## AMBIGUOUS RISK MEASURES AND OPTIMAL ROBUST PORTFOLIOS\*

## GIUSEPPE C. CALAFIORE<sup>†</sup>

**Abstract.** This paper deals with a problem of guaranteed (robust) financial decision-making under model uncertainty. An efficient method is proposed for determining optimal robust portfolios of risky financial instruments in the presence of ambiguity (uncertainty) on the probabilistic model of the returns. Specifically, it is assumed that a *nominal* discrete return distribution is given, while the *true* distribution is only known to lie within a distance *d* from the nominal one, where the distance is measured according to the Kullback–Leibler divergence. The goal in this setting is to compute portfolios that are worst-case optimal in the mean-risk sense, that is, to determine portfolios that minimize the maximum with respect to all the allowable distributions of a weighted risk-mean objective. The analysis in the paper considers both the standard variance measure of risk and the absolute deviation measure.

 ${\bf Key}$  words. worst-case financial risk, portfolio selection, asset allocation, statistical ambiguity, robust optimization

AMS subject classifications. 91B28, 90B50, 90C25, 90C47, 90C51, 90C90

DOI. 10.1137/060654803

1. Introduction. A classical problem in computational finance is that of optimally selecting a portfolio of finitely many risky assets so as to maximize the expected return of the investment while keeping "risk" under control. In the mainstream approach, dating back to the seminal work of Markowitz [25], risk is measured according to the variance of the portfolio return, and the determination of an optimal portfolio amounts to the solution of a convex quadratic programming problem. Since the introduction of this basic mean-variance model for portfolio selection, however, many criticisms have been raised on its practical relevance, especially in regard to the sensitivity of the optimal portfolios with respect to the statistical errors in the parameters (the estimated expected returns and covariances of the assets), and possible remedies have been proposed. An in-depth overview of this literature is out of the scope of the present work, but the interested reader could find some useful pointers in [3, 6, 8, 26].

More recently, the issue of model uncertainty in portfolio optimization has been the subject of study from different groups of researchers; see, for instance, [13, 18, 24, 34, 37]. Many of these recent contributions propose ideas and computational tools derived from the robust convex optimization field [4, 14]. The general approach in this setting is to model the uncertain market parameters (expected returns and covariances) as deterministic unknown-but-bounded quantities, and then take a worstcase approach where an optimal portfolio is sought that minimizes the worst risk that the investor may face as the market parameters vary in any possible way inside their admissible domains. These deterministic models are practically and theoretically sound, since they either are naturally derived from confidence regions around the leastsquares estimates of the market parameters (see [18]) or may reflect an analyst feeling

<sup>\*</sup>Received by the editors March 22, 2006; accepted for publication (in revised form) February 5, 2007; published electronically October 4, 2007. This work was supported by MIUR under the FIRB project "Learning, Randomization and Guaranteed Predictive Inference for Complex Uncertain Systems."

http://www.siam.org/journals/siopt/18-3/65480.html

<sup>&</sup>lt;sup>†</sup>Dipartimento di Automatica e Informatica, Politecnico di Torino, 10129 Torino, Italy (giuseppe. calafiore@polito.it).

of the reliability of the parameter estimates. In this latter case, the uncertainty model typically takes the form of elementwise bounds on some or all entries of the expected return vector and covariance matrix; see [13, 34]. Specifically, El Ghaoui, Oks, and Oustry in [13] consider the problem of computing and optimizing the worst-case valueat-risk of a portfolio, under bounded uncertainty on the mean and covariance matrix of the returns, and show how the computation can be efficiently performed by recasting the problem in the form of a semidefinite optimization program [33, 35]. Goldfarb and Iyengar in [18] develop a robust factor model for the returns, show how the uncertainty description can be naturally obtained from confidence regions of standard statistical estimation techniques, and pose the corresponding robust allocation problem in the form of a convex second order cone program (SOCP); see [23]. Tütüncü and Koenig in [34] propose the use of an interval uncertainty model for the return mean and covariance and solve the resulting worst-case Markowitz problem via an ad hoc saddle-point algorithm.

While the mentioned approaches are specific to portfolio selection problems, more general models dealing with uncertainty in the underlying probability measures have a long history and have been studied in different fields, such as economics, finance, and stochastic optimization; see, e.g., [7, 11, 12, 32, 36]. Uncertainty in the probabilistic model is usually referred to as *ambiguity* in the decision theory literature. The recent work from Erdoğan and Iyengar [15] discusses ambiguous chance-constrained problems and employs the Prohorov metric to describe the uncertainty "ball" of admissible distributions. As we shall see in section 2.1, in this paper we adopt a similar approach for describing the "ambiguity" set around a nominal distribution and employ the Kullback–Leibler divergence function as a distance measure among distributions. This distance measure has nice invariance and convexity properties, and the degree of ambiguity in this metric can be estimated from samples; see [20].

The main goal of this paper is to present an efficient computational framework for robust portfolio selection in the situation of asset returns described by an ambiguous discrete joint probability distribution. We consider two risk measures (see [2, 30]) given by composite objectives of the form  $\rho(x,\pi) - \gamma \mu(x,\pi)$ , where  $\rho(x,\pi)$  is either the variance or the expected absolute deviation of the portfolio,  $\mu(x,\pi)$  is the portfolio expected return, and  $\gamma$  is a nonnegative parameter. Here, x denotes the portfolio mix and  $\pi$  the discrete distribution of the returns (see section 2 for precise definitions and notation). The measure based on the expected absolute deviation is (for  $\gamma \geq 2$ ) a *coherent* measure of risk in the sense of Artzner et al. [2]. The measure based on the variance is instead not coherent, since it violates a monotonicity condition; see [30]. However, the use of this latter measure is justified by both historical reasons and its wide popularity.

In the nominal case—i.e., when the probability distribution  $\pi$  is known and given—minimizing the above objectives is equivalent either to a standard Markowitz problem (in the case of the variance-based risk measure) or to the absolute deviation problem, discussed, for instance, in [21, 31]. It is well known that, in this latter case, the optimal portfolio can be found by solving a linear programming problem.

The key point in this paper is to consider the return distribution  $\pi$  to be imprecisely known. In particular, we assume that a nominal value  $\eta$  for the distribution is given, but that the actual  $\pi$  is only known to lie in a region at distance no larger than d from its nominal value, where d is a user-definable parameter that quantifies the (lack of) confidence in the nominal probability (the "index of ambiguity," in the terminology of [20]). To measure the distance among distributions, we use the standard metric given by the Kullback–Leibler divergence. In this setting, we define the worst-case risk of a portfolio x as the supremum of  $\rho(x, \pi) - \gamma \mu(x, \pi)$  for  $\pi$  that ranges over its uncertainty set. An optimal worst-case portfolio is a composition vector x that minimizes this worst-case risk.

We detail in the paper two numerical schemes that permit us to efficiently evaluate and optimize the worst-case risk in both the variance and the absolute deviation cases. For the variance-based risk measure, the worst-case optimal portfolio can be determined using an interior-point barrier method, in conjunction with an analytic center cutting plane technique. The absolute deviation-based risk measure poses a slight additional complication, due to nonconcavity in  $\pi$  of this function. This issue is here resolved by adding a suitable line search to the algorithm.

The paper is organized as follows. Section 2 sets the stage by providing the basic definitions and introducing the distribution uncertainty model. Section 3 discusses a barrier method for computing the worst-case variance-based risk of a given portfolio, whereas section 4.1 describes the overall cutting plane algorithm for optimizing the worst-case risk over the portfolio composition. Section 5 extends the methodology to the absolute deviation-based risk measure. Some numerical examples are presented in section 6, and conclusions are finally drawn in section 7. To improve readability, some of the technical details have been relegated to appendices.

**1.1. Notation.** Whenever useful for notational compactness, we use MATLABlike notation for operations on vectors. If x, y are two vectors of compatible dimensions, relational operators such as  $>, \ge$ , etc., are to be intended elementwise (e.g., x > y means that all entries of vector x - y are positive). Similarly, powers and operators +, -, \*, / work elementwise, and the same holds for standard functions. For example,  $\log x/y$  denotes a vector whose *i*th entry is  $\log x_i/y_i$ .

**2.** Preliminaries. Consider a collection of assets or asset classes  $a_1, \ldots, a_n$  and let

$$r \doteq \begin{bmatrix} r_1 & \cdots & r_n \end{bmatrix}^\top$$

be a random vector describing the returns of the considered assets over a fixed period of time. Let  $r(1), \ldots, r(T)$  be T possible scenarios for the outcomes of the random return vector r, and let  $\pi_k$  be the probability associated to the scenario r(k), with the obvious properties that

$$\pi_k \ge 0, \quad k = 1, \dots, T$$

$$\sum_{k=1}^T \pi_k = 1.$$

Defining the probability vector

$$\pi \doteq [\pi_1 \cdots \pi_T]^{\top},$$

the two previous conditions are simply rewritten as  $\pi \ge 0$ ,  $\mathbf{1}^{\top} \pi = 1$ , where **1** denotes a vector of ones of suitable dimensions. Now let

$$x \doteq \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$$

be a vector such that  $x_i$  represents the fraction of an investor portfolio that is invested in asset  $a_i$ . We shall refer to x as the "portfolio composition," or "portfolio mix." The portfolio composition can be subject to various kinds of constraints, which we assume to be representable by the condition

$$x \in \mathcal{X}$$
, where  $\mathcal{X}$  is a given polytope.

