, \Sigma})$ and $\text{IF}(x; \Sigma, P_{\mu, \Sigma})$ given by the formulas (18) and (19) in the case $\alpha = 0$. In order to compute DIM, the true parameters values μ , Σ , p^* have to be known. In practice, these parameters should be estimated in a robust way, such that DIM is not affected by the outlying observations it is supposed to detect.

In Fig 5 (left hand side) we represent the influence of each of the 172 observations on the estimator of the optimal portfolio weights based on maximum likelihood estimators of μ and Σ . Since DIM is related to a specific portfolio on the efficient frontier, we made a choice, namely the level of the portfolio variance has been set to 0.005. This choice corresponds to $\lambda = 3.85$. The necessary robust estimates of μ , Σ , p^* have been obtained with minimum pseudodistance estimators corresponding to $\alpha = 0.2$. The most influential observations as detected by DIM correspond to negative economic events associated with known financial crisis periods: 1998 Russian financial crisis (August 1998), “dot-com crash” of 2000–2002 and 2007–2012 global financial crisis. On the other hand, the influence of these observations is substantially reduced when using robust procedures. This fact can be seen in the right hand side of Fig 5 where we represent the influence of each of the 172 observations on the robust estimator of the optimal portfolio weights based on the minimum pseudodistance estimators of μ and Σ corresponding to $\alpha = 0.2$. Reducing the influence of outlying observations leads to optimal portfolios with higher returns and smaller variances.

Out-of-sample evaluation

In this subsection, we illustrate the out-of-sample stability of the proposed portfolios. For this purpose we use the same empirical data set as in the preceding subsection. The evaluated portfolios are: optimal mean-variance portfolios using the classical MLE or minimum pseudodistance estimators (MPE) or S-estimators, minimum-variance portfolios using the classical MLE or minimum pseudodistance estimators or S-estimators, as well as the equally-weighted portfolio. The S-estimators are well known robust estimators of multivariate location and covariance, allowing the flexibility of choosing the breakdown point ε^* , which is the amount of data deviating from the reference model that an estimator can accept while giving meaningful information (see for example [32]). We considered in our study S-estimators based on the Tukey’s biweight function with breakdown point $\varepsilon^* = 0.25$ or $\varepsilon^* = 0.5$. For the optimal mean-variance portfolios we used $\lambda = 3.85$, as in the preceding section, but similar results can be obtained for other finite values of λ , too. The minimum-variance portfolios correspond to $\lambda = \infty$, as it is known. We consider both cases, when short selling is allowed or not.

