Stochastic Efficiency and Inefficiency in Portfolio Optimization with Incomplete Information: A Set-Valued Probability Approach

D. La Torre[1], F. Mendivil[2]

July 8, 2020

[1] PRISM Research Center, SKEMA Business School and Université Côte d'Azur, Sophia Antipolis Campus, France
[2] Acadia University, Department of Mathematics and Statistics, Wolfville, Nova Scotia, Canada.

Abstract

In this paper we extend the notion of stochastic efficiency and inefficiency in portfolio optimization to the case of incomplete information by means of set-valued probabilities. The notion of set-valued probability models the concept of incomplete information about the underlying probability space and the probability associated with each scenario. Unlike other approaches in literature, our notion of inefficiency is introduced by means of the Monge-Kantorovich metric. We provide some numerical examples to illustrate this approach.

Keywords: Portfolio Optimization, Set-valued Optimization, Stochastic Efficiency, Stochastic Inefficiency.

1 Introduction

In Portfolio Optimization the notion of optimality is based on the notion of stochastic dominance and on a knowledge of the underlying probability space. However, often the probabilities associated with each scenario are not completely known. This leads to analyzing portfolio problems with incomplete information [5, 6, 12, 22, 23, 32, 33]. The notion of set-valued probability seems to be the right tool to describe this lack of information. In this setting, each event is associated with a compact and convex set that models the uncertainty; as a result, the comparison of uncertainties can be based on a partial order and many such partial orders can be used. The consequence is a rich framework for modeling uncertainty.

In this paper we focus on the problem of evaluating whether a given portfolio is stochastically efficient in the presence of a lack of information. More

specifically, we consider the case where the Decision Maker (DM) knows the set of all possible scenarios or the underlying space of events but he/she does not exactly know the probability distribution of each event and so we describe this uncertainty with a set-valued probability.

We extend the classical notion of inefficiency measure by introducing a point-to-set distance between a given set-valued probability and the nearest probabilities that can make a given portfolio to be stochastically efficient for some admissible utility function. To introduce this distance we rely on an extended notion of the Monge-Kantorovich distance between set-valued probabilities introduced in [21].

The paper is organized as follows. Section 2 recalls some basic definitions and constructions in the theory of convex sets, Section 3 presents background in set-valued analysis, and Section 4 focuses on the definition of set-valued probabilities and the Monge-Kantorovich distance between set-valued probabilities. Section 5 provides a brief comparison of set-valued probability with other generalizations of classical probability, focusing specifically on a comparison with *imprecise probability*. Section 6 presents the notion of efficiency and inefficiency; our notion of inefficiency is introduced by means of the Monge-Kantorovich distance between probability measures. Section 7 extends the notion of stochastic efficiency and inefficiency to the case of set-valued probabilities. Finally Section 8 presents a numerical example and Section 9 concludes.

2 Preliminaries on Sets and Probabilities

In this section we present some basic facts related to sets and set-valued functions. More details can be found in [2, 3]. In the sequel we denote by $\mathcal{K}$ the collection of all nonempty compact and convex subsets of $\mathbb{R}^d$. Addition of sets and scalar multiplication ($\lambda \in \mathbb{R}$) for $\mathcal{K}$ are defined by

$$A + B := \{a + b : a \in A, b \in B\} \text{ and } \lambda A = \{\lambda a : a \in A\}.$$

For $A \in \mathcal{K}$, we say that A *is nonnegative* ($A \geq 0$) if $0 \in A$. Given $A \in \mathcal{K}$ the *support function* $\mathrm{spt}(\cdot, A) : \mathbb{R}^d \to \mathbb{R}$ is defined by

$$\mathrm{spt}(p, A) = \sup\{p \cdot a : a \in A\}.$$

The support function completely defines A since

$$A = \bigcap_{\|p\|=1} \{x : x \cdot p \leq \mathrm{spt}(p, A)\}. \tag{2.1}$$

Furthermore, $A \subseteq B$ if and only if $\mathrm{spt}(p, A) \leq \mathrm{spt}(p, B)$ for all $p \in S^1 = \{p : \|p\| = 1\}$. The function $\mathrm{spt}(,)$ also satisfies the following properties: For all $\lambda \geq 0$ and $A, B \in \mathcal{K}$,

$$\mathrm{spt}(p, \lambda A + B) = \lambda\, \mathrm{spt}(p, A) + \mathrm{spt}(p, B),\ \mathrm{spt}(p, -B) = \mathrm{spt}(-p, B)$$

but it is usually the case that $\operatorname{spt}(p, -A) \neq -\operatorname{spt}(p, A)$.

For any $A \in \mathcal{K}$, we can also define the *norm* of A using the support function as follows

$$\|A\| := \sup\{\|x\| : x \in A\} = \sup_{\|p\|=1} \operatorname{spt}(p, A).$$

This definition satisfies all of the classical property of a norm. There is a nice connection between the support function and the Hausdorff distance [7]: for $A, B \in \mathcal{K}$

$$d_H(A, B) = \sup_{\|p\|=1} |\operatorname{spt}(p, A) - \operatorname{spt}(p, B)|.$$

It is also the case that both addition and scalar multiplication on $\mathcal{K}$ are continuous in the Hausdorff distance.

A set $A \subset \mathbb{R}^d$ is *balanced* if $\lambda A \subseteq A$ for all $|\lambda| \leq 1$. For us a *unit ball* in $\mathbb{R}^d$ is any balanced $\mathbb{B} \in \mathcal{K}$ with $0 \in int(\mathbb{B})$. Any such unit ball defines a norm on $\mathbb{R}^d$ via the Minkowski functional

$$\|x\| = \sup\{\lambda \geq 0 : \lambda x \in \mathbb{B}\}.$$

Given a unit ball $\mathbb{B}$, the *dual sphere* is defined as

$$\mathbb{S}^* = \{y : \sup\{y \cdot x : x \in \mathbb{B}\} = 1\} \subset \mathbb{R}^d$$

and is also a nonempty compact set. Notice that since $\mathbb{B}$ is compact, for each $y \in \mathbb{S}^*$, there is some $x \in \mathbb{B}$ with $y \cdot x = 1$.

