Università degli Studi "G. D'Annunzio"
Dipartimento di Scienze

Statistical Dominance Over Pareto Optimal Mean-variance Portfolios

Giacomo di Tollo Pamela Peretti
Mauro Birattari Eliseo Ferrante
Thomas Stuetzle

February 2, 2010

Technical Report no. R-2010-001 *Research Series*

Statistical Dominance Over Pareto Optimal Mean-variance Portfolios

Giacomo di Tollo [1] Pamela Peretti [1] Mauro Birattari [2]
Eliseo Ferrante [2] Thomas Stuetzle [2]

[1]
Dipartimento di Scienze, Universitá di Chieti-Pescara
Viale Pindaro, 42 Pescara, Italy
ditollo@sci.unich.it, peretti@sci.unich.it
[2]
IRIDIA,ULB
50, Av. F. Roosevelt, CP 194/6,Brussels, Belgium
mbiro@ulb.ac.be, eferrante@iridia.ulb.ac.be, stuetzle@ulb.ac.be

February 2, 2010

Abstract. Portfolios belonging to the Mean-Variance Efficient set are Pareto-Efficient. In this work we outline a procedure to understand if this assertion is always true in practical sense, or if some portfolios belonging to the frontier can be considered as dominated in a probabilistic way.

Keywords: *Statistical Dominance, Efficient Set, Portfolio Selection*

Contents

Introduction: The Mean Variance Portfolio Selection

Portfolio selection is one of the most studied topics in finance: the problem (referred to as PSP), in its basic formulation, is concerned with selecting the portfolio of assets that minimizes the risk, given a certain level of returns. The basic model is formulated in the seminal work by Markowitz [13], in which the problem is stated as the minimisation of the variance (as a risk measure) for a given level of return r_p:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j \tag{1}$$

$$\sum_{i=1}^{n} r_i x_i \geq r_p \quad \sum_{i=1}^{n} x_i = 1 \quad x_i \in [0, 1] \quad i, j = 1 \dots n \tag{2}$$

where x_i is the quantity invested in asset i, σ_{ij} represents covariance between assets i and j, r_p is the minimum desired return rate and r_i is the (actual or forecasted) return rate of asset i. Note that portfolios are modeled as sets of assets whose weights sum up to one and can assume any value in the range [0, 1]. This means that short selling is not allowed.

In this formulation, the problem is solvable with exact methods. By solving the problem for each level of return, we obtain a Pareto Efficient Frontier composed of non dominated points. This means that a rational investor should use an external criterion in order to choose a portfolio out of the set at hand. The goal of this work is to define a basic methodology in order to understand *if* portfolios over the pareto frontier are trully non dominated, and if there exists some mechanism in order to drive the investor choice amongst such portfolios: Our purpose is to check whether some of the point in the frontier are dominated by others in a *probabilistic* sense. The need for this control arises as the two objectives we are taking into account are not unrelated quantities, but rather the mean and the variance of the same random variable: Asset returns. In this sense, it is easy to devise a procedure able to show that two points are not non-dominated. Please notice that in the following we try to devise a method just relying on statistical and analytical procedures, so we do not take into account more complex portfolio theory such as utility analysis and statistical dominance [17, 2].

1 An analytical approach to Mean Variance

In a bi-dimensional space of risk and return, a solution s is said to be efficient (Pareto-optimal) if there is no other solution s_1 such that $return(s_1) > return(s)$ and $risk(s_1) \leq risk(s)$ or $return(s_1) \geq return(s)$ and $risk(s_1) < risk(s)$; furthermore, as seen before, minimizing the variance for several levels of return leads us obtain a frontier composed of (purportedly) non dominated points.

Let us take two portfolios (points) over this frontier and let us denote the first point as h and the second as l. We must be aware that these points are computed from hystorical data, but we can conceive them as the state representing the future outcome of a portfolio: This outcome could be described as a normal distribution with a given mean (expected return) and a given standard deviation (risk), so we can describe the two points as follows[1]:

$$h \sim N(\mu_h, \sigma_h^2) \tag{3}$$

$$l \sim N(\mu_l, \sigma_l^2) \tag{4}$$

Applying the basic distribution properties (see tab 1), we can draw that the true return of portfolio l is supposed to be in the interval

$$[\mu_l - k \cdot \sigma_l, \quad \mu_l + k \cdot \sigma_l] \tag{5}$$

with Confidence(k) likelihood. For example, the true return of the portfolio has 99.9% likelihood to stay in the interval