Short selling not allowed

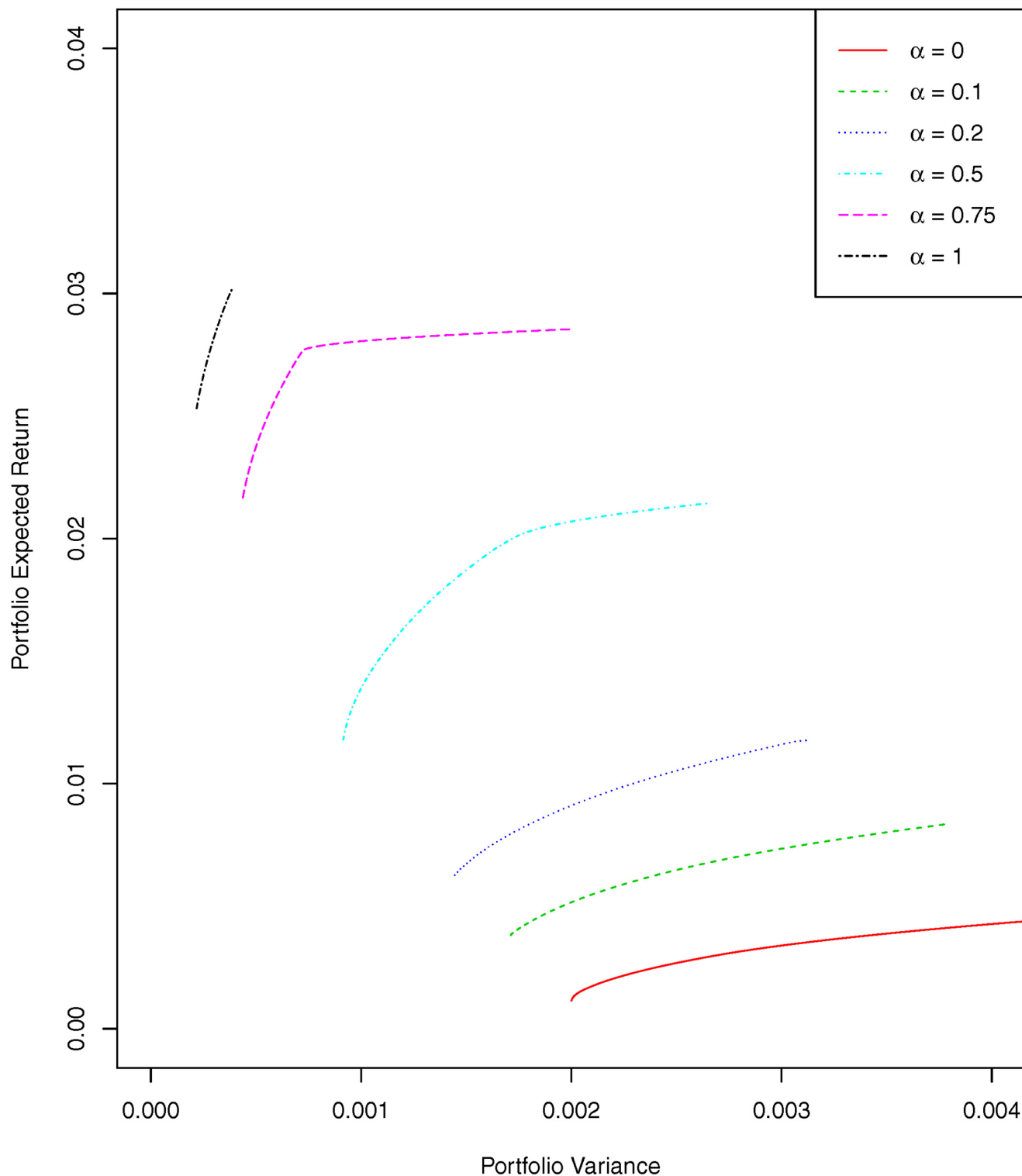


Fig 4. Mean-variance efficient frontiers.

doi:10.1371/journal.pone.0140546.g004

Table 5. Turnover, out-of-sample portfolio variance and Sharpe ratio for optimal mean-variance portfolios (Mean-var) and minimum-variance portfolios (Min-var), corresponding to different estimators of mean and covariance, in the case “short selling allowed”.

Estimators	Mean-var			Min-var		
	Turnover	$(\hat{\sigma}^k)^2$	\widehat{SR}^k	Turnover	$(\hat{\sigma}^k)^2$	\widehat{SR}^k
MLE	1.5665	0.0105	0.0368	0.2507	0.0021	-0.0378
MPE, $\alpha = 0.1$	1.4717	0.0161	0.0727	0.2109	0.0020	-0.0282
MPE, $\alpha = 0.2$	1.9083	0.0292	0.0880	0.1838	0.0019	-0.0322
MPE, $\alpha = 0.25$	2.4381	0.0400	0.0911	0.1753	0.0019	-0.0367
SE, $\epsilon^* = 0.25$	1.4119	0.0126	0.0631	0.2283	0.0020	-0.0291
SE, $\epsilon^* = 0.5$	1.5506	0.0218	0.0878	0.1966	0.0019	-0.0302

doi:10.1371/journal.pone.0140546.t005

The weights $p_{j,t+}^k$ are computed using the formula

$$p_{j,t+}^k = \frac{1 + X_j^{t+1}}{1 + (X^{t+1})^\top p_t^k} \cdot p_{j,t}^k$$

X^{t+1} representing the data at the time $t + 1$. The out-of-sample return at time $t + 1$, corresponding to the strategy k , is defined as $(p_t^k)^\top X^{t+1}$. For each strategy k , using these out-of-sample returns, the out-of-sample variance is defined by

$$(\hat{\sigma}^k)^2 = \frac{1}{T - \tau - 1} \sum_{t=\tau}^{T-1} ((p_t^k)^\top X^{t+1} - \hat{\mu}^k)^2 \quad \text{with} \quad \hat{\mu}^k = \frac{1}{T - \tau} \sum_{t=\tau}^{T-1} (p_t^k)^\top X^{t+1}$$

and the Sharpe ratio is defined by $\widehat{SR}^k = \frac{\hat{\mu}^k}{\hat{\sigma}^k}$.

