Robust Estimation for the Single Index Model Using Pseudodistances

Aida Toma [1,2,*] and Cristinca Fulga [1]

[1] Department of Applied Mathematics, Bucharest Academy of Economic Studies, 010374 Bucharest, Romania; cristinca.fulga@gmail.com
[2] "Gh. Mihoc - C. Iacob" Institute of Mathematical Satistics and Applied Mathematics, Romanian Academy, 010071 Bucharest, Romania
[*] Correspondence: aida_toma@yahoo.com

Abstract: For portfolios with a large number of assets, the single index model allows for expressing the large number of covariances between individual asset returns through a significantly smaller number of parameters. This avoids the constraint of having very large samples to estimate the mean and the covariance matrix of the asset returns, which practically would be unrealistic given the dynamic of market conditions. The traditional way to estimate the regression parameters in the single index model is the maximum likelihood method. Although the maximum likelihood estimators have desirable theoretical properties when the model is exactly satisfied, they may give completely erroneous results when outliers are present in the data set. In this paper, we define minimum pseudodistance estimators for the parameters of the single index model and using them we construct new robust optimal portfolios. We prove theoretical properties of the estimators, such as consistency, asymptotic normality, equivariance, robustness, and illustrate the benefits of the new portfolio optimization method for real financial data.

Keywords: minimum divergence methods; robustness; single index model

1. Introduction

The problem of portfolio optimization in the mean-variance approach depends on a large number of parameters that need to be estimated on the basis of relatively small samples. Due to the dynamics of market conditions, only a short period of market history can be used for estimation of the model's parameters. In order to reduce the number of parameters that need to be estimated, the single index model proposed by Sharpe (see [1,2]) can be used. The traditional estimators for parameters of the single index model are based on the maximum likelihood method. These estimators have optimal properties for normally distributed variables, but they may give completely erroneous results in the presence of outlying observations. Since the presence of outliers in financial asset returns is a frequently occurring phenomenon, robust estimates for the parameters of the single index model are necessary in order to provide robust and optimal portfolios.

Our contribution to robust portfolio optimization through the single index model is based on using minimum pseudodistance estimators.

The interest on statistical methods based on information measures and particularly on divergences has grown substantially in recent years. It is a known fact that, for a wide variety of models, statistical methods based on divergence measures have some optimal properties in relation to efficiency, but especially in relation to robustness, representing viable alternatives to the classical methods. We refer to the monographs of Pardo [3] and Basu et al. [4] for an excellent presentation of such methods, for their importance and applications.

We can say that the minimum pseudodistance methods for estimation go to the same category as the minimum divergence methods. The minimum divergence estimators are defined by minimizing some appropriate divergence between the assumed theoretical model and the true model corresponding to the data. Depending on the choice of the divergence, minimum divergence estimators can afford considerable robustness with a minimal loss of efficiency. The classical minimum divergence methods require nonparametric density estimation, which imply some difficulties such as the bandwidth selection. In order to avoid the nonparametric density estimation in minimum divergence estimation methods, some proposals have been made in [5–7] and robustness properties of such estimators have been studied in [8,9].

The pseudodistances that we use in the present paper were originally introduced in [6], where they are called "type-0" divergences, and corresponding minimum divergence estimators have been studied. They are also obtained (using a cross entropy argument) and extensively studied in [10] where they are called γ-divergences. They are also introduced in [11] in the context of decomposable pseudodistances. By its very definition, a pseudodistance satisfies two properties, namely the nonnegativity and the fact that the pseudodistance between two probability measures equals to zero if and only if the two measures are equal. The divergences are moreover characterized by the information processing property, i.e., by the complete invariance with respect to statistically sufficient transformations of the observation space (see [11], p. 617). In general, a pseudodistance may not satisfy this property. We adopted the term pseudodistance for this reason, but in the literature we can also meet the other terms above. The minimum pseudodistance estimators for general parametric models have been presented in [12] and consist of minimization of an empirical version of a pseudodistance between the assumed theoretical 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 this paper, we define minimum pseudodistance estimators for the parameters of the single index model and using them we construct new robust optimal portfolios. We study properties of the estimators, such as, consistency, asymptotic normality, robustness and equivariance and illustrate the benefits of the proposed portfolio optimization method through examples for real financial data.

We mention that we define minimum pseudodistance estimators, and prove corresponding theoretical properties, for the parameters of the simple linear regression model (35), associated with the single index model. However, in a very similar way, we can define minimum pseudodistance estimators and obtain the same theoretical results for the more general linear regression model $Y_j = X_j^T \beta + e_j, j = 1, \ldots, n$, where the errors e_j are i.i.d. normal variables with mean zero and variance σ^2, $X_j = (X_{j1}, \ldots, X_{jp})^T$ is the vector of independent variables corresponding to the j-th observation and $\beta = (\beta_1, \ldots, \beta_p)^T$ represents the regression coefficients.

The rest of the paper is organized as follows. In Section 2, we present the problem of robust estimation for some portfolio optimization models. In Section 3, we present the proposed approach. We define minimum pseudodistance estimators for regression parameters corresponding to the single index model and obtain corresponding estimating equations. Some asymptotic properties and equivariance properties of these estimators are studied. The robustness issue for estimators is considered through the influence function analysis. Using minimum pseudodistance estimators, new optimal portfolios are defined. Section 4 presents numerical results illustrating the performance of the proposed methodology. Finally, the proofs of the theorems are provided in the Appendix A.

2. The Single Index Model

Portfolio selection represents the problem of allocating a given capital over a number of available assets in order to maximize the return of the investment while minimizing the risk. We consider a portfolio formed by a collection of N assets. The returns of the assets are given by the random vector $X := (X_1, \ldots, X_N)^T$. Usually, it is supposed that X follows a multivariate normal distribution $\mathcal{N}_N(\mu, \Sigma)$, with μ being the vector containing the mean returns of the assets and $\Sigma = (\sigma_{ij})$ the

covariance matrix of the assets returns. Let $w := (w_1, \ldots, w_N)^T$ be the vector of weights associated with the portfolio, where w_i is the proportion of capital invested in the asset i. Then, the total return of the portfolio is defined by the random variable

$$w^T X = w_1 X_1 + \cdots + w_N X_N. \tag{1}$$

The mean and the variance of the portfolio return are given by

$$R(w) \quad := \quad w^T \mu, \tag{2}$$
$$S(w) \quad := \quad w^T \Sigma w. \tag{3}$$

A classical approach for portfolio selection is the mean-variance optimization introduced by Markowitz [13]. For a given investor's risk aversion $\lambda > 0$, the mean-variance optimization gives the optimal portfolio w^*, solution of the problem

$$\arg \max_w \{ R(w) - \frac{\lambda}{2} S(w) \}, \tag{4}$$

with the constraint $w^T e_N = 1$, e_N being the N-dimensional vector of ones. The solution of the optimization problem (4) is explicit, the optimal portfolio weights for a given value of λ being

$$w^* = \frac{1}{\lambda} \Sigma^{-1} (\mu - \eta e_N), \tag{5}$$

where

$$\eta = \frac{e_N^T \Sigma^{-1} \mu - \lambda}{e_N^T \Sigma^{-1} e_N}. \tag{6}$$

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 w_i are positive.

Another classical approach for portfolio selection is to minimize the portfolio risk defined by the portfolio variance, under given constraints. This means determining the optimal portfolio w^* as a solution of the optimization problem

$$\arg \min_w S(w), \tag{7}$$

subject to $R(w) = w^T \mu \geq \mu_0$, for a given value μ_0 of the portfolio return.

However, the mean-variance analysis has been criticized for being sensitive to estimation errors of the mean and the covariance of the assets returns. For both optimization problems above, estimations of the input parameters μ and Σ are necessary. The quality and hence the usefulness of the results of the portfolio optimization problem critically depend on the quality of the statistical estimates for these input parameters. The mean vector and the covariance matrix of the returns are in practice estimated by the maximum likelihood estimators under the multivariate normal assumption. When the model is exactly satisfied, the maximum likelihood estimators have optimal properties, being the most efficient. On the other hand, in the presence of outlying observations, these estimators may give completely erroneous results and consequently the weights of the corresponding optimal portfolio may be completely misleading. It is a known fact that outliers frequently occur in asset returns, where an outlier is defined to be an unusually large value well separated from the bulk of the returns. Therefore, robust alternatives to the classical approaches need to be carefully analyzed.

For an overview on the robust methods for portfolio optimization, using robust estimators of the mean and covariance matrix in the Markowitz's model, we refer to [14]. We also cite the methods proposed by Vaz-de Melo and Camara [15], Perret-Gentil and Victoria-Feser [16], Welsch and Zhou [17], DeMiguel and Nogales [18], and Toma and Leoni-Aubin [19].

On the other hand, in portfolio analysis, one is sometimes faced with two conflicting demands. Good quality statistical estimates require a large sample size. When estimating the covariance matrix, the sample size must be larger than the number of different elements of the matrix. For example, for a portfolio involving 100 securities, this would mean observations from 5050 trading days, which is about 20 years. From a practical point of view, considering such large samples is not adequate for the considered problem. Since the market conditions change rapidly, very old observations would lead to irrelevant estimates for the current or future market conditions. In addition, in some situations, the number of assets could even be much larger than the sample size of exploitable historical data. Therefore, estimating the covariance matrix of asset returns is challenging due to the high dimensionality and also to the heavy-tailedness of asset return data. It is a known fact that extreme events are typical in financial asset prices, leading to heavy-tailed asset returns. One way to treat these problems is to use the single index model.

