

Direct Data-Driven Portfolio Optimization with Guaranteed Shortfall Probability

*Original*

Direct Data-Driven Portfolio Optimization with Guaranteed Shortfall Probability / Calafiore, Giuseppe Carlo. - In: AUTOMATICA. - ISSN 0005-1098. - STAMPA. - 49:(2013), pp. 370-380. [10.1016/j.automatica.2012.11.012]

*Availability:*

This version is available at: 11583/2503507 since:

*Publisher:*

Elsevier

*Published*

DOI:10.1016/j.automatica.2012.11.012

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

# Direct Data-Driven Portfolio Optimization with Guaranteed Shortfall Probability <sup>★</sup>

Giuseppe Carlo Calafiore <sup>a</sup>

<sup>a</sup>*Dipartimento di Automatica e Informatica, Politecnico di Torino, Italy*

---

## 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.

*Key words:* Scenario design, Portfolio optimization, Asset allocation, Data-driven methods, Random convex programs, Value-at-Risk.

---

## 1 Preliminaries

Consider a collection of assets  $a_1, \dots, a_n$ , 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

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

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  $\{r(k)\}_{k=1,2,\dots}$  form an iid (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  $\Delta \subseteq \mathbb{R}^n$ .* ★

---

★ This work is supported by PRIN grant 20087W5P2K from Italian Ministry of University and Research.

Email address: [giuseppe.calafiore@polito.it](mailto:giuseppe.calafiore@polito.it)  
(Giuseppe Carlo Calafiore).

Besides assuming that  $\{r(k)\}$  is an iid 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 Markets Hypothesis (EMH), see Chapter 2 of [8]. Although this hypothesis is debated (for example, in Behavioral Finance, see, e.g., [21]), 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 iid hypothesis since it is widely assumed in most of the existing computational models for portfolio optimization.

A *portfolio* of assets  $a_1, \dots, a_n$  is defined by a vector  $x \in \mathbb{R}^n$  whose entry  $x_i$ ,  $i = 1, \dots, n$ , 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 ex-

posure 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)  $\mathcal{X}$ . 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.* ★

Under Assumption 1, the portfolio return over any period of duration  $\Delta$  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., [1,12]. 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., [2,3,14,15] 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 [3], or by incorporating several layers of robustness corresponding to uncertainty sets of different sizes, as done in [22]). 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 [13] 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 [13].

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 [5–7,10], 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 [16] in the context of Conditional Value-at-Risk (CVaR) portfolio optimization. The key point, however, is that the approach of [16] 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  $N$ :  $r(1), r(2), \dots, r(N)$ , and collect these return vectors by rows in a matrix  $R_N$ :

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

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  $\Delta^N$ . 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  $k = 1, \dots, N$ :

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

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

## 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  $\mathcal{I} = \{1, \dots, N\}$ , 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  $\gamma^*, x^*$  denote the optimal solutions of the following optimization problem:

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

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  $\gamma^*$  which is the largest possible lower bound for *all* the returns  $\varrho_i(x^*)$ ,  $i = 1, \dots, N$ . Such a return level  $\gamma^*$  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  $q = 0$ ,  $\gamma^*$  is the optimal level of the following min/max game:

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

It seems then reasonable to look for a return level  $\gamma$  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  $\gamma^*$  is the largest lower bound over a (suitably selected, see Remark 1) subset of  $N - q$  returns, while  $q$  of the returns fall below  $\gamma^*$ . Obviously, the advantage of doing so is to obtain a return level  $\gamma^*$  which is generally larger than the level obtained for  $q = 0$ . Figure 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  $\gamma$ , with  $x$  fixed.

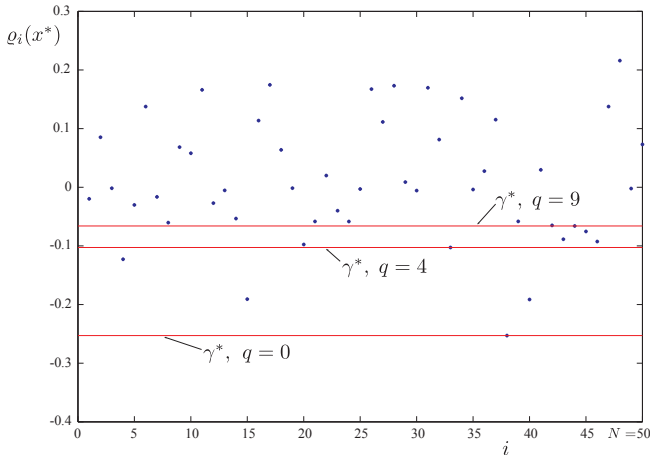


Fig. 1. Illustration of  $\gamma^*$  levels for  $q = 0, q = 4, q = 9$ , for a fixed portfolio  $x^*$ .

As it will be made rigorously clear in the next section, we are in the presence of a fundamental tradeoff here: while

level  $\gamma^*$  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  $\gamma^*$  decreases. This fact should not come too much as a surprise, since level  $\gamma^*$  can also be interpreted as the empirical  $(q/N)$ -quantile of the return sequence  $\{\varrho_i(x^*)\}_{i=1, \dots, N}$ .

**Remark 1** (*On the implementation of the selection rule*) Our subsequent results hold for *any* selection rule that fulfills the requirements (a), (b), (c) above. There are indeed several ways to define a suitable selection rule; some of these possibilities are briefly described next.

(i) **Optimal selection rule.** One possibility is to remove those  $q$  returns that provide the best possible improvement of the  $\gamma^*$  level in problem (1). We call this rule the *optimal* selection rule. From a computational point of view, implementing the optimal selection rule may be hard numerically, since it corresponds in principle to a combinatorial problem: among all subsets of  $\{1, \dots, N\}$  of cardinality  $q$ , select the one subset that provides the largest value in  $\gamma^*$ . Finding the optimal portfolio  $x^*$  and the corresponding level  $\gamma^*$  under the optimal selection rule may be cast in the form of a mixed-integer linear program as follows:

$$\begin{aligned} \max_{x \in \mathcal{X}, \gamma, s_i \in \{0,1\}} \quad & \gamma \\ \text{subject to: } \quad & Ms_i + \varrho_i(x) \geq \gamma, i = 1, \dots, N \\ & \sum_{i=1}^N s_i = q, \end{aligned} \quad (2)$$

where  $M$  is some “large” positive number (e.g., one may take  $M = 1$ , if all  $\varrho_i(x)$  are known to be all smaller than one), and  $s_i, i = 1, \dots, N$ , are additional 0/1 variables. Such problems can usually be solved quite effectively, for moderate sizes, using numerical packages for mixed-integer linear optimization, such as IBM CPLEX.

(ii) ***m-at-a-time rule.*** An alternative, suboptimal, rule for return removal can be implemented as described next. The idea is that although problem (2) is theoretically hard, it turns out in practice that it can be solved quite rapidly, if the number of suppressed constraints is small. Therefore, while it can be prohibitive to remove all  $q$  constraints at once (as it is prescribed by the optimal rule), it is usually doable to remove  $m \ll q$  of them iteratively. The “*m-at-a-time rule*” rule thus simply prescribes to suppress iteratively  $1 \leq m \ll q$  returns at a time, by solving repeatedly a problem of the form (2) with  $m$  instead of  $q$ , until all  $q$  constraints have been removed. In principle, this approach is suboptimal, and may not yield the same result as the optimal rule. However, it usually gives good results in practice.

(iii) **Lagrange multiplier-based rule.** Another possibility (useful if one does not have a mixed-integer solver available) is to prune the returns sequentially (one by one, or  $m \geq 1$  at a time) according to their impact on objective sensitivity. With this approach, one first solves the LP with all returns in place, then removes the  $1 \leq m \leq \min(n, q)$  returns that yield the best local improvement in the objective, then solves again the LP, and so on, until all  $q$  returns are removed. Suitable implementation of such a technique provides a valid selection rule. At each iteration, the returns to be removed can be determined by looking at the values  $\lambda_i$  of the Lagrange multipliers (dual variables) associated to the surviving constraints  $\varrho_i(x) \geq \gamma$ . It is indeed well known (see, e.g., [4]) that a positive Lagrange multiplier  $\lambda_i$  represents the sensitivity of the optimal objective value to variations in the  $i$ -th constraint, hence the locally-best choice is to remove the  $m$  constraints corresponding to the  $m$  largest  $\lambda_i$ , since this would induce (to first order approximation) the largest improvement in objective value. A distinctive advantage of this selection rule is that the optimal portfolio allocation problem in (1) is solved efficiently by solving a sequence of standard linear programming problems.