Table 6. Turnover, out-of-sample portfolio variance and Sharpe ratio for optimal mean-variance portfolios (Mean-var) and minimum-variance portfolios (Min-var), corresponding to different estimators of mean and covariance, in the case “short selling not allowed”.

Estimators	Mean-var			Min-var		
	Turnover	$(\hat{\sigma}^k)^2$	\widehat{SR}^k	Turnover	$(\hat{\sigma}^k)^2$	\widehat{SR}^k
MLE	0.2068	0.0054	0.0032	0.0519	0.0025	-0.0304
MPE, $\alpha = 0.1$	0.1510	0.0060	0.0186	0.0405	0.0025	-0.0264
MPE, $\alpha = 0.2$	0.1475	0.0059	0.0170	0.0370	0.0025	-0.0220
MPE, $\alpha = 0.25$	0.1562	0.0059	0.0138	0.0364	0.0026	-0.0199
MPE, $\alpha = 0.3$	0.1594	0.0059	0.0129	0.0364	0.0026	-0.0184
MPE, $\alpha = 0.5$	0.1084	0.0058	0.0254	0.0363	0.0026	-0.0148
MPE, $\alpha = 0.75$	0.0235	0.0060	0.0319	0.0511	0.0028	-0.0050
MPE, $\alpha = 1$	0.0143	0.0060	0.0345	0.1219	0.0028	-0.0259
SE, $\epsilon^* = 0.25$	0.1771	0.0058	0.0125	0.0435	0.0025	-0.0273
SE, $\epsilon^* = 0.5$	0.1320	0.0059	0.0173	0.0384	0.0026	-0.0224

doi:10.1371/journal.pone.0140546.t006

In Tables 5 and 6 we report the turnover, the out-of-sample variance and the Sharpe ratio for optimal mean-variance portfolios (with $\lambda = 3.85$) and for minimum-variance portfolios in the case “short selling allowed”, respectively in the case “short selling not allowed”. For estimation of the mean and covariance we considered the classical MLE, minimum pseudodistance estimators for several values of α , as well as S-estimators. We note that the portfolios that minimize the variance have better turnover than the portfolios that optimize the trade-off between mean and variance, confirming the fact that the policies that ignore estimates of the expected returns lead to better stability results in the out-of-sample analysis. The robust portfolios using minimum pseudodistance estimators attain higher Sharpe ratios and lower turnover than the traditional portfolios using MLE, the influence of asset returns that deviate from normality being more reduced.

When short selling is allowed, the mean-variance portfolios with the best turnover are those corresponding to the S-estimators with $\varepsilon^* = 0.25$, respectively to the minimum pseudodistance estimators with $\alpha = 0.1$, the Sharpe ratios of these portfolios being very close. When short selling is not allowed, the best mean-variance portfolios in terms of turnover and Sharpe ratio are those corresponding to the minimum pseudodistance estimators with $\alpha = 0.75$ and $\alpha = 1$. In the case of minimum-variance portfolios, the best turnover is obtained for minimum pseudodistance estimators ($\alpha = 0.2, 0.25, 0.3, 0.5$) and the best Sharpe ratio is obtained for the minimum pseudodistance estimators with $\alpha = 0.75$, both in the case “short selling not allowed”.

We also compare our results with those corresponding to the equally-weighted portfolio which is known to have a good out-of-sample behavior, even in the case of contaminated data. For our data set and this portfolio, we obtained: Turnover = 0.02234, $(\hat{\sigma}_k)^2 = 0.0025$ and $\hat{SR}^k = -0.0272$. Note that, among all the considered portfolios, the equally-weighted portfolio has the smallest turnover. In the meantime, all the minimum-variance portfolios using minimum pseudodistance estimators, in the case “short selling not allowed”, have better Sharpe ratio than the equally-weighted portfolio. For example, the minimum-variance portfolio using minimum pseudodistance estimators with $\alpha = 0.5$ represents a good choice in terms of turnover and Sharpe ratio.