The single index model (see [1]) allows us to express the large number of covariances between the returns of the individual assets through a significantly smaller number of parameters. This is possible under the hypothesis that the correlation between two assets is strictly given by their dependence on a common market index. The return of each asset i is expressed under the form

$$X_i = \alpha_i + \beta_i X_M + e_i, \tag{8}$$

where X_M is the random variable representing the return of the market index, e_i are zero mean random variables representing error terms and α_i, β_i are new parameters to be estimated. It is supposed that the e_i's are independent and also that the e_is are independent of x_M. Thus, $E(e_i) = 0$, $E(e_i e_j) = 0$ and $E(e_i x_M) = 0$ for all i and all $j \neq i$.

The intercept in Equation (35) represents the asset's expected return when the market index return is zero. The slope coefficient β_i represents the asset's sensitivity to the index, namely the impact of a unit change in the return of the index. The error e_i is the return variation that cannot be explained by the index.

The following notations are also used:

$$\sigma_i^2 \quad := \quad Var(e_i), \quad \mu_M := E(X_M), \quad \sigma_M^2 := Var(X_M).$$

Using Equation (35), the components of the parameters μ and Σ from the models (4) and (7) are given by

$$\mu_i \quad = \quad \alpha_i + \beta_i \mu_M, \tag{9}$$
$$\sigma_{ii} \quad = \quad \beta_i^2 \sigma_M^2 + \sigma_i^2, \tag{10}$$
$$\sigma_{ij} \quad = \quad \beta_i \beta_j \sigma_M^2. \tag{11}$$

Both variances and covariances are determined by the assets' betas and sigmas and by the standard deviation of the market index. Thus, the $N(N+1)/2$ different elements of the covariance matrix Σ can be expressed by $2N + 1$ parameters $\beta_i, \sigma_i, \sigma_M$. This is a significant reduction of the number of parameters that need to be estimated.

The traditional estimators for parameters of the single index model are based on the maximum likelihood method. These estimators have optimal properties for normally distributed variables, but they may give completely erroneous results in the presence of outlying observations. Therefore, robust estimates for the parameters of the single index model are necessary in order to provide robust and optimal portfolios.

3. Robust Estimators for the Single Index Model and Robust Portfolios

3.1. Definitions of the Estimators

Consider the linear regression model

$$X = \alpha + \beta X_M + e. \tag{12}$$

Suppose we have i.i.d. two-dimensional random vectors $Z_j = (X_{Mj}, X_j)$, $j = 1, \ldots, n$, such that $X_j = \alpha + \beta X_{Mj} + e_j$. The random variables e_j, $j = 1, \ldots, n$, are i.i.d. with $\mathcal{N}(0, \sigma)$ and independent on the X_{Mj}, $j = 1, \ldots, n$.

The classical estimators for the unknown parameters α, β, σ of the linear regression model are the maximum likelihood estimators (MLE). The classical MLE estimators perform well if the model hypotheses are satisfied exactly and may otherwise perform poorly. It is well known that the MLE are not robust, since a small fraction of outliers, even one outlier may have an important effect inducing significant errors on the estimates. Therefore, robust alternatives of the MLE should be considered, in order to propose robust estimates for the single index model, leading then to robust portfolio weights.

In order to robustly estimate the unknown parameters α, β, σ, suppressing the outsized effects of outliers, we use the approach based on pseudodistance minimization.

For two probability measures P, Q admitting densities p, respectively, q with respect to the Lebesgue measure, we consider the following family of pseudodistances (also called γ-divergences in some articles) of orders $\gamma > 0$

$$R_\gamma(P, Q) := \frac{1}{(1+\gamma)} \ln \left(\int p^\gamma dP \right) + \frac{1}{\gamma(1+\gamma)} \ln \left(\int q^\gamma dQ \right) - \frac{1}{\gamma} \ln \left(\int p^\gamma dQ \right), \tag{13}$$

satisfying the limit relation

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

Note that $R_0(P, Q)$ is the well-known modified Kullback–Leibler divergence. Minimum pseudodistance estimators for parametric models, using the family (13), have been studied by [6,10,11]. We also mention that pseudodistances (13) have also been used for defining optimal robust M-estimators with the Hampel's infinitesimal approach in [20].

For the linear regression model, we consider the joint distribution of the entire data, the explanatory variable X_M being random together with the response variable X, and write a pseudodistance between a theoretical model and the data. Let P_θ, with $\theta =: (\alpha, \beta, \sigma)$, be the probability measure associated with the theoretical model given by the random vector (X_M, X), where $X = \alpha + \beta X_M + e$ with $e \sim \mathcal{N}(0, \sigma)$, e independent on X_M, and Q the probability measure associated with the data. Denote by p_θ, respectively, q, the corresponding densities. For $\gamma > 0$, the pseudodistance between P_θ and Q is defined by

$$R_\gamma(P_\theta, Q) := \frac{1}{(1+\gamma)} \ln \left(\int p_\theta^\gamma(x_M, x) dP_\theta(x_M, x) \right) + \frac{1}{\gamma(1+\gamma)} \ln \left(\int q^\gamma(x_M, x) dQ(x_M, x) \right) - \frac{1}{\gamma} \ln \left(\int p_\theta^\gamma(x_M, x) dQ(x_M, x) \right). \tag{14}$$

Using the change of variables $(x_M, x) \to (u, v) := (x_M, x - \alpha - \beta x_M)$ and taking into account that $f(u, v) := p_\theta(u, v + \alpha + \beta u)$ is the density of (X_M, e), since X_M and e are independent, we can write

$$\int p_\theta^\gamma(x_M, x)dP_\theta(x_M, x) = \int p_M^{\gamma+1}(u)du \cdot \int \phi_\sigma^{\gamma+1}(v)dv, \tag{15}$$

$$\int p_\theta^\gamma(x_M, x)dQ(x_M, x) = \int p_M^\gamma(x_M) \cdot \phi_\sigma^\gamma(x - \alpha - \beta x_M)dQ(x_M, x), \tag{16}$$

where p_M is the density of X_M and ϕ_σ is the density of the random variable $e \sim \mathcal{N}(0, \sigma)$. Then,

$$\begin{aligned} R_\gamma(P_\theta, Q) = {} & \frac{1}{(1+\gamma)} \ln\left(\int p_M^{\gamma+1}(u)du\right) + \frac{1}{(1+\gamma)} \ln\left(\int \phi_\sigma^{\gamma+1}(v)dv\right) \\ & + \frac{1}{\gamma(1+\gamma)} \ln\left(\int q^\gamma(x_M, x)dQ(x_M, x)\right) \\ & - \frac{1}{\gamma} \ln\left(\int p_M^\gamma(x_M) \cdot \phi_\sigma^\gamma(x - \alpha - \beta x_M)dQ(x_M, x)\right). \end{aligned}$$

Notice that the first and the third terms in the pseudodistance $R_\gamma(P_\theta, Q)$ do not depend on θ and hence are not included in the minimization process. The parameter $\theta_0 := (\alpha_0, \beta_0, \sigma_0)$ of interest is then given by

$$\begin{aligned} (\alpha_0, \beta_0, \sigma_0) := {} & \arg\min_{\alpha, \beta, \sigma} R_\gamma(P_\theta, Q) \\ = {} & \arg\min_{\alpha, \beta, \sigma} \left\{ \frac{1}{(1+\gamma)} \ln\left(\int \phi_\sigma^{\gamma+1}(v)dv\right) - \frac{1}{\gamma} \ln\left(\int p_M^\gamma(x_M) \cdot \phi_\sigma^\gamma(x - \alpha - \beta x_M)dQ(x_M, x)\right) \right\}. \end{aligned} \tag{17}$$

Suppose now that an i.i.d. sample $Z_1, \ldots, Z_n$ is available from the true model. For a given $\gamma > 0$, we define a minimum pseudodistance estimator of $\theta_0 = (\alpha_0, \beta_0, \sigma_0)$ by minimizing an empirical version of the objective function in Equation (17). This empirical version is obtained by replacing $p_M(x_M)$ with the empirical density function $\widehat{p}_M(x_M) = \frac{1}{n} \sum_{i=1}^n \delta(x_M - X_{Mi})$, where $\delta(\cdot)$ is the Dirac delta function, and Q with the empirical measure corresponding to the sample. More precisely, we define $\widehat{\theta} := (\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma})$

$$\begin{aligned} (\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma}) := {} & \arg\min_{\alpha, \beta, \sigma} \left\{ \frac{1}{(1+\gamma)} \ln\left(\int \phi_\sigma^{\gamma+1}(v)dv\right) - \frac{1}{\gamma} \ln\left(\int \widehat{p}_M^\gamma(x_M) \cdot \phi_\sigma^\gamma(x - \alpha - \beta x_M)dP_n(x_M, x)\right) \right\} \\ = {} & \arg\min_{\alpha, \beta, \sigma} \left\{ \frac{1}{(1+\gamma)} \ln\left(\int \phi_\sigma^{\gamma+1}(v)dv\right) - \frac{1}{\gamma} \ln\left(\frac{1}{n^{\gamma+1}} \sum_{j=1}^n \phi_\sigma^\gamma(X_j - \alpha - \beta X_{Mj})\right) \right\}, \end{aligned} \tag{18}$$

or equivalently

$$\begin{aligned} (\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma}) = {} & \arg\max_{\alpha, \beta, \sigma} \sum_{j=1}^n \frac{\phi_\sigma^\gamma(X_j - \alpha - \beta X_{Mj})}{[\int \phi_\sigma^{\gamma+1}(v)dv]^{\gamma/(\gamma+1)}} \\ = {} & \arg\max_{\alpha, \beta, \sigma} \sum_{j=1}^n \sigma^{-\gamma/(\gamma+1)} \exp\left(-\frac{\gamma}{2}\left(\frac{X_j - \alpha - \beta X_{Mj}}{\sigma}\right)^2\right). \end{aligned}$$