For example, a typical form for the set  $\mathcal{X}$  is

(2.1) 
$$\mathcal{X} = \left\{ x : \sum_{i=1}^{n} x_i = 1, \ x_i \ge 0 \text{ for } i = 1, \dots, n \right\},$$

which reflects the standard situation where the investor cannot hold a negative amount of an asset (i.e., short-selling is not allowed). However, the results in this paper are not restricted to the specific admissible portfolios set in (2.1) and apply to the general polytopic case.

With the positions above, the investor's total return at the end of the investment period is represented by the random variable

(2.2) 
$$w \doteq r^{\top} x,$$

whose expected value is

$$\mu(x,\pi) \doteq \mathbf{E}\left\{r^{\top}x\right\} = \sum_{k=1}^{T} \pi_k r^{\top}(k) x = \left(\sum_{k=1}^{T} \pi_k r^{\top}(k)\right) x = \hat{r}^{\top}(\pi) x,$$

where  $\hat{r}(\pi) \doteq \mathbf{E} \{r\} = \sum_{k=1}^{T} \pi_k r(k)$ . The portfolio risk is quantified as a measure of variability of w around its expectation. A classical measure of variability (see, e.g., [25]) is given by the variance

(2.3) 
$$\rho_2(x,\pi) \doteq \mathbf{E}\left\{\left(r^\top x - \mathbf{E}\left\{r^\top x\right\}\right)^2\right\} = x^\top \Sigma(\pi) x,$$

where

$$\Sigma(\pi) \doteq \mathbf{E}\left\{ (r - \hat{r}(\pi))(r - \hat{r}(\pi))^{\top} \right\} = \sum_{k=1}^{T} \pi_k (r(k) - \hat{r}(\pi))(r(k) - \hat{r}(\pi))^{\top}$$

is the covariance matrix of r. In this paper, we also consider an alternative measure of risk, which is based on the expected absolute deviation, and whose justification in the portfolio selection context is discussed, for instance, in [21, 31]:

$$\rho_1(x,\pi) \doteq \mathbf{E}\left\{ \left| r^{\top} x - \mathbf{E}\left\{ r^{\top} x\right\} \right| \right\} = \sum_{k=1}^T \pi_k |r^{\top}(k)x - \mu(x,\pi)|.$$

Following a mean-risk approach, we introduce an objective function which represents a tradeoff between risk (variance or expected absolute deviation) and expected return of the portfolio. Specifically, for given  $\gamma \geq 0$ , we define an objective based on the variance measure

(2.4) 
$$\Upsilon_2(x,\pi) \doteq \rho_2(x,\pi) - \gamma \mu(x,\pi)$$

and one based on the absolute deviation measure

(2.5) 
$$\Upsilon_1(x,\pi) \doteq \rho_1(x,\pi) - \gamma \mu(x,\pi).$$

Notice that if the probability distribution  $\pi$  is known and given, then minimizing  $\Upsilon_2(x,\pi)$  over  $x \in \mathcal{X}$  is a well-known Markowitz problem, whose solution can be obtained by solving numerically a convex quadratic programming problem. Minimizing  $\Upsilon_1(x,\pi)$  in this same situation amounts instead to solving a linear programming problem; see, for instance, [21, 31].

The point of this paper is to propose computationally efficient schemes for determining optimal worst-case portfolios, when the probability distribution  $\pi$  is not precisely known. To this end, we introduce in the next section an uncertainty model for  $\pi$  and define the related robust risk functions.

2.1. Distribution ambiguity and robust measures of risk. Assume that a nominal return probability distribution  $\eta$  is given, for instance, as a result of estimation from samples. Then the Kullback-Leibler (KL) divergence (see [22]) represents a natural measure of the expected amount of information in a sample from the unknown distribution for discriminating against  $\eta$  [19], and it is a frequently used information-theoretic "distance" measure between probability distributions; see, e.g., [1, 10]. If  $\pi$ ,  $\eta$  are two probability vectors in  $\mathbb{R}^T$ , with  $\eta > 0$  describing the nominal probability, the KL distance between  $\pi$  and  $\eta$  is defined as

$$\mathrm{KL}(\pi, \eta) \doteq \sum_{k=1}^{T} \pi_k \log \frac{\pi_k}{\eta_k}$$

We shall henceforth assume that the "true" probability  $\pi$  is only known to lie within KL distance  $d \ge 0$  from  $\eta$ , i.e.,  $\pi \in \mathcal{K}(\eta, d)$ , where

$$\mathcal{K}(\eta, d) \doteq \{ \pi \in \Pi : \mathrm{KL}(\pi, \eta) \le d \},\$$

 $\Pi$  being the probability simplex  $\Pi = \{\pi : \pi \ge 0, \mathbf{1}^\top \pi = 1\}.$ 

 $\mathcal{K}(\eta, d)$  thus represents the *ambiguity set* for the return distribution, and  $d \ge 0$  is the uncertainty level (radius of ambiguity). The risk functions (2.4), (2.5), with  $\pi \in \mathcal{K}(\eta, d)$ , are *ambiguous* risk functions. The nominal distribution  $\eta$  and the ambiguity level may either be assigned by expert advice or estimated from data; see, for instance, [20]. In what follows we shall not investigate further the issue of determination of  $\eta$ and d and shall assume that these quantities are given data.

Remark 1 (domain, range, and convexity of  $\text{KL}(\pi, \eta)$ ). Notice that, since  $\pi_k \log \pi_k$  is a convex function over the domain  $\pi_k \geq 0$ ,<sup>1</sup> then  $\text{KL}(\pi, \eta)$  is a convex function in  $\pi$  over  $\Pi$ , and hence the uncertainty set  $\mathcal{K}(\eta, d)$  is convex.

The function  $\operatorname{KL}(\pi, \eta)$ , with  $\pi \in \Pi$ , takes values in the interval  $[0, \log 1/\eta_{\min}]$ , where  $\eta_{\min} = \min_{k=1,\dots,T} \eta_k$ . The lower end of the interval is attained for  $\pi = \eta$ , whereas the higher end is attained for  $\pi = e_i$ , where *i* is the index of the smallest element in  $\eta$ , and  $e_i$  is the *i*th vector in the standard basis of  $\mathbb{R}^T$ .

Figure 2.1 gives a pictorial idea of the shape of the set  $\mathcal{K}(\eta, d)$  in a threedimensional example where  $\eta$  is assumed to be the uniform distribution.

We pursue a worst-case approach in dealing with ambiguity in the risk functions. Notice that it is known (see [30, Theorem 2]) that the risk functions (2.4), (2.5), for a fixed probability  $\pi$ , can be represented in dual form as the result of a maximization

$$\Upsilon(x,\pi) = \max_{\zeta \in A(\pi)} \left[ \langle \zeta, r^{\top} x \rangle_{\pi} - \Upsilon^*(\zeta) \right],$$

<sup>&</sup>lt;sup>1</sup>It is assumed by continuity that  $0 \log 0 = 0$ .



FIG. 2.1. A visualization of the subsets  $\mathcal{K}(\eta, d)$  of the probability simplex in a three-dimensional example with  $\eta = [1/3 \ 1/3 \ 1/3]$  and  $d = 0.01, \ 0.05, \ 0.1, \ 0.2$ .

where  $A(\pi)$  is a closed convex set of measures,  $\Upsilon^*$  is the conjugate of  $\Upsilon$ , and  $\langle \zeta, w \rangle_{\pi} = \sum_i \zeta_i w_i \pi_i$ . In the the worst-case approach that we follow in this paper, the "robust-ness" of the nominal risk functions is improved by adding a second level of maximization over a set of admissible probabilities  $\mathcal{K}$ . That is, we shall consider robust risk functions of the form

$$\Upsilon_{\rm wc}(x) = \max_{\pi \in \mathcal{K}} \Upsilon(x, \pi) = \max_{\pi \in \mathcal{K}} \max_{\zeta \in A(\pi)} \left[ \langle \zeta, r^{\top} x \rangle_{\pi} - \Upsilon^*(\zeta) \right].$$

Specifically, given the ambiguity model  $\mathcal{K}(\eta, d)$  for the return distribution, we define the following worst-case (or robust) measures of risk for a portfolio with composition x:

(2.6) 
$$\Upsilon_{\mathrm{wc2}}(x) \doteq \max_{\pi \in \mathcal{K}(\eta, d)} \rho_2(x, \pi) - \gamma \mu(x, \pi)$$

for the variance-based measure, and

(2.7) 
$$\Upsilon_{\mathrm{wc1}}(x) \doteq \max_{\pi \in \mathcal{K}(\eta, d)} \rho_1(x, \pi) - \gamma \mu(x, \pi)$$

for the absolute deviation-based measure.

The distribution  $\pi_{wc}$  that attains the supremum in the above optimization problems is named the *worst-case distribution*, and the corresponding value function  $\Upsilon_{wc}(x)$  is the *worst-case risk* (to uncertainty level d) of the portfolio x. In the next section we provide an efficient numerical scheme for solving (2.6). We anticipate that the existence of a polynomial-time algorithm for computing the worst-case variancebased risk is due to the fact that we can construct a self-concordant barrier for the convex domain  $\mathcal{K}(\eta, d)$ . Successively, in section 4, we develop a polynomial-time algorithm that permits us to further optimize  $\Upsilon_{wc2}(x)$  with respect to x, and hence to find an optimal portfolio mix that minimizes the worst-case variance-based risk. In section 5, we describe a similar approach for dealing with the absolute deviation-based objective (2.7).