$$[\mu_l - 3 \cdot \sigma_l, \quad \mu_l + 3 \cdot \sigma_l] \tag{6}$$

Given that, we can introduce a new normally distributed variable d, defined as the difference between the two portfolios and represented as follows:

$$d \sim N(\mu_h - \mu_l, \sigma_h^2 + \sigma_l^2) \tag{7}$$

K	Confidence(k)
1	84.2 %
2	97.6 %
3	99.7 %

Table 1. *Basic Distribution Properties*

And we can state that portfolio h statistically dominates portfolio l at confidence level $Confidence(k)$ if

$$\mu_d - k \cdot \sigma_d \geq 0 \Longrightarrow \mu_d \geq k \cdot \sigma_d \tag{8}$$

This becomes

$$\mu_h - \mu_l \geq k \cdot \sqrt{\sigma_h^2 + \sigma_l^2} \tag{9}$$

[1]It is not necessary to assume normality of return distribution in order to carry on with our discussion, as the Chebyshev inequality holds for any distribution. Anyhow, following the well established assumption, we will think of returns as normally distributed.

$$\mu_h \geq \mu_l + k \cdot \sqrt{\sigma_h^2 + \sigma_l^2} \tag{10}$$

and eventually

$$\mu_h - \mu_l \geq k \cdot \sqrt{\sigma_h^2 + \sigma_l^2} \tag{11}$$

This means that, picking two portfolios h and l over the Mean Variance Space, we can state that portfolio h probabilistically dominates l iif equation 11 holds.

2 Experimental Analysis

We used five instances in order to see if the aforementioned formula holds over some efficient portfolios. As instance set we used a group of five instances taken from the repository `ORlib`[2]. These instances have been used by [3, 1, 14, 15] and are referring to five well-known stock exchange indices. The following table shows our benchmark size.

Inst.	Origin	assets
1	Hong Kong	31
2	Germany	85
3	UK	89
4	USA	98
5	Japan	225

Table 2. *The benchmark instances.*

For each instance we have the Unconstraint Efficient Frontier, as provided by the very same repository, showing, for each portfolio, its mean return and its return variance. So we have firstly applied this test to the Unconstraint Case over the 5 instances at hand. Then, we added constraints to the formulation in order to analyse the behaviour of different efficient frontiers. To this goal, two cases were taken into account:

- Loose Cardinality Constraint: Introducing a binary variable z_i (equal to 1 if asset i is in the portfolio and 0 otherwise) the constraint can be expressed as follows:

$$\sum_{i=1}^{n} z_i \leq k \tag{12}$$

being imposed to facilitate the portfolio management and to reduce its management costs. Most works on Portfolio Selection Problems introduce this

[2] available at the URL `http://mscmga.ms.ai.ac.uk/~jeb/orlib/portfolio.html`

constraint [7, 9, 16, 6, 11]. It has been experimentally shown that when the cardinality constraint is imposed the ACEF tends to tightly approximate the UEF for high values of k [10, 4], so we took into account several values of k, starting from 2 up to the maximum possible number[3];

- Strict Cardinality Constraint: The same constraint as before, but imposing a fixed number of stocks rather than a maximum

$$\sum_{i=1}^{n} z_i = k \tag{13}$$

This latter formulation can be introduced to investigate the effect of the marginal cardinality increase[12, 8]

Quantity constraint has been added in top of the cardinality, imposing minimum and maximum proportions (ε_i and δ_i respectively) to be held for each asset, so that

$$\varepsilon_i z_i \leq x_i \leq \delta_i z_i \tag{14}$$

Ceiling constraints (i.e., upper bound constraints) are introduced to avoid excessive exposure to a specific asset and in some case are imposed by law; Floor constraint (i.e., lower bound) is used to avoid the cost of administrating very small portions of assets. In our case, we set $\varepsilon_i = 0.01$ and $\delta_i = 1 \quad \forall i$. Please notice that this constraint impose implicitly a loose cardinality constraint, as the maximum achievable asset number k will happen to be $\frac{1}{\varepsilon}$. In our case, the maximum k number will be 100 $(= \frac{1}{0.01})$ The computation was made through a Quadratic Programming solver[4] already used by [8]. Test implementation has been developed with Matlab and for

each frontier compares the *highest return - highest risk* and *lowest return - lowest risk* points ,i.e. point with maximum non-dominance likelihood.