Differentiating with respect to α, β, σ, the estimators $\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma}$ are solutions of the system

$$\sum_{j=1}^n \exp\left(-\frac{\gamma}{2}\left(\frac{X_j - \alpha - \beta X_{Mj}}{\sigma}\right)^2\right)\left(\frac{X_j - \alpha - \beta X_{Mj}}{\sigma}\right) = 0, \tag{19}$$

$$\sum_{j=1}^n \exp\left(-\frac{\gamma}{2}\left(\frac{X_j - \alpha - \beta X_{Mj}}{\sigma}\right)^2\right)\left(\frac{X_j - \alpha - \beta X_{Mj}}{\sigma}\right)X_{Mj} = 0, \tag{20}$$

$$\sum_{j=1}^n \exp\left(-\frac{\gamma}{2}\left(\frac{X_j - \alpha - \beta X_{Mj}}{\sigma}\right)^2\right)\left[\left(\frac{X_j - \alpha - \beta X_{Mj}}{\sigma}\right)^2 - \frac{1}{\gamma+1}\right] = 0. \tag{21}$$

Note that, for $\gamma = 0$, the solution of this system is nothing but the maximum likelihood estimator of (α, β, σ). Therefore, the estimating Equations (19)–(21) are generalizations of the maximum likelihood score equations. The tuning parameter γ associated with the pseudodistance controls the trade-off between robustness and efficiency of the minimum pseudodistance estimators.

We can also write that $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma})$ is a solution of

$$\sum_{j=1}^{n} \Psi(Z_j, \widehat{\theta}) = 0 \quad \text{or} \quad \int \Psi(z, \widehat{\theta}) dP_n(z) = 0, \tag{22}$$

where

$$\Psi(z, \theta) = \left(\phi\left(\frac{x - \alpha - \beta x_M}{\sigma} \right), \phi\left(\frac{x - \alpha - \beta x_M}{\sigma} \right) x_M, \chi\left(\frac{x - \alpha - \beta x_M}{\sigma} \right) \right)^T, \tag{23}$$

with $z = (x_M, x)$, $\theta = (\alpha, \beta, \sigma)$, $\phi(t) = \exp(-\frac{\gamma}{2} t^2) t$ and $\chi(t) = \exp(-\frac{\gamma}{2} t^2)[t^2 - \frac{1}{\gamma + 1}]$.

When the measure Q corresponding to the data pertain to the theoretical model, hence $Q = P_{\theta_0}$, it holds that

$$\int \Psi(z, \theta_0) dP_{\theta_0}(z) = 0. \tag{24}$$

Thus, we can consider $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma})$ as a Z-estimator of $\theta_0 = (\alpha_0, \beta_0, \sigma_0)$, which allows for adapting in the present context asymptotic results from the general theory of Z-estimators (see [21]).

Remark 1. *In the case when the density p_M is known, by replacing Q with the empirical measure P_n in Equation (17), a new class of estimators of $(\alpha_0, \beta_0, \sigma_0)$ can be obtained. These estimators can also be written under the form of Z-estimators, using the same reasoning as above. The results of Theorems 1–4 below could be adapted for these new estimators, and moreover all the influence functions of these estimators would be redescending bounded. However, in practice, the density of the index return is not known. Therefore, we will work with the class of minimum pseudodistance estimators as defined above.*

3.2. Asymptotic Properties

In order to prove the consistency of the estimators, we use their definition (22) as Z-estimators.

3.2.1. Consistency

Theorem 1. *Assume that, for any $\varepsilon > 0$, the following condition for the separability of solution holds*

$$\inf_{\theta \in M} \left\| \int \psi(z, \theta) dP_{\theta_0}(z) \right\| > 0 = \left\| \int \psi(z, \theta_0) dP_{\theta_0}(z) \right\|, \tag{25}$$

where $M := \{\theta \text{ s.t. } \|\theta - \theta_0\| \geq \varepsilon\}$. Then, $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma})$ converges in probability to $\theta_0 = (\alpha_0, \beta_0, \sigma_0)$.

3.2.2. Asymptotic Normality

Assume that $Z_1, \ldots, Z_n$ are i.i.d. two-dimensional random vectors having the common probability distribution P_{θ_0}. For $\gamma > 0$ fixed, let $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma})$ be a sequence of estimators of the unknown parameter $\theta_0 = (\alpha_0, \beta_0, \sigma_0)$, solution of

$$\sum_{j=1}^{n} \Psi(Z_j, \widehat{\theta}) = 0, \tag{26}$$

where

$$\Psi(z, \theta) = \left(\sigma^2 \phi\left(\frac{x - \alpha - \beta x_M}{\sigma} \right), \sigma^2 \phi\left(\frac{x - \alpha - \beta x_M}{\sigma} \right) x_M, \sigma^2 \chi\left(\frac{x - \alpha - \beta x_M}{\sigma} \right) \right)^T, \tag{27}$$

with $z = (x_M, x)$, $\theta = (\alpha, \beta, \sigma)$, $\phi(t) = \exp(-\frac{\gamma}{2}t^2)t$ and $\chi(t) = \exp(-\frac{\gamma}{2}t^2)[t^2 - \frac{1}{\gamma+1}]$. Note that the estimators $\widehat{\theta} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma})$ defined by Equations (19)–(21), or equivalently by (22), are also solutions of the system (26). Using the function (27) for defining the estimators allows for obtaining the asymptotic normality, only imposing the consistency condition of the estimators, without other supplementary assumptions that are usually imposed in the case of Z-estimators.

Theorem 2. *Assume that $\widehat{\theta} \to \theta_0$ in probability. Then,*

$$\sqrt{n}(\widehat{\theta} - \theta_0) \to \mathcal{N}_3(0, B^{-1}A(B^{-1})^T) \tag{28}$$

in distribution, where $A = E(\Psi(Z, \theta_0)\Psi(Z, \theta_0)^T)$ and $B = E(\dot{\Psi}(Z, \theta_0))$, with Ψ defined by (27), $\dot{\Psi}$ being the matrix with elements $\dot{\Psi}_{ik} = \frac{\partial \Psi_i}{\partial \theta_k}$.

After some calculations, we obtain the asymptotic covariance matrix of $\widehat{\theta}$ having the form

$$\sigma_0^2 \frac{(\gamma+1)^3}{(2\gamma+1)^{3/2}} \begin{pmatrix} \dfrac{\mu_M^2 + \sigma_M^2}{\sigma_M^2} & \dfrac{-\mu_M}{\sigma_M^2} & 0 \\ \dfrac{-\mu_M}{\sigma_M^2} & \dfrac{1}{\sigma_M^2} & 0 \\ 0 & 0 & \dfrac{3\gamma^2 + 4\gamma + 2}{4(2\gamma+1)} \end{pmatrix}.$$

It follows that $\widehat{\beta}$ and $\widehat{\sigma}$ are asymptotically independent; in addition, $\widehat{\alpha}$ and $\widehat{\sigma}$ are asymptotically independent.

3.3. Influence Functions

In order to describe stability properties of the estimators, we use the following well-known concepts from the theory of robust statistics. A map T, defined on a set of probability measures and parameter space valued, is a statistical functional corresponding to an estimator $\widehat{\theta}$ of the parameter θ, if $\widehat{\theta} = T(P_n)$, P_n being the empirical measure pertaining to the sample. The influence function of T at P_θ is defined by

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

where $\widetilde{P}_{\varepsilon z} := (1 - \varepsilon)P_\theta + \varepsilon\delta_z$, δ_z being the Dirac measure putting all mass at z. As a consequence, the influence function describes the linearized asymptotic bias of a statistic under a single point contamination of the model P_θ. An unbounded influence function implies an unbounded asymptotic bias of a statistic under single point contamination of the model. Therefore, a natural robustness requirement on a statistical functional is the boundedness of its influence function.

For $\gamma > 0$ fixed and a given probability measure P, the statistical functionals $\alpha(P)$, $\beta(P)$ and $\sigma(P)$, corresponding to the minimum pseudodistance estimators $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\sigma}$, are defined by the solution of the system

$$\int \Psi(z, T(P))dP(z) = 0, \tag{29}$$

with Ψ defined by (23) and $T(P) := (\alpha(P), \beta(P), \sigma(P))$, whenever this solution exists.

When $P = P_\theta$ corresponds to the considered theoretical model, the solution of system (29) is $T(P_\theta) = \theta = (\alpha, \beta, \sigma)$.