3. Computing the worst-case variance-based risk. Let the portfolio composition x be fixed, let  $w(k) = r^{\top}(k)x$ , k = 1, ..., T, and define

$$\mathbf{w} \doteq \begin{bmatrix} w(1) & \cdots & w(T) \end{bmatrix}^\top.$$

Then from (2.2)-(2.3) we have

$$\rho_2(x,\pi) = \mathbf{E}\left\{(w - \mathbf{E}\left\{w\right\})^2\right\} = \mathbf{E}\left\{w^2\right\} - \mathbf{E}^2\{w\} = \pi^\top \mathbf{w}^2 - \pi^\top \Omega \pi,$$

where  $\Omega \doteq \mathbf{w}\mathbf{w}^{\top}$ . Since  $\Omega$  is symmetric positive semidefinite, it follows that  $\rho_2(x,\pi)$  is a concave function of the probability vector  $\pi$ . Therefore, the objective function (2.4),

$$\Upsilon_2(x,\pi) = \rho_2(x,\pi) - \gamma \mu(x,\pi) = -\pi^{\top} \Omega \pi - \pi^{\top} (\gamma \mathbf{w} - \mathbf{w}^2),$$

is also concave in  $\pi$ ; hence problem (2.6) can be written in the equivalent form of a convex minimization problem as follows:

(3.1) 
$$\Upsilon_{wc2} = -\min \pi^{\top} \Omega \pi + \pi^{\top} (\gamma \mathbf{w} - \mathbf{w}^2)$$

(3.2) subject to 
$$\operatorname{KL}(\pi, \eta) \leq d$$
,

$$(3.3) \pi \ge 0,$$

$$\mathbf{1}^{ op}\pi=1.$$

We next develop an interior-point barrier method for solving problem (3.1).

**3.1.** A logarithmic barrier method. For a fixed portfolio x, we solve problem (3.1) by solving a sequence of equality constrained problems of the form

(3.4) 
$$\min_{\pi} f(\pi) \doteq t f_0(\pi) + \phi(\pi)$$

(3.5) subject to 
$$\mathbf{1}^{\top}\pi = 1$$

for increasing values of  $t \ge 0$ , where

$$f_0(\pi) \doteq \pi^\top \Omega \pi + \pi^\top (\gamma \mathbf{w} - \mathbf{w}^2)$$

is the objective function of (3.1),

(3.6) 
$$b(\pi) \doteq \sum_{k=1}^{T} \pi_k \log \pi_k - \sum_{k=1}^{T} \pi_k \log \eta_k - d,$$

and

(3.7) 
$$\phi(\pi) \doteq -\log(-b(\pi)) - \sum_{k=1}^{T} \log \pi_k$$

is a logarithmic barrier for the inequality constraints (3.2), (3.3). For fixed  $t \ge 0$ , we denote by  $\pi^*(t)$  the corresponding optimal solution of (3.4). The *central path* is the curve  $\pi^*(t)$  obtained varying t from 0 to  $\infty$ . A standard implementation of a barrier method is hence the following.

Algorithm 1 (barrier method [5]).

Given strictly feasible  $\pi$ , set t = t(0) > 0,  $\rho > 1$ , tolerance  $\epsilon > 0$ . repeat

- 1. Centering step:
- Compute  $\pi^*(t)$  by solving (3.4) using the Newton method, starting at  $\pi$ .
- 2. Update:  $\pi = \pi^*(t)$ .

3. Stopping criterion: quit if  $(T+1)/t < \epsilon$ .

4. <u>Increase t</u>:  $t = \rho t$ .

Notice that since  $\eta > 0$ , an initial feasible point for the algorithm is simply given by  $\pi = \eta$ . For this algorithm,  $\pi^*(t)$  tends to the optimal solution of problem (3.1) as  $t \to \infty$ . The convergence properties of the method are analyzed in terms of the number of *outer iterations* (centering steps) needed to reach a solution with the desired accuracy  $\epsilon$  and the number of *inner iterations* (i.e., the iterations required by the Newton method to compute each center, up to accuracy  $\epsilon_{nw}$ ). A standard result states that the number of outer iterations (centering steps) is given exactly by (see [5, section 11.3.3])

$$1 + \left\lceil \frac{\log(T+1)/(\epsilon t(0))}{\log \varrho} \right\rceil$$

The analysis of complexity of each centering step relies on the property of *self*-concordance of the objective function in (3.4), which is discussed next.

**3.2. Centering step and self-concordance.** In their seminal work [28], Nesterov and Nemirovskii provided a key condition under which the complexity of the Newton method could be analyzed, that is, self-concordance of the objective function. We next show that the objective function in (3.4) is indeed self-concordant and provide a bound on the number of Newton steps required in each centering phase.

We start with some definitions. A function of a scalar variable  $\psi(z) : \mathbb{R} \to \mathbb{R}$  is self-concordant if it is convex and

$$|\psi^{(3)}(z)| \le k\psi^{(2)}(z)^{3/2}$$

for all z in the domain of  $\psi$ , where  $\psi^{(2)}, \psi^{(3)}$  denote the second and the third derivatives of  $\psi$ , respectively, and k is a positive constant. A function  $\psi(z)$  of vector variable  $z \in \mathbb{R}^n$  is self-concordant if it is self-concordant along any line in its domain, i.e., if the function of scalar variable  $\tilde{\psi}(\alpha) \doteq \psi(z + \alpha v)$  is a self-concordant function of  $\alpha \in \mathbb{R}$ for all z in the domain of  $\psi$  and for all v.

The following proposition on the self-concordance of the barrier function (3.7) holds; see Appendix A for a proof.

PROPOSITION 3.1. The function  $\phi(\pi)$  in (3.7) is a self-concordant barrier for the domain

$$\{(\pi, d): \pi > 0, \varphi(\pi) < d\}.$$

Since the sum of self-concordant functions is self-concordant, and since convex quadratic functions are obviously self-concordant (they have zero third derivative; see also some standard rules of "self-concordant calculus" in [5]), we deduce from

Proposition 3.1 that the objective function in (3.4) is indeed self-concordant. It then follows that using an (equality-constrained) Newton method with backtracking linesearch, each center can be computed up to accuracy  $\epsilon_{nw}$  in at most (see, for instance, [5, section 11])

$$\frac{(T+1)(\varrho-1-\log \varrho)}{\ell} + \log_2 \log_2 \frac{1}{\epsilon_{\mathrm{nv}}}$$

Newton steps, where  $\ell$  is a constant that depends on two technical parameters used in the line-search phase of the algorithm. We conclude this section by reporting explicitly the gradient and Hessian of the function  $f(\pi)$  in (3.4). We have

$$\nabla f(\pi) = t \nabla f_0(\pi) + \nabla \phi(\pi),$$
  
$$\nabla^2 f(\pi) = t \nabla^2 f_0(\pi) + \nabla^2 \phi(\pi)$$

with

$$\nabla f_0(\pi) = 2\Omega \pi + (\gamma \mathbf{w} - \mathbf{w}^2),$$
  
$$\nabla \phi(\pi) = \frac{1}{-b(\pi)} \nabla b(\pi) - \pi^{-1}; \quad \nabla b(\pi) = \mathbf{1} + \log \frac{\pi}{\eta}$$

and

$$\nabla^2 f_0(\pi) = 2\Omega,$$
  

$$\nabla^2 \phi(\pi) = \frac{1}{-b(\pi)} \operatorname{diag}(\pi^{-1}) + \frac{1}{b^2(\pi)} \nabla b(\pi) \nabla^\top b(\pi) + \operatorname{diag}(\pi^{-2}).$$

4. Minimizing the worst-case variance-based risk. In the previous section, we described a numerically efficient technique for computing the worst-case risk of a given portfolio mix x, i.e., for evaluating the function  $\Upsilon_{wc2}(x)$  in (2.6). We now elaborate on this technique and develop an efficient algorithm for determining a portfolio mix that minimizes the worst-case risk. That is, we now aim at solving the portfolio design problem

(4.1) 
$$\min_{x \in \mathcal{X}} \Upsilon_{\mathrm{wc2}}(x).$$

We shall do so by employing an analytic center cutting plane technique, which is described in the next section. Notice preliminarily that function  $\Upsilon_2(x, \pi)$  in (2.4),

$$\Upsilon_2(x,\pi) = x^T \Sigma(\pi) x - \gamma \hat{r}^\top(\pi) x,$$

is convex (and quadratic) in x for any given  $\pi$ , whence the function  $\Upsilon_{wc2}(x)$ , which is defined as the pointwise maximum of  $\Upsilon_2(x,\pi)$  over  $\pi$ , is also convex in x. At any given  $\pi$ , the gradient of  $\Upsilon_2(x,\pi)$  with respect to x is

(4.2) 
$$\nabla_x \Upsilon_2(x,\pi) = 2\Sigma(\pi)x - \gamma \hat{r}(\pi).$$

The gradient defines a supporting hyperplane for the epigraph of  $\Upsilon_2(x,\pi)$ , i.e.,

(4.3) 
$$\Upsilon_2(z,\pi) \ge \Upsilon_2(x,\pi) + \left[\nabla_x \Upsilon_2(x,\pi)\right]^\top (z-x) \quad \forall z \in \mathcal{X}.$$