Other constraint removal heuristics may be devised, besides the described ones. It is important, however, to stress again the fact that the theory that is presented in this paper *does not depend* on the specific selection rule and, in particular, does not need implementation of the optimal rule (which can be hard to compute). The results in this paper hold for *any* selection rule that satisfies the requirements (a), (b), (c) specified at the beginning of Section 2.1.  $\star$

## 2.2 The short-fall probability

Given a selection rule  $\mathcal{S}_q$ , the optimal portfolio allocation strategy  $x^*$  that we propose is a solution of the following LP

$$\gamma^* = \max_{x \in \mathcal{X}, \gamma} \gamma \quad (3)$$

subject to:  $r^\top(i)x \geq \gamma, i \in \mathcal{I}_{N-q}$ ,

where, by the definition of the selection rule, it holds that  $\varrho_i(x^*) = r^\top(i)x^* < \gamma^*$ , for all  $i \in \mathcal{I}_q$ . It is important to underline that we here take an a-priori point of view that is, a priori, the return vectors  $r(i)$ ,  $i = 1, \dots, N$ , are random variables, hence also the optimal solutions  $\gamma^*$ ,  $x^*$  of (3) are random variables, which are functions of the random data of the problem (i.e., of  $r(i)$ ,  $i = 1, \dots, N$ , which are collected in the random matrix  $R_N$ ). Events involving  $\gamma^*$ ,  $x^*$  are thus measured by the product probability  $\mathbb{P}^N$ . The problem under study belongs to the class of so-called random convex programs (RCP), or scenario-based optimization, see [5,7,10]. In particular,

we here build upon the technique of random programs with violated constraints described in [5] (see also [9] for an earlier application of this idea) in order to derive the desired probabilistic bounds.

If we *observe* an actual *realization* of the returns (for example, by looking at a stream of  $N$  historical data for the returns), then the observed return sequence becomes deterministic, and (3) would return a deterministic vector  $x^*$  and a deterministic level  $\gamma^*$ . However, *before* we look at the actual realization, these two variables remain uncertain and random. We are interested in providing an a-priori probabilistic characterization on the optimal solution of (3). To this end, we introduce a further assumption and a definition.

**Assumption 3 (Uniqueness)** *Assumption 2 is verified and, with probability one, the optimal solution  $x^*$ ,  $\gamma^*$  of (3) is unique.*  $\star$

**Remark 2** Assumption 2 guarantees that the set  $\mathcal{X}$  is compact and nonempty, which implies that problem (3) is feasible and it attains an optimal solution; this assumption is generally fulfilled in portfolio optimization problems, hence it is not restrictive in practice. Assumption 3 further postulates that the optimal solution is uniquely identified, i.e., that the optimum of the LP is attained at a vertex. This is usually the case for LP constraints in “general position” (e.g., excluding cases of two or more identical returns, which happen with zero probability under continuous distributions). Moreover, an infinitesimal perturbation of the constraints, or introduction of a strictly convex regularization term in the objective would always make the optimal solution unique. Assumption 3 is thus made for technical reasons, and it is not restrictive from a practical point of view in the present context.  $\star$

For a *fixed* portfolio  $x \in \mathcal{X}$  and return level  $\gamma \in \mathbb{R}$ , we define the short-fall probability as

$$V(x, \gamma) = \mathbb{P}\{r : r^\top x < \gamma\}.$$

Such a probability is a *number* in  $[0, 1]$ . However, if we now ask about the short-fall probability relative to the optimal solution of (3), we have

$$V^* \doteq V(x^*, \gamma^*) = \mathbb{P}\{r : r^\top x^* < \gamma^*\}, \quad (4)$$

and this is, a priori, a random variable, since  $x^*$ ,  $\gamma^*$  are so. Indeed, for each different realization of the random returns  $R_N$  we will get different  $x^*$ ,  $\gamma^*$ , hence a different  $V^*$ . Therefore,  $V^*$  is a random variable with support  $[0, 1]$ , and events related to  $V^*$  are measured by the product probability  $\mathbb{P}^N$ . It is then natural to consider as a measure of “riskiness” of the optimal portfolio the expected value (with respect to  $\mathbb{P}^N$ ) of the short-fall probability  $V^*$ . This leads to the following definition.

**Definition 1 (Expected short-fall probability)**

The expected short-fall probability of the optimal portfolio resulting from (3) is defined as

$$\mathbb{E}_{\mathbb{P}^N}\{V^*\} = \mathbb{E}_{\mathbb{P}^N}\{\mathbb{P}\{r^\top x^* < \gamma^*\}\}. \quad (5)$$

Our key result concerns a quantification of an upper bound on the expected short-fall probability of the optimal portfolio. This is developed in the next section.

**3 Short-fall probability of the optimal data-driven portfolio**

The first result we report concerns an upper bound on the upper tail of the distribution of  $V^*$ . This results follows directly from Theorem 4.1 and Corollary 4.2 in [5], considering that the problem (3) we are dealing with is precisely a random convex program with Helly's dimension upper bounded by  $n + 1$ , which is the number of decision variables in problem (3); see [5] for further details and definitions.

**Lemma 1 (Upper-tail bound on  $V^*$ )** *Let Assumptions 1–3 hold, and let  $x^*, \gamma^*$  be the optimal solution of problem (3), under any given selection rule satisfying properties (a), (b), (c) specified in Section 2.1. Let  $V^*$  be defined as in (4). Then it holds that*

$$\mathbb{P}^N\{V^* > z\} \leq \binom{q+n}{q} \bar{\Phi}(z; q+n, N), \quad (6)$$

where

$$\bar{\Phi}(z; q+n, N) \doteq \sum_{j=0}^{q+n} \binom{N}{j} z^j (1-z)^{N-j}, \quad (7)$$

and

$$\binom{q+n}{q} = \frac{(q+n)!}{q!n!} = \prod_{i=1}^q \frac{n+i}{i}.$$

★

Note that  $\bar{\Phi}$  in (7) can be expressed in terms of the regularized incomplete beta function  $I(z; a, b)$  (in Matlab: `betacdf(z, a, b)`) as follows

$$\begin{aligned} \bar{\Phi}(z; a, b) &= I(1-z; b-a, a+1) \\ &= 1 - I(z; a+1, b-a). \end{aligned} \quad (8)$$

An important consequence of Lemma 1 is that, for given level  $z$  and a suitable choice of the time window  $N$  and of the removal cardinality  $q$ , we can make the upper tail bound (6) as small as desired, so that with practical

certainty<sup>1</sup> the optimization will provide a  $V^*$  such that  $V^* \leq z$ .

We next state a new result which provides an explicit and efficient upper bound on the expected short-fall probability.

**Lemma 2 (Upper bound on the expected short-fall probability)** *Let Assumptions 1–3 hold, and let  $x^*, \gamma^*$  be the optimal solution of problem (3), under any given selection rule satisfying properties (a), (b), (c) specified in Section 2.1. Let  $V^*$  be defined as in (4). Then it holds that*

$$\mathbb{E}_{\mathbb{P}^N}\{V^*\} \leq \frac{q}{N} + \left( \frac{n}{N} + \frac{\omega(n, q)}{2\sqrt{N}} \right), \quad (9)$$

where  $\omega(n, q) = \mathcal{O}(\sqrt{2n \ln(q+n)})$  and, more precisely,

$$\omega(n, q) = \frac{2n(1 + \ln(q+n) - \ln n) - 2 \ln 2 + 1}{\sqrt{2n(1 + \ln(q+n) - \ln n) - 2 \ln 2}}.$$

★

**Remark 3** Equation (9) has the following interpretation:  $\eta \doteq q/N$  is the *empirical* short-fall probability, i.e. the short-fall probability of the optimal data-driven portfolio on the data  $R_N$  that are used for the optimization. In other words,  $\eta$  is the in-sample short-fall empirical probability. The extra term in (9)

$$\epsilon(N, q, n) = \frac{n}{N} + \frac{\omega(n, q)}{2\sqrt{N}} \quad (10)$$

represents the *excess* short-fall probability due to the fact that the data upon which the optimal portfolio is built (the  $R_N$ ) are themselves random. Notice that (9) is an over-bounding of  $\mathbb{E}_{\mathbb{P}^N}\{V^*\}$ : there is room for improvement of the bound. First we outline a proof for Lemma 2, then we provide in Lemma 3 a tighter, albeit somewhat less explicit, bound on  $\mathbb{E}_{\mathbb{P}^N}\{V^*\}$ .

