Depth-based portfolio selection

Data depth per la selezione di portafoglio

Giuseppe Pandolfo, Carmela Iorio and Antonio D'Ambrosio

Abstract The estimation of multivariate location and scatter is the cornerstone of the classical multivariate statistical methods widely used in portfolio selection problems. However, they are not robust. We propose to use an alternative non-parametric approach based on the weighted L^P depth as robust location and scatter estimator in order to deal with extreme events in asset returns analysis. We first review weighted L^P depth along with its main properties and then discuss its application to portfolio selection through a small simulation study.

Abstract *Le depth functions La stima dei parametri di location e dispersione rappresentano la pietra miliare dei classici metodi statistici multivariati utilizzati nella selezione di portafoglio. Tuttavia questi metodi non sono robusti. Proponiamo, quindi, l'utilizzo di un metodo non parametrico basato sulla weighted L^P depth come stimatore robusto di location e dispersione nel caso di eventi estremi nell'analisi dei rendimenti. Forniamo la definizione e le proprietlla weighted L^P depth*, e mostriamo la sua applicabilitla selezione di portafoglio attraverso uno studio di simulazione.

Key words: Data depth, Robust estimation, Financial assets, Contaminated model.

Giuseppe Pandolfo
Department of Industrial Engineering, University of Naples Federico II, Napoli, e-mail: giuseppe.pandolfo@unina.it

Carmela Iorio
Department of Industrial Engineering, University of Naples Federico II, Napoli, e-mail: carmela.iorio@unina.it

Antonio D'Ambrosio
Department of Economics and Statistics, University of Naples Federico II, Napoli, e-mail: antdambr@unina.it

1 Introduction

Portfolio selection concerns how to allocate capital over a number of assets to maximize the "return" on the investment and minimizing the "risk". The first mathematical model for portfolio selection was introduced by Markowitz (1952). In this model, the "return" on a portfolio is measured by the expected value of the random portfolio return, and the associated "risk" is given by the variance of the portfolio return. This mean-variance model has had a great impact on the economic modelling of financial markets and represents the milestone of the modern portfolio theory. Despite the success of the mean-variance model, its solutions are often very sensitive to perturbations in the parameters of the problem. Indeed, when the sample distribution deviates even slightly from the assumed distribution, the efficiency of classical estimators may be drastically reduced. Robust estimators, on the other hand, are not as efficient as maximum likelihood estimators when the underlying model is correct, but their properties are not as sensitive to deviations from the assumed distribution. In order to reduce the sensitivity of the Markowitz-optimal several techniques were proposed in the literature. Among the proposals we recall Vaz-de Melo and Camara (2005) who used the M-estimators, Perret-Gentil and Victoria-Feser (2005) who adopted a translated biweight S-estimator, while Welsch and Zhou (2007) used the minimum covariance determinant estimator and the winsorization. DeMiguel and Nogales (2009) proposed a class of policies that are constructed using both on M- and S-estimators. For an overview on the robust methods used for in portfolio selection we refer to Fabozzi, Huang and Zhou (2010).

In recent years, attention on portfolio selection strategies based on non-parametric and semi-parametric techniques have been also shown to exist (see e.g., Ben Salah et al., 2018 and Iorio et al., 2018)

Despite the growing interest on non-parametric and/or robust methods for portfolio selection, to the best of authors knowledge there is no literature on the exploitation of data depth functions to this purpose. Hence, in this paper we propose to adopt a non-parametric approach based on the weighted L^p depth to perform robust location and scatter estimation in financial applications.

In the following we first recall the definition of data depth function and review the notion of L^p depth, then the results of a small simulation are offered to the reader.

2 Data depth concept

Many statistical techniques in multivariate analysis assume normality of the distribution of the data. This assumption is often disputable and thus other, non-parametric, approaches are worth to be considered. One is based on the so called *data depth*, which is a way to measure the depth or outlyingness of a given point with respect to a multivariate data cloud or its underlying distribution. This concept was originally introduced to generalize the concepts of the median and the quantiles to a multivariate framework. The principle is very simple. For a distribution F in $\mathbb{R}^d$, a depth function, denoted by $D(x;F)$, provides a certer-outward ordering of points in $x \in \mathbb{R}^d$.

Zuo and Serfling (2000) provided general notions of depth function on $\mathbb{R}^d$ and presented four reasonable properties that a depth function (bounded and non-negative) should possess, that is:

P1 **Affine invariance**. The depth of a point $x \in \mathbb{R}^d$ should not be dependent on the underlying coordinate system or, in particular, on the scales of the underlying measurements.

