Automatica 49 (2013) 370–380

Contents lists available at SciVerse ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Direct data-driven portfolio optimization with guaranteed shortfall probability[☆]

Giuseppe Carlo Calafiore [1]

Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy

ARTICLE INFO

Article history:
Received 7 October 2011
Received in revised form
31 August 2012
Accepted 5 September 2012
Available online 6 December 2012

Keywords:
Scenario design
Portfolio optimization
Asset allocation
Data-driven methods
Random convex programs
Value-at-Risk

ABSTRACT

This paper proposes a novel methodology for optimal allocation of a portfolio of risky financial assets. Most existing methods that aim at compromising between portfolio performance (e.g., expected return) and its risk (e.g., volatility or shortfall probability) need some statistical model of the asset returns. This means that: (i) one needs to make rather strong assumptions on the market for eliciting a return distribution, and (ii) the parameters of this distribution need be somehow estimated, which is quite a critical aspect, since optimal portfolios will then depend on the way parameters are estimated. Here we propose instead a direct, data-driven, route to portfolio optimization that avoids both of the mentioned issues: the optimal portfolios are computed directly from historical data, by solving a sequence of convex optimization problems (typically, linear programs). Much more importantly, the resulting portfolios are theoretically backed by a guarantee that their expected shortfall is no larger than an a-priori assigned level. This result is here obtained assuming efficiency of the market, under no hypotheses on the shape of the joint distribution of the asset returns, which can remain unknown and need not be estimated.

1. Preliminaries

Consider a collection of assets $a_1, \ldots, a_n$, and let $p_i(k)$ be the market price of a_i at time kT, where T is a fixed period of time, e.g., say, one minute, one day, one month, or one year. The simple return of an investment in asset i over the k-th period from $(k-1)T$ to kT is

$$r_i(k) = \frac{p_i(k) - p_i(k-1)}{p_i(k-1)}, \quad i = 1, \ldots, n;\ k = 1, 2, \ldots.$$

We denote with $r(k) \doteq [r_1(k) \cdots r_n(k)]^\top$ the vector of assets' returns over the k-th period, and we make the following standard working assumption.

Assumption 1. The returns $\{r(k)\}_{k=1,2,\ldots}$ form an i.i.d. (independent, identically distributed) random sequence. In particular, each $r(k)$ is distributed according to the same and possibly unknown probability distribution $\mathbb{P}$ having support $\Delta \subseteq \mathbb{R}^n$. □

Besides assuming that $\{r(k)\}$ is an i.i.d. sequence, we shall make no further assumption on the probability distribution $\mathbb{P}$, and all subsequent results do not require $\mathbb{P}$ to be known. Assumption 1 is compatible with the classical Efficient Market Hypothesis (EMH), see, e.g., Chapter 2 of Campbell, Lo, and MacKinlay (1996). Although this hypothesis is debated (for example, in Behavioral Finance, see, e.g., Shleifer, 2000), it still remains, in practice, the foundation of modern portfolio theory (MPT). Entering in such a discussion, however, is far beyond the scope of the present paper: here we take a pragmatic position and accept the i.i.d. hypothesis since it is widely assumed in most of the existing computational models for portfolio optimization.

A *portfolio* of assets $a_1, \ldots, a_n$ is defined by a vector $x \in \mathbb{R}^n$ whose entry x_i, $i = 1, \ldots, n$, describes the (signed) fraction of an investor's wealth invested in asset a_i, where $x_i \geq 0$ denotes a "long" position, and $x_i < 0$ denotes a "short" position. We assume the portfolio to be self-financing, hence $\sum_{i=1}^n x_i = 1$, a condition that we shall write more compactly as $\mathbf{1}^\top x = 1$, where $\mathbf{1}$ is an n-vector of ones. The portfolio vector may have additional constraints. For example, if short-selling is not allowed then it must be $x \geq 0$ (element-wise inequality). Other common constraints on x include minimum and maximum exposure in an individual asset, or limits in the exposure over classes of assets, etc. In this paper, we shall treat the problem in reasonable generality by assuming that the portfolio vector is constrained in a polytope (a bounded polyhedron) $\mathcal{X}$. For instance, the classical Markowitz case is given by the conditions $\mathbf{1}^\top x = 1$, $x \geq 0$ (no short-selling), in which case $\mathcal{X}$ is the standard simplex.