**Proof of Lemma 2:** see Appendix A.1.

A tighter bound on  $\mathbb{E}_{\mathbb{P}^N}\{V^*\}$  is given in the next lemma.

**Lemma 3** *Under the hypotheses and notation in Lemma 2, we have that*

$$\mathbb{E}_{\mathbb{P}^N}\{V^*\} \leq z_1 + \frac{C}{N+1} \sum_{j=0}^{q+n} I(1-z_1; N-j+1, j+1), \quad (11)$$

<sup>1</sup> We informally define as “practically certain,” in this financial context, an event that occurs with probability larger than, say,  $1 - 10^{-6}$ .

where  $I(x; a, b)$  is the regularized incomplete beta function,  $C = \binom{n+q}{n}$ , and

$$z_1 = \frac{q}{N} + \frac{1}{N} \left( n + \ln C + \sqrt{\ln^2 C + 2(q+n) \ln C} \right).$$

★

**Proof of Lemma 3:** see Appendix A.2.

Lemma 2 and Lemma 3 provide us with explicit upper bounds on the expected short-fall probability, for given data length  $N$  and removal cardinality  $q$ . The formulas in (9), (11), as well as the formula in (6), can be “inverted,” at least numerically, in order to find suitable  $N$  and  $q$ , given assigned levels of tolerable short-fall probability or of expected short-fall probability. These fundamental tradeoffs are highlighted in the following corollary.

**Corollary 1 (Explicit conditions on  $N$  and  $q$ )** *Let  $\beta \in (0, 1)$  be a very small probability level chosen by the user (e.g., set  $\beta = 10^{-6}$ , or lower, for “practical certainty”). Let  $z_{\text{tol}} \in (0, 1)$  be a desired tolerable short-fall probability level, let  $z_{\text{exp}} \in (0, 1)$  be a desired expected short-fall probability level, and let  $q \leq N - n - 1$ . Then, the following statements are true:*

(1) *If*

$$N \geq \frac{2}{z_{\text{tol}}} \ln \beta^{-1} + \frac{4}{z_{\text{tol}}} (q + n), \quad (12)$$

*then  $\{V^* \leq z_{\text{tol}}\}$  holds with probability larger than  $1 - \beta$  (i.e., with practical certainty).*

(2) *If*

$$N \geq 4 \frac{q+n}{z_{\text{exp}}} + \frac{(c+1/c)^2}{4z_{\text{exp}}^2}, \quad (13)$$

*with  $c \doteq \sqrt{2n + 2n \ln \frac{n+q}{q} - 2 \ln 2}$ , then it holds that  $\mathbb{E}_{\mathbb{P}^N} \{V^*\} \leq z_{\text{exp}}$ . For “large”  $q$ , bound (13) simplifies approximately to*

$$N \geq 4 \frac{q+n}{z_{\text{exp}}} + \frac{2n + 2n \ln \frac{n+q}{q} - 2 \ln 2}{4z_{\text{exp}}^2}.$$

★

**Proof of Corollary 1:** see Appendix A.3.

**Remark 4 (Practical use of conditions (12), (13))**

A long-standing issue in classical portfolio optimization concerns the choice of the depth  $N$  of the time interval used to estimate the parameters (typically, the expected

returns and covariance) to be then employed for portfolio optimization. It is well known that changing the estimation window  $N$  may change considerably the estimated parameters and hence the optimal allocations. Partial remedies to reduce the estimation sensitivity include the use of shrinkage estimators (see, e.g., [11] and the references therein) or exponential discounting of the observations (exponential forgetting implies an “equivalent” look-back window length  $N$ , hence the choice of the forgetting factor is essentially equivalent to the choice of the length of  $N$ ). On the other hand, to the best of these authors’ knowledge, there is no *rigorous* recipe to help the user choose the right estimation window (or the right forgetting factor). This choice thus remains a kind of an “art,” which is left to case-by-case judgement and empirical experimentation.

One contribution of Corollary 1 goes in the direction of filling this gap. Indeed, albeit relying on iid and stationarity hypotheses on the return process, equations (12), (13) do provide a rigorous quantification of the tradeoff between acceptable risk and the length of the data window used for optimization. The practical use of these equations is illustrated next; to fix ideas we concentrate on design based on the expected short-fall probability (eq. (13)), the discussion on (12) being analogous. We make four observations.

(i) For fixed  $N$  and for a given desired level of expected short-fall probability  $z_{\text{exp}}$ , it is the investor’s interest to make  $\gamma^*$  as large as possible. On a given realization of the returns, level  $\gamma^*$  increases if we increase the number  $q$  of suppressed returns, hence we want to make  $q$  as large as possible. However, if one increases  $q$  too much, then the resulting portfolio will fail to satisfy the expected short-fall probability requirement. The right-hand-side of (13) is increasing in  $q$ , hence this term tells us precisely how large  $q$  can be made, while satisfying the requirement  $\mathbb{E}_{\mathbb{P}^N} \{V^*\} \leq z_{\text{exp}}$ : we choose  $q$  such that the right-hand-side of (13) is the largest integer that does not exceed  $N$ . Actually, since (13) is only a bound (which can introduce some conservatism), we can also use the more accurate estimation given by equation (11): the maximal allowable  $q$  is the largest integer that maintains the right-hand side of (11) no larger than  $z_{\text{exp}}$ .

(ii) If  $N$  is not fixed (e.g., we are free to decide what the historical data length to be used for optimization should be), then eq. (13) (or, better, the tighter bound in (11)) can be used to plot a tradeoff set on an  $(N, q)$  plane, for the given desired level  $z_{\text{exp}}$ . Any pair  $N, q$  in the admissible set is a valid pair guaranteeing that the result of the optimization will satisfy  $\mathbb{E}_{\mathbb{P}^N} \{V^*\} \leq z_{\text{exp}}$ . To this end, one sweeps over a range of values for  $N$  and, for each fixed  $N$ , numerically evaluates (11) to find the largest  $q$  such that the right-hand side of (11) is below the assigned threshold  $z_{\text{exp}}$ , see Figure 2.

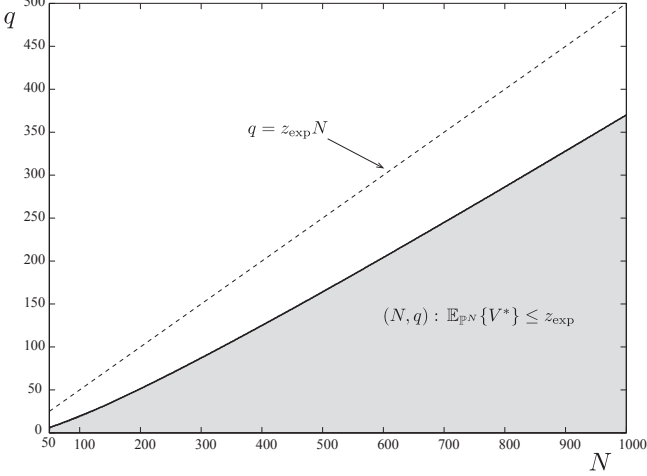


Fig. 2. The shaded region represents pairs  $(N, q)$  such that  $\mathbb{E}_{\mathbb{P}^N}\{V^*\} \leq z_{\text{exp}}$ . This example is plotted for  $n = 3$  and  $z_{\text{exp}} = 0.5$ , using numerical evaluation of the formula in (11).