P2 **Maximality at center.** For a distribution having a uniquely defined “center”, the depth function should attain maximum value at this center.

P3 **Monotonicity relative to deepest point.** As a point $x \in \mathbb{R}^d$ moves away from the “deepest point” (the point at which the depth function attains maximum value) along any fixed ray through the center, the depth at x should decrease monotonically.

P4 **Vanishing at infinity.** The depth of a point x should approach zero as $\|x\|$ approaches infinity.

Let $\mathscr{P}$ denote the class of distributions on Borel sets on $\mathbb{R}^d$, while F_X denote the distribution of a given random vector X belonging to the class of random vectors X

There are several notions of data depth in the literature.

Definition 1. Let the mapping $D(\cdot,\cdot) : \mathbb{R}^d \times \mathscr{P} \to \mathbb{R}_+$ satisfy P1, P2, P3 and P4. That is, assume:

(i) $D(Ax+b, F_{AX+b}) = D(x, F_X)$ holds for any random vector $X \in \mathbb{R}^d$ and any $d \times d$ nonsingular matrix A, and any d dimensional vector b.

(ii) $D\left(\theta, F\right) = \sup_{x \in \mathbb{R}^d} D(x, F)$ holds for any $F \in \mathscr{P}$ having centre θ.

(iii) For any $F \in \mathscr{P}$ having deepest point θ, $D(x,F) \leq D(\theta + \alpha(x?\theta), F)$ holds for $\alpha \in [0,1]$; and

(iv) $D(x,F) \to 0$ as $\|x\| \to \infty$, for each $F \in \mathscr{P}$.

Then $D(\cdot;F)$ is called a statistical depth function.

The sample version of a depth function $D(\cdot;F)$ is denoted by $D(\cdot;F_n)$, where F is replaced with an empirical measure F_n, computed on a sample $X_n = \{x_1, \ldots, x_n\}$.

There are several notions of data depth function available in the literature. The halfspace, simplicial, Mahalanobis and L^p depths are some of the most popular ones. The notion of data depth has been also extended to the functional space (see e.g., Lopez-Pintado and Romo, 2009) and on the spheres (see e.g., Liu and Singh, 1992 and Pandolfo et al., 2017).

In this paper, we adopt the notion of weighted L^p depth introduced by Zuo (2004) because of its ease of computation and (local and global) robustness properties.

2.1 The weighted L^p depth

Zuo and Serfling (2000) defined the L^p depth based on the L^p-norm. Different distances (norms) were used with equal weights. However, in practice, the importance (weight, cost, penalty, or incentive) may not be the same for different distances (norms). This motivates to adopt this notion of depth, that is define as follows:

$$WL^pD(x;F) = \frac{1}{1 + Ew\left(\|x - X\|_p\right)},$$

where w is a weight function on $[0,1)$, $X \sim F$ and $\|\cdot\|$ denotes the L^p-norm (when $p = 2$ we have the Euclidean norm), w is assumed to be non-decreasing and continuous on $[0,\infty)$. The weighted L^p depth possesses some desirable properties of depth functions. It is translation invariant (can be affine invariant for $p = 2$ under some modification), maximized at the center of a (centrally) symmetric distribution for convex w, decreasing when a point moves along a ray stemming from the deepest point, and vanishing at infinity. For more related discussions see Zuo and Serfling (2000).

The weighted L^p depth-induced medians (multivariate location estimator) are globally robust with the highest breakdown point for any reasonable estimator. The weighted L^p medians are also locally robust with bounded influence functions for suitable weight functions. Unlike other existing depth functions and multivariate medians, the weighted L^p depth and medians are computationally feasible and easy to calculate in high dimensions. The price to be paid is the lack of affine invariance.

3 Simulation study

In this section, we present a small Montecarlo simulation to investigate the performance of the weighted L^p depth-based estimators of the mean and covariance matrix, for both contaminated and non-contaminated simulated data.

Following Toma and Leoni-Aubin (2015), and DeMiguel and Nogales (2009), we use simulations to generate asset returns data following a distribution that deviates slightly from the normal distribution. Specifically, we considered the multivariate normal distribution $F \sim N(\mu_F, \Sigma_F)$ with mean $\mu_F = 0$ and Σ_F a $N \times N$ covariance matrix with variances equal to 1 and covariances all equal to 0.2, with N denoting the total number of assets. We generated samples of size $T = 100$ according to the following contaminated model:

$$F_{\varepsilon} = (1-\varepsilon)F + \varepsilon G, \tag{1}$$

where G is a contaminating distribution and ε is the fraction of the data that follows the contaminating distribution $G \sim N(\mu_G, \Sigma_G)$ with $mu_G = -4$ and $\Sigma_G = 4\Sigma_F$. We considered $N \in \{2, 5, 10, 20\}$ and three different contamination levels $\varepsilon \in \{0\%, 2.5\%, 5\%\}$ to investigate how the estimates change when the asset returns deviate from normality.