The boxplots of portfolio weights give a graphical representation of the stability of the different portfolio policies. By applying the “rolling-horizon” procedure, we obtain $T - \tau$ portfolio weight vectors for each strategy. In Figs 6 and 7, each boxplot represent the variability of the weight assigned to a particular asset in a minimum-variance portfolio. As it can be seen, the robust minimum-variance portfolios based on minimum pseudodistance estimators are characterized by a better out-of-sample stability than the classical minimum-variance portfolios. We choose for these examples $\alpha = 0.25$, but we obtained similar results for other values of α , too. Also, in this example, the weights of the minimum-variance portfolio based on minimum pseudodistance estimators are more stable than the weights of the minimum-variance portfolio based on S-estimators (although the out-of-sample behaviors of the portfolios based on these two type of estimators are very close).

Our theoretical and numerical results show that the optimal portfolios based on minimum pseudodistance estimators are much more stable to extreme events than those obtained by plugging-in the MLEs and compare well with other optimal robust portfolios. When α is not far from 0, the minimum pseudodistance estimators of μ and Σ combine robustness with high efficiency and these qualities are transferred to the portfolio weights estimator. The numerical results based on simulations or real data show that $\alpha = 0.2, 0.25$ represent good choices in terms of robustness and efficiency, but also higher values of α lead to good results in some situations. All these aspects recommend the new procedure as a viable alternative to existing robust portfolio selection methods.