(iii) An interesting feature that is captured by the present theory (see, e.g., eq. (11)) is that, all the other parameters being the same, the expected short-fall probability bound increases as  $n$  (the number of securities in the portfolio) increases. The reason for this lies at the fundamental tradeoff between the complexity of the random optimization model (here, the number of variables,  $n + 1$ ) and the out-of-sample *reliability* of the model: the more complex the model is (i.e., the larger  $n$  is), the more training data we need for achieving a given reliability level (i.e., the larger  $N$  needs to be, see eq. (13)). A financial interpretation of this phenomenon is that high diversification of a portfolio (large  $n$ ) needs a large number  $N$  of data in order to provide meaningful portfolios. This fact is often overlooked in standard portfolio optimization, where one may be lead to think that more diversification leads to less risk: not necessarily so, since the larger  $n$  the more accurate the parameter estimation must be, for otherwise the actual decrease in risk level may just be illusory. ★

## 4 Numerical tests

In this section we illustrate the results obtained on two series of numerical experiments. The first test is conducted on a simple three assets allocation model, using synthetic iid data generated according to realistic fat-tailed distributions. The second test is conducted on real market data, and concerns allocation over six asset classes.

### 4.1 A three assets model

We consider  $n = 3$  synthetic assets having increasing return and risk. The return data of the three assets are generated independently, according to  $t$ -distributions with

low degrees of freedom. Fat-tailed distributions, like the Pareto or the  $t$ -distribution, generally provide more realistic models of asset returns than the classical Normal distribution, see, e.g., [19,20]. The parameters of these distributions are shown in Table 4.1, where  $\mu$  is the location parameter of the distribution,  $s$  is the scale parameter,  $\nu$  is the degrees of freedom parameter, and  $\xi$  is the skewness. Incidentally, these three distributions do fit historical daily returns of (1) the XTXE tracking IBOXX global inflation-linked total return index, (2) the XGIN tracking ITRAXX crossover 5-year total return index, and (3) the DAX stock index, from January 28, 2003 to July 29, 2011.

Table 1  
 $t$ -distribution parameters for asset returns

Asset	$\mu$	$s$	$\nu$	$\xi$
1 (XTXE)	0.0000798	0.00233	4	1
2 (XGIN)	0.0003416	0.00620	4	1
3 (DAX)	0.0004699	0.01538	3	0.9421

**4.1.0.1 Testing a-posteriori statistics.** The numerical test is conducted as follows: we first fix a value of  $N$  (e.g.,  $N = 200$ ,  $N = 400$ ,  $N = 800$ ,  $N = 1000$ ) and a desired value of  $z_{\text{exp}}$ . Next, we determine  $q$  as the largest integer such that the right-hand-side of (11) does not exceed  $z_{\text{exp}}$  (notice that this computation can be easily performed numerically via bisection on  $q$ ). Then, we generate  $N$  iid samples of returns for the three assets, and we use this data for computing an optimal data-driven portfolio  $x^*$  with  $q$  suppressed constraints, using the suboptimal technique based on Lagrange multipliers described in Remark 1. Finally, the resulting portfolio is tested a posteriori on  $N_{\text{test}}$  newly generated random returns. The feasible portfolio set  $\mathcal{X}$ , in all the following simulations, is assumed to be the standard simplex  $\mathcal{X} = \{x \in \mathbb{R}^n : x \geq 0, \sum_{i=1}^n x_i = 1\}$ , hence the optimization problems involved are simple linear programs.

Table 4.1.0.1 reports the results of such a simulation, for  $z_{\text{exp}} = 0.5$ ,  $N_{\text{test}} = 3000$ , and for various values of  $N$ . Each column in the table corresponds to a  $(N, q)$  pair for which  $\mathbb{E}_{\mathbb{P}^N}\{V^*\} \leq z_{\text{exp}}$  is satisfied. For each of such pairs, we computed the optimal data-driven portfolio  $x^*$  and corresponding level  $\gamma^*$ , based on  $N$  generated random returns. Then, we tested this solution a-posteriori on an out-of-sample batch of  $N_{\text{test}}$  random returns, and we evaluated the following a-posteriori statistics:  $\hat{\gamma}$  is the empirical  $z_{\text{exp}}$ -quantile of the optimal portfolio returns, that is the portfolio return level that is underperformed with at most  $z_{\text{exp}}$  empirical frequency;  $\hat{V}^*$  is the empirical out-of-sample shortfall probability, that is the empirical frequency of the event  $r^\top x^* < \gamma^*$ . Table 4.1.0.1 shows analogous results for the case  $z_{\text{exp}} = 0.05$ . We observe that the out-of-sample empirical shortfall  $\hat{V}^*$  sometimes gets above the in-sample shortfall  $q/N$ , but always remains below the  $z_{\text{exp}}$  threshold.



Table 2  
Statistics for  $z_{\text{exp}} = 0.5$

	$N = 200$ $q = 51$ $q/N\% = 25.5\%$	$N = 400$ $q = 125$ $q/N\% = 31.25\%$	$N = 800$ $q = 286$ $q/N\% = 35.75\%$	$N = 1000$ $q = 370$ $q/N\% = 37\%$
$\gamma^*$	-0.13%	-0.07%	-0.05%	-0.04%
$\hat{\gamma}$	0.01%	0.01%	0.01%	0.01%
$\hat{V}^*$	21.5%	32.5%	37.0%	37.9%

Table 3  
Statistics for  $z_{\text{exp}} = 0.05$

	$N = 200$ $q = 1$ $q/N\% = 0.5\%$	$N = 400$ $q = 4$ $q/N\% = 1\%$	$N = 800$ $q = 14$ $q/N\% = 1.75\%$	$N = 1000$ $q = 19$ $q/N\% = 1.9\%$
$\gamma^*$	-0.88%	-1.06%	-0.52%	-0.50%
$\hat{\gamma}$	-0.57%	-0.54%	-0.36%	-0.34%
$\hat{V}^*$	1.43%	0.70%	1.50%	1.47%

**4.1.0.2 Sliding-horizon implementation.** Next, we simulated the application of the proposed portfolio optimization technique in a sliding horizon fashion: at each time instant a history batch of  $N$  observations is used to compute the optimal data-driven portfolio, then this portfolio is tested on a new next-day return datum. The history window is next moved one day forward, and the process is repeated. Figure 3 shows the averaged portfolio weights obtained in this way. Figure 4 plots the portfolio weights over time. It is apparent that the weights become smoother as  $N$  increases.

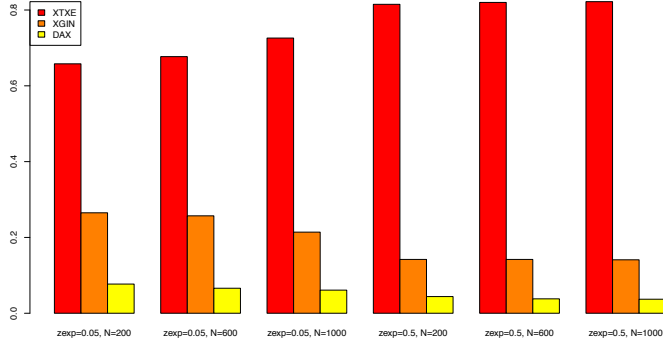


Fig. 3. Average portfolio weights for data-driven method.