The estimates of location and scatter were obtained through the weighted L^2, with all the observations having same weight. More in detail, in case of location, the estimator is defined as:

$$L(F) = \int x w_1 \left(L^2 D(x,F)\right) dF(x) \Big/ \int w_1 \left(L^2 D(x,F)\right) dF(x),$$

then a weighted L^2 depth scatter estimator is defined as

$$S(F) = \frac{\int (x - L(F))(x - L(F))^T w_2 (D(x,F)) \, dF(x)}{\int w_2 (D(x,F)) \, dF(x)}$$

where w_2 are suitable weight function that can be different from w_1.

For each setting, we generated $R = 250$ samples and for each sample we computed the depth estimates of location and scatter. The performances are evaluated through the empirical mean squared error (EMSE) given by

$$EMSE = \frac{1}{R} \sum_{i=1}^{R} \left\| \hat{\theta}_i - \theta_0 \right\|^2$$

where $\theta_0 = (\mu_F, vech(\Sigma_F))'$ and $\hat{\theta}_i = \left(\hat{\mu}_i, vech\left(\hat{\Sigma}_i\right)\right)'$ is an estimate corresponding to the i-th sample, while $vech(\Sigma)$ 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.

Results are presented in Table 1. The mean squared errors generally increase along with the number of assets (i.e., the sample size). However, for low dimensions ($N = 2$ and 5), the L^2 depth-based method estimates appear to be less affected by the contamination.

4 Final comments

In this paper we suggest to exploit the use of the weighted L^p depth function to perform robust estimation in financial settings. The very first results obtained through simulations are promising. Further research are needed to determine how to assign (depth-)weights to the observations, and to investigate the behaviour in real data applications.

Table 1: Empirical mean squared errors of WL^2 depth estimates.

N	ε		
	0%	2.5%	5%
2	0.10	0.12	0.48
5	0.71	0.95	2.57
10	2.89	3.76	9.46
20	9.23	12.58	32.50

References

1. Ben Salah, H., Chaouch, M., Gannoun, A., de Peretti, C., Trabelsi, A.: Mean and median-based nonparametric estimation of returns in mean-downside risk portfolio frontier. Annals of Operations Research **262**, 653–681 (2018)
2. DeMiguel, V., Nogales, J.F.: Portfolio selection with robust estimation. Operations Research **57**, 560–577 (2009)
3. Fabozzi, F.J., Huang, D., Zhou, G.: Robust portfolios: contributions from operations research and finance. Annals of Operations Research **176**, 191–220 (2010)
4. Iorio, C., Frasso, G., D'Ambrosio, A., Siciliano, R.: A P-spline based clustering approach for portfolio selection. Expert Systems With Applications **104**, 88–103 (2018)
5. Liu, R.Y., Singh, K.: Ordering directional data: Concepts of data depth on circles and spheres. The Annals of Statistics **20**, 1468–1484 (1992)
6. Lopez-Pintado, S., Romo, J.: On the concept of depth for functional data. The Annals of Statistics **104**, 718–734 (2009)
7. Markowitz, H.: Portfolio selection. The Journal of Finance **7**, 77–91 (1952)
8. Pandolfo, G., Paindaveine, D., Porzio, G.C.: Distance-based depths for directional data. Working Papers ECARES 2017 – 35 (2017) Available at `https://ideas.repec.org/p/eca/wpaper/2013-258549.html`
9. Perret-Gentil, C., Victoria-Feser, M.P.: Robust mean-variance portfolio selection. FAME research paper 140 (2005) Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=721509`
10. Toma, A., Leoni-Aubin, S.: Robust portfolio optimization using pseudodistances. PLoS ONE (2015) Available at `e0140546.https://doi.org/10.1371/journal.pone.0140546`
11. Vaz-de Melo, B., Camara R.P.: Robust multivariate modeling in finance. International Journal of Managerial Finance **4**, 12–23 (2005)
12. Welsch, R.E., Zhou, X.: Application of robust statistics to asset allocation models. REVSTAT – Statistical Journal **5**, 97–114 (2007)
13. Zuo, Y., Serfling, R.: General notions of statistical depth functions. Annals of Statistics textbf28, 461–482 (2000)
14. Zuo, Y.: Robustness of weighted L^p-depth and L^p-median. Allgemeines Statistisches Archiv textbf28, 215–234 (2004)