Now let x be a given point and let  $\pi^*(x)$  be the probability vector that attains the optimal value in problem (2.6) (such an optimal argument is attained, since the feasible set is compact), so that

(4.4) 
$$\Upsilon_{\mathrm{wc2}}(x) = \Upsilon_2(x, \pi^*(x)).$$

Evaluating (4.3) in  $\pi = \pi^*(x)$ , we get

$$\Upsilon_2(z,\pi^*(x)) \ge \Upsilon_2(x,\pi^*(x)) + \left[\nabla_x \Upsilon_2(x,\pi^*(x))\right]^{\perp} (z-x) \quad \forall z \in \mathcal{X}.$$

Since  $\Upsilon_{wc2}(z) \ge \Upsilon_2(z, \pi^*(x))$ , continuing the previous inequality on the left and using (4.4), we obtain

(4.5) 
$$\Upsilon_{\mathrm{wc2}}(z) \ge \Upsilon_{\mathrm{wc2}}(x) + [\nabla_x \Upsilon_2(x, \pi^*(x))]^\top (z-x) \quad \forall z \in \mathcal{X},$$

which means that  $\nabla_x \Upsilon_2(x, \pi^*(x))$  is a subgradient of  $\Upsilon_{wc2}(x)$  at the point x. Notice that each time we solve problem (2.6)—or its equivalent formulation (3.1)—for a given x, we get both the value of  $\Upsilon_{wc2}(x)$  and the worst-case probability vector  $\pi^*(x)$ , and hence (evaluating (4.2) for  $\pi = \pi^*(x)$ ) a subgradient of  $\Upsilon_{wc2}(x)$  at x.

**4.1.** An analytic center cutting plane algorithm for optimizing the portfolio mix. We now briefly describe an analytic center cutting plane (ACCP) method for solving problem (4.1). An overview of ACCP techniques for convex optimization can be found, for instance, in [29].

Let initially  $\mathcal{P}_1 = \mathcal{X}$ , and compute the analytic center  $x^{(1)}$  of  $\mathcal{P}_1$ . The analytic center of a polytope can be efficiently computed by minimizing a logarithmic barrier via a Newton-type algorithm; see, for instance, [17]. Then solve problem (2.6) to get  $\Upsilon_{wc2}(x^{(1)})$ , along with the worst-case probability  $\pi^*(x^{(1)})$  and a subgradient

$$g_1 \doteq \nabla_x \Upsilon_2(x^{(1)}, \pi^*(x^{(1)}))$$

of  $\Upsilon_{wc2}(x)$  at  $x = x^{(1)}$ . Using inequality (4.5), notice next that for all points in the hyperplane

$$\{z: g_1^\top (z - x^{(1)}) > 0\}$$

we have that  $\Upsilon_{wc2}(z) > \Upsilon_{wc2}(x^{(1)})$ , hence all such points are worse than the current point  $x^{(1)}$  in terms of the objective value that we are trying to minimize. Therefore, the optimal point should lie in the complementary hyperplane

$$\mathcal{H}_1 \doteq \{ z : g_1^\top (z - x^{(1)}) \le 0 \}.$$

Hence, we update the current polytope by adding the constraint  $\mathcal{H}_1$  to  $\mathcal{P}_1$ , i.e., we set

$$\mathcal{P}_2 = \mathcal{P}_1 \cap \mathcal{H}_1$$

and iterate the whole process (compute the analytic center  $x^{(2)}$  of  $\mathcal{P}_2$ , etc.).

The convergence of this method relies on the fact that the polytopes  $\mathcal{P}_k$  shrink at each iteration, thus eventually localizing the optimal solution  $x^*$ . The ACCP method converges to a solution in polynomial time. A precise assessment of the numerical complexity of the ACCP method and some of its variants has been discussed in several papers; see, for instance, [16, 17].

In a practical implementation of the method, we may terminate the iterations if either  $||x^{(k)} - x^{(k-1)}||$  goes below a given threshold  $\epsilon_{ac}$  or the Chebyshev radius of  $\mathcal{P}_k$  becomes sufficiently small.<sup>2</sup>

 $<sup>^{2}</sup>$ The Chebyshev radius of a polytope is defined as the radius of the largest Euclidean hypersphere contained in the polytope. Computing the Chebyshev radius amounts to solving a linear programming problem; therefore checking the exit condition based on the Chebyshev radius requires some additional numerical effort.

A schematic implementation of an algorithm that permits us to solve (up to a numerical tolerance) the robust portfolio design problem is given next.

Algorithm 2 (ACCP).

Given the exit tolerance  $\epsilon_{ac} > 0$ , and the initial polytope  $\mathcal{X}$ , set  $\mathcal{P} = \mathcal{X}$ . repeat

- 1. <u>Centering step</u>:
  - $\overline{\text{Compute the analytic center } x \text{ of } \mathcal{P}.$
- 2. Solve subproblem (2.6):
- Compute  $\Upsilon_{wc2}(x)$ ,  $\pi^*(x)$ , and a subgradient g of  $\Upsilon_{wc2}$  at x. 3. Stopping criterion:
  - If Chebyshev-radius( $\mathcal{P}$ ) <  $\epsilon_{ac}$ , then quit.
- 4. Update the polytope: Set  $\mathcal{P} = \mathcal{P} \cap \{z : g^{\top}(z-x) \leq 0\}.$

5. The worst-case absolute deviation-based risk. The robust design approach outlined in the previous section for the variance-based measure can be extended to the absolute deviation-based measure (2.7). In this section, we mainly discuss how to evaluate the worst-case absolute deviation risk of a given portfolio x,

(5.1) 
$$\Upsilon_{\mathrm{wc1}}(x) \doteq \max_{\pi \in \mathcal{K}(\eta, d)} \Upsilon_1(x, \pi),$$

and then hint at how to minimize  $\Upsilon_{wc1}(x)$  over  $x \in \mathcal{X}$  in section 5.1.1. This latter process is completely analogous to the one described for the variance-based risk function.

The main technical difference with respect to the case considered previously is that, contrary to the variance function  $\rho_2(x, \pi)$ , the absolute deviation function  $\rho_1(x, \pi)$  is *not* concave in  $\pi$ ; see, e.g., Figure 5.1, and Appendix B for a proof. Therefore, the inner problem (5.1) cannot be solved by directly using a logarithmic barrier method such as the one described in section 3.1. However, we show in the next section that (5.1) can still be solved efficiently to any given accuracy by using the barrier method in conjunction with a one-dimensional search.

5.1. Evaluating the worst-case absolute deviation risk. We assume the portfolio composition x to be fixed, and use the notation introduced in section 3. The absolute measure and objective function in (5.1) are then

(5.2) 
$$\rho_1(\pi) \doteq \rho_1(x,\pi) = \sum_{k=1}^T \pi_k |w(k) - \mathbf{w}^\top \pi|,$$

(5.3) 
$$\Upsilon_1(\pi) \doteq \Upsilon_1(x,\pi) = \sum_{k=1}^T \pi_k |w(k) - \mathbf{w}^\top \pi|, -\gamma \mathbf{w}^\top \pi.$$

As we mentioned before, the function  $\rho_1(x,\pi)$ , and hence  $\Upsilon_1(x,\pi)$ , is not concave (nor convex) over  $\pi \in \Pi$ . However,  $\Upsilon_1(x,\pi)$  is concave (and actually linear) in  $\pi$  if we fix the value of the expected value:  $\mathbf{w}^{\top}\pi \stackrel{!}{=} \mu$ .

The solution idea is therefore the following one. First, determine the extreme feasible values for the portfolio mean return:

(5.4) 
$$\mu_{\min} = \min_{\pi \in \mathcal{K}(\eta, d)} \mathbf{w}^{\top} \pi; \quad \mu_{\max} = \max_{\pi \in \mathcal{K}(\eta, d)} \mathbf{w}^{\top} \pi$$



FIG. 5.1. Projection of the superlevel sets of (5.2) on the  $\pi_1, \pi_2$  plane in an example with T = 3 and  $\mathbf{w} = [0.1 \ 0.2 \ 0.3]^{\top}$ .

(we show in Appendix C that these two values can be computed very quickly via a scalar bisection algorithm). Then, for  $\mu \in [\mu_{\min}, \mu_{\max}]$ , define

(5.5) 
$$\varphi(\mu) \doteq \max_{\pi \in \mathcal{K}(\eta, d)} \sum_{k=1}^{T} \pi_{k} |w(k) - \mu| - \gamma \mu$$
subject to  $\mathbf{w}^{\top} \pi = \mu$ .

Clearly, we have that

(5.6) 
$$\Upsilon_{\mathrm{wc1}}(x) = \max_{\pi \in \mathcal{K}(\eta, d)} \Upsilon_1(x, \pi) = \max_{\mu \in [\mu_{\mathrm{min}}, \ \mu_{\mathrm{max}}]} \varphi(\mu).$$

In practice, we divide the interval  $[\mu_{\min}, \mu_{\max}]$  into N grid points  $\mu_1, \ldots, \mu_N$ , where N is chosen in accordance to the desired solution accuracy. For  $i = 1, \ldots, N$ , computing  $\varphi(\mu_i)$  is a convex optimization program that can be solved efficiently using, for instance, a barrier method such as the one described in section 3.1. An approximate solution to (5.1) is hence given by

$$\Upsilon_{\mathrm{wc1}}(x) \simeq \max_{i=1,\dots,N} \varphi(\mu_i).$$

Notice that the main difficulty in (5.5) is due to the presence of the KL constraint. It is instructive to detail the solution of (5.5) in the particular situation when this constraint is not present, since a closed-form solution is obtained in this case. The solution in (5.7) is computed basically at no cost, and it is optimal for problem (5.5) if the constraint  $KL(\pi, \eta) \leq d$  happens to be inactive.