Theorem 3. *The influence functions corresponding to the estimators $\widehat{\alpha}, \widehat{\beta}$ and $\widehat{\sigma}$ are respectively given by*

$$IF(x_{M0}, x_0; \alpha, P_\theta) = \sigma(\gamma + 1)^{3/2} \phi\left(\frac{x_0 - \alpha - \beta x_{M0}}{\sigma}\right)\left[1 - \frac{(x_{M0} - E(X_M))E(X_M)}{Var(X_M)}\right], \quad (30)$$

$$IF(x_{M0}, x_0; \beta, P_\theta) = \sigma(\gamma + 1)^{3/2} \phi\left(\frac{x_0 - \alpha - \beta x_{M0}}{\sigma}\right)\frac{x_{M0} - E(X_M)}{Var(X_M)}, \quad (31)$$

$$IF(x_{M0}, x_0; \sigma, P_\theta) = \frac{\sigma(\gamma + 1)^{5/2}}{2}\chi\left(\frac{x_0 - \alpha - \beta x_{M0}}{\sigma}\right). \quad (32)$$

Since χ is redescending, $\widehat{\sigma}$ has a bounded influence function and hence it is a redescending B-robust estimator. On the other hand, $IF(x_{M0}, x_0, \alpha, P)$ and $IF(x_{M0}, x_0, \beta, P)$ will tend to infinity only when x_{M0} tends to infinity and $\left|\frac{x_0 - \alpha - \beta x_{M0}}{\sigma}\right| \leq k$, for some k. Hence, these influence functions are bounded with respect to partial outliers or leverage points (outlying values of the independent variable). This means that large outliers with respect to x_M, or with respect to x, will have a reduced influence on the estimates. However, the influence functions are clearly unbounded for $\gamma = 0$, which corresponds to the non-robust maximum likelihood estimators.

3.4. Equivariance of the Regression Coefficients' Estimators

If an estimator is equivariant, it means that it transforms "properly" in some sense. Rousseeuw and Leroy [22] (p. 116) discuss three important equivariance properties for a regression estimator: regression equivariance, scale equivariance and affine equivariance. These are desirable properties since they allow one to know how the estimates change under different types of transformations of the data. Regression equivariance means that any additional linear dependence is reflected in the regression vector accordingly. The regression equivariance is routinely used when studying regression estimators. It allows for assuming, without loss generality, any value for the parameter (α, β) for proving asymptotic properties or describing Monte-Carlo studies. An estimator being scale equivariant means that the fit produced by it is independent of the choice of measurement unit for the response variable. The affine equivariance is useful because it means that changing to a different co-ordinate system for the explanatory variable will not affect the estimate. It is known that the maximum likelihood estimator of the regression coefficients satisfies all these three properties. We show that the minimum pseudodistance estimators of the regression coefficients satisfy all the three equivariance properties, for all $\gamma > 0$.

Theorem 4. *For all $\gamma > 0$, the minimum pseudodistance estimators $(\widehat{\alpha}, \widehat{\beta})^T$ of the regression coefficients $(\alpha, \beta)^T$ are regression equivariant, scale equivariant and affine equivariant.*

On the other hand, the objective function in the definition of the estimators depends on data only through the summation

$$\sum_{j=1}^{n} \sigma^{-\gamma/(\gamma+1)} \exp\left(-\frac{\gamma}{2}\left(\frac{X_j - \alpha - \beta X_{Mj}}{\sigma}\right)^2\right), \quad (33)$$

which is permutation invariant. Thus, the corresponding estimators of the regression coefficients and of the error standard deviation are permutation invariant, therefore the ordering of data does not affect the estimators.

The minimum pseudodistance estimators are also equivariant with respect to reparametrizations. If $\theta = (\alpha, \beta, \sigma)$ and the model is reparametrized to $Y = Y(\theta)$ with a one-to-one transformation, then the minimum pseudodistance estimator of Y is simply $\widehat{Y} = Y(\widehat{\theta})$, in terms of the minimum pseudodistance estimator $\widehat{\theta}$ of θ, for the same γ.

3.5. Robust Portfolios Using Minimum Pseudodistance Estimators

The robust estimation of the parameters $\alpha_i, \beta_i, \sigma_i$ from the single index model given by (35), using minimum pseudodistance estimators, together with the robust estimation of μ_M and σ_M lead to robust estimates of μ and Σ, on the basis of relations (9)–(11). Since we do not model the explanatory variable X_M in a specific way, we estimate μ_M and the standard deviation σ_M using as robust estimators the median, respectively the median absolute deviation. Then, the portfolio weights, obtained as solutions of the optimization problems (4) or (7) with input parameters robustly estimated, will also be robust. This methodology leads to new optimal robust portfolios. In the next section, on the basis of real financial data, we illustrate this new methodology and compare it with the traditional method based on maximum likelihood estimators.

4. Applications

4.1. Comparisons of the Minimum Pseudodistance Estimators with Other Robust Estimators for the Linear Regression Model

In order to illustrate the performance of the minimum pseudodistance estimators for the simple linear regression model, we compare them with the least median of squares (LMS) estimator (see [22,23]), with S-estimators (SE) (see [24]) and with the minimum density power divergence (MDPD) estimators (see [25]), estimators that are known to have a good behavior from the robustness point of view.

We considered a data set that comes from astronomy, namely the data from the Hertzsprung–Russell diagram of the star clusters CYG OB1 containing 47 stars in the direction of Cygnus. For these data, the independent variable is the logarithm of the effective temperature at the surface of the star and the dependent variable is the logarithm of its light intensity. The data are given in Rousseeuw and Leroy [22] (p. 27), who underlined that there are two groups of points: the majority, following a steep band, and four stars clearly forming a separate group from the rest of the data. These four stars are known as giants in astronomy. Thus, these outliers are not recording errors, but represents leverage points coming from a different group.

The estimates of the regression coefficients and of error standard deviation obtained with minimum pseudodistance estimators for several values of γ are given in Table 1 and some of the fitted models are plotted in Figure 1. For comparison, in Table 1, we also give estimates obtained with S-estimators based on the Tukey biweighted function, these estimates being taken from [24], as well as estimations obtained with minimum density power divergence methods for several values of the tuning parameter, and estimates obtained with the least median of squares method, all these estimates being taken from [25]. The MLE estimates, given on the first line of Table 1, are significantly affected by the four leverage points. On the other hand, like the robust least median of squares estimator, the robust S-estimators and some minimum density power divergence estimators, the minimum pseudodistance estimators with $\gamma \geq 0.32$ can successfully ignore outliers. In addition, the minimum pseudodistance estimators with $\gamma \geq 0.5$ give robust fits that are closer to the fits generated by the least median of squares estimates or by the S-estimates than the fits generated by the minimum density power divergence estimates.

Table 1. The parameter estimates for the linear regression model for the Hertzsprung–Russell data using several minimum pseudodistance (MP) methods, several minimum density power divergence (MDPD) methods, the least median of squares (LMS) method, S-estimators and the MLE method. γ represents tuning parameter.

| MLE Estimates			
	α	β	σ
	6.79	−0.41	0.55
MP Estimates			
γ	α	β	σ
0.01	6.79	−0.41	0.55
0.1	6.81	−0.41	0.56
0.25	6.86	−0.42	0.58
0.3	6.88	−0.42	0.59
0.31	6.89	−0.43	0.59
0.32	−6.81	2.66	0.39
0.35	−7.16	2.74	0.38
0.4	−7.62	2.85	0.38
0.5	−8.17	2.97	0.37
0.75	−8.65	3.08	0.38
1	−8.84	3.12	0.39
1.2	−8.94	3.15	0.40
1.5	−9.08	3.18	0.41
2	−9.31	3.23	0.43
MDPD Estimates			
γ	α	β	σ
0.1	6.78	−0.41	0.60
0.25	−5.16	2.30	0.42
0.5	−7.22	2.76	0.40
0.8	−7.89	2.91	0.40
1	−8.03	2.95	0.41
S-Estimates			
	α	β	σ
	−9.59	3.28	−
LMS Estimates			
	α	β	σ
	−12.30	3.90	−
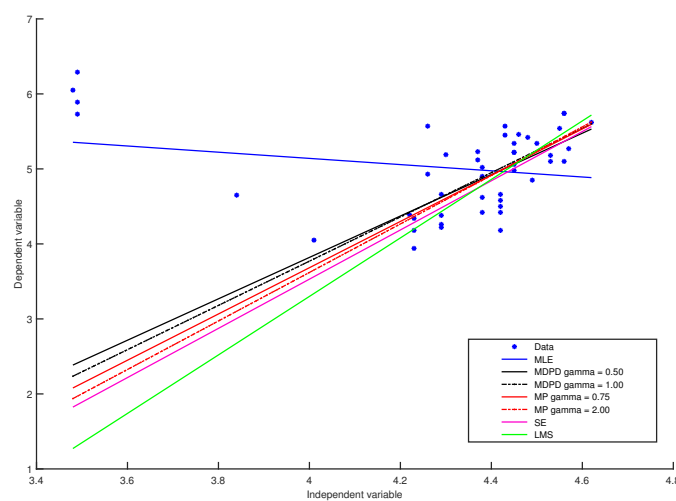
Figure 1. Plots of the Hertzsprung–Russell data and fitted regression lines using MLE, minimum density power divergence (MDPD) methods for several values of γ, minimum pseudodistance (MP) methods for several values of γ, S-estimators (SE) and the least median of squares (LMS) method.

4.2. Robust Portfolios Using Minimum Pseudodistance Estimators

In order to illustrate the performance of the proposed robust portfolio optimization method, we considered real data sets for the Russell 2000 index and for 50 stocks from its components. The stocks are listed in Appendix B. We selected daily return data for the Russell 2000 index and for all these stocks from 2 January 2013 to 30 June 2016. The data were retrieved from Yahoo Finance.