Given a set Ω and a σ-algebra $\mathcal{A}$ on Ω a *probability measure* on $(\Omega, \mathcal{A})$ with values in $[0, 1]$ is a function $\Phi : \mathcal{A} \to [0, 1]$ such that $\Phi(\emptyset) = 0$, $\Phi(\Omega) = 1$, and

$$\Phi\left(\bigcup_i A_i\right) = \sum_i \Phi(A_i) \tag{2.2}$$

for any sequence of disjoint sets $A_i \in \mathcal{A}$. Similarly one can define the notion of *vector-valued probability measure* on $(\Omega, \mathcal{A})$. This is a function $\Phi : \mathcal{A} \to [0, 1]^s$, where $s \in \mathbb{N}$, such that $\Phi(\emptyset) = (0, ..., 0) \in \mathbb{R}^s$, $\Phi(\Omega) = (1, ..., 1) \in \mathbb{R}^s$, and

$$\Phi\left(\bigcup_i A_i\right) = \sum_i \Phi(A_i) \tag{2.3}$$

for any sequence of disjoint sets $A_i \in \mathcal{A}$. This last property is meant to be satisfied componentwise. Note that with this definition a vector-valued probability measure is simply a vector of probability measures.

3 Set-valued Functions

A *set-valued function* or *multifunction* taking compact and convex values is a map from $\mathbb{R}^n$ to $\mathcal{K}$. For a given set-valued function $f : \mathbb{R}^n \to \mathcal{K}$ and measure μ,

we can define the integral of f with respect to μ as an element of $\mathcal{K}$ via support functions using the property (see [3])

$$\operatorname{spt}\left(q, \int_{\mathbb{R}^n} f(x)\ d\mu(x)\right) = \int_{\mathbb{R}^n} \operatorname{spt}(q, f(x))\ d\mu(x),$$

which defines the set as in (2.1). For more results on set-valued analysis see [3].

Given a compact subset Θ of $\mathbb{R}^n$ and a set-valued function $f : \Theta \subseteq \mathbb{R}^n \to \mathcal{K}$, consider the optimization problem

$$\min_{x \in \Theta} f(x). \tag{3.4}$$

We say that $x_0 \in \Theta$ is a *global minimizer for f over* Θ if for any $x \in \Theta$ we have $f(x_0) \subseteq f(x)$. Notice that we are using the natural ordering of sets given by inclusion.

Let us now recall that a set-valued function $f : \mathbb{R} \to \mathcal{K}$ is *increasing* if

$$f(x) \subseteq f(y) \tag{3.5}$$

for $x, y \in \mathbb{R}$, $x \leq y$; moreover, $f : \mathbb{R} \to \mathcal{K}$ is *concave* if

$$tf(x) + (1-t)f(y) \subseteq f(tx + (1-t)y) \tag{3.6}$$

for $x, y \in \mathbb{R}^n$, $t \in [0, 1]$.

Using the support function, we have that f is concave if and only if the function $\operatorname{spt}(p, f(x))$ is concave for all $\|p\| = 1$.

4 The Notion of Set-Valued Probability

We provide only basic definitions and those properties of multimeasures that we will need; for more information and proofs see [1, 2, 3, 14, 15, 16, 29, 31]. A *set-valued measure* or *multimeasure* on $(\Omega, \mathcal{A})$ with values in $\mathcal{K}$, where $\mathcal{A}$ is a σ-algebra on the set Ω, is a function $\Phi : \mathcal{A} \to \mathcal{K}$ such that $\Phi(\emptyset) = \{0\}$ and

$$\Phi\left(\bigcup_i A_i\right) = \sum_i \Phi(A_i) \tag{4.7}$$

for any sequence of disjoint sets $A_i \in \mathcal{A}$. Convergence of the infinite sum in (4.7) is given in the Hausdorff distance.

A multimeasure Φ is *nonnegative* if $\Phi(A) \geq 0$ (i.e., $0 \in \Phi(A)$) for all A. This condition implies monotonicity of the measure since if $A \subseteq B$ then $\Phi(A) = \{0\} + \Phi(A) \subseteq \Phi(B \setminus A) + \Phi(A) = \Phi(B)$. Thus nonnegative multimeasures a nice generalization of (nonnegative) scalar measures. The *total variation* of a multimeasure Φ is defined in the usual way as

$$|\Phi|(A) = \sup \sum_i \|\Phi(A_i)\|,$$

where the supremum is taken over all finite measurable partitions of $A \in \mathcal{A}$. The set-function $|\Phi|$ defined in this fashion is a (nonnegative and scalar) measure on Ω. If $|\Phi|(\Omega) < \infty$ then Φ is of *bounded variation*.

If Φ is a multimeasure and $p \in \mathbb{R}^d$ then the *scalarization* Φ^p defined by

$$\Phi^p(A) = \mathrm{spt}(p, \Phi(A)) \tag{4.8}$$

is a signed measure on Ω and is a measure if Φ is nonnegative.