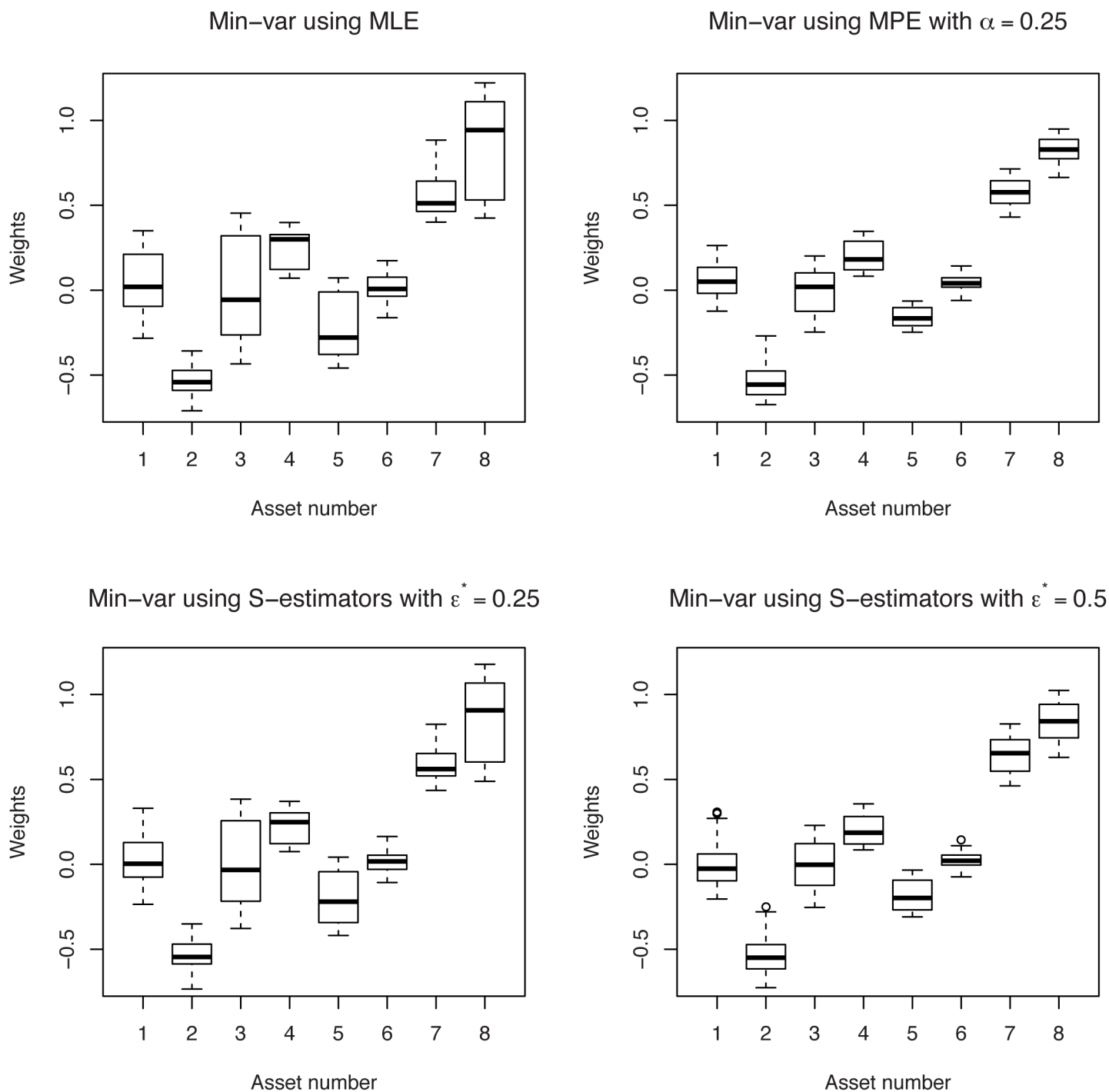


Fig 6. Boxplots of the weights corresponding to minimum-variance portfolios. Different estimators of the covariance matrix are used, in the case “short selling allowed”.