The data has been divided by quarter, in total 14 quarters for index and each stock. For each quarter, on the basis of data corresponding to the index, we estimated μ_M and the standard deviation σ_M using as robust estimators the median (MED), respectively the median absolute deviation (MAD) defined by

$$MAD := \frac{1}{0.6745} \cdot MED(|X_i - MED(X_i)|). \tag{34}$$

We also estimated μ_M and σ_M classically, using sample mean and sample variance. Then, for each quarter and each of the 50 stocks, we estimated α, β and σ from the regression model using robust minimum pseudodistance estimators, respectively the classical MLE estimators. Then, on the basis of relations (9), (10) and (11), we estimated μ and Σ first using the robust estimates and then the classical estimates, all being previously computed.

Once the input parameters for the portfolio optimization procedure were estimated, for each quarter, we determined efficient frontiers, for both robust estimates and classical estimates. In both cases, the efficient frontier is determined as follows. Firstly, the range of returns is determined as the interval comprised between the return of the portfolio of global minimum risk (variance) and the maximum value of the return of a feasible portfolio, where the feasible region is

$$X = \left\{ w \in \mathbb{R}^N \,\middle|\, w^T e_N = 1, w_k \geq 0,\, k \in \{1, \ldots, 50\} \right\}$$

and $N = 50$. We trace each efficient frontier in 100 points; therefore, the range of returns is divided, in each case, in ninety-nine sub-intervals with

$$\mu_1 < \mu_2 < \cdots < \mu_{100},$$

where μ_1 is the return of the portfolio of global minimum variance and μ_{100} is the maximum return for the feasible region X. We determined μ_1 and μ_{100} using robust estimates of μ and Σ (for the robust frontier) and then using classical estimates (for the classical frontier). In each case, 100 optimization problems are solved:

$$\arg \min_{w \in \mathbb{R}^N} S(w)$$
$$w_k \geq 0,\, k \in \{1, \ldots, 50\}$$
$$w^T e_N = 1$$
$$R(w) \geq \mu_i,$$

where $i \in \{1, \ldots, 100\}$.

In Figure 2, for eight quarters (the first four quarters and the last four quarters), we present efficient frontiers corresponding to the optimal minimum variance portfolios based on the robust minimum pseudodistance estimates with $\gamma = 0.5$, respectively based on the classical estimates. Thus, on the ox-axis, we consider the portfolio risk (given by the portfolio standard deviation) and, on the oy-axis, we represent the portfolio return. We notice that, in comparison with the classical method based on MLE, the proposed robust method provides optimal portfolios that have higher returns for the same level of risk (standard deviation). Indeed, for each quarter, the robust frontier is situated above the classical one, the standard deviations of the robust portfolios being smaller compared with those of the classical portfolios. We obtained similar results for the other quarters and for other choices of the tuning parameter γ, corresponding to the minimum pseudodistance estimators, too.

We also illustrate the empirical performance of the proposed optimal portfolios through an out-of-sample analysis, by using the Sharpe ratio as out-of-sample measure. For this analysis, we apply a "rolling-horizon" procedure as presented in [18]. First, we choose a window over which to perform the estimation. We denote the length of the estimation window by $\tau < T$, where T is the size of the entire data set. Then, using the data in the first estimation window, we compute the weights for the considered portfolios. We repeat this procedure for the next window, by including the data for the next day and dropping the data for the earliest day. We continue doing this until the end of the data set is reached. At the end of this process, we have generated $T - \tau$ portfolio weight vectors for each strategy, which are the vectors w_t^k for $t \in \{\tau, \ldots, T - 1\}$, k denoting the strategy. For a strategy k, w_t^k has the components $w_{j,t}^k$, where $w_{j,t}^k$ denotes the portfolio weight in asset j chosen at the time t.

The out-of-sample return at the time $t + 1$, corresponding to the strategy k, is defined as $(w_t^k)^T X_{t+1}$, $X_{t+1} := (X_{1,t+1}, \ldots, X_{N,t+1})^T$ representing the data at the time $t + 1$. For each strategy k, using these out-of-sample returns, the out-of-sample mean and the out-of-sample variance are defined by

$$\widehat{\mu}^k = \frac{1}{T - \tau} \sum_{t=\tau}^{T-1} (w_t^k)^T X_{t+1} \quad \text{and} \quad (\widehat{\sigma}^k)^2 = \frac{1}{T - \tau - 1} \sum_{t=\tau}^{T-1} ((w_t^k)^T X_{t+1} - \widehat{\mu}^k)^2 \tag{35}$$

and the out-of-sample Sharpe ratio is defined by

$$\widehat{SR}^k = \frac{\widehat{\mu}^k}{\widehat{\sigma}^k}. \tag{36}$$

In this example, we considered the data set corresponding to the quarters 13 and 14. The size of the entire data set was $T = 126$ and the length of the estimation window was $\tau = 63$ points. For the data from the first window, classical and robust efficient frontiers were traced, following all the steps that we explained in the first part of this subsection. More precisely, we considered the classical efficient frontier corresponding to the optimal minimum variance portfolios based on MLE and three robust frontiers, corresponding to the optimal minimum variance portfolios using robust minimum pseudodistance estimations with $\gamma = 1$, $\gamma = 1.2$ and $\gamma = 1.5$, respectively. Then, on each frontier, we chose the optimal portfolio associated with the maximal value of the ratio between the portfolio return and portfolio standard deviation. These four optimal portfolios represent the strategies that we compared in the out-of-sample analysis. For each of these portfolios, we computed the out-of-sample returns for the next time (next day). Then, we repeated all these procedures for the next window, and so on until the end of the data set has been reached. In the spirit of [18] Section 5, using (35) and (36), we computed out-of-sample means, out-of-sample variances and out-of-sample Sharpe ratios for each strategy. The out-of-sample means and out-of-sample variances were annualized, and we also considered a benchmark rate of 1.5 %. In this way, we obtained the following values for the out-of-sample Sharpe ratio: $\widehat{SR} = 0.22$ for the optimal portfolio based on MLE, $\widehat{SR} = 0.74$ for the optimal portfolio based on minimum pseudodistance estimations with $\gamma = 1$, $\widehat{SR} = 0.71$ for the optimal portfolio based on minimum pseudodistance estimations with $\gamma = 1.2$ and $\widehat{SR} = 0.29$ for the optimal portfolio based on minimum pseudodistance estimations with $\gamma = 1.5$. In Figure 3, we illustrate efficient frontiers for the windows 7 and 8, as well as the optimal portfolios chosen on each frontier.