Example 4.1 Given a vector-valued probability $\mu = (\mu_1, ..., \mu_s)$ there is a natural way to define an associated set-valued measure as

$$\Phi(A) = [-\mu_1(A), \mu_1(A)] \times ... \times [-\mu_s(A), \mu_s(A)].$$

It is easy to check that Φ satisfies all properties that characterize a set-valued measure. Notice that here we have $\Phi(\Omega) = [-1, 1]^s$ is the l^∞ unit ball. In this case for $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_s)$ with $\|\lambda\|_1 = 1$ we have

$$\Phi^\lambda(A) = \sum_i |\lambda_i| \mu_i(A)$$

is just a convex combination of the components μ_i of μ.

Thus the case of vector-valued probability can be easily analyzed using the associated set-valued probability (to be discussed below).

One simple way to construct a multimeasure is by integrating a *set-valued density function* F with respect to a measure μ:

$$\Phi(A) = \int_A F(x)\ d\mu(x). \tag{4.9}$$

This integral can be defined in several different ways (see [3]). If the set-valued function F is nonnegative (that is, $0 \in F(x)$ for all x), then the resulting multimeasure will also be nonnegative. In addition, if $0 \leq f(x) \leq g(x)$ are scalar functions and Φ is a positive multimeasure, then

$$\int f(x)\ d\Phi(x) \subseteq \int g(x)\ d\Phi(x),$$

the convexity of the values of Φ is crucial.

Definition 4.2 Let $\mathbb{B} \in \mathcal{K}$ be a unit ball. A $\mathbb{B}$ *set-valued probability* or *probability multimeasure (or* pmm) on $(\Omega, \mathcal{A})$ is a nonnegative multimeasure Φ with $\Phi(\Omega) = \mathbb{B}$.

A pmm Φ defines a parameterized family, Φ^p for $p \in \mathbb{S}^*$, of probability measures. However, in general Φ^p and Φ^q are related and the relationship can be quite complicated (the main constraint on this relationship is that $p \mapsto \Phi^p(A)$ is convex).

We can construct a pmm by integrating an appropriate density F against a finite measure μ, as in (4.9). For this to define a pmm we need F to satisfy some properties. The simplest conditions are to assume that $F(x) \in \mathcal{K}$ is balanced for each x, $\|F(x)\| \leq C$ for some C and all x, and

$$0 \in int \int_\Omega F(x)\ d\mu = int(\mathbb{B}).$$

Given a specific $\mathbb{B}$, it is difficult to find a density F which will give $\mathbb{B}$; it is better to use the integral of the density to define $\mathbb{B}$.

As usual, by a *random variable* on $(\Omega, \mathcal{A})$ we mean a Borel measurable function $X : \Omega \to \mathbb{R}$ and its *expectation* with respect to a pmm Φ is defined in the usual way as

$$\mathbb{E}_\Phi(X) = \int_\Omega X(\omega)\ d\Phi(\omega). \tag{4.10}$$

This integral can be constructed using support functions (that is, using the Φ^p) and each part of the decomposition $X = X^+ - X^-$ separately (since support functions work best with nonnegative scalars); see [16] for another approach. Since $0 \in \Phi(A)$ for each A, it is easy to see that $0 \in \mathbb{E}_\Phi(X)$ as well.

Example 4.3 Let $s \in 2\mathbb{N}$ and $q_i \in \mathbb{R}^d$ be given with $\|q_i\| = 1$ and $q_{2i} = -q_{2i+1}$. Then for any s probability measures μ_i, the function Φ defined by

$$\Phi(A) = \overline{\text{co}}(\{\mu_i(A)q_i\})$$

is a probability multimeasure with total mass the polytope in $\mathbb{R}^d$ with vertices q_i. (Here $\overline{\text{co}}(S)$ is the closed convex hull of $S \subset \mathbb{R}^d$.)

This example is a generalization of Example 4.1 but where the different directions are not independent and thus the component measures μ_i "interact" in the resulting pmm Φ.

Example 4.4 (Finite Ω) When $\Omega = \{\omega_1, \omega_2, \ldots, \omega_s\}$ is a finite set, a pmm Φ on Ω is defined by a collection of set-valued probabilities $\{P_1, P_2, \ldots, P_s\} \subset \mathcal{K}$ such that $\sum_i P_i = \mathbb{B}$ and $0 \in \bigcap_i P_i$. For a random variable X, we have $\mathbb{E}_\Phi(X) = \sum_i X(\omega_i)P_i$.

Unlike in the standard (scalar-valued probability) case, we generally cannot use the P_i to rank the events ω_i from "most probable" (greatest P_i) to "least probable" (least P_i). This is the strength of using set-valued probabilities as it allows us to express more complicated uncertainty relationships among the elementary events.

There are extensions of the strong law of Large Numbers, the Glivenko-Cantelli, and the Central Limit Theorems. More details on these can be found in [23]. We mention that the notion of an i.i.d. sample is fundamentally different in the set-valued case since such a sample must necessarily contain some set-valued information if there is to be any chance of recovering the expectation, which is a set in this context. However, the standard limiting results (just mentioned) allow one to use familiar tools of statistical estimation. The resulting sequence of set-valued estimates converge in the Hausdorff distance to the true value.

5 Imprecise vs Set-Valued Probability

The notion of imprecise probability has been widely investigated in the literature as it represents a quite natural way to extend the traditional notion of probability. While in the tradition approach associated with each event there is a number in $[0, 1]$, in the theory of imprecise improbability this is replaced by a subset $[p_l, p_u]$, where p_l and p_u are the so-called lower and upper probabilities. The notion of imprecise probability aims at quantifying, at the same time, the aleatory and the uncertainty to provide precise values of probability measures.

Imprecise probability has been used in several contexts and different applications have been proposed as well. Some classical examples in this theory are the Dempster-Shafer evidence theory ([9, 30]), coherent lower prevision theory ([35]), probability bound analysis ([13]), F-probability ([37]), and possibility theory ([11]).