Assumption 2. Portfolio composition constraints are expressed by the condition $x \in \mathcal{X}$, where $\mathcal{X}$ is a given nonempty polytope. □

☆ This work is supported by PRIN grant 20087W5P2K from Italian Ministry of University and Research. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Editor Berç Rüstem.

E-mail address: giuseppe.calafiore@polito.it.

[1] Tel.: +39 011 090 7071; fax: +39 011 090 7071.

0005-1098/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2012.11.012

Under Assumption 1, the portfolio return over any period of duration Δ is described by a random variable $\varrho(x) = r^\top x$, where r is distributed according to $\mathbb{P}$. Much of portfolio optimization literature is concerned with determining x so that the probability distribution of $\varrho(x)$ has some desirable shape. This is an "easy" problem, if the distribution of r (or some relevant characteristic of it) is known. In reality, however, this is actually a very hard problem, since the distribution of r is *not* known. For example, classical portfolio theory assumes that one knows the expected value and the covariance matrix of r. Under this hypothesis, a sensible optimization problem can be defined and easily solved, seeking an optimal tradeoff between expected return (the more the better) and risk quantified by the variance of $\varrho(x)$ (the less the better). However, the expected value and the covariance of r are *not* known in practice. Of course, they can be *estimated* by using both prior knowledge and historical data, but then the use of the estimated quantities in place of the true (unknown) ones in the optimization problem rises a wealth of theoretical (and practical) issues, since the result of optimization will be sensitive to errors in the estimation and thus potentially unreliable, see, e.g., Bawa, Brown, and Klein (1979) and Chopra and Ziemba (1993). Interestingly, a new wave of literature recently emerged, trying to cope with the problem of unreliable data via the technology of robust optimization. In robust portfolio optimization, uncertainties in the data are taken into account on a worst-case basis, see, e.g., Ben-Tal, Margalit, and Nemirovski (2000), Bertsimas and Sim (2004), El Ghaoui, Oks, and Oustry (2003), Goldfarb and Iyengar (2003) and the references therein. While effective in some cases, the robust optimization approach to portfolio allocation suffers from two drawbacks: first, being a worst-case approach, it tends to be conservative and to yield overly pessimistic results that may be useless in practice (this issue may be mitigated to some extent by flexibly adjusting the level of conservatism of the robust solutions in terms of probabilistic bounds of constraint violations, as proposed in Bertsimas & Sim, 2004, or by incorporating several layers of robustness corresponding to uncertainty sets of different sizes, as done in Zymler, Rustem, & Kuhn, 2011). Second, the robust approach is an *indirect* approach: observed data is first used to compute nominal estimates of the distribution parameters, along with some regions of confidence around them, then a suitable robust optimization program is solved by taking into account this information. Each of these steps may involve restrictive assumptions and various degrees of conservatism. In the same direction, a learning-theoretic approach is used in Delage and Ye (2010) to precisely quantify the uncertainty set in estimation of mean and covariance from finite data, and hence to derive a robust portfolio optimization model, see Sections 3 and 4 in Delage and Ye (2010).

In this work, we take a radically different route to reliable portfolio optimization. Our route is *direct* in that it does not rely on a two phases (estimation/optimization) approach. Rather, we use directly the observed data to construct the optimal allocation. Then, we leverage on the theory of random convex programs (RCP), see Calafiore (2010), Calafiore and Campi (2005, 2006) and Campi and Garatti (2008), to attach to the computed portfolio a precisely guaranteed level of shortfall probability, under no additional hypotheses beyond Assumption 1. Among other benefits, the proposed approach makes fully transparent the fundamental link between the depth of the historical data upon which the optimal allocation is computed (look-back period), and the resulting reliability of the computed portfolio. A similar direct approach has been recently proposed in Jabbour, Pena, Vera, and Zuluaga (2008) in the context of Conditional Value-at-Risk (CVaR) portfolio optimization. The key point, however, is that the approach of Jabbour et al. (2008) does not guarantee theoretically the out-of-sample (i.e., future) behavior of the computed portfolio, which is instead the main feature of the methodology developed here.