This example shows that the optimal minimum variance portfolios based on robust minimum pseudodistance estimations in the single index model may attain higher Sharpe ratios than the traditional optimal minimum variance portfolios given by the single index model using MLE.
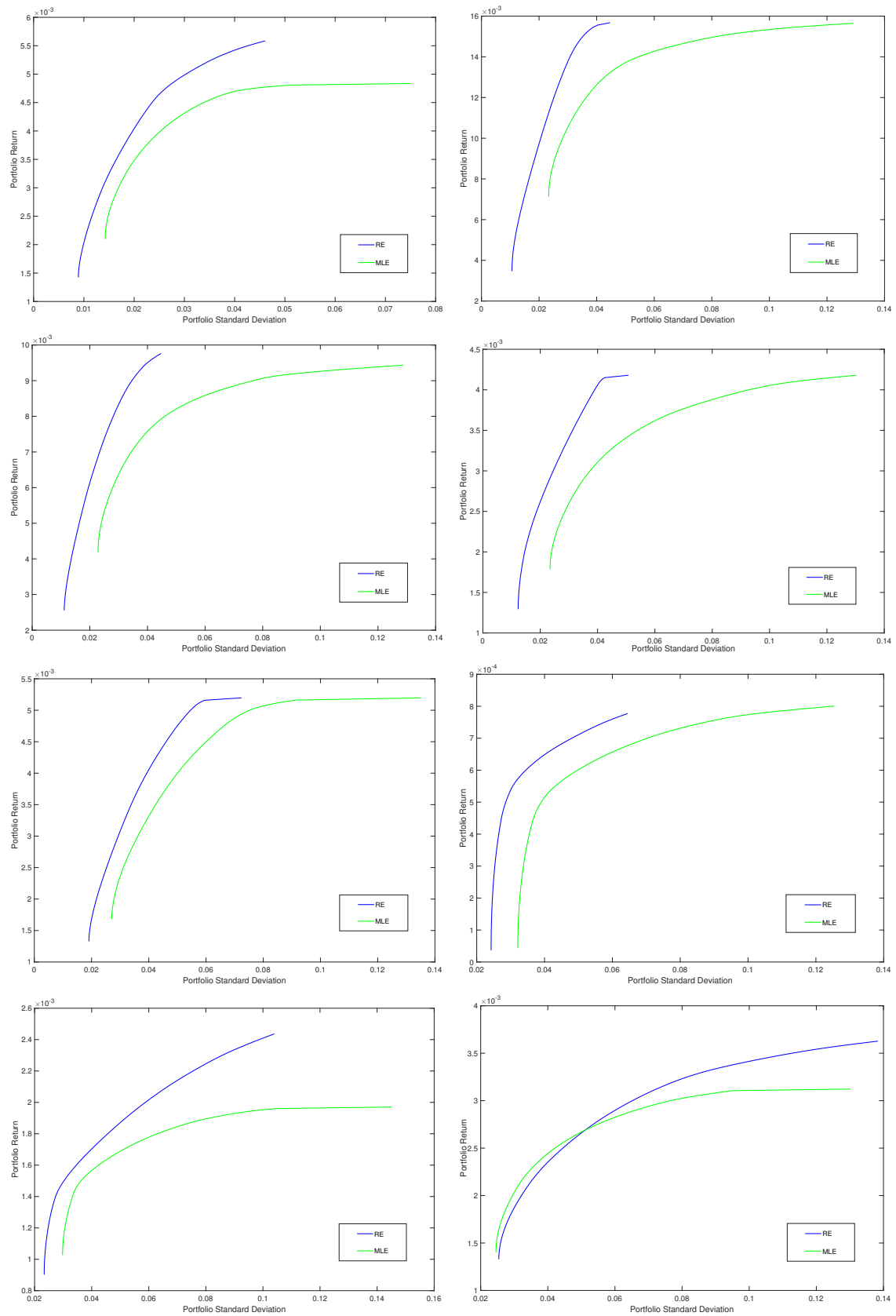
Figure 2. Efficient frontiers, classical (MLE) vs. robust corresponding to $\gamma = 0.5$ (RE), for eight quarters (the first four quarters and the last four quarters).
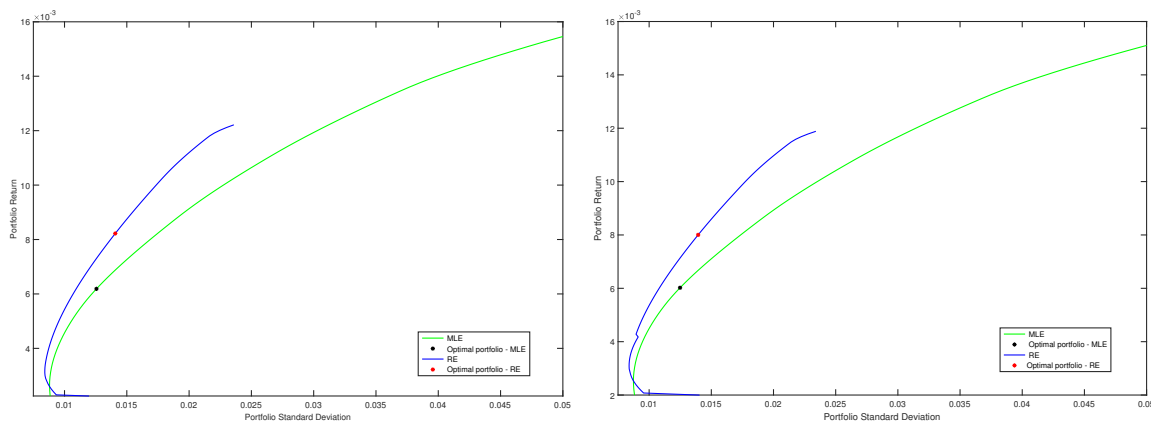
Figure 3. Efficient frontiers, classical (MLE) vs. robust corresponding to $\gamma = 1$ (RE), and optimal portfolios chosen on frontiers, for the windows 7 (**left**) and 8 (**right**).

The obtained numerical results show that, for the single index model, the presented robust technique for portfolio optimization yields better results than the classical method based on MLE, in the sense that it leads to larger returns for the same value of risk in the case when outliers or atypical observations are present in the data set. The considered data sets contain such outliers. This is often the case for the considered problem, since outliers frequently occur in asset returns data. However, when there are no outliers in the data set, the classical method based on MLE is more efficient than the robust ones and therefore may lead to better results.

5. Conclusions

When outliers or atypical observations are present in the data set, the new portfolio optimization method based on robust minimum pseudodistance estimates yields better results than the classical single index method based on MLE estimates, in the sense that it leads to larger returns for smaller risks. In literature, there exist various methods for robust estimation in regression models. In the present paper, we proposed the method based on the minimum pseudodistance approach, which suppose to solve a simple optimization problem. In addition, from a theoretical point of view, these estimators have attractive properties, such as being redescending robust, consistent, equivariant and asymptotically normally distributed. The comparison with other known robust estimators of the regression parameters, such as the least median of squares estimators, the S-estimators or the minimum density power divergence estimators, shows that the minimum pseudodistance estimators represent an attractive alternative that may be considered in other applications too.

Author Contributions: A.T. designed the methodology, obtained the theoretical results and wrote the paper. A.T. and C.F. conceived the application part. C.F. implemented the methods in MATLAB and obtained the numerical results. Both authors have read and approved the final manuscript.

Appendix A. Proof of the Results

Proof of Theorem 1. Since the functions ϕ and χ are redescending bounded functions, for a compact neighborhood N_{θ_0} of θ_0, it holds

$$\int \sup_{\theta \in N_{\theta_0}} \|\Psi(z,\theta)\| dP_{\theta_0}(z) < \infty. \tag{A1}$$

Since $\theta \mapsto \Psi(z,\theta)$ is continuous, by the uniform law of large numbers, (A1) implies

$$\sup_{\theta \in N_{\theta_0}} \left\| \int \psi(z,\theta) dP_n(z) - \int \psi(z,\theta) dP_{\theta_0}(z) \right\| \to 0 \tag{A2}$$

in probability.

Then, (A2) together with assumption (25) assure the convergence in probability of $\widehat{\theta}$ toward θ_0. The arguments are the same as those from van der Vaart [21], Theorem 5.9, p. 46. $\square$

Proof of Theorem 2. First, note that Ψ defined by (27) is twice differentiable with respect to θ with bounded derivatives. The matrix $\dot{\Psi}(z,\theta)$ has the form

$$\begin{pmatrix} -\sigma\phi'\left(\frac{x-\alpha-\beta x_M}{\sigma}\right) & -\sigma\phi'\left(\frac{x-\alpha-\beta x_M}{\sigma}\right)x_M & 2\sigma\phi\left(\frac{x-\alpha-\beta x_M}{\sigma}\right) - \sigma\phi'\left(\frac{x-\alpha-\beta x_M}{\sigma}\right)\left(\frac{x-\alpha-\beta x_M}{\sigma}\right) \\ -\sigma\phi'\left(\frac{x-\alpha-\beta x_M}{\sigma}\right)x_M & -\sigma\phi'\left(\frac{x-\alpha-\beta x_M}{\sigma}\right)x_M^2 & 2\sigma\phi\left(\frac{x-\alpha-\beta x_M}{\sigma}\right)x_M - \sigma\phi'\left(\frac{x-\alpha-\beta x_M}{\sigma}\right)\left(\frac{x-\alpha-\beta x_M}{\sigma}\right)x_M \\ -\sigma\chi'\left(\frac{x-\alpha-\beta x_M}{\sigma}\right) & -\sigma\chi'\left(\frac{x-\alpha-\beta x_M}{\sigma}\right)x_M & 2\sigma\chi\left(\frac{x-\alpha-\beta x_M}{\sigma}\right) - \sigma\chi'\left(\frac{x-\alpha-\beta x_M}{\sigma}\right)\left(\frac{x-\alpha-\beta x_M}{\sigma}\right) \end{pmatrix}$$

with $\phi'(t) = [1 - \gamma t^2]\exp(-\frac{\gamma}{2}t^2)$ and $\chi'(t) = [\frac{3\gamma+2}{\gamma+1}t - \gamma t^3]\exp(-\frac{\gamma}{2}t^2)$. Since $\phi(t), \chi(t), \phi'(t), \chi'(t)$ are redescending bounded functions, for $\theta = \theta_0$, it holds

$$|\ddot{\Psi}_{ik}(z,\theta_0)| \le K(z) \text{ with } E(K(Z)) < \infty. \tag{A3}$$

In addition, a simple calculation shows that each component $\frac{\partial \Psi_i}{\partial \theta_k \partial \theta_l}$ is a bounded function, since it can be expressed through the functions $\phi(t), \chi(t), \phi'(t), \chi'(t), \phi''(t), \chi''(t)$, which are redescending bounded functions. In addition, bounds that can be established for each component $\frac{\partial \Psi_i}{\partial \theta_k \partial \theta_l}$ do not depend on the parameter θ.

For each i, call $\ddot{\Psi}_i$ the matrix with elements $\frac{\partial \Psi_i}{\partial \theta_k \partial \theta_l}$ and $C_n(z,\theta)$ the matrix with its i-th raw equal to $(\widehat{\theta} - \theta_0)^T \ddot{\Psi}_i(z,\theta)$. Using a Taylor expansion, we get

$$0 = \sum_{j=1}^{n} \Psi(Z_j, \widehat{\theta}) = \sum_{j=1}^{n} \{\Psi(Z_j, \theta_0) + \dot{\Psi}(Z_j, \theta_0)(\widehat{\theta} - \theta_0) + \frac{1}{2}C_n(Z_j, \theta_j)(\widehat{\theta} - \theta_0)\}. \tag{A4}$$

Therefore,

$$0 = A_n + (B_n + \overline{C}_n)(\widehat{\theta} - \theta_0) \tag{A5}$$

with

$$A_n = \frac{1}{n}\sum_{j=1}^{n}\Psi(Z_j,\theta_0), \ B_n = \frac{1}{n}\sum_{j=1}^{n}\dot{\Psi}(Z_j,\theta_0), \ \overline{C}_n = \frac{1}{2n}\sum_{j=1}^{n}C_n(Z_j,\theta_j) \tag{A6}$$

i.e., $\overline{C}_n$ is the matrix with its i-th raw equal to $(\widehat{\theta} - \theta_0)^T \ddot{\Psi}_i^-$, where

$$\ddot{\Psi}_i^- = \frac{1}{2n}\sum_{j=1}^{n}\ddot{\Psi}_i(Z_j,\theta_j), \tag{A7}$$

which is bounded by a constant that does not depend on θ, according to the arguments mentioned above. Since $\widehat{\theta} - \theta_0 \to 0$ in probability, this implies that $\overline{C}_n \to 0$ in probability.