The research literature also contains other approaches that involve extensions to set-valued objects. Among these are the notions of fuzzy randomness ([4]), random sets ([26]), clouds ([27]), and the notion of imprecise probability based on the generalized interval been presented in [36].

Set-valued measures were first introduced for the needs of mathematical economics in [34] where it was used to study equilibria in exchange economies in which coalitions correspond to measurable sets and are the primary economic units (see also [8]). Moreover, the study of set-valued measures has been developed extensively because of its applications in other fields such as optimization and optimal control.

Imprecise probability can be modeled using set-valued probability. In fact, given a sample space Ω and a sigma-algebra $\mathcal{F}$, for any set $A \in \mathcal{F}$ the imprecise probability associated with A is a generalized interval $[p_l(A), p_u(A)] \subset [0, 1]$ where p_l and p_U are classical probabilities (and we don't insist that $p_l(A) \leq p_u(A)$). As such this structure is encoded in a pair of classical probabilities (which can also be thought of as a vector-valued measure $\mathbf{p} : \mathcal{F} \to [0, 1]^2$).

The notion of set-valued probability cannot be seen immediately as a generalization of the notion of imprecise probability, as the notion of set-valued probability is implicitly connected with a given notion of ordering (and thus also of positivity). In this paper we assume that the ordering is based on set-inclusion and, therefore, the notion of positivity is based on the inclusion of the zero element (this is forced since $\emptyset \subseteq A$ and thus $\{0\} = \Phi(\emptyset) \subseteq \Phi(A)$). However, if we define the set-valued probability $\Phi : \mathcal{F} \to [0, 1]$ as $\Phi(A) = [-p_l(A), p_u(A)]$ this definition satisfies all the properties that characterize a set-valued probability. In particular, one can easily prove that:

- $0 \in \Phi(A)$, for any $A \in \mathcal{F}$,
- if $A \subseteq B$ then $\Phi(A) \subseteq \Phi(B)$,
- $\Phi\left(\bigcup_{i=1}^{+\infty}\right) = \sum_{i=1}^{+\infty} \Phi\left(A_i\right)$ for any sequence of disjoint sets A_i, $i \in \mathbb{N}$.

The mapping $[p_l(A), p_u(A)] \mapsto [-p_l(A), p_u(A)]$ provides the correspondence between this particular set-valued probability and the imprecise probability given in the generalized interval form (mentioned above).

6 Stochastic Efficiency and Inefficiency

Let $(\Omega, \mathcal{F}, p)$ be a probability space, where Ω is the space of all events, $\mathcal{F}$ is the sigma-algebra of all measurable events, and p is a probability defined on $\mathcal{F}$. We use $\mathcal{P}$ to denote the set of all probability measures defined on $(\Omega, \mathcal{F})$. Let $X_j : \Omega \to \mathbb{R}_+$ be random variables, $j = 1...N$, with outcome $X_j(\omega)$ for any possible scenario $\omega \in \Omega$ (as usual, the X_j represent the assets). Let

$$\Lambda = \{\lambda \in \mathbb{R}_+^N : \sum_{i=1}^{N} \lambda_i = 1\}$$

so that $Y_\lambda = \sum_{j=1}^{N} \lambda_j X_j$ is a portfolio (a convex combination of the assets X_j) for $\lambda \in \Lambda$.

Additionally, we introduce the space $\mathcal{U}$ of all utility functions defined as:

$$\mathcal{U} = \{u : \mathbb{R} \to \mathbb{R}, u \in C^1, u \text{ is non-decreasing and concave}\}.$$

The concept of utility is used to represent consumer's preference ordering over a choice set and it assigns a real number to each alternative in such a way that, if consumer prefers alternative A to alternative B, then alternative A is assigned a number greater than alternative B. If $u \in \mathcal{U}$ then the preferences described by u are weakly monotone and convex. The following theorem shows that $\mathcal{U}$ can be equipped with a distance.

Proposition 6.1 *$(\mathcal{U}, d^*_{\sup})$ is a complete metric space where, as usual, $d^*_{\sup}$ is defined as:*

$$d^*_{\sup}(u_1, u_2) = d_{\sup}(u_1, u_2) + d_{\sup}(u_1', u_2') = \sup_{x \in \mathbb{R}} |u_1(x) - u_2(x)| + \sup_{x \in \mathbb{R}} |u_1'(x) - u_2'(x)|$$

for any u_1 and u_2 in $\mathcal{U}$.

Proof. The proof is quite straightforward. To prove that d is a distance is standard. The completeness, instead, follows by noticing that the $d^*_{\sup}$ metric induces the uniform convergence of any sequence of functions u_n and their derivatives u_n', and this is enough the preserve differentiability, monotonicity, and concavity. ■

Definition 6.2 Given $\hat{\lambda} \in \Lambda$, we say that $Y_{\hat{\lambda}}$ is *stochastically efficient with respect to $p \in \mathcal{P}$* if

$$\hat{\lambda} = \text{argmax}_{\lambda \in \Lambda} \mathbb{E}_p(u(Y_\lambda)) = \text{argmax}_{\lambda \in \Lambda} \int_\Omega u(Y_\lambda(\omega)) dp(\omega) \tag{6.11}$$

for some utility function $u \in \mathcal{U}$.

In other words, $Y_{\hat{\lambda}}$ is stochastically efficient with respect to p if there exists a utility function $u \in \mathcal{U}$ such that

$$\mathbb{E}_p(u(Y_\lambda)) \leq \mathbb{E}_p(u(Y_{\hat{\lambda}})) \tag{6.12}$$

for any $\lambda \in \Lambda$. Other definitions of stochastic efficiency could also be considered which include, for instance, an upper bound for the portfolio variance. This represents an open research avenue for further analysis in this area.