Consider now a sequence of returns of finite length N : $r(1), r(2), \ldots, r(N)$, and collect these return vectors by rows in a matrix R_N:

$$R_N^\top = [r(1)\, r(2) \cdots r(N)] \in \mathbb{R}^{n,N}.$$

Notice that R_N is a random matrix, with each column independently distributed according to the unknown distribution $\mathbb{P}$; events related to R_N are thus measured by the product probability measure $\mathbb{P}^N$, having support Δ^N. If $x \in \mathcal{X}$ is a given portfolio vector, then such a portfolio would produce the following random sequence of returns over the time interval $k = 1, \ldots, N$:

$$\rho_N(x) = R_N x = [\varrho_1(x)\, \varrho_2(x) \cdots \varrho_N(x)]^\top \in \mathbb{R}^N,$$

where $\varrho_i(x) \doteq r^\top(i)x$, $i = 1, \ldots, N$.

2. Portfolio allocation as a random optimization problem

2.1. The observation selection rule

Let $q \leq N-n-1$ be a given nonnegative integer. We introduce a *rule* $\mathcal{S}_q$ for selecting a subset of cardinality $N-q$ of the returns in R_N. Rule $\mathcal{S}_q$ takes as input the matrix R_N and returns a partition $\mathcal{I}_q$, $\mathcal{I}_{N-q}$ of the set of indices $\mathcal{I} = \{1, \ldots, N\}$, such that, with probability one, the following properties are satisfied:

(a) $|\mathcal{I}_q| = q$, $|\mathcal{I}_{N-q}| = N - q$, and $\mathcal{I}_q \cup \mathcal{I}_{N-q} = \mathcal{I}$, $\mathcal{I}_q \cap \mathcal{I}_{N-q} = \emptyset$;
(b) the partition is independent of the order of the columns in R_N;
(c) Let γ^*, x^* denote the optimal solutions of the following optimization problem:

$$\max_{x\in\mathcal{X},\gamma} \gamma \text{ subject to: } \varrho_i(x) \geq \gamma, \quad i \in \mathcal{I}_{N-q}. \tag{1}$$

Then, it must be $\varrho_i(x^*) < \gamma^*$, for all $i \in \mathcal{I}_q$.

The rationale behind the introduction of such a selection rule is explained next. Suppose first that $q = 0$, then the optimization problem in (1) would determine an optimal portfolio x^* and an optimal return level γ^* which is the largest possible lower bound for *all* the returns $\varrho_i(x^*)$, $i = 1, \ldots, N$. Such a return level γ^* would however be typically low and uninteresting from an investment point of view, since it is the minimum return in the sequence $\{\varrho_i(x^*)\}$. Indeed, for $q = 0$, γ^* is the optimal level of the following min/max game:

$$\gamma^* = \max_{x\in\mathcal{X}} \min_{i=1,\ldots,N} \varrho_i(x).$$

It seems then reasonable to look for a return level γ such that $\varrho_i(x) \geq \gamma$ for *many*, albeit not all, of the $\varrho_i(x)$, while allowing the requirement $\varrho_i(x) \geq \gamma$ to be *violated* on q of the returns in the sequence. This is precisely what the selection rule does: it selects q returns in the sequence $\{\varrho_i(x)\}$ such that γ^* is the largest lower bound over a (suitably selected, see Remark 1) subset of $N - q$ returns, while q of the returns fall below γ^*. Obviously, the advantage of doing so is to obtain a return level γ^* which is generally larger than the level obtained for $q = 0$. Fig. 1 illustrates this idea on a simplified situation where portfolio x^* is held fixed, i.e., where the maximization in (1) is performed only on γ, with x fixed.

As it will be made rigorously clear in the next section, we are in the presence of a fundamental tradeoff here: while level γ^* increases by increasing q, intuitively this level also becomes less and less *reliable* that is, informally, the probability of the actual portfolio return $\varrho(x^*)$ being larger than γ^* decreases. This fact should not come too much as a surprise, since level γ^* can also be interpreted as the empirical (q/N)-quantile of the return sequence $\{\varrho_i(x^*)\}_{i=1,\ldots,N}$.