**4.1.0.3 Comparison with VaR-optimal portfolios.** The proposed data-driven portfolio selection method has a close relation with Value-at-Risk (VaR) portfolio optimization, see [17]. Assuming a continuous return distribution, the  $\alpha$ -VaR of a given portfolio  $x$  can be defined as

$$\gamma_\alpha(x) = \sup \gamma : \mathbb{P}\{r^\top x < \gamma\} \leq \alpha, \quad \alpha \in (0, 1).$$

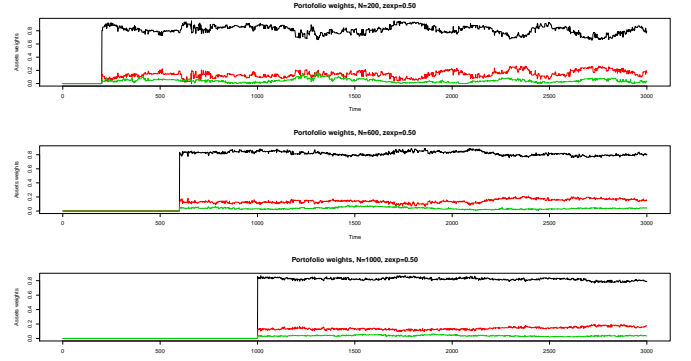


Fig. 4. Time plots of portfolio weights. XTXE (black, top), XGIN (red, middle), DAX (green, bottom).  $z_{\text{exp}} = 0.50$ ,  $N = 200; 600; 1000$ .

That is, the VaR  $\gamma_\alpha(x)$  is the largest return level guaranteeing that the probability of the portfolio return be lower than this level is at most  $\alpha$ . For given  $\alpha$ , one can look for the portfolio  $x$  that maximizes  $\gamma_\alpha(x)$ , which is obtained by solving

$$\gamma_\alpha^* = \max_{x, \gamma} \gamma_\alpha(x) = \max_{x, \gamma} \gamma : \mathbb{P}\{r^\top x < \gamma\} \leq \alpha. \quad (14)$$

To solve the optimization problem (14) it is of course necessary to elicit some distribution for  $r$ . The classical and most well-known approach assumes that  $r$  is Normal with covariance  $\Sigma$  and expected value  $\mu$ . Under such an assumption, for  $\alpha < 0.5$ , the above problem can be rewritten explicitly as a convex second-order cone program, [18]

$$\gamma_\alpha^* = \max_x \Phi^{-1}(\alpha) \sqrt{x^\top \Sigma x} + \mu^\top x,$$

where  $\Phi^{-1}$  is the inverse of the standard Normal cumulative distribution function. It can then be observed that the data-driven method proposed in this paper is a kind of empirical version of the VaR optimization problem in (14), with  $\alpha = z_{\text{exp}}$ . The key problem in applying the classical VaR method is that one has to *assume* Normality of the returns (in this example, for instance, returns are actually non-Normal), which is often a very strong and unrealistic assumption (see, e.g., [19]) and, moreover, one has to *estimate* the parameters  $\Sigma$  and  $\mu$ , which is a delicate problem in his own right. Robust approaches exist that allow for consideration of uncertainty in the  $\Sigma$  and  $\mu$  parameters, as well as in the return distribution, see, e.g., [14]. However, these approaches require correct estimation of the uncertainty bounds on the distribution parameters, which is a further critical issue, hence they are not further discussed in the present experiments. The data-driven method, however, bypasses both issues, since it does not need a probabilistic model of the returns.

We thus compared the performances of the data-driven method with those of a traditional VaR system, in a sliding-horizon simulation. For given look-back period  $N$ , the covariance  $\Sigma$  and expected returns  $\mu$  are estimated empirically from data, then a VaR-optimal portfolio is computed and next it is tested against a new random datum, simulating the next-day return. The process is iterated in a sliding-window fashion. The average weights for the VaR portfolios are reported in Figure 5, for  $\alpha = 0.05$ . Comparing these data with those in Figure 3 one may observe that the VaR weights are more “cautious” (that is, they put less weight on equity) than the data-driven ones.

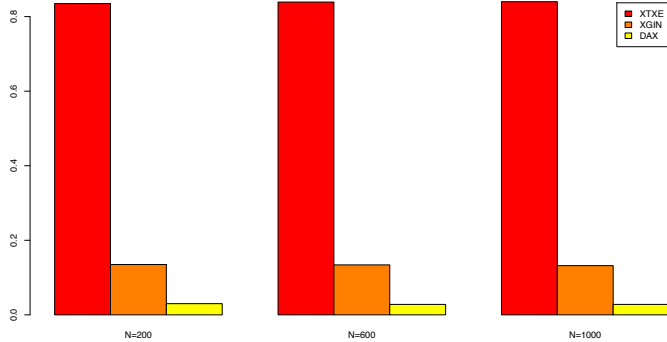


Fig. 5. Average portfolio weights for VaR-optimal method,  $\alpha = 0.05$ .

Table 4.1.0.3 shows the out-of-sample shortfall frequency  $\hat{V}^*$  and average level of  $\gamma^*$  (or  $\gamma_\alpha^*$ , for VaR), obtained by implementing the two strategies repeatedly in time. Table 4.1.0.3 shows the gross cumulative return at the end of the simulation period, for the two strategies.

We remark that the out-of-sample empirically-observed shortfall probability of the data-driven portfolios is lower than the one of the VaR-optimal portfolios. This might

suggest that the data driven method entails some conservatism. However, it should be underlined that the data-driven bound is distribution independent, that is it would hold no matter what the return distribution is. Hence, it is natural that the observed results, obtained from simulations with *one fixed* return distribution, display some gap between the desired bound and the empirical one. Moreover, at the  $\alpha = z_{\text{exp}} = 0.01$  level, the shortfall probability bound remains satisfied by the data-driven portfolios, while it is *violated* by the VaR-optimal portfolios; see the last two rows in Table 4.1.0.3. Furthermore, it is worth to notice that although the data-driven portfolios are apparently “safer” than the VaR-optimal portfolios, the gross return performance of the formers is consistently better than that of the latters, see Table 4.1.0.3.

Table 4

Comparison of average  $\gamma^*$  or  $\gamma_\alpha^*$  level and of out-of-sample probability  $\hat{V}^*$ , for VaR and data-driven portfolios. Table reports  $(\hat{V}^*, \gamma^*)$  for data-driven, or  $(\hat{V}^*, \gamma_\alpha^*)$  for VaR

	$\alpha, z_{\text{exp}}$	$N = 1000$	$N = 1400$
VaR-optimal	5%	(4.3; -0.34)%	(3.7; -0.34)%
data-driven	5%	(1.7; -0.52)%	(1.8; -0.49)%
VaR-optimal	1%	(1.7; -0.49)%	(1.4; -0.49)%
data-driven	1%	(0.4; -0.94)%	(0.3; -1.01)%

Table 5

Gross returns of VaR and data-driven portfolios, for  $\alpha = z_{\text{exp}} = 0.05$

	$N = 200$ $q = 1$	$N = 600$ $q = 8$	$N = 1000$ $q = 19$	$N = 1400$ $q = 31$
VaR-opt.	1.47	1.36	1.30	1.30
data-dr.	1.73	1.64	1.45	1.44

#### 4.2 Allocation with real securities

We next performed some experiments using real market data. We considered six assets: three bond indexes, one stock index, one cross currency and one commodity index. The assets are:

- (1) the IBOXX Euro Germany 1-3 Total Return Index;
- (2) the IBOXX Euro Germany 7-10 Total Return Index;
- (3) the IBOXX Liquid High Yield bonds index;
- (4) the cross currency Euro - Japanese Yen, JPYEUR;
- (5) the DAX stock index;
- (6) the S&P GSCI commodity index, GSG.