Due to the hypotheses on the utility function, the optimization problem

$$\max_{\lambda \in \Lambda} \ \mathbb{E}_p \left(u \left(\sum_{j=1}^{N} \lambda_j X_j \right) \right)$$

is a concave optimization program on the polyhedral set Λ and, therefore, it has at least a global maximizer (not unique, in general, because u is only concave). The problem is also well-posed, in the sense that if the sequence of portfolios $X^k = (X_1^k, ..., X_N^k)$ converges pointwise to the portfolio $X = (X_1, ..., X_N)$, for a.e. $\omega \in \Omega$, and $\lambda^k \in \Lambda$ is the solution to the sequence of problems:

$$\max_{\lambda \in \Lambda} \ \mathbb{E}_p \left(u \left(\sum_{j=1}^{N} \lambda_j X_j^k \right) \right)$$

then $\lambda^k \to \lambda$ and $\lambda \in \Lambda$ is the optimal solution to the problem

$$\max_{\lambda \in \Lambda} \ \mathbb{E}_p \left(u \left(\sum_{j=1}^{N} \lambda_j X_j \right) \right).$$

Remark 6.3 If the notion of classical probability is replaced by a vector-valued probability, the above notion of efficiency has to be understood in the Pareto sense. However, because we have already noticed that the vector-valued case can be included in the set-valued one by introducing an associated set-valued probability, this case is subsumed by the more general set-valued one which we discuss in detail in the next section.

The following result has been proved in [28] for the case of discrete space events. Here we extend this result to the case of any arbitrary probability space.

Proposition 6.4 *A given portfolio $Y_{\hat{\lambda}}$ is optimal for given $u \in \mathcal{U}$ and $p \in \mathcal{P}$ and if and only if it obeys the following first-order optimality conditions:*

$$\int_\Omega u'(Y_{\hat{\lambda}}(\omega))(X_i(\omega) - Y_{\hat{\lambda}}(\omega))dp(\omega) \leq 0$$

for all $i = 1...N$.

Proof. The proof of this result is quite straightforward and it follows by applying the classical first order optimality conditions for a set-constrained problem to the function $q(\lambda) = \mathbb{E}(u(Y_\lambda)) = \int_\Omega u(Y_\lambda(\omega))dp(\omega)$. ■

When the space is discrete, namely $\Omega = \{\omega_1, ..., \omega_S\}$, then the previous condition reduces to:

$$\sum_{s=1}^{S} p(\omega_s)u'(Y(\omega_s))(X_i(\omega_s) - Y_{\hat{\lambda}}(\omega_s)) \leq 0$$

for all $i = 1...N$ as in [28].

Suppose that Ω is also a metric space with respect to a distance d. Given two probability measures $p, q \in \mathcal{P}$, the Monge-Kantorovich distance between p and q is given by

$$d_{MK}(p, q) = \sup_{f \in Lip_1} \left\{ \int_\Omega f(\omega)dp(\omega) - \int_\Omega f(\omega)dq(\omega) \right\}$$

where $Lip_1(\Omega)$ is defined to be the set of all Lipschitz functions $f : \Omega \to \mathbb{R}$ such that

$$|f(\nu) - f(\xi)| \leq d(\nu, \xi)$$

for any $\nu, \xi \in \Omega$. If we define the point-to-set distance $d'(q, \Xi)$ as

$$d'_{MK}(q, \Xi) := \inf_{p \in \Xi} d_{MK}(q, p)$$

then the measure of inefficiency we use can be written as:

$$SIM(q) = d'_{MK}(q, \mathcal{P}^*)$$

where

$$\mathcal{P}^* = \{p \in \mathcal{P} : \exists u \in \mathcal{U} \textit{ such that } \int_\Omega u'(Y(\omega))(X_i(\omega) - Y(\omega))dp(\omega) \leq 0, i = 1...N\}.$$

In other words $SIM(q)$ is the distance point-to-set between q and the set $\mathcal{P}^*$. A given portfolio Y_λ is stochastically efficient relative to a given $p \in \mathcal{P}$ if and only if $SIM(p) = 0$. If $SIM(p) > 0$ we say the portfolio Y is stochastically inefficient.

If $\Omega \subseteq \mathbb{R}$, then the Monge-Kantorovich distance $d_{MK}(p, q)$ between two probability measures p and q can be rewritten in terms of the cumulative functions F_p and F_q as

$$d_{MK}(p, q) = \int_\Omega |F_p(x) - F_q(x)| \, dx.$$

If Ω is discrete set, namely $\Omega = \{\omega_1, ..., \omega_S\}$, the Monge-Kantorovich distance becomes

$$d_{MK}(p, q) = \max \left\{ \sum_{s=1}^{S} f(\omega_s)(p(\omega_s) - q(\omega_s)) \, : \, \|Af\|_\infty \leq 1 \right\},$$

where A is the edge-adjacency matrix for the weighted graph that describes the geometry of Ω (see [25] for details and discussion). Notice that this is a linear programming problem and so it can be solved by classical methods. The special structure also allows one to easily find approximate solutions. For estimating d_{MK} from empirical data one can see [17, 18, 19].

7 Stochastic Efficiency and Inefficiency with Set-Valued Probabilities

In this section we extend the notion of portfolio efficiency to the case of probability multimeasures. As before we let $(\Omega, \mathcal{F}, \Phi)$ be a probability space, where Ω and $\mathcal{F}$ are defined as in the previous section but now Φ is a set-valued probability according to the definition presented in Section 4. We also introduce the space of all $\mathbb{B}$ set-valued probabilities on $\mathcal{F}$, defined as

$$\mathbb{P} = \{\Phi : \Omega \to \mathcal{K} : \Phi(\Omega) = \mathbb{B}\}. \tag{7.13}$$

We comment that the particular choice of $\mathbb{B}$ does not affect the theory we present.