3 Results Achieved

As stated so far, we performed our test over three classes of instances:

- Unconstraint Efficient Frontier;
- Loose Cardinality Efficient Frontier
- Strict Cardinality Efficient Frontier

[3] W.r.t. the quantity constraint, see afterwards.
[4] Publicly available at `http://www.diegm.uniud.it/digaspero/index.php?page=software`.

As result of the test, we found out that the hypothesis

$$H_0 = h \ \ dominates \ \ l \tag{15}$$

cannot be accepted even with k values lower than 1.

This simple test provide us with evidence about the Markowitz model skill in capturing the basic properties of the problem: It gives evidence to Mean Variance Portfolios being not dominated and encourages investors in defining their strategies to develop their preferences. But it is worthwile to notice here that Mean Variance theory suffers from several drawbacks:

1. The estimation of return and covariance (used for defining the risk) from historical data is very sensitive to measurement errors[5];

2. The model is nowadays considered too simplistic for practical purposes, because it does not incorporate non-negligible aspects of real-world trading, such as maximum size of portfolio, minimum lots, transaction costs, preferences over assets, management costs, etc

This two considerations suggest two further developments of our testing-approach: First, we should introduce more uncertainty about the model: So far, we introduced uncertainty only over the space of return (i.e., just over one of the two dimensions) stating that the actual return could be considered as belonging to an interval rather than being a point. Next step will be the introduction of uncertainty over the remaining dimension;
Latter, more experiments embedding other risk measures, constraint settings and temporal horizon are needed. In principle, different observations over years can be witnessing different financial-economic cycles and conditions. It is not clear if these different conditions have impact over our test or not.

References

[1] R. Armañanzas and J. A. Lozano. A multiobjective approach to the portfolio optimization problem. In *The 2005 IEEE Congress on Evolutionary Computation (CEC 2005)*. Springer Verlag, 2005.

[2] P. Artzner, F. Delbaen, J.M Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–208, 1999.

[3] T.-J. Chang, N. Meade, J. E. Beasley, and Y. M. Sharaiha. Heuristics for cardinality constrained portfolio optimisation. *Computer Operation Research*, 27(13):1271–1302, 2000.

[4] T.J. Chang, N. Meade, J.E. Beasley, and Y.M. Sharaiha. Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, 27(13):1271–1302, 2000.

[5] V.K. Chopra and W.T. Ziemba. The effect of errors in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19:6–11, 1993.

[6] Y. Crama and M. Schyns. Simulated annealing for complex portfolio selection problems. *European Journal of Operational Research*, 150:546–571, 2003.

[7] Ulrich Derigs and Nils-H. Nickel. On a local-search heuristic for a class of tracking error minimization problems in portfolio management. *Annals of Operations Research*, 131(1-4):45–77, October 2004.

[8] L. De Gaspero, G. di Tollo, A. Roli, and A. Schaerf. Hybrid local search for constrained financial portfolio selection problems. In *Proceedings of CPAIOR*, 2007.

[9] Manfred Gilli and Evis Kllezi. The threshold accepting heuristic for index tracking. In Panos Pardalos and Vassilis Tsitsiringos, editors, *Financial engineering, E-commerce, and supply chain*. Kluwer Academic Press, Boston, Mass., 2002.

[10] J. Matatko J.E. Fieldsend and M. Peng. Cardinality constrained portfolio optimisation. In *Proceedings of the Fifth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'04), Lecture Notes in Computer Science (LNCS 3177)*, pages 788–793. Springer, 2004.

[11] H. Kellerer and D. Maringer. Optimization of cardinality constrained portfolios with a hybrid local search algorithm. *OR Spectrum*, 25(4):481–495, 2003.

[12] Dietmar Maringer and Olufemi Oyewumi. Index tracking with constrained portfolios. *Intelligent Systems in Accounting, Finance and Management*, 15:57–71, 2007.

[13] H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

[14] R. Moral-Escudero, R. Ruiz-Torrubiano, and A. Suárez. Selection of optimal investment with cardinality constraints. In *Proceedings of the IEEE World Congress on Evolutionary Computation (CEC 2006)*, pages 2382–2388, 2006.

[15] A. Schaerf. Local search techniques for constrained portfolio selection problems. *Computational Economics*, 20(3):177–190, 2002.

[16] A. Schaerf. Local search techniques for constrained portfolio selection problems. *Comput. Econ.*, 20(3):177–190, 2002.

[17] G.A. Whitmore and M.C. Findlay (eds.). *Stochastic dominance: An approach to Decision-making under risk.* D.C.Heath, 1978.