The IBOXX 1-3, represents the 1-3 years maturity Euro sovereign debt issued by the German government. Within the index, each bond is weighted according to its outstanding amount. Identical rules apply to the 7-10

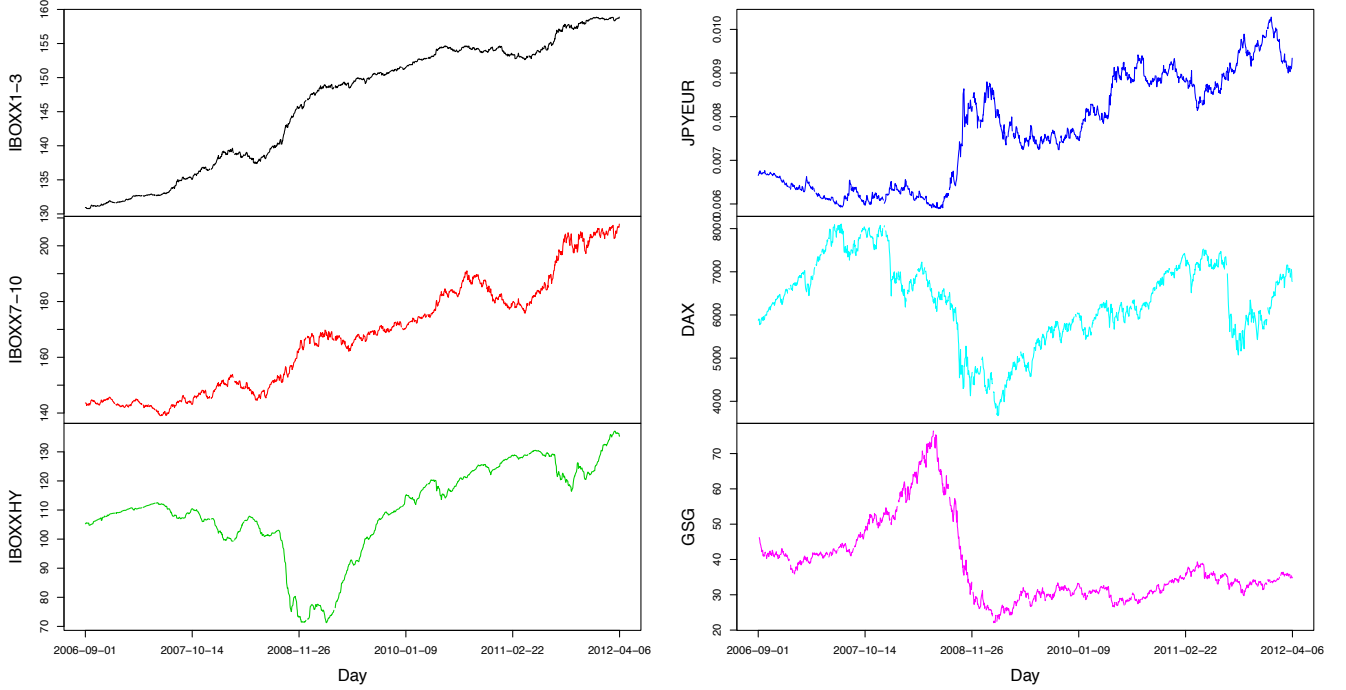


Fig. 6. Time plots of indexes daily close quotes, from 9/1/06 to 4/06/12.

years index. The IBOXX Liquid High Yield bonds index consists of liquid Euro sub-investment grade rated bonds.

All indices are calculated on a total return basis, which means that the payments from coupons and dividends (when applicable) are reinvested in the index. For all the securities the prices are in Euro. Figure 6 shows the daily close prices of the indices, from September 1, 2006 to April 6, 2012.

Figure 7 shows a plot of the sample autocorrelation function for three of the considered assets. Serial correlations at nonzero lags are obviously non null. However, correlations are essentially below the two standard error limit, hence they do not necessarily invalidate the use of the iid hypothesis as a reasonable working assumption in practice.

We then proceeded with a sliding-horizon implementation of the data-driven optimization technique, as illustrated in the previous section, also using VaR-optimal portfolios for comparison. The data-driven and VaR-optimal average portfolio compositions are very similar (see Figure 8), with a very slight bias of the data-driven portfolios towards more volatile but profitable assets.

We further compared the actual (empirical, out-of-sample) shortfall frequency and average levels  $\gamma^*$ ,  $\gamma_\alpha^*$  of the data-driven and VaR-optimal methods, see Table 4.2. It may be remarked that the shortfall frequency of the VaR-optimal method is greater than expected

for both  $\alpha = 0.1$  and  $\alpha = 0.05$ , while the data-driven bound is always satisfied out of sample.

## 5 Conclusions

In this paper, we presented a novel data-driven approach for computing optimal portfolio compositions directly from historical data. The proposed approach is based on iid and stationarity hypotheses on the returns process, but avoids assumptions on the cross-sectional distribution model of the returns, and does not need estimation of distribution parameters. The key feature of the method is that the optimal portfolios come with a rigorously established probability tag, guaranteeing that their out-of-sample expected short-fall probability is no larger than an a-priori assigned level. Computationally, the method is effective, in that it typically requires the solution of a sequence of linear programming problems. Numerical tests with both synthetic and real financial data seem to support the practical effectiveness of the proposed methodology.

## Acknowledgments

I sincerely thank Dr. Bruno Monastero for his precious help in selecting, collecting and organizing the financial data used in Section 4, and for his contribution in performing part of the numerical simulations.

Table 6

Empirical out-of-sample shortfall frequency and average  $\gamma^*$  levels for the data-driven and the VaR-optimal method

	$\alpha, z_{\text{exp}}$	$N = 400$	$N = 600$	$N = 800$	$N = 1000$
VaR-optimal	25%	(22.7; -0.03)%	(20.7; -0.03)%	(19.9; -0.04)%	(20.6; -0.04)%
data-driven	25%	(10.2; -0.08)%	(10.2; -0.07)%	(10.4; -0.06)%	(13.1; -0.06)%
VaR-optimal	10%	(10.1; -0.08)%	(8.1; -0.08)%	(7.3; -0.08)%	(10.8; -0.08)%
data-driven	10%	(2.2; -0.17)%	(2.0; -0.16)%	(1.2; -0.15)%	(1.8; -0.14)%
VaR-optimal	5%	(7.3; -0.11)%	(4.6; -0.11)%	(2.7; -0.11)%	(3.6; -0.11)%
data-driven	5%	(1.5; -0.24)%	(0.9; -0.27)%	(0.2; -0.21)%	(0.3; -0.19)%

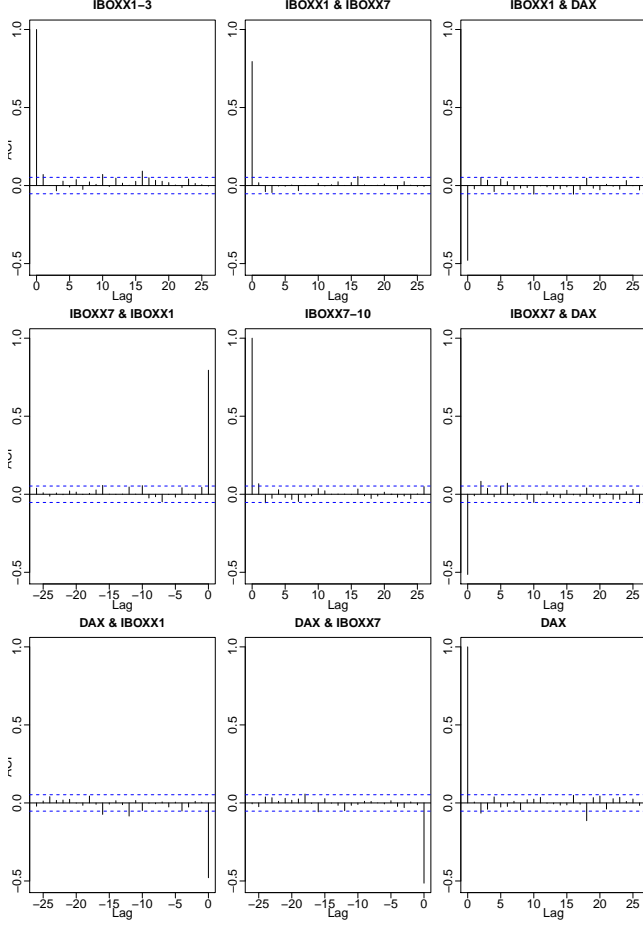


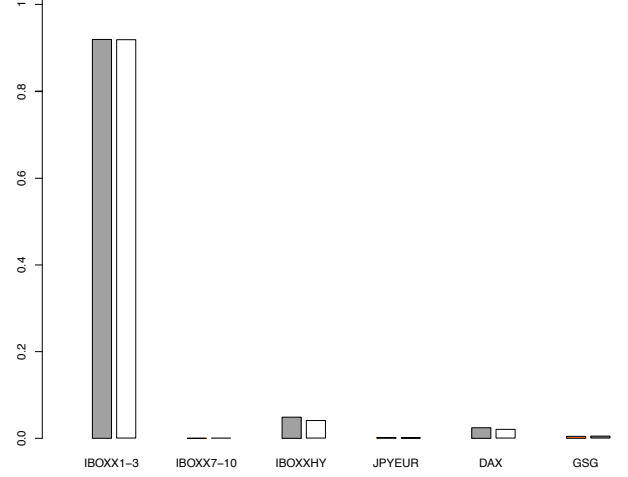
Fig. 7. Sample autocorrelation function for daily returns from September 1, 2006 to April 6, 2012. The horizontal dashed lines indicate two standard error limits (95% confidence interval).