We have

$$\sqrt{n}(\widehat{\theta} - \theta_0) = -(B_n + \overline{C}_n)^{-1}\sqrt{n}A_n. \tag{A8}$$

Note that, for $j = 1, \ldots, n$, the vectors $\Psi(Z_j, \theta_0)$ are i.i.d. with mean zero and the covariance matrix A, and the matrices $\dot{\Psi}(Z_j, \theta_0)$ are i.i.d. with mean B. Hence, when $n \to \infty$, using (A3), the law of

large numbers implies that $B_n \to B$ in probability, which implies $B_n + \overline{C}_n \to B$ in probability, which is nonsingular. Then, the multivariate central limit theorem implies $\sqrt{n}A_n \to \mathcal{N}_3(0, A)$ in distribution. Then,

$$\sqrt{n}(\hat{\theta} - \theta_0) \to \mathcal{N}_3(0, B^{-1}A(B^{-1})^T) \tag{A9}$$

in distribution, according to the multivariate Slutzki's Lemma. $\square$

Proof of Theorem 3. The system (29) can be written as

$$\int \phi \left(\frac{x - \alpha(P) - \beta(P)x_M}{\sigma(P)} \right) dP(x_M, x) = 0,$$

$$\int \phi \left(\frac{x - \alpha(P) - \beta(P)x_M}{\sigma(P)} \right) x_M dP(x_M, x) = 0,$$

$$\int \chi \left(\frac{x - \alpha(P) - \beta(P)x_M}{\sigma(P)} \right) dP(x_M, x) = 0.$$

We consider the contaminated model $\widetilde{P}_{\varepsilon, x_{M0}, x_0} := (1 - \varepsilon)P_\theta + \varepsilon\delta_{(x_{M0}, x_0)}$, where $\delta_{(x_{M0}, x_0)}$ is the Dirac measure putting all mass in the point (x_{M0}, x_0), which we simply denote here by $\widetilde{P}_\varepsilon$. Then, it holds

$$(1 - \varepsilon) \int \phi \left(\frac{x - \alpha(\widetilde{P}_\varepsilon) - \beta(\widetilde{P}_\varepsilon)x_M}{\sigma(\widetilde{P}_\varepsilon)} \right) dP_\theta(x_M, x) + \varepsilon\phi \left(\frac{x_0 - \alpha(\widetilde{P}_\varepsilon) - \beta(\widetilde{P}_\varepsilon)x_{M0}}{\sigma(\widetilde{P}_\varepsilon)} \right) = 0, \tag{A10}$$

$$(1 - \varepsilon) \int \phi \left(\frac{x - \alpha(\widetilde{P}_\varepsilon) - \beta(\widetilde{P}_\varepsilon)x_M}{\sigma(\widetilde{P}_\varepsilon)} \right) x_M dP_\theta(x_M, x) + \varepsilon\phi \left(\frac{x_0 - \alpha(\widetilde{P}_\varepsilon) - \beta(\widetilde{P}_\varepsilon)x_{M0}}{\sigma(\widetilde{P}_\varepsilon)} \right) x_{M0} = 0, \tag{A11}$$

$$(1 - \varepsilon) \int \chi \left(\frac{x - \alpha(\widetilde{P}_\varepsilon) - \beta(\widetilde{P}_\varepsilon)x_M}{\sigma(\widetilde{P}_\varepsilon)} \right) dP_\theta(x_M, x) + \varepsilon\chi \left(\frac{x_0 - \alpha(\widetilde{P}_\varepsilon) - \beta(\widetilde{P}_\varepsilon)x_{M0}}{\sigma(\widetilde{P}_\varepsilon)} \right) = 0. \tag{A12}$$

Derivating the first equation with respect to ε and taking the derivatives in $\varepsilon = 0$, we obtain

$$\int \phi' \left(\frac{x - \alpha - \beta x_M}{\sigma} \right) \left[\frac{1}{\sigma}(-IF(x_{M0}, x_0, \alpha, P_\theta) - x_M IF(x_{M0}, x_0, \beta, P_\theta)) \right.$$
$$\left. - \frac{x - \alpha - \beta x_M}{\sigma^2} IF(x_{M0}, x_0, \sigma, P_\theta) \right] dP_\theta(x_M, x) + \phi \left(\frac{x_0 - \alpha - \beta x_{M0}}{\sigma} \right) = 0.$$

After some calculations, we obtain the relation

$$-\frac{1}{\sigma(\gamma + 1)^{3/2}} IF(x_{M0}, x_0, \alpha, P_\theta) - \frac{1}{\sigma(\gamma + 1)^{3/2}} IF(x_{M0}, x_0, \beta, P_\theta)E(X_M) + \phi \left(\frac{x_0 - \alpha - \beta x_{M0}}{\sigma} \right) = 0. \tag{A13}$$

Similarly, derivating with respect to ε Equations (A11) and (A12) and taking the derivatives in $\varepsilon = 0$, we get

$$-\frac{1}{\sigma(\gamma + 1)^{3/2}} E(X_M)IF(x_{M0}, x_0, \alpha, P_\theta) - \frac{1}{\sigma(\gamma + 1)^{3/2}} E(X_M^2)IF(x_{M0}, x_0, \beta, P_\theta)$$
$$+ \phi \left(\frac{x_0 - \alpha - \beta x_{M0}}{\sigma} \right) x_{M0} = 0 \tag{A14}$$

and

$$-\frac{2}{\sigma(\gamma + 1)^{5/2}} IF(x_{M0}, x_0, \sigma, P_\theta) + \chi \left(\frac{x_0 - \alpha - \beta x_{M0}}{\sigma} \right) = 0. \tag{A15}$$

Solving the system formed with the Equations (A13)–(A15), we find the expressions for the influence functions. $\square$

Proof of Theorem 4. In the following, we simply denote by X_{Mj} the vector $(1, X_{Mj})^T$. Then,

$$(\widehat{\alpha}, \widehat{\beta})^T(\{(X_{Mj}, X_j) : j = 1, \ldots, n\})$$

$$= \arg_{(\alpha,\beta)^T} \max_{(\alpha,\beta,\sigma)} \sum_{j=1}^n \sigma^{-\gamma/(\gamma+1)} \exp\left(-\frac{\gamma}{2}\left(\frac{X_j - X_{Mj}^T(\alpha,\beta)^T}{\sigma}\right)^2\right).$$

For any two-dimensional column vector v, we have

$$(\widehat{\alpha}, \widehat{\beta})^T(\{(X_{Mj}, X_j + X_{Mj}^T v) : j = 1, \ldots, n\})$$

$$= \arg_{(\alpha,\beta)^T} \max_{(\alpha,\beta,\sigma)} \sum_{j=1}^n \sigma^{-\gamma/(\gamma+1)} \exp\left(-\frac{\gamma}{2}\left(\frac{X_j + X_{Mj}^T v - X_{Mj}^T(\alpha,\beta)^T}{\sigma}\right)^2\right)$$

$$= \arg_{(\alpha,\beta)^T} \max_{(\alpha,\beta,\sigma)} \sum_{j=1}^n \sigma^{-\gamma/(\gamma+1)} \exp\left(-\frac{\gamma}{2}\left(\frac{X_j - X_{Mj}^T((\alpha,\beta)^T - v)}{\sigma}\right)^2\right)$$

$$= \arg_{((\alpha,\beta)^T - v)} \max_{((\alpha,\beta)^T - v)^T, \sigma)} \sum_{j=1}^n \sigma^{-\gamma/(\gamma+1)} \exp\left(-\frac{\gamma}{2}\left(\frac{X_j - X_{Mj}^T((\alpha,\beta)^T - v)}{\sigma}\right)^2\right) + v$$

$$= (\widehat{\alpha}, \widehat{\beta})^T(\{(X_{Mj}, X_j) : j = 1, \ldots, n\}) + v,$$

which show that $(\widehat{\alpha}, \widehat{\beta})^T$ is regression equivariant.