As in the previous section, X_j are random variables which represent the possible assets and $Y_\lambda = \sum_i \lambda_i X_i$ is a portfolio with $\lambda \in \Lambda$ and again

$$\Lambda = \{\lambda \in \mathbb{R}_+^N : \sum_{i=1}^N \lambda_i = 1\}.$$

By using the notion of the expected value with respect to a set-valued probability Φ, we can define the extend the notion of stochastic portfolio efficiency.

Definition 7.1 Given $\hat{\lambda} \in \Lambda$, we say that $Y_{\hat{\lambda}}$ is stochastically efficient with respect to the set-valued probability Φ if

$$\hat{\lambda} = \text{argmax}_{\lambda \in \Lambda} \mathbb{E}_\Phi(u(Y_\lambda)) = \text{argmax}_{\lambda \in \Lambda} \int_\Omega u(Y_\lambda(\omega) d\Phi(\omega) \tag{7.14}$$

for some utility function $u \in \mathcal{U}$.

Here the maximum has to be understood in the sense of set-inclusion ordering as previously discussed. The portfolio $Y_{\hat{\lambda}}$ is stochastically efficient with respect to Φ if there exists an utility function $u \in \mathcal{U}$ such that

$$\mathbb{E}_\Phi(u(Y_\lambda)) \subseteq \mathbb{E}_\Phi(u(Y_{\hat{\lambda}})) \tag{7.15}$$

for any $\lambda \in \Lambda$.

The following result characterizes the concavity of the optimization problem $\max_{\lambda \in \Lambda} \mathbb{E}_\Phi(u(Y_\lambda))$.

Proposition 7.2 *The function* $\xi(\lambda) := \mathbb{E}_\Phi\left(u\left(\sum_{j=1}^N \lambda_j X_j\right)\right)$ *is a concave set-valued map.*

Proof. To prove it, let us take $\lambda_1, \lambda_2 \in \Lambda$ and $\alpha \in (0,1)$. Then the following inequality holds:

$$u(Y_{\alpha\lambda_1+(1-\alpha)\lambda_2})) \geq \alpha u(Y_{\lambda_1}) + (1-\alpha)u(Y_{\lambda_2})$$

For any direction $p \in S^1$, the expected value with respect to the probability measure Φ^p reads as

$$\mathbb{E}_{\Phi^p}(u(Y_{\alpha\lambda_1+(1-\alpha)\lambda_2})) = \int_\Omega u(Y_{\alpha\lambda_1+(1-\alpha)\lambda_2}(\omega))d\Phi^p(\omega) \geq$$

$$\alpha\int_\Omega u(Y_{\lambda_1}(\omega))d\Phi^p(\omega) + (1-\alpha)\int_\Omega u(Y_{\lambda_2}(\omega))d\Phi^p(\omega) =$$

$$\alpha\mathbb{E}_{\Phi^p}(u(Y_{\lambda_1})) + (1-\alpha)\mathbb{E}_{\Phi^p}(u(Y_{\lambda_2}))$$

which now implies the (set-valued) concavity of the map $\xi(\lambda)$. ■

The concavity property implies that any local solution to the program

$$\max_{\lambda\in\Lambda} \mathbb{E}(u(Y_\lambda))$$

is a global solution and that it is well-posed.

We first observe that if a given portfolio is stochastically efficient with respect to Φ then it is stochastically efficient for any scalarized probability measure Φ^p. The converse is also true. That is, if a portfolio Y is stochastically efficient with respect to all the scalarized probability measures Φ^p then it is stochastically efficient with respect to Φ.

Proposition 7.3 *A given portfolio* $Y_{\hat\lambda}$ *is optimal for given utility function* $u \in \mathcal{U}$ *and a set-valued probability* Φ *if and only if it satisfies the following first-order optimality conditions:*

$$\int_\Omega u'(Y_{\hat\lambda}(\omega))X_i(\omega)d\Phi \subseteq \int_\Omega u'(Y_{\hat\lambda}(\omega))Y_{\hat\lambda}(\omega)d\Phi$$

for $i = 1...N$.

Proof. Because Y is stochastically efficient with respect to any probability Φ^p, we have

$$\int_\Omega u'(Y_{\hat\lambda}(\omega))(X_i(\omega) - Y_{\hat\lambda}(\omega))d\Phi^p(\omega) \leq 0, \quad i = 1...N$$

which is equivalent to

$$\int_\Omega u'(Y_{\hat\lambda}(\omega))X_i(\omega)d\Phi^p(\omega) \leq \int_\Omega u'(Y_{\hat\lambda}(\omega))Y_{\hat\lambda}(\omega)d\Phi^p(\omega), \quad i = 1...N$$

But now this can be written as

$$\text{spt}\left(\int_\Omega u'(Y_{\hat{\lambda}}(\omega))X_i(\omega)d\Phi, p\right) \leq \text{spt}\left(\int_\Omega u'(Y_{\hat{\lambda}}(\omega))Y_{\hat{\lambda}}(\omega)d\Phi, p\right)$$

for any $p \in S^1$, which is equivalent to

$$\int_\Omega u'(Y_{\hat{\lambda}}(\omega))X_i(\omega)d\Phi \subseteq \int_\Omega u'(Y_{\hat{\lambda}}(\omega))Y_{\hat{\lambda}}(\omega)d\Phi$$

for $i = 1...N$. ■

Given two set-valued probabilities $\Phi, \Xi \in \mathbb{P}$, the Monge-Kantorovich distance between them is defined as

$$\mathbf{d_{MK}}(\Phi, \Xi) = \sup_{q \in S^1} d_{MK}(\Phi^q, \Xi^q).$$

Several properties of the space $(\mathbb{P}, \mathbf{d_{MK}})$, including its completeness, are presented in [17, 21]. Define the space $\mathbb{P}^*$ as:

$$\mathbb{P}^* = \{\Phi \in \mathbb{P} : \exists u \in \mathcal{U}\ s.t. \int_\Omega u'(Y(\omega))X_i(\omega)d\Phi \subseteq \int_\Omega u'(Y(\omega))Y(\omega)d\Phi, i = 1...N\} \tag{7.16}$$