PROPOSITION 5.1. Assume without loss of generality that the values  $w(1), \ldots, w(T)$  are arranged in increasing order, and let  $\pi$  be such that

(5.7) 
$$\pi_1 = \frac{w(T) - \mu}{w(T) - w(1)}, \quad \pi_2 = 0, \dots, \ \pi_{T-1} = 0, \quad \pi_T = \frac{\mu - w(1)}{w(T) - w(1)}.$$

If  $KL(\pi, \eta) \leq d$ , then  $\pi$  is an optimal solution for problem (5.5), with corresponding optimal value function

(5.8) 
$$\varphi(\mu) = 2 \frac{-\mu^2 + \mu(w(1) + w(T)) - w(1)w(T)}{w(T) - w(1)} - \gamma \mu$$

A proof of this proposition is given in Appendix D.

**5.1.1. Optimizing over the portfolio composition.** The procedure described in the previous section can further be wrapped by a cutting plane scheme, similarly to the one described in section 4, in order to optimize  $\Upsilon_{wc1}(x)$  over the portfolio mix x.

Notice that the absolute deviation measure (2.5) is convex in x for any given  $\pi$ . A subgradient of  $\Upsilon_1(x,\pi)$  at point x is given by

$$g_x(x,\pi) = \sum_{k=1}^{T} \pi_k (r(k) - \hat{r}(\pi)) s_k - \gamma \hat{r}(\pi),$$

where

$$s_k \doteq \begin{cases} 1 & \text{if } (r(k) - \hat{r})x \ge 0, \\ -1 & \text{otherwise.} \end{cases}$$

Now let  $\pi^*(x)$  denote the probability of attaining the optimum in problem (5.6): following steps similar to (4.2)–(4.5) we have that  $g_x(x, \pi^*(x))$  is a subgradient of  $\Upsilon_{wc1}(x)$ at x. This subgradient can be used in the cutting plane scheme of section 4.1, thus providing an overall polynomial-time method to solve the worst-case design problem  $\min_{x \in \mathcal{X}} \Upsilon_{wc1}(x)$ .

6. Numerical examples. We considered a financial allocation problem over five asset classes, where each class is represented by a sector index. We used the following indices to represent the classes: (1) Russell 1000 Large Cap Growth Index (RKGR), (2) Russell 1000 Large Cap Value Index (RKVA), (3) Russell 2000 Small Cap Growth Index (R2KGR), (4) Russell 2000 Small Cap Value Index (R2KVA), (5) Merrill Lynch Intermediate Bond Index (MACTX), with historical data of daily logarithmic returns collected over the period from July 14, 2004 to December 30, 2005 (T = 371 scenarios). We assumed that the return on the next day after the observed period can take on any of the historical values, with equal probability. This amounts to choosing a uniform nominal distribution  $\eta$  on the scenarios, which also conforms to the approach undertaken in [21, 31].

Return-risk analysis. We first analyze a fixed portfolio  $x_{\text{fix}}$  which allocates 30% of the wealth in bonds, and the rest equally distributed among the remaining assets. The nominal expected return for this portfolio is  $\mu(x_{\text{fix}}, \eta) = 3.8782 \times 10^{-4}$ , the nominal



FIG. 6.1. Upper and lower limits for the expected return of portfolio  $x_{\text{fix}}$  as a function of the percent uncertainty level (%u.l.).

variance is  $\rho_2(x_{\text{fix}}, \eta) = 3.3307 \times 10^{-5}$ , and the nominal absolute deviation risk is  $\rho_1(x_{\text{fix}}, \eta) = 0.0046$ .

We can perform various worst-case analyses on this portfolio. First, we computed the range of variation of the expected return, using the technique discussed in Appendix C. The results are shown in Figure 6.1 (the uncertainty level in the plots is expressed in percent units of the maximum allowable value of d, i.e.,  $\% u.l. = 100 \frac{d}{\log 1/\eta_{\min}}$ ).

Next, we evaluated the worst-case variance-based risk (2.6) of portfolio  $x_{\text{fix}}$ , with  $\gamma = 0.1$  and for increasing values of the percent uncertainty level; see Figure 6.2. An analogous plot, obtained from the absolute deviation-based risk measure (2.7) is instead shown in Figure 6.3. Notice from these plots that relatively low uncertainty levels may induce significant variations in the risk measure, with respect to the nominal (no uncertainty) situation.

*Return-risk optimization.* We next tested the ACCP algorithm described in section 4.1 for optimizing the worst-case variance-based risk. We computed worst-case optimal portfolios at different levels of uncertainty, which resulted in the plot shown in Figure 6.4. The composition of the worst-case optimal portfolios is shown in Figure 6.5.

Numerical performance. In the previous numerical tests, based on nonoptimized codes run under MATLAB 7.2 on an AMD Opteron 280 workstation, we experienced times of less than one minute to compute a worst-case optimal portfolio to an  $\epsilon_{\rm ac} = 10^{-5}$  accuracy.

As a further example, we considered the 30 assets composing the Dow Jones Industrial Average Index (DJI) and collected T = 138 historical daily return scenarios from March 24, 2006 to October 9, 2006. We ran the variance-based ACCP algorithm on these data, setting  $\gamma = 0.2$ , % u.l. = 0.1, and exit accuracy  $\epsilon_{\rm ac} = 10^{-5}$ . Algorithm 2 executed 82 iterations before returning the optimal portfolio. The total execution time was 23.9 seconds. Figure 6.6 shows the reduction in the Chebyshev radius of the



FIG. 6.2. Worst-case variance-based risk measure of portfolio  $x_{\rm fix}$  as a function of %u.l.



FIG. 6.3. Worst-case absolute deviation-based risk measure of portfolio  $x_{\rm fix}$  as a function of %u.l.



FIG. 6.4. Dashed line: risk measure of portfolio  $x_{\rm fix}$  (same as in Figure 6.2). Solid line: risk measure for optimal portfolios.



 ${\rm Fig.}~6.5.~Composition~of~worst-case~optimal~portfolios.$ 



FIG. 6.6. Chebyshev radius of the localization polytope versus iteration count for computing an optimal robust portfolio in the DJI example.



FIG. 6.7. Nominal efficient frontier for DJI data (dashed), and robust version (solid) for ambiguity level % u.l. = 0.01.

localization polytope versus the iteration count.

Finally, Figure 6.7 shows a discrete approximation of a portion of the nominal and robust efficient frontiers for the DJI data. The frontiers have been computed at 20 discretized values of  $\gamma \in [0.005, 0.3]$ . The dashed curve plots the expected return and standard deviation of efficient portfolios under the nominal distribution  $\eta$  (uniform). The solid curve plots the expected return and standard deviation of efficient portfolios (each value of  $\gamma$  results in a different worst-case distribution) for percent ambiguity level % u.l. = 0.01. The worst-case curve was obtained in about 8.5 minutes.

## GIUSEPPE C. CALAFIORE

7. Conclusions. Following the recent stream in the literature dealing with statistical model uncertainty (ambiguity) in asset allocation problems, this work explores the case where the ambiguity level in a discrete return distribution is measured according to the Kullback–Leibler divergence. A methodology is proposed for assessing and optimizing the worst-case risk of a portfolio under this type of uncertainty. Two standard risk measures (expected return composed with variance or absolute deviation) are examined, and polynomial-time algorithms are developed for solving efficiently the ensuing problems.

The proposed algorithms are based on interior-point barrier methods for convex optimization, in conjunction with a cutting plane technique. Although it is known in general that cutting plane methods have quite a high iterations-per-digit ratio, they do provide polynomial-time guaranteed convergence to the global optimum to any given practical accuracy. Moreover, in the specific application area discussed in this paper, they permit one to decouple a portfolio analysis phase (step 2 in Algorithm 2) from a mix optimization one (step 1). The numerical experiments show that worstcase optimal portfolios can be computed in reasonable time on a modern computer, and suggest that the proposed methods may be potentially useful in practice for controlling analytically the effects of model uncertainty on financial risk.