For any constant $c \neq 0$, we have

$$(\widehat{\alpha}, \widehat{\beta})^T(\{(X_{Mj}, cX_j) : j = 1, \ldots, n\})$$

$$= \arg_{(\alpha,\beta)^T} \max_{(\alpha,\beta,\sigma)} \sum_{j=1}^n \sigma^{-\gamma/(\gamma+1)} \exp\left(-\frac{\gamma}{2}\left(\frac{cX_j - X_{Mj}^T(\alpha,\beta)^T}{\sigma}\right)^2\right)$$

$$= \arg_{(\alpha,\beta)^T} \max_{(\alpha,\beta,\sigma)} \sum_{j=1}^n c^{-\gamma/(\gamma+1)}(\sigma/c)^{-\gamma/(\gamma+1)} \exp\left(-\frac{\gamma}{2}\left(\frac{X_j - X_{Mj}^T((\alpha,\beta)^T/c)}{(\sigma/c)}\right)^2\right)$$

$$= c \cdot \arg_{(\alpha/c,\beta/c)^T} \max_{(\alpha/c,\beta/c,\sigma/c)} \sum_{j=1}^n c^{-\gamma/(\gamma+1)}(\sigma/c)^{-\gamma/(\gamma+1)} \exp\left(-\frac{\gamma}{2}\left(\frac{X_j - X_{Mj}^T((\alpha,\beta)^T/c)}{(\sigma/c)}\right)^2\right)$$

$$= c \cdot (\widehat{\alpha}, \widehat{\beta})^T(\{(X_{Mj}, X_j) : j = 1, \ldots, n\}).$$

This implies that the estimator $(\widehat{\alpha}, \widehat{\beta}) = (\widehat{\alpha}, \widehat{\beta})(\{(X_{Mj}, X_j) : j = 1, \ldots, n\})$ is scale equivariant. Now, for any two-dimensional square matrix A, we get

$$(\widehat{\alpha}, \widehat{\beta})^T(\{(A^T X_{Mj}, X_j) : j = 1, \ldots, n\})$$

$$= \arg_{(\alpha,\beta)^T} \max_{(\alpha,\beta,\sigma)} \sum_{j=1}^n \sigma^{-\gamma/(\gamma+1)} \exp\left(-\frac{\gamma}{2}\left(\frac{X_j - X_{Mj}^T A(\alpha,\beta)^T}{\sigma}\right)^2\right)$$

$$= A^{-1} \arg_{A(\alpha,\beta)^T} \max_{((\alpha,\beta)A^T,\sigma)} \sum_{j=1}^n \sigma^{-\gamma/(\gamma+1)} \exp\left(-\frac{\gamma}{2}\left(\frac{X_j - X_{Mj}^T(A(\alpha,\beta)^T)}{\sigma}\right)^2\right)$$

$$= A^{-1} \cdot (\widehat{\alpha}, \widehat{\beta})^T(\{(X_{Mj}, X_j) : j = 1, \ldots, n\}),$$

which show the affine equivariance of the estimator $(\widehat{\alpha}, \widehat{\beta}) = (\widehat{\alpha}, \widehat{\beta})(\{(X_{Mj}, X_j) : j = 1, \ldots, n\})$. □

Appendix B. The 50 Stocks and Their Abbreviations

1. Asbury Automotive Group, Inc. (ABG)
2. Arctic Cat Inc. (ACAT)
3. American Eagle Outfitters, Inc. (AEO)
4. AK Steel Holding Corporation (AKS)
5. Albany Molecular Research, Inc. (AMRI)
6. The Andersons, Inc. (ANDE)
7. ARMOUR Residential REIT, Inc. (ARR)
8. BJ's Restaurants, Inc. (BJRI)
9. Brooks Automation, Inc. (BRKS)
10. Caleres, Inc. (CAL)
11. Cincinnati Bell Inc. (CBB)
12. Calgon Carbon Corporation (CCC)
13. Coeur Mining, Inc. (CDE)
14. Cohen & Steers, Inc. (CNS)
15. Cray Inc. (CRAY)
16. Cirrus Logic, Inc. (CRUS)
17. Covenant Transportation Group, Inc. (CVTI)
18. EarthLink Holdings Corp. (ELNK)
19. Gray Television, Inc. (GTN)
20. Triple-S Management Corporation (GTS)
21. Getty Realty Corp. (GTY)
22. Hecla Mining Company (HL)
23. Harmonic Inc. (HLIT)
24. Ligand Pharmaceuticals Incorporated (LGND)
25. Louisiana-Pacific Corporation (LPX)
26. Lattice Semiconductor Corporation (LSCC)
27. ManTech International Corporation (MANT)
28. MiMedx Group, Inc. (MDXG)
29. Medifast, Inc. (MED)
30. Mentor Graphics Corporation (MENT)
31. Mistras Group, Inc. (MG)
32. Mesa Laboratories, Inc. (MLAB)
33. Meritor, Inc. (MTOR)
34. Monster Worldwide, Inc. (MWW)
35. Nektar Therapeutics (NKTR)
36. Osiris Therapeutics, Inc. (OSIR)
37. PennyMac Mortgage Investment Trust (PMT)
38. Paratek Pharmaceuticals, Inc. (PRTK)
39. Repligen Corporation (RGEN)
40. Rigel Pharmaceuticals, Inc. (RIGL)
41. Schnitzer Steel Industries, Inc. (SCHN)
42. comScore, Inc. (SCOR)
43. Safeguard Scientifics, Inc. (SFE)
44. Silicon Graphics International (SGI)
45. Sagent Pharmaceuticals, Inc. (SGNT)
46. Semtech Corporation (SMTC)
47. Sapiens International Corporation N.V. (SPNS)
48. Sarepta Therapeutics, Inc. (SRPT)
49. Take-Two Interactive Software, Inc. (TTWO)
50. Park Sterling Corporation (PSTB)

References

1. Sharpe, W.F. A simplified model to portfolio analysis. *Manag. Sci.* **1963**, *9*, 277–293. [CrossRef]
2. Alexander, G.J.; Sharpe, W.F.; Bailey, J.V. *Fundamentals of Investments*; Prentice-Hall: Upper Saddle River, NJ, USA, 2000.
3. Pardo, L. *Statistical Inference Based on Divergence Measures*; Chapman & Hall: Boca Raton, FL, USA, 2006.
4. Basu, A.; Shioya, H.; Park, C. *Statistical Inference: the Minimum Pseudodistance Approach*; CRC Press: Boca Raton, FL, USA, 2011.
5. Basu, A.; Harris, I.R.; Hjort, N.L.; Jones, M.C. Robust and efficient estimation by minimizing a density power divergence. *Biometrika* **1998**, *85*, 549–559. [CrossRef]
6. Jones, M.C.; Hjort, N.L.; Harris, I.R.; Basu, A. A comparison of related density-based minimum divergence estimators. *Biometrika* **2001**, *88*, 865–873. [CrossRef]
7. Broniatowski, M.; Keziou, A. Parametric estimation and tests through divergences and the duality technique. *J. Multivar. Anal.* **2009**, *100*, 16–36. [CrossRef]
8. Toma, A.; Leoni-Aubin, S. Robust tests based on dual divergence estimators and saddlepoint approximations. *J. Multivar. Anal.* **2010**, *101,* 1143–1155. [CrossRef]
9. Toma, A.; Broniatowski, M. Dual divergence estimators and tests: Robustness results. *J. Multivar. Anal.* **2011**, *102*, 20–36. [CrossRef]
10. Fujisawa, H. Eguchi, S. Robust parameter estimation with a small bias against heavy contamination. *J. Multivar. Anal.* **2008**, *99*, 2053–2081. [CrossRef]
11. Broniatowski, M.; Vajda, I. Several applications of divergence criteria in continuous families. *Kybernetica* **2012**, *48*, 600–636.
12. Broniatowski, M.; Toma, A.; Vajda, I. Decomposable pseudodistances and applications in statistical estimation. *J. Stat. Plan. Inference* **2012**, *142*, 2574–2585. [CrossRef]
13. Markowitz, H.M. Mean-variance analysis in portfolio choice and capital markets. *J. Finance* **1952**, *7*, 77–91.
14. Fabozzi, F.J.; Huang, D.; Zhou, G. Robust portfolios: contributions from operations research and finance. *Ann. Oper. Res.* **2010**, *176*, 191–220. [CrossRef]
15. Vaz-de Melo, B.; Camara, R.P. Robust multivariate modeling in finance. *Int. J. Manag. Finance* **2005**, *4*, 12–23. [CrossRef]
16. Perret-Gentil, C.; Victoria-Feser, M.P. *Robust Mean-Variance Portfolio Selection*; FAME Research Paper, No. 140; 2005. Available online: papers.ssrn.com/sol3/papers.cfm?abstract_id=721509 (accessed on 28 February 2018).
17. Welsch, R.E.; Zhou, X. Application of robust statistics to asset allocation models. *Revstat. Stat. J.* **2007**, *5*, 97–114.
18. DeMiguel, V.; Nogales, F.J. Portfolio selection with robust estimation. *Oper. Res.* **2009**, *57*, 560–577. [CrossRef]
19. Toma, A.; Leoni-Aubin, S. Robust portfolio optimization using pseudodistances. *PLoS ONE* **2015**, *10*, 1–26. [CrossRef] [PubMed]
20. Toma, A.; Leoni-Aubin, S. Optimal robust M-estimators using Renyi pseudodistances. *J. Multivar. Anal.* **2013**, *115*, 359–373. [CrossRef]
21. Van der Vaart, A. *Asymptotic Statistics*; Cambridge University Press: New York, NY, USA, 1998.
22. Rousseeuw, P.J.; Leroy, A.M. *Robust Regression and Outlier Detection*; John Wiley & Sons: Hoboken, NJ, USA, 2005.
23. Andersen, R. *Modern Methods for Robust Regression*; SAGE Publications, Inc.: Los Angeles, CA, USA, 2008.
24. Rousseeuw, P.J.; Yohai, V. Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis*; Franke, J., Hardle, W., Martin, D., Eds.; Springer: New York, NY, USA, 1984; pp. 256–272, ISBN 978-0-387-96102-6.
25. Ghosh, A.; Basu, A. Robust estimations for independent, non-homogeneous observations using density power divergence with applications to linear regression. *Electron. J. Stat.* **2013**, *7*, 2420–2456. [CrossRef]