Using this we extend the notion of ineffiency as to the set-valued case as follows:

$$SIM(\Phi) = \inf_{\Xi \in \mathbb{P}^*} \mathbf{d_{MK}}(\Phi, \Xi). \tag{7.17}$$

A given portfolio Y is efficient if $SIM(\Phi) = 0$. We say that it is inefficient if $SIM(\Phi) > 0$. We can estimate (7.17) by scalarizing along a chosen set of directions. For finite Ω this leads to a finite set of linear programs.

8 Numerical Example

In this section we present an extended example that illustrates one possible use of set-valued probability and suggests a way to construct Φ in a practical context. We take the set of possible events to be finite, $\Omega = \{\omega_1, ..., \omega_s\}$. This means that in order to construct Φ on Ω we need a finite set of set-valued probabilities $\{P_1, \ldots, P_s\}$. Again we use X_j, $j = 1, 2, \ldots, N$, as the random variables which represent the assets.

Suppose that the investment decision must be taken now at time T, and that historical data ω_j^t are available for $t \in [0, T)$. In other words there is a complete knowledge about the different scenarios that have occurred over the continuous interval $[0, T)$. We will use a combination of the ideas from Examples 4.3 and 4.4 to construct Φ.

Rather than using a number (a probability) to describes the occurrence of each scenario over the interval $[0, T)$, let us instead take $2W$ different time-windows (which could be overlapping). Associated with each time-window there is a direction $q_w \in \mathbb{R}^2$ with $\|q_w\| = 1$ and $q_{2w} = -q_{2w+1}$ (as in Example 4.3). We

use the samples from the time-windows to estimate a set of classical probabilities $\{\hat{p}^1(\omega_s), \hat{p}^2(\omega_s), ..., \hat{p}^{2W}(\omega_s)\}$, one set for each $\omega_s \in \Omega$. Then we construct Φ be setting

$$\Phi(\omega_s) = \overline{\text{co}}(\{\hat{p}^w(\omega_s)q_w : w = 1, 2, \ldots, 2N\}).$$

It is easy to see that $\Phi(\Omega) = \overline{\text{co}}(q_w)$ is the polygon in $\mathbb{R}^2$ with vertices q_w. This polygon is a unit ball by our assumption on the directions q_w. By increasing the number of windows W, the polygon converges in the Hausdorff distance to a circle with radius equal to 1. The distribution induced in each direction q_w by the $\hat{p}^w(\omega_s)$s gives an estimate of the uncertainty within a given time-window. In this way we can model any nonstationarity of the uncertainty of the performance of the assets. Other mechanisms can also be used for constructing the set-valued probabilities in order to capture different features of the data.

The expected value of the portfolio $Y_\lambda = \sum_{j=1}^N \lambda_j X_j$ with respect to a given utility $u \in \mathcal{U}$, is given by

$$\mathbb{E}(u(Y_\lambda)) = \sum_{s=1}^{S} u(Y_\lambda(\omega_s))\Phi(\omega_s).$$

If we take each direction q_w and calculate the support of $\mathbb{E}(u(Y_\lambda))$ along the direction q_w, this leads to

$$\text{spt}(\mathbb{E}(u(Y_\lambda)), q_w) = \sum_{s=1}^{S} u(Y_\lambda(\omega_s))p^w(\omega_s).$$

The notion of stochastic dominance can be checked using the scalarized expected value with respect to any direction q_w. $Y_{\hat{\lambda}}$ is stochastically efficient with respect to Φ if

$$\hat{\lambda} = \text{argmax}_{\lambda \in \Lambda} \sum_{s=1}^{S} u(Y_\lambda(\omega_s))p^w(\omega_s).$$

for some utility function $u \in \mathcal{U}$ and for any $w = 1, 2, \ldots, 2W$.

9 Conclusion

In this paper we have extended the notion of stochastic efficiency and inefficiency to the financial context in which there is lack of information about the probability of each scenario. These situations can be found when financial data are missing, or corrupted, or noised, and there is only a partial information about some historical data. We propose to tackle this problem by means of the notion of set-valued probability. This object models any sort of possible vagueness and incompleteness related to the probability of a certain event by assuming that, instead of a positive number between 0 and 1, there is a compact and convex set associated with each event. The notion of set-valued probability possesses several properties that are similar to those held in the classical case. This is a first attempt to extend the notion of portfolio efficiency using

set-valued probabilities and it represents a further development of the material presented in [22, 23].

Acknowledgments

The second author (FM) was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Discovery Grant (2019-05237).

References

[1] Z. Artstein, Set-valued measures, *Transactions of the AMS*, **165** (1972), 103-125.

[2] Z. Artstein, On the calculus of closed set-valued functions, *Indiana Univeristy Mathematical Journal*, **24** (1974), 433-441.

[3] J.P. Aubin, H. Frankowska, *Set-valued analysis*, Birkhäuser, Boston, 1990.

[4] M. Beer (2004), Fuzzy Randomness Uncertainty in Civil Engineering and Computational Mechanics (Engineering Online Library), Springer.

[5] F. Ben Abdelaziz, H. Masri, Stochastic programming with fuzzy linear partial information on probability distribution, *European Journal of Operational Research*, **162**, (2005), no. 3, 619-629.