Appendix A. Proof of Proposition 3.1. Consider the function

$$\vartheta(\pi) = \sum_{k=1}^{T} \pi_k \log \pi_k - \sum_{k=1}^{T} \pi_k \log \eta_k,$$

which is convex and three times differentiable over the domain  $\pi > 0$ , and let  $D^2 \vartheta(\pi)[h,h], D^3 \vartheta(\pi)[h,h,h]$  denote, respectively, the second and third differentials of  $\vartheta(\pi)$ , taken at  $\pi$  along the direction  $h \in \mathbb{R}^T$ . That is,

$$D^{2}\vartheta(\pi)[h,h] = \left.\frac{\mathrm{d}^{2}}{\mathrm{d}t^{2}}\vartheta(\pi+th)\right|_{t=0} = \sum_{k=1}^{T}\frac{h_{k}^{2}}{\pi_{k}},$$
$$D^{3}\vartheta(\pi)[h,h,h] = \left.\frac{\mathrm{d}^{3}}{\mathrm{d}t^{3}}\vartheta(\pi+th)\right|_{t=0} = \sum_{k=1}^{T}\frac{-h_{k}^{3}}{\pi_{k}^{2}}$$

We have that

$$\begin{split} |D^{3}\vartheta(\pi)[h,h,h]| &\leq \sum_{k=1}^{T} \frac{|h_{k}|^{3}}{\pi_{k}^{2}} = \sum_{k=1}^{T} \frac{h_{k}^{2}}{\pi_{k}} \cdot \frac{|h_{k}|}{\pi_{k}} \\ &\leq \sqrt{\sum_{k=1}^{T} \left(\frac{h_{k}^{2}}{\pi_{k}}\right)^{2}} \cdot \sqrt{\sum_{k=1}^{T} \left(\frac{|h_{k}|}{\pi_{k}}\right)^{2}} \\ &\leq \sum_{k=1}^{T} \frac{h_{k}^{2}}{\pi_{k}} \cdot \sqrt{\sum_{k=1}^{T} \left(\frac{|h_{k}|}{\pi_{k}}\right)^{2}} \\ &= D^{2}\vartheta(\pi)[h,h] \cdot \sqrt{\sum_{k=1}^{T} \frac{h_{k}^{2}}{\pi_{k}^{2}}}, \end{split}$$

where the first inequality in the chain is the triangle inequality, the second is Hölder's inequality, and the third follows from the inequality  $||x||_2 \leq ||x||_1$  between the  $\ell_2$  and

 $\ell_1$  norms. Summarizing, the relation

$$\left| D^{3}\vartheta(\pi)[h,h,h] \right| \leq D^{2}\vartheta(\pi)[h,h] \cdot \sqrt{\sum_{k=1}^{T} \frac{h_{k}^{2}}{\pi_{k}^{2}}}$$

holds for all  $\pi > 0$  and  $h \in \mathbb{R}^T$ . We now apply a known result on logarithmic barriers: Due to the previous inequality, function  $\vartheta(\pi)$  satisfies the hypotheses of Lemma 2 of [9], from which it follows that the function

$$-\log(d - \vartheta(\pi)) - \sum_{k=1}^{T} \log \pi_k$$

(which coincides with function  $\phi(\pi)$  defined in (3.7)) is a self-concordant barrier for the domain  $\{(\pi, d) : \pi > 0, \vartheta(\pi) < d\}$ , thus concluding the proof.  $\Box$ 

Appendix B. Nonconcavity of the absolute deviation function. Assume without loss of generality that the data w(k) are arranged in increasing order. The function

$$\rho_1(\pi) = \sum_k \pi_k |w(k) - \mathbf{w}^\top \pi|$$

is not only nonconcave over the simplex, but it is nonconcave also on the restricted domains

$$R_i \doteq \{ \pi \in \Pi : \mathbf{w}^\top \pi \in (w(i), w(i+1)) \}.$$

For  $\pi \in R_i$ , we have that  $w(k) - \mathbf{w}^\top \pi > 0$  for  $k \in K_+ \doteq \{i + 1, \dots, T\}$  and  $w(k) - \mathbf{w}^\top \pi < 0$  for  $k \in K_- \doteq \{1, \dots, i\}$ . Hence, for  $\pi \in R_i$ , we may write

$$\rho_{1}(\pi) = \sum_{k \in K_{+}} \pi_{k}(w(k) - \mathbf{w}^{\top}\pi) - \sum_{k \in K_{-}} \pi_{k}(w(k) - \mathbf{w}^{\top}\pi)$$
$$= \sum_{k \in K_{+}} \pi_{k}w(k) - \sum_{k \in K_{-}} \pi_{k}w(k) - \mathbf{w}^{\top}\pi \left(\sum_{k \in K_{+}} \pi_{k} - \sum_{k \in K_{-}} \pi_{k}\right).$$

From the latter expression, since  $\sum_{k \in K_-} \pi_k = 1 - \sum_{k \in K_+} \pi_k$  and  $\sum_{k \in K_-} \pi_k w(k) = \mathbf{w}^\top \pi - \sum_{k \in K_+} \pi_k w(k)$ , we further have

$$\rho_1(\pi) = 2 \sum_{k \in K_+} \pi_k(w(k) - \mathbf{w}^\top \pi).$$

Now let  $\pi^a, \pi^b \in R_i$  and consider

$$\begin{split} \rho_1 \left( \frac{1}{2} (\pi^a + \pi^b) \right) &= 2 \sum_{k \in K_+} \frac{\pi_k^a + \pi_k^b}{2} \left( w(k) - \mathbf{w}^\top \frac{\pi^a + \pi^b}{2} \right) \\ &= \frac{1}{2} \sum_{k \in K_+} \pi_k^a (w(k) - \mathbf{w}^\top \pi^a) + \frac{1}{2} \sum_{k \in K_+} \pi_k^b (w(k) - \mathbf{w}^\top \pi^b) \\ &\quad + \frac{1}{2} \sum_{k \in K_+} \pi_k^a (w(k) - \mathbf{w}^\top \pi^b) + \frac{1}{2} \sum_{k \in K_+} \pi_k^b (w(k) - \mathbf{w}^\top \pi^a) \end{split}$$

$$\begin{split} &= \frac{1}{2} \sum_{k \in K_{+}} \pi_{k}^{a}(w(k) - \mathbf{w}^{\top} \pi^{a}) + \frac{1}{2} \sum_{k \in K_{+}} \pi_{k}^{b}(w(k) - \mathbf{w}^{\top} \pi^{b}) \\ &+ \frac{1}{2} \sum_{k \in K_{+}} \pi_{k}^{a}(w(k) - \mathbf{w}^{\top} \pi^{a}) + \frac{1}{2} (\mathbf{w}^{\top} \pi^{a} - \mathbf{w}^{\top} \pi^{b}) \sum_{k \in K_{+}} \pi_{k}^{a} \\ &+ \frac{1}{2} \sum_{k \in K_{+}} \pi_{k}^{b}(w(k) - \mathbf{w}^{\top} \pi^{b}) - \frac{1}{2} (\mathbf{w}^{\top} \pi^{a} - \mathbf{w}^{\top} \pi^{b}) \sum_{k \in K_{+}} \pi_{k}^{b} \\ &= \frac{1}{2} \rho_{1}(\pi^{a}) + \frac{1}{2} \rho_{1}(\pi^{b}) + \frac{1}{2} (\mathbf{w}^{\top} \pi^{a} - \mathbf{w}^{\top} \pi^{b}) \sum_{k \in K_{+}} (\pi_{k}^{a} - \pi_{k}^{b}). \end{split}$$

The last term in the sum is sign-indefinite (in particular, it does not hold in general that  $\mathbf{w}^{\top}\pi^{a} - \mathbf{w}^{\top}\pi^{b} > 0$  implies  $\sum_{k \in K_{+}} \pi_{k}^{a} - \sum_{k \in K_{+}} \pi_{k}^{b} > 0$ ; hence  $\rho_{1}$  is not concave over the domain  $R_{i}$ . To see this latter point, take, for instance,  $\pi^{a}$  all zero, except for  $\pi_{i}^{a} = \pi_{i+1}^{a} = 1/2$ , and  $\pi^{b}$  all zero, except for  $\pi_{i-1}^{b} = 1/2 - \epsilon$ ,  $\pi_{i+1}^{b} = 1/2 + \epsilon$ , with

$$\max\left(0, \, \epsilon_{\max} - \frac{1}{2} \frac{w(i+1) - w(i)}{w(i+1) - w(i-1)}\right) < \epsilon < \epsilon_{\max}; \quad \epsilon_{\max} \doteq \frac{1}{2} \frac{w(i) - w(i-1)}{w(i+1) - w(i-1)}.$$

Then one can check by direct inspection that  $\pi^a, \pi^b \in R_i, \mathbf{w}^{\top} \pi^a - \mathbf{w}^{\top} \pi^b > 0$ , but  $\sum_{k \in K_+} \pi^a_k - \sum_{k \in K_+} \pi^b_k < 0.$ 

**Appendix C. Computing the range for the expected return.** We discuss here an efficient technique for determining the extreme values (5.4) of the expected return of a given portfolio. Consider the problem

(C.1) 
$$\mu_{\min} \doteq \min_{\pi} \quad \mathbf{w}^{\top} \pi$$
subject to  $\pi \in \mathcal{K}(\eta, d)$ .