## A Appendix: proofs

### A.1 Proof of Lemma 2

Since  $V^*$  is a nonnegative random variable (it has support on the interval  $[0, 1]$ ), we have that

$$\mathbb{E}_{\mathbb{P}^N}\{V^*\} = \int_0^1 (1 - F(z)) dz, \quad (\text{A.1})$$

Fig. 8. Average portfolio weights for the data-driven (gray bars on the left) and the VaR (empty bars) methods. Portfolios computed for  $z_{\text{exp}} = 0.1$ ,  $N = 1000$ .

where  $F(z)$  denotes the cumulative distribution of  $V^*$ , that is

$$F(z) = \mathbb{P}^N\{V^* \leq z\}, \quad z \geq 0.$$

Define

$$\bar{F}(z) = 1 - F(z) = \mathbb{P}^N\{V^* > z\}.$$

Then, from Lemma 1 we have that

$$\bar{F}(z) \leq C\bar{\Phi}(z; a, N), \quad (\text{A.2})$$

where  $a \doteq q + n$ ,  $C \doteq \binom{a}{q}$ , and

$$\bar{\Phi}(z; a, N) = \sum_{j=0}^a \binom{N}{j} z^j (1-z)^{N-j}.$$

Our objective is to find an upper bound for  $\bar{F}(z)$  that can be used in (A.1) to obtain an upper bound for the expected short-fall probability. To this end, we notice

that (A.2) is meaningful only when its right-hand-side is smaller than one. Our strategy is hence the following: we determine a point  $z_1 \in (0, 1)$  such that  $C\bar{\Phi}(z; a, N) \leq 1$  for all  $z \geq z_1$ , and then bound  $\bar{F}(z)$  as follows

$$\bar{F}(z) \leq \begin{cases} 1 & \text{for } z \in [0, z_1] \\ C\bar{\Phi}(z; a, N) & \text{for } z \in (z_1, 1]. \end{cases} \quad (\text{A.3})$$

In order to find a suitable abscissa  $z_1$  we use the Hoeffding bound on the Binomial tail, valid for  $Nz \geq a$ :

$$\begin{aligned} \bar{\Phi}(z; a, N) &= \sum_{j=0}^a \binom{N}{j} z^j (1-z)^{N-j} \\ &\leq \frac{1}{2} \exp\left(-2\frac{(Nz-a)^2}{N}\right) \\ &= \frac{\sqrt{2\pi}}{4\sqrt{N}} \mathcal{N}(z; \mu, \sigma), \end{aligned} \quad (\text{A.4})$$

where  $\mathcal{N}(z; \mu, \sigma)$  denotes the Normal probability density function with mean  $\mu$  and variance  $\sigma^2$ :

$$\mathcal{N}(z; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right), \quad (\text{A.5})$$

where  $\mu \doteq a/N$  and  $\sigma \doteq 1/2\sqrt{N}$ . We thus look for  $z$  such that

$$C\bar{\Phi}(z; a, N) \leq C\frac{1}{2} \exp\left(-2\frac{(Nz-a)^2}{N}\right) = 1.$$

Taking logarithms we obtain the second-order equation  $(Nz-a)^2 = \frac{N}{2} \ln(C/2)$ , which has a root for  $Nz \geq a$  at

$$z_1 = \frac{a}{N} + \frac{\sqrt{2\ln(C/2)}}{2\sqrt{N}} = \mu + \sigma\sqrt{2\ln(C/2)}. \quad (\text{A.6})$$

Now, substituting (A.3) in (A.1) we obtain

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N}\{V^*\} &= \int_0^1 \bar{F}(z) dz \leq \int_0^{z_1} dz + C \int_{z_1}^1 \bar{\Phi}(z; a, N) dz \\ &\leq z_1 + C \frac{\sqrt{2\pi}}{4\sqrt{N}} \int_{z_1}^1 \mathcal{N}(z; \mu, \sigma) dz, \\ &\leq z_1 + C \frac{\sqrt{2\pi}}{4\sqrt{N}} \int_{z_1}^{\infty} \mathcal{N}(z; \mu, \sigma) dz, \end{aligned} \quad (\text{A.7})$$

where the last inequality is obtained by extending the upper limit of integration from one to infinity. Next, we bound the integral in (A.7) as follows: note that  $z -$

$\mu \geq z_1 - \mu$ , for  $z$  in the domain of integration, hence  $(z - \mu)/(z_1 - \mu) \geq 1$ , and therefore

$$\begin{aligned} \int_{z_1}^{\infty} \mathcal{N}(z; \mu, \sigma) dz &\leq \int_{z_1}^{\infty} \frac{z - \mu}{z_1 - \mu} \mathcal{N}(z; \mu, \sigma) dz \\ &= \frac{1}{\sigma(z_1 - \mu)\sqrt{2\pi}} \int_{z_1}^{\infty} (z - \mu) \exp\left(-\frac{(z - \mu)^2}{2\sigma^2}\right) dz \\ &= \frac{1}{\sigma(z_1 - \mu)\sqrt{2\pi}} \int_{z_1 - \mu}^{\infty} \xi \exp\left(-\frac{\xi^2}{2\sigma^2}\right) d\xi \\ &= \frac{1}{\sigma(z_1 - \mu)\sqrt{2\pi}} \cdot \sigma^2 \exp\left(-\frac{(z_1 - \mu)^2}{2\sigma^2}\right). \end{aligned}$$

Thus, continuing the chain of inequalities in (A.7), we obtain

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N}\{V^*\} &\leq z_1 + C \frac{\sqrt{2\pi}}{4\sqrt{N}} \frac{\sigma}{(z_1 - \mu)\sqrt{2\pi}} \exp\left(-\frac{(z_1 - \mu)^2}{2\sigma^2}\right) \\ [\text{using (A.6)}] &= \mu + \sigma\sqrt{2\ln(C/2)} + \frac{1}{2\sqrt{N}\sqrt{2\ln(C/2)}} \\ [\text{using (A.5)}] &= \frac{a}{N} + \frac{1}{2\sqrt{N}}(\zeta + 1/\zeta), \end{aligned} \quad (\text{A.8})$$

where  $\zeta \doteq \sqrt{2\ln(C/2)}$ . Further, recalling a binomial identity and inequality

$$C = \binom{n+q}{q} = \binom{n+q}{n} \leq \left(\frac{e(n+q)}{n}\right)^n, \quad (\text{A.9})$$

we have that  $\zeta \leq \sqrt{2n(1 + \ln(q+n) - \ln n) - 2\ln 2}$ , and, since  $\zeta + 1/\zeta$  is increasing for  $\zeta \geq 1$ , we continue and conclude the proof with

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N}\{V^*\} &\leq \frac{q}{N} + \left(\frac{n}{N} + \frac{\omega(n, q)}{2\sqrt{N}}\right), \\ \omega(n, q) &\doteq \frac{2n(1 + \ln(q+n) - \ln n) - 2\ln 2 + 1}{\sqrt{2n(1 + \ln(q+n) - \ln n) - 2\ln 2}}. \end{aligned}$$

Notice finally that, for “large”  $n, q$ , the “+1” term in the numerator of  $\omega(n, q)$  becomes negligible, hence  $\omega(n, q) \simeq \sqrt{2n(1 + \ln(q+n) - \ln n) - 2\ln 2}$ , which grows as  $\mathcal{O}(\sqrt{2n \ln(q+n)})$  for large  $n, q$ .  $\square$