[6] F. Ben Abdelaziz, H. Masri, A compromise solution for the multiobjective stochastic linear programming under partial uncertainty, *European Journal of Operational Research*, **202**, (2010), 55-59.

[7] G. Beer, *Topologies on closed and closed convex sets*, Kluwer, Netherlands, 1993.

[8] G. Debreu, D. Schmeidler (1967) The Radon-Nikodym derivative of a correspondence. In: Lecam, L., Neyman, J., Scott, E.L. (eds.) Proceedings of Fifth Berkeley Symposium Math.Stat. and Probability. II, Part 1, pp. 351372. Berkely: University of California Press.

[9] Dempster, A. P (1966) New approaches for reasoning towards posterior distributions based on sample data. *Ann. Math. Statist.* 37 355374.

[10] Dempster A.P. (2008) Upper and Lower Probabilities Induced by a Multivalued Mapping. In: Yager R.R., Liu L. (eds) Classic Works of the Dempster-Shafer Theory of Belief Functions. *Studies in Fuzziness and Soft Computing*, vol 219. Springer, Berlin, Heidelberg.

[11] D. Dubois (1988) *Possibility theory an approach to computerized processing of uncertainty*, Plenum Press, New York.

[12] Y. Ermoliev, A. Gaivoronski, Stochastic optimization problems with incomplete information on distribution functions, *SIAM Journal on Control and Optimization*, **23**, (1985), no. 5, 697–716.

[13] Ferson, S., V. Kreinovich, L. Ginzburg, D.S. Myers, and K. Sentz. (2003). Constructing Probability Boxes and Dempster–Shafer Structures Archived 22 July 2011 at the Wayback Machine. SAND2002-4015. Sandia National Laboratories, Albuquerque, NM.

[14] C. Hess, Set-valued integration and set-valued probability theory: an overview, in *Handbook of Measure Theory, vols 1, II*, North-Holland, Amsterdam (2002).

[15] F. Hiai, Radon-Nikodým theorems for set-valued measures, *J. Multivariate Anal.* **8** (1978), no. 1, 96-118.

[16] D. Kandilakis, On the extension of multimeasures and integration with respect to a multimeasure, *Proceedings of the AMS*, **116**, (1992), no.1, 85-92.

[17] H.Kunze, D. La Torre, F.Mendivil, E.R.Vrscay, *Fractal-based methods in analysis*, Springer, New York, 2012.

[18] S. Iacus, D. La Torre, Approximating distribution functions by iterated function systems, *Journal of Applied Mathematics and Decision Sciences*, **17**, (2005), no.1, pp. 33-46

[19] S. Iacus, D. La Torre, A comparative simulation study on the IFS distribution function estimator, *Nonlinear Analysis: Real World Applications*, **6**, (2005), no5., pp. 858-873.

[20] D. La Torre, F. Mendivil, Iterated function systems on multifunctions and inverse problems *Journal of Mathematical Analysis and Applications*, **340** (2007), no. 2, 1469–1479.

[21] D. La Torre, F. Mendivil, The Monge–Kantorovich metric on multimeasures and self-similar multimeasures, *Set-Valued and Variational Analysis* **23** (2015), no. 2, 319-331.

[22] D. La Torre, F. Mendivil, Portfolio optimization under partial uncertainty and incomplete information: a probability multimeasure-based approach, *Annals of Operations Research*, **267** (2018), 267–279.

[23] D. La Torre, F. Mendivil, Stochastic linear optimization under partial uncertainty and incomplete information using the notion of probability multimeasure, *Journal of the Operational Research Society*, **69** (2018), no. 10, 1549–1556.

[24] H. Markowitz, Portfolio Selection, *The Journal of Finance* **7** (1952), 77-91.

[25] F. Mendivil, Computing the Monge-Kantorovich distance, *Comp. Appl. Math.*, **36**(3) (2017), 1989-1402.

[26] I. Molchanov (2005), Theory of Random Sets (Probability and its Applications), Springer.

[27] A. Neumaier (2004), Clouds, Fuzzy Sets, and Probability Intervals, *Reliable Computing* 10, 249272.

[28] T. Post, Empirical Tests for Stochastic Dominance Efficiency, *Journal of Finance*, **58** (2003), 19051932.

[29] R.T. Rockafellar, R. J-B. Wets, *Variational analysis*, Springer, New York, 1998.

[30] G. Shafer (1976), *A mathematical theory of evidence*, Princeton University Press in Princeton, NJ.

[31] M. Stojaković, Set valued probability and its connection with set valued measure, *Statist. Probab. Lett.* **82** (2012), 1043-1048.

[32] B. Urli, R. Nadeau, Stochastic MOLP with incomplete information: An interactive approach with recourse, *Journal Operational Research Society*, **41**, (1990), no. 12, 1143–1152.

[33] B. Urli, R. Nadeau, PROMISE/scenarios: An interactive method for multiobjective stochastic linear programming under partial uncertainty, *European Journal Operational Research*, **155**, (2004), 361–372.

[34] K. Vind (1964), Edgeworth-allocations in an exchange economy with many traders, *Int. Econ. Rev.* 5(2), 165177.

[35] P. Walley (1991), *Statistical reasoning with imprecise probabilities*, Chapman and Hall, London, New York.

[36] Y. Wang (2008) Imprecise probabilities with a generalized interval form. In R.L. Muhanna and R.L. Mullen, eds., Proc. 3rd Int. Workshop on Reliability Engineering Computing (REC'08), Savannah, Georgia, 45-59.

[37] K. Weichselberger (2000), The theory of interval-probability as a unifying concept for uncertainty, International Journal of Approximate Reasoning, 24 (23), pp. 149-170.