The Lagrangian for this problem is

$$\mathcal{L}(\pi,\lambda_{(\pi)},\lambda_{(\mathrm{kl})},\nu_{(1)}) = \mathbf{w}^{\top}\pi - \lambda_{(\pi)}^{\top}\pi + \lambda_{(\mathrm{kl})}(\mathrm{KL}(\pi,\eta) - d) + \nu_{(1)}(\mathbf{1}^{\top}\pi - 1),$$

where  $\lambda_{(\pi)}, \lambda_{(kl)}, \nu_{(1)}$  are Lagrange multipliers (dual variables). We assume henceforth that  $\lambda_{(kl)}$  is strictly positive.<sup>3</sup> The Lagrange dual function is

$$g(\lambda_{(\pi)}, \lambda_{(\mathrm{kl})}, \nu_{(1)}) = \inf_{\pi} \mathcal{L}(\pi, \lambda_{(\pi)}, \lambda_{(\mathrm{kl})}, \nu_{(1)})$$
  
=  $-\lambda_{(\mathrm{kl})}d - \nu_{(1)} + \inf_{\pi} \left( \mathbf{q}^{\top}\pi + \lambda_{(\mathrm{kl})} \operatorname{KL}(\pi, \eta) \right)$ 

where

$$\mathbf{q} \doteq \mathbf{w} - \lambda_{(\pi)} + \nu_{(1)} \mathbf{1}.$$

Observe now that

$$\nabla_{\pi} \left( \mathbf{q}^{\top} \pi + \lambda_{(\mathrm{kl})} \operatorname{KL}(\pi, \eta) \right) = \mathbf{q} + \lambda_{(\mathrm{kl})} (\mathbf{1} + \log \pi / \eta) = 0$$

for

$$\pi_k = \eta_k \mathrm{e}^{-\mathbf{q}_k/\lambda_{(\mathrm{kl})}-1}$$

<sup>&</sup>lt;sup>3</sup>When the KL constraint is inactive at optimum, the optimal value of the dual variable  $\lambda_{(kl)}$  is zero, due to the complementary slackness condition. However, in this case the solution to problem (C.1) is trivially given by min<sub>k</sub> w(k).

which yields

$$\inf_{\pi} \left( \mathbf{q}^{\top} \pi + \lambda_{(\mathrm{kl})} \operatorname{KL}(\pi, \eta) \right) = -\lambda_{(\mathrm{kl})} \sum_{k} \eta_{k} \mathrm{e}^{-\mathbf{q}_{k}/\lambda_{(\mathrm{kl})}-1},$$

and hence the dual function

$$g(\lambda_{(\pi)}, \lambda_{(\mathrm{kl})}, \nu_{(1)}) = -\lambda_{(\mathrm{kl})}d - \nu_{(1)} - \lambda_{(\mathrm{kl})}\sum_{k} \eta_{k} \mathrm{e}^{-\mathbf{q}_{k}/\lambda_{(\mathrm{kl})}-1}.$$

The dual to problem (C.1) is therefore

$$\begin{split} \mu_{\min} &\doteq \max -\lambda_{(\mathrm{kl})} d - \nu_{(1)} - \lambda_{(\mathrm{kl})} \sum_{k} \eta_{k} \mathrm{e}^{-\mathbf{q}_{k}/\lambda_{(\mathrm{kl})} - 1} \\ \mathrm{subject \ to} \ \lambda_{(\mathrm{kl})} > 0, \\ \lambda_{(\pi)} &\geq 0 \end{split}$$

or, equivalently,

(C.2) 
$$\mu_{\min} \doteq -\min \lambda_{(kl)} d + \nu_{(1)} + \lambda_{(kl)} \sum_{k} \eta_{k} e^{-\mathbf{q}_{k}/\lambda_{(kl)}-1}$$
subject to  $\lambda_{(kl)} > 0$ ,  
 $\lambda_{(\pi)} \ge 0$ .

Notice that, for fixed  $\lambda_{(kl)}, \lambda_{(\pi)}$ , at the optimum the derivative of (C.2) with respect to  $\nu_{(1)}$  must be zero, i.e.,

$$1 + \lambda_{(\mathrm{kl})} \sum_{k} \eta_{k} \mathrm{e}^{-\mathbf{q}_{k}/\lambda_{(\mathrm{kl})}-1} \frac{-1}{\lambda_{(\mathrm{kl})}} \frac{\partial \mathbf{q}_{k}}{\partial \nu_{(1)}} = 1 - \sum_{k} \eta_{k} \mathrm{e}^{-\mathbf{q}_{k}/\lambda_{(\mathrm{kl})}-1} = 0,$$

from which we obtain

$$e^{-\nu_{(1)}/\lambda_{(kl)}} \sum_{k} \eta_{k} e^{(\lambda_{(\pi),k} - w(k))/\lambda_{(kl)} - 1} = 1$$

$$\downarrow$$

$$\nu_{(1)} = \lambda_{(kl)} \left( \log \sum_{k} \eta_{k} e^{(\lambda_{(\pi),k} - w(k))/\lambda_{(kl)}} - 1 \right).$$

Substituting this latter expression into (C.2), we write the dual in the following reduced form:

$$\begin{split} &-\mu_{\min} \doteq \min \lambda_{(\mathrm{kl})} d + \lambda_{(\mathrm{kl})} \log \sum_{k} \eta_{k} \mathrm{e}^{(\lambda_{(\pi),\mathrm{k}} - w(k))/\lambda_{(\mathrm{kl})}} \\ &\text{subject to } \lambda_{(\mathrm{kl})} > 0, \\ &\lambda_{(\pi)} \ge 0. \end{split}$$

Observe further that, for any given  $\lambda_{(kl)} > 0$ , the optimal choice for  $\lambda_{(\pi)}$  is zero, which finally yields the dual in the form of a univariate convex problem:

$$\begin{split} -\mu_{\min} &\doteq \min f_l(\lambda_{(\mathrm{kl})}) \doteq \lambda_{(\mathrm{kl})} d + \lambda_{(\mathrm{kl})} \log \sum_k \eta_k \mathrm{e}^{-\frac{w(k)}{\lambda_{(\mathrm{kl})}}} \\ &\text{subject to } \lambda_{(\mathrm{kl})} > 0. \end{split}$$

An identical reasoning can be applied for computing the upper limit  $\mu_{max}$  of the mean range. In this case, we obtain

(C.3) 
$$\mu_{\max} \doteq \min f_u(\lambda_{(kl)}) \doteq \lambda_{(kl)} d + \lambda_{(kl)} \log \sum_k \eta_k e^{\frac{w(k)}{\lambda_{(kl)}}}$$
subject to  $\lambda_{(kl)} > 0$ .

Both problems can be readily solved via a bisection scheme (such as the one sketched in Algorithm 3 below), given the gradient of the objective function

$$g_l(\lambda) \doteq \frac{\partial f_l}{\partial \lambda} = \frac{1}{\lambda} \left( f_l(\lambda) + \frac{\sum_k w(k) \eta_k e^{-w(k)/\lambda}}{\sum_k \eta_k e^{-w(k)/\lambda}} \right)$$

for the lower mean limit, and

$$g_u(\lambda) \doteq \frac{\partial f_u}{\partial \lambda} = \frac{1}{\lambda} \left( f_u(\lambda) - \frac{\sum_k w(k) \eta_k e^{-w(k)/\lambda}}{\sum_k \eta_k e^{w(k)/\lambda}} \right)$$

for the upper limit.

Algorithm 3 (bisection).

Given initial  $\lambda_l = 0$ ,  $\lambda_r = 1$  and tolerance  $\epsilon > 0$ 

- 1. while  $g(\lambda_r) < 0$ , let  $\lambda_r = 2\lambda_r$ , end while;
- 2. while  $\lambda_r \lambda_l > \epsilon$ 
  - 2.a let  $\lambda = \frac{1}{2}(\lambda_r + \lambda_l)$
- 2.b if  $g(\lambda) > 0$ , let  $\lambda_r = \lambda$ ; else let  $\lambda_l = \lambda$ , end if 3. end while
- 4. return  $\lambda$ .

A perhaps interesting observation is that the function appearing in (C.3)

$$\lambda_{\rm (kl)} \log \sum_k \eta_k {\rm e}^{\frac{w(k)}{\lambda_{\rm (kl)}}}$$

tends to  $\max_k w(k)$  as  $\lambda_{(kl)} \to 0$ . Indeed, for  $\lambda_{(kl)} > 0$ , this function is a "uniform smooth approximation" of the max function, using the terminology introduced in [27].

Appendix D. Proof of Proposition 5.1. Consider problem (5.5) without the KL constraint, and with the values w(k) arranged in increasing order (see Figure D.1):

(D.1) 
$$\begin{aligned} \varphi(\mu) \doteq \max_{\pi} \sum_{k=1}^{T} \pi_{k} |w(k) - \mu| \\ \text{subject to} \quad \mathbf{w}^{\top} \pi = \mu, \\ \mathbf{1}^{\top} \pi = 1, \\ \pi \ge 0. \end{aligned}$$

It is immediate by inspection that (5.7) is a feasible solution for (D.1). We next show that this solution is actually optimal.

Let  $d(k) \doteq |w(k) - \mu|$ . The Lagrangian for problem (D.1) is

$$\mathcal{L}(\pi,\lambda,\nu) = -\sum_{k=1}^{T} \pi_k d(k) - \sum_{k=1}^{T} \lambda_k \pi_k + \nu_1 \left(\sum_{k=1}^{T} \pi_k w(k) - \mu\right) + \nu_2 \left(\sum_{k=1}^{T} \pi_k - 1\right).$$



FIG. D.1. Data arranged in increasing order. The d(k) are the distances from the mean:  $d(k) \doteq |w(k) - \mu|$ .