doi:10.1371/journal.pone.0140546.g006

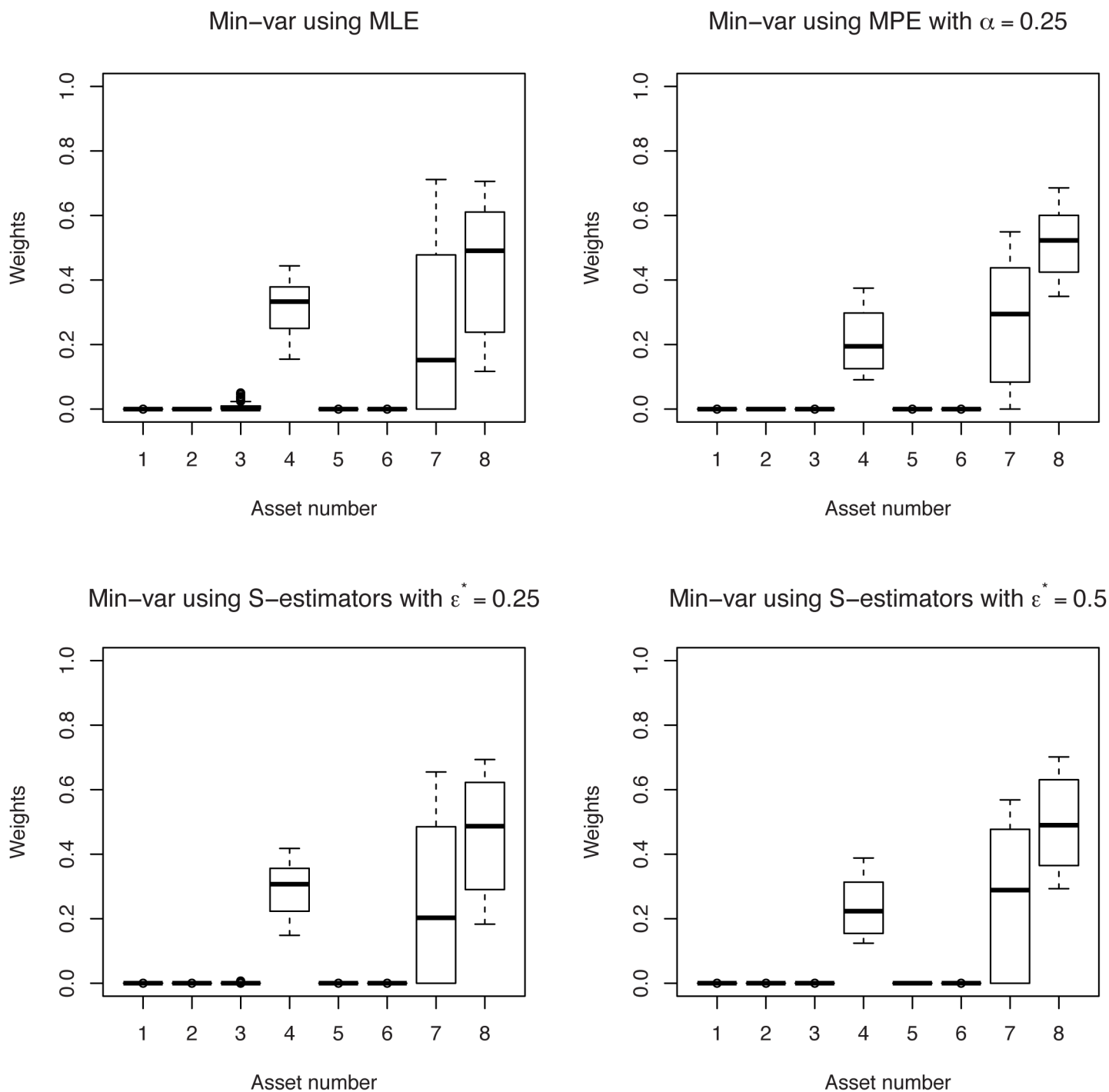


Fig 7. Boxplots of the weights corresponding to minimum-variance portfolios. Different estimators of the covariance matrix are used, in the case “short selling not allowed”.

doi:10.1371/journal.pone.0140546.g007

Supporting Information

S1 Appendix. Appendix. Codes of the Figures.
(DOC)

Author Contributions

Conceived and designed the experiments: AT. Performed the experiments: AT SLA. Analyzed the data: AT SLA. Contributed reagents/materials/analysis tools: AT SLA. Wrote the paper: AT SLA.

References

1. Markowitz HM. Mean-variance analysis in portfolio choice and capital markets. *J Finance*. 1952; 7: 77–91.
2. Michaud R. The Markowitz optimization enigma: is optimized optimal? *Financ Anal J*. 1989; 45: 31–42. doi: [10.2469/faj.v45.n1.31](https://doi.org/10.2469/faj.v45.n1.31)
3. Perret-Gentil C, Victoria-Feser MP. Robust mean-variance portfolio selection; 2005. Preprint. Available: FAME Research Paper, no. 140. papers.ssrn.com/sol3/papers.cfm?abstract_id=721509
4. Fabozzi FJ, Huang D, Zhou G. Robust portfolios: contributions from operations research and finance. *Ann Oper Res*. 2010; 176: 191–220. doi: [10.1007/s10479-009-0515-6](https://doi.org/10.1007/s10479-009-0515-6)
5. Vaz-de Melo B, Camara RP. Robust multivariate modeling in finance. *Int J Manag Finance*. 2005; 4: 12–23.
6. Welsch RE, Zhou X. Application of robust statistics to asset allocation models. *Revstat Stat J*. 2007; 5: 97–114.
7. DeMiguel V, Nogales FJ. Portfolio selection with robust estimation. *Oper Res*. 2009; 57: 560–577. doi: [10.1287/opre.1080.0566](https://doi.org/10.1287/opre.1080.0566)
8. Ferrari D, Paterlini S. Efficient and robust estimation for financial returns: an approach based on q -entropy; 2010. Preprint. Available: *Materiali di discussione*, no. 623, Università degli Studi di Modena e Reggio Emilia. www.dep.unimore.it/materiali_discussione/0623.pdf
9. Basu A, Harris IR, Hjort NL, Jones MC. Robust and efficient estimation by minimizing a density power divergence. *Biometrika*. 1998; 85: 549–559. doi: [10.1093/biomet/85.3.549](https://doi.org/10.1093/biomet/85.3.549)
10. Jones MC, Hjort NL, Harris IR, Basu A. A comparison of related density-based minimum divergence estimators. *Biometrika*. 2001; 88: 865–873. doi: [10.1093/biomet/88.3.865](https://doi.org/10.1093/biomet/88.3.865)
11. Broniatowski M, Keziou A. Parametric estimation and tests through divergences and the duality technique. *J Multivar Anal*. 2009; 100: 16–36. doi: [10.1016/j.jmva.2008.03.011](https://doi.org/10.1016/j.jmva.2008.03.011)
12. Toma A, Leoni-Aubin S. Robust tests based on dual divergence estimators and saddlepoint approximations. *J Multivar Anal*. 2010; 101: 1143–1155. doi: [10.1016/j.jmva.2009.11.001](https://doi.org/10.1016/j.jmva.2009.11.001)
13. Toma A, Broniatowski M. Dual divergence estimators and tests: Robustness results. *J Multivar Anal*. 2011; 102: 20–36. doi: [10.1016/j.jmva.2010.07.010](https://doi.org/10.1016/j.jmva.2010.07.010)
14. Fujisawa H, Eguchi S. Robust parameter estimation with a small bias against heavy contamination. *J Multivar Anal*. 2008; 99: 2053–2081. doi: [10.1016/j.jmva.2008.02.004](https://doi.org/10.1016/j.jmva.2008.02.004)
15. Basu A, Shioya H, Park C. Statistical inference: the minimum pseudodistance approach. Florida: CRC Press, Boca Raton; 2011.
16. Broniatowski M, Toma A, Vajda I. Decomposable pseudodistances and applications in statistical estimation. *J Stat Plan Inference*. 2012; 142: 2574–2585. doi: [10.1016/j.jspi.2012.03.019](https://doi.org/10.1016/j.jspi.2012.03.019)
17. Toma A, Leoni-Aubin S. Portfolio selection using minimum pseudodistance estimators. *Econ Comput Econ Cybern Stud Res*. 2013; 47: 97–112.
18. Pardo L. Statistical inference based on divergence measures. CRC Press, Taylor & Francis; 2005.
19. Vonta F, Karagrigoriou A. Generalized measures of divergence in survival analysis and reliability. *J Appl Probab*. 2010; 47: 216–234. doi: [10.1239/jap/1269610827](https://doi.org/10.1239/jap/1269610827)
20. Preda V, Dedu S, Sheraz M. New measure selection for Hunt-Devolder semi-Markov regime switching interest rate models. *Physica A*. 2014; 407: 350–359. doi: [10.1016/j.physa.2014.04.011](https://doi.org/10.1016/j.physa.2014.04.011)
21. Preda V, Dedu S, Gheorghe C. New classes of Lorenz curves by maximizing Tsallis entropy under mean and Gini equality and inequality constraints. *Physica A*. 2015; 436: 925–932. doi: [10.1016/j.physa.2015.05.092](https://doi.org/10.1016/j.physa.2015.05.092)