## A.2 Proof of Lemma 3

Consider the Chernoff bound on the Binomial tail:

$$\bar{\Phi}(z; a, N) = \sum_{j=0}^a \binom{N}{j} z^j (1-z)^{N-j}$$

$$\begin{aligned} &\leq \exp\left(-\frac{(Nz-a)^2}{2Nz}\right), \quad Nz \geq a \\ &= \exp\left(-\frac{N(z-\mu)^2}{2z}\right), \quad \mu = \frac{a}{N}. \end{aligned}$$

Similar to what we did in the proof of Lemma 2, we use the Chernoff bound (instead of the Hoeffding bound, as we did in the proof of Lemma 2) for finding a  $z_1$  such that  $C\bar{\Phi}(z; a, N) \leq 1$  for all  $z \geq z_1$ , obtaining the condition

$$C \exp\left(-\frac{(Nz-a)^2}{2Nz}\right) = 1, \quad a = q + n,$$

and, taking logarithms,  $2Nz \ln C - (Nz-a)^2 = 0$ , that is

$$(Nz)^2 - 2(\ln C + a)Nz + a^2 = 0,$$

which, under the condition  $Nz \geq a$ , has a root at

$$Nz = a + \ln C + \sqrt{\ln^2 C + 2a \ln C},$$

and hence

$$z_1 = \frac{a}{N} + \frac{1}{N} \left( \ln C + \sqrt{\ln^2 C + 2a \ln C} \right). \quad (\text{A.10})$$

The advantage here is that all terms in  $z_1$  scale as  $1/N$ . Now, we consider the integral (A.7), and integrate directly  $\bar{\Phi}$ :

$$\begin{aligned} \mathbb{E}_{\mathbb{P}^N}\{V^*\} &\leq \int_0^{z_1} dz + C \int_{z_1}^1 \bar{\Phi}(z; a, N) dz \\ &= z_1 + C \int_{z_1}^1 \sum_{j=0}^a \binom{N}{j} z^j (1-z)^{N-j} dz \\ &= z_1 + C \sum_{j=0}^a \binom{N}{j} \int_{z_1}^1 z^j (1-z)^{N-j} dz \\ &= z_1 + C \sum_{j=0}^a \binom{N}{j} \mathcal{B}(1-z_1; N-j+1, j+1), \end{aligned} \quad (\text{A.11})$$

where  $\mathcal{B}(x; a, b)$  is the incomplete beta function

$$\mathcal{B}(x; a, b) = \int_0^x z^{a-1} (1-z)^{b-1} dz,$$

thus

$$\begin{aligned} \int_x^1 z^{a-1} (1-z)^{b-1} dz &= \int_0^{1-x} z^{b-1} (1-z)^{a-1} dz \\ &= \mathcal{B}(1-x; b, a). \end{aligned}$$

Further, the regularized incomplete beta function is  $I(x; a, b) = \frac{\mathcal{B}(x; a, b)}{\mathcal{B}(a, b)}$ , where  $\mathcal{B}(a, b)$  is the beta function which, for integer arguments, is such that  $\mathcal{B}(N-j+1, j+1) = \frac{1}{N+1} \binom{N}{j}^{-1}$ . Therefore, substituting in (A.11) we conclude the proof.  $\square$

### A.3 Proof of Corollary 1

Equation (12) comes directly from Corollary 5.1 in [5]. We next prove equation (13). To this end, we start from equation (A.8) and impose that  $\mathbb{E}_{\mathbb{P}^N}\{V^*\} \leq z_{\text{exp}}$ , that is

$$\mathbb{E}_{\mathbb{P}^N}\{V^*\} \leq \frac{a}{N} + \frac{1}{2\sqrt{N}}p \leq z_{\text{exp}},$$

where  $a = q + n$ ,  $p = (\zeta + 1/\zeta)$ ,  $\zeta = \sqrt{2 \ln(C/2)}$ . The previous condition is equivalent to

$$N - 2\frac{1}{4} \frac{p}{z_{\text{exp}}} \sqrt{N} - \frac{a}{z_{\text{exp}}} \geq 0 \quad (\text{A.12})$$

which is satisfied for  $\sqrt{N} \geq \frac{1}{4} \frac{p}{z_{\text{exp}}} + \sqrt{\left(\frac{1}{4} \frac{p}{z_{\text{exp}}}\right)^2 + \frac{a}{z_{\text{exp}}}}$ . Since the right hand side of this equation is  $\leq 2\sqrt{\left(\frac{1}{4} \frac{p}{z_{\text{exp}}}\right)^2 + \frac{a}{z_{\text{exp}}}}$ , we have that (A.12) is satisfied, if  $N \geq \frac{1}{4} \frac{p^2}{z_{\text{exp}}^2} + 4 \frac{q+n}{z_{\text{exp}}}$ . Since  $p = (\zeta + 1/\zeta)$  is increasing in  $\zeta$  for  $\zeta \geq 1$ , using (A.9) we have that

$$\zeta \leq c \Rightarrow p \leq (c + 1/c), \quad c \doteq \sqrt{2n + 2n \ln \frac{n+q}{q} - 2 \ln 2},$$

from which the statement in (13) follows. Further, for “large”  $q$  it holds that  $p = (\zeta + 1/\zeta) \simeq \zeta$ , hence  $p^2 \simeq 2 \ln(C/2)$  and, using (A.9),  $p^2 \simeq 2 \ln(C/2) \leq c^2$ , which proves the last statement in the corollary.  $\square$

### References

- [1] V.S. Bawa, S.J. Brown, and R.W. Klein. *Estimation Risk and Optimal Portfolio Choice*. North-Holland, Amsterdam, 1979.
- [2] A. Ben-Tal, T. Margalit, and A. Nemirovski. Robust modeling of multi-stage portfolio problems. In H. Frenk, K. Roos, T. Terlaky, and S. Zhang, editors, *High Performance Optimization*, pages 303–328. Kluwer Academic Publishers, 2000.
- [3] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.
- [4] D. Bertsimas and J.N. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, 1997.
- [5] G.C. Calafiore. Random convex programs. *SIAM Journal on Optimization*, 20(6):3427–3464, 2010.

- [6] G.C. Calafiore and M.C. Campi. Uncertain convex programs: randomized solutions and confidence levels. *Mathematical Programming*, 102(1):25–46, 2005.
- [7] G.C. Calafiore and M.C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.
- [8] J.Y. Campbell, A.W. Lo, and A.C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1996.
- [9] M.C. Campi, G.C. Calafiore, and S. Garatti. Interval predictor models: Identification and reliability. *Automatica*, 45(2):382–392, 2009.
- [10] M.C. Campi and S. Garatti. The exact feasibility of randomized solutions of uncertain convex programs. *SIAM Journal on Optimization*, 19(3):1211–1230, 2008.
- [11] Y. Chen, A. Wiesel, Y.C. Eldar, and A.O. Hero. Shrinkage algorithms for MMSE covariance estimation. *IEEE Transactions on Signal Processing*, 58(10):5016–5029, 2010.
- [12] V.K. Chopra and W.T. Ziemba. The effect of errors in means, variances and covariances on optimal portfolio choice. *Journal of Portfolio Management*, (Winter):6–11, 1993.
- [13] E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Journal of Operations Research*, 58(3):595–612, 2010.
- [14] L. El Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research*, 51(4):543–556, July 2003.
- [15] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.
- [16] C. Jabbour, J.F. Pena, J.C. Vera, and L.F. Zuluaga. An estimation-free, robust conditional value-at-risk portfolio allocation model. *The Journal of Risk*, 11(1):1–22, 2008.
- [17] P. Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, 2006.
- [18] M.S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebrete. Applications of second-order cone programming. *Linear Algebra Appl.*, 284:193–228, 1998.
- [19] S.T. Rachev, F.J. Fabozzi, and C. Menn. *Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing*. Wiley, 2005.
- [20] D. Ruppert. *Statistics and Data Analysis for Financial Engineering*. Springer-Verlag, 2011.
- [21] A. Shleifer. *Inefficient Markets: An Introduction to Behavioral Finance*. Clarendon Lectures in Economics, 2000.
- [22] S. Zymmler, B. Rustem, and D. Kuhn. Robust portfolio optimization with derivative insurance guarantees. *European Journal of Operations Research*, 210(2):410–424, 2011.