For a feasible  $\pi$  to be optimal, the Karush–Kuhn–Tucker (KKT) condition

(D.2) 
$$\nabla_{\pi} \mathcal{L}(\pi, \lambda, \nu) = -\begin{bmatrix} d(1) \\ \vdots \\ d(T) \end{bmatrix} - \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_T \end{bmatrix} + \nu_1 \begin{bmatrix} w(1) \\ \vdots \\ w(T) \end{bmatrix} + \nu_2 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = 0$$

must hold for some  $\nu_1, \nu_2$  and  $\lambda_k \ge 0$ , along with the complementary slackness condition

(D.3) 
$$\lambda_k \pi_k = 0, \quad k = 1, \dots, T.$$

We now show that for the solution in (5.7) we can find  $\nu_1, \nu_2$  and  $\lambda_k \ge 0$  such that (D.2), (D.3) hold. From (D.2) we have

(D.4) 
$$\lambda_k = \nu_1 w(k) + \nu_2 - d(k), \quad k = 1, \dots, T.$$

Since  $\pi_1 > 0, \pi_T > 0$ , (D.3) implies  $\lambda_1 = \lambda_T = 0$ , and hence

$$\nu_1 w(1) + \nu_2 - d(1) = 0,$$
  
$$\nu_1 w(T) + \nu_2 - d(T) = 0,$$

from which we obtain

(D.5) 
$$\nu_1 = \frac{d(T) - d(1)}{w(T) - w(1)}, \quad \nu_2 = \frac{d(1)[w(T) - w(1)] - w(1)[d(T) - d(1)]}{w(T) - w(1)}.$$

Substituting into (D.4), for  $k = 2, \ldots, T - 1$ , we get

(D.6) 
$$\lambda_k = \frac{[d(T) - d(1)][w(k) - w(1)] + [w(T) - w(1)][d(1) - d(k)]}{w(T) - w(1)}$$

We now verify that these  $\lambda_k$  are nonnegative. Define  $K_- \doteq \{k : w(k) \leq \mu\}$  and  $K_+ \doteq \{k : w(k) > \mu\}$ . For  $k \in K_-$  we have that (see Figure D.1)

$$w(k) - w(1) = d(1) - d(k); \quad w(T) - w(1) = d(1) + d(T),$$

which, once substituted in (D.6), give

$$\lambda_k = \frac{[d(T) - d(1)][d(1) - d(k)] + [d(1) + d(T)][d(1) - d(k)]}{w(T) - w(1)}$$
$$= \frac{2d(T)[d(1) - d(k)]}{w(T) - w(1)} \ge 0, \quad k \in K_-.$$

Similarly, for  $k \in K_+$  we have that

$$w(k) - w(1) = d(1) + d(k); \quad w(T) - w(1) = d(1) + d(T),$$

which, once substituted in (D.6), give

$$\lambda_k = \frac{[d(T) - d(1)][d(1) + d(k)] + [d(1) + d(T)][d(1) - d(k)]}{w(T) - w(1)}$$
$$= \frac{2d(1)[d(T) - d(k)]}{w(T) - w(1)} \ge 0, \quad k \in K_+.$$

Overall, we have that the primal feasible solution (5.7), together with the dual feasible variables (D.5), (D.6), satisfies the KKT conditions (D.2), (D.3), and hence (5.7) is optimal for problem (D.1). If this solution satisfies the constraint  $KL(\pi, \eta) \leq d$ , then clearly the solution is also optimal for the original problem (5.5).

Substituting (5.7) into the objective (D.1), we easily obtain the optimal value function in (5.8), which concludes the proof.  $\Box$ 

Acknowledgments. I sincerely thank Fabrizio Dabbene, Constantino Lagoa, Arkadi Nemirovski, and Andrzej Ruszczyński for their useful discussions and correspondence. A special thanks goes to the Associate Editor and to the anonymous reviewers for their precious comments and suggestions.

## REFERENCES

- S. M. ALI AND S. D. SILVEY, A general class of coefficients of divergence of one distribution from another, J. Royal Statist. Soc., 28 (1966), pp. 131–142.
- [2] P. ARTZNER, F. DELBAEN, J.-M. EBER, AND D. HEATH, Coherent measures of risk, Math. Finance, 9 (1999), pp. 203–228.
- [3] V. S. BAWA, S. J. BROWN, AND R. W. KLEIN, Estimation Risk and Optimal Portfolio Choice, North-Holland, Amsterdam, 1979.
- [4] A. BEN-TAL AND A. NEMIROVSKII, Robust convex optimization, Math. Oper. Res., 23 (1998), pp. 769–805.
- [5] S. BOYD AND L. VANDENBERGHE, Convex Optimization, Cambridge University Press, Cambridge, UK, 2004.
- [6] M. BROADIE, Computing efficient frontiers using estimated parameters, Ann. Oper. Res., 45 (1993), pp. 21–58.
- Z. CHEN AND L. EPSTEIN, Ambiguity, risk, and asset returns in continuous time, Econometrica, 70 (2002), pp. 1403–1443.
- [8] V. K. CHOPRA AND W. T. ZIEMBA, The effect of errors in means, variances and covariances on optimal portfolio choice, J. Portfolio Management, 19 (2) (1993), pp. 6–11.
- [9] D. DEN HERTOG, F. JARRE, C. ROOS, AND T. TERLAKY, A sufficient condition for selfconcordance, with application to some classes of structured convex programming problems, Math. Programming, 69 (1995), pp. 75–88.
- [10] R. M. DUDLEY, Distance of probability measures and random variables, Ann. Math. Statist., 39 (1968), pp. 1563–1572.
- J. DUPAČOVÁ, The minimax approach to stochastic program and illustrative application, Stochastics, 20 (1987), pp. 73–88.
- [12] J. DUPAČOVÁ, Stochastic Programming: Minimax Approach, in Encyclopedia of Optimization, Vol. 5, Kluwer Academic, Dordrecht, The Netherlands, 2001, pp. 327–330.
- [13] L. EL GHAOUI, M. OKS, AND F. OUSTRY, Worst-case value-at-risk and robust portfolio optimization: A conic programming approach, Oper. Res., 51 (2003), pp. 543-556.
- [14] L. EL GHAOUI, F. OUSTRY, AND H. LEBRET, Robust solutions to uncertain semidefinite programs, SIAM J. Optim., 9 (1998), pp. 33–52.
- [15] E. ERDOĞAN AND G. IYENGAR, Ambiguous chance constrained problems and robust optimization, Math. Program. Ser. B, 107 (2006), pp. 37–61.
- [16] J.-L. GOFFIN, Z.-Q. LUO, AND Y. YE, Complexity analysis of an interior cutting plane method for convex feasibility problems, SIAM J. Optim., 6 (1996), pp. 638–652.

- [17] J.-L. GOFFIN AND J.-P. VIAL, Convex nondifferentiable optimization: A survey focused on the analytic center cutting plane method, Optim. Methods Softw., 17 (2002), pp. 805–867.
- [18] D. GOLDFARB AND G. IYENGAR, Robust portfolio selection problems, Math. Oper. Res., 28 (2003), pp. 1–38.
- [19] P. HALL, On Kullback-Leibler loss and density estimation, Ann. Statist., 15 (1987), pp. 1491– 1519.
- [20] M. HENRY, Generalized Entropy Measure of Ambiguity and Its Estimation, preprint, Columbia University, New York, 2005.
- [21] H. KONNO AND H. YAMAZAKI, Mean absolute deviation portfolio optimization model and its application to Tokio stock market, Management Sci., 37 (1991), pp. 519–531.
- [22] S. KULLBACK, Information Theory and Statistics, Wiley, New York, 1959.
- [23] M. S. LOBO, L. VANDENBERGHE, S. BOYD, AND H. LEBRET, Applications of second-order cone programming, Linear Algebra Appl., 284 (1998), pp. 193–228.
- [24] P. J. MAENHOUT, Robust portfolio rules and asset pricing, Rev. Financial Stud., 17 (2004), pp. 951–983.
- [25] H. M. MARKOWITZ, Portfolio selection, J. Finance, 7 (1952), pp. 77–91.
- [26] R. O. MICHAUD, The Markowitz optimization enigma: Is optimized optimal?, Financial Analysts J., 45 (1989), pp. 31–42.
- [27] YU. NESTEROV, Smooth Minimization of Nonsmooth Functions, CORE Tech. Report, 2003.
- [28] YU. NESTEROV AND A. S. NEMIROVSKII, Interior-Point Polynomial Algorithms in Convex Programming, SIAM, Philadelphia, 1994.
- [29] O. PÉTON, The Homogeneous Analytic Center Cutting Plane Method, Ph.D. Thesis, Université de Genève, 2002.
- [30] A. RUSZCZYŃSKI AND A. SHAPIRO, Optimization of risk measures, in Probabilistic and Randomized Methods for Design under Uncertainty, G. Calafiore and F. Dabbene, eds., Springer-Verlag, London, 2006, pp. 119–157.
- [31] A. RUSZCZYNSKI AND R. J. VANDERBEI, Frontiers of stochastically nondominated portfolios, Econometrica, 71 (2003), pp. 1287–1297.
- [32] A. SHAPIRO AND A. J. KLEYWEGT, Minimax analysis of stochastic problems, Optim. Methods Softw., 17 (2002), pp. 523–542.
- [33] M. J. TODD, Semidefinite optimization, Acta Numer., 10 (2001), pp. 515–560.
- [34] R. H. TÜTÜNCÜ AND M. KOENIG, Robust asset allocation, Ann. Oper. Res., 132 (2004), pp. 157–187.
- [35] L. VANDENBERGHE AND S. BOYD, Semidefinite programming, SIAM Rev., 38 (1996), pp. 49–95.
- [36] J. ŽÁČKOVÁ, On minimax solutions of stochastic linear programming problems, Čas. Pěst. Mat., 91 (1966), pp. 423–430.
- [37] Z. WANG, A shrinkage approach to model uncertainty and asset allocation, Rev. Financial Stud., 18 (2005), pp. 673–705.