22. Toma A, Leoni-Aubin S. Optimal robust M-estimators using Renyi pseudodistances. *J Multivar Anal.* 2013; 115: 359–373. doi: [10.1016/j.jmva.2012.10.003](https://doi.org/10.1016/j.jmva.2012.10.003)
23. Hampel FR, Ronchetti E, Rousseeuw PJ, Stahel W. Robust statistics: the approach based on influence functions. New York: Wiley; 1986.
24. Huber PJ. Robust covariances. In: Gupta S, Moore D, editors. Statistical decision theory and related topics II. New York: Academic Press; 1977. pp. 165–191.
25. Jaupi L, Saporta G. Using the influence function in robust principal component analysis. In: Morgenthaler S, Ronchetti E, Stahel WA, editors. New directions in statistical data analysis and robustness. Verlag: Birkhäuser; 1993. pp. 147–156.
26. Gervini D. The influence function of the Stahel-Donoho estimator of multivariate location and scatter. *Stat Probab Lett.* 2002; 60: 425–435. doi: [10.1016/S0167-7152\(02\)00315-2](https://doi.org/10.1016/S0167-7152(02)00315-2)
27. Serfling R. Asymptotic relative efficiency in estimation. In: Lovric M, editor. International encyclopedia of statistical sciences. New-York: Springer; 2011. pp. 68–72.
28. Schöttle K, Werner R. Consistency of robust portfolio estimators. *OR and Management Sciences.* 2006; pp. 1–23.
29. Bank B, Guddat J, Klatte D, Kummer B, Tammer K. Non-linear parametric optimization. Basel: Birkhäuser; 1983.
30. Arslan O. Convergence behavior of an iterative reweighting algorithm to compute multivariate M-estimates for location and scatter. *J Stat Plan Inference.* 2004; 118: 115–128. doi: [10.1016/S0378-3758\(02\)00402-0](https://doi.org/10.1016/S0378-3758(02)00402-0)
31. www.msci.com
32. Davies PL. Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices. *Ann Stat.* 1987; 15: 1269–1292. doi: [10.1214/aos/1176350505](https://doi.org/10.1214/aos/1176350505)