ISSN No. 2454 – 1427

CDE
March 2016

The Generalized Arrow-Pratt Coefficient

Sudhir A. Shah
Email: sudhir@econdse.org
Department of Economics
Delhi School of Economics

Working Paper No. 254

http://www.cdedse.org/working-paper-frameset.htm

CENTRE FOR DEVELOPMENT ECONOMICS
DELHI SCHOOL OF ECONOMICS
DELHI 110007

The generalized Arrow-Pratt coefficient

Sudhir A. Shah*

February 22, 2016

Abstract

We define vector-valued generalized Arrow-Pratt (GAP) coefficients for a utility defined on a Hilbert outcome space. Given risk averse, increasing and twice differentiable utilities on such outcome spaces, comparisons of their risk aversion using GAP coefficients are congruent to comparisons using well-founded decision-theoretic criteria. The Hilbert space setting admits risks embodied in a significant class of random processes, especially second-order processes. We also provide a theoretically well-founded and computationally tractable method for estimating the realized GAP coefficient from observed data when the outcome space is a reproducing kernel Hilbert space. We use the GAP coefficients to predict the effect of differences in risk aversion on an asset portfolio when assets are specified by dividend processes. Finally, we show a duality between utility functions on Euclidean spaces and GAP coefficients.

JEL classification: C44, C63, D81, G11

Key words: comparative risk aversion; equivalence results; generalized Arrow-Pratt coefficients; random processes; reproducing kernel Hilbert spaces; duality; eikonal equation

1 Introduction

1.1 Outline and motivation

Consider a risk averse, increasing and twice differentiable (Bernoulli-von Neumann-Morgenstern) utility function defined on an open interval in the real line. The function mapping each point in the domain to the Arrow-Pratt (henceforth, AP) coefficient of absolute risk aversion (Arrow [2], Pratt [18]) at that point may be called the AP function associated with the given utility.

Although random vector outcomes are commonplace in economic models, there is hitherto no satisfactory definition of a generalized AP function derived from a utility defined on a vector outcome space. We provide such a definition and explore its ramifications.

*Department of Economics, Delhi School of Economics, University of Delhi, Delhi 110007, India. Tel: +91-9899453559. Fax: +91-1127667159. E-mail: sudhir@econdse.org

The claim of a satisfactory definition, of course, begs the question regarding our notion of "satisfactory". We begin by describing our desiderata.

Suppose we are given two risk averse, increasing and twice differentiable utilities defined on a real interval and their associated AP functions. Given the notion that an AP coefficient at a point is a local measure of risk aversion, if one AP function pointwise dominates the other AP function over the interval, then it is plausible to interpret this as implying that the preference over lotteries underlying the former is more risk averse than the preference underlying the latter. The substantive justifications for this interpretation are the following well-known facts:

1. The AP function is determined by the underlying cardinal preference as it is uniform across all equivalent representations of the preference.

2. The partial order generated on the set of risk averse, increasing and twice differentiable utilities by comparing the associated AP functions is congruent to the partial order generated by various well-founded decision-theoretically compelling criteria for comparing risk aversion.[1]

We regard the exact replication of these properties as necessary for any satisfactory generalization of the AP function to a vector outcomes setting.

Random vector outcomes in an economic theory model usually are sample paths of a random process (henceforth, process). If a process has an infinite time-domain, as is usually the case, then its sample paths belong to an infinite dimensional path-space, i.e., a vector space of real-valued functions on the time-domain.

Therefore, in order to facilitate applications, an additional desirable property of a generalized AP function is:

3. It should be defined for a rich class of infinite dimensional path-spaces that allow the representation of a significant class of processes as admissible risks in the canonical form, i.e., as lotteries on the path-spaces.

In Section 4, equation (1) defines a vector-valued generalized Arrow-Pratt (henceforth, GAP) coefficient that is derived from a utility defined on a Hilbert space. It is easily verified that the resulting GAP function replicates property (1) and Theorem 4.5 shows that it also replicates property (2).

We show in Section 5 that lotteries on a reproducing kernel Hilbert (henceforth, RKH) path-space are decision-theoretically equivalent to processes whose sample paths belong to that space. This result is sharpened to an equivalence between second-order processes with an RKH path-space and lotteries on such spaces that satisfy a square-integrability property. These dualities show that the GAP function satisfies property (3), i.e., there is

[1]Proposition 6.C.2 in Mas-Colell et al. [15] is an omnibus version of this result.

a rich class of path-spaces that serve as domains for GAP functions and allow the representation of a significant class of processes by lotteries. In Appendix A, we verify that some familiar, and some less familiar but very useful, path-spaces are RKH spaces and characterize the second-order processes that can be represented by lotteries on these spaces.

RKH path-spaces play another important role. The definition of a GAP coefficient implies that, if the outcome space is a subset of an RKH space, then the realized GAP coefficient (contingent on the realized random outcome) is a real-valued function belonging to that RKH space. Suppose this function is unknown, but we have a finite set of observations from its graph. The question arises: Can this data be used to estimate the unknown GAP coefficient in some well-founded way? We show in Section 6 that this can be done systematically by modeling the estimate as the optimal trade-off between the competing concerns of goodness-of-fit and regularity.

It is a well-known application of AP functions that, given a stock and a risk-free bond with real-valued returns, optimal investment in the stock is inversely related to risk aversion when comparative risk aversion is modeled using AP functions. Now suppose the asset returns are generated by processes whose sample paths belong to a Hilbert path-space. Theorem 7.2 shows that the scalar comparative statics result can be generalized appropriately using the GAP function to model comparative risk aversion.

It is also well-known that there is a duality, i.e., a bijection, between appropriately defined sets of utility functions on the real line and AP functions. This means that an AP function embodies all the preference-related information that is embodied in its dual utility. Theorem 8.10 generalizes this duality by showing that utilities and GAP functions are dual objects in the general Euclidean setting. This result hinges on there being one and only solution of a system of non-linear partial differential equations associated with a GAP function. We show that (a) this seemingly intractable problem can be reduced to that of showing the existence of a unique solution of a single eikonal partial differential equation, and (b) solving the reduced problem subject to an auxiliary condition.

Apart from the sections outlined above, we shall state our technical conventions in Section 2, state our formal model in Section 3 and conclude in Section 9. Proofs of all the propositions in Sections 4-8 are collected in Appendix B.

1.2 The literature

Given a utility and a lottery over a Euclidean outcome space, Duncan [5] derived an "approximate vector risk premium" in terms of the "absolute risk aversion matrix" and Hellwig [7] and Kihlstrom and Mirman [9] have defined real-valued generalizations of AP coefficients. As comparisons of these constructs across preferences have not been characterized in terms of other well-founded decision-theoretic criteria, these constructs do not replicate property (2), which we regard as a *sine qua non* for a satisfactory generalization of an AP coefficient. Given their Euclidean setting, they also cannot satisfy criterion (3).

Since they are linked to this paper by property (2), we cite here the papers that compare risk aversion in vector outcome settings using criteria based on increasing concave transformations of utilities, acceptance sets, sets of risk premia (relations $\succeq_C$, $\succeq_A$ and $\succeq_\pi$, respectively, in Definition 4.2), sets of certainty equivalent outcomes and certainty equivalent probabilities (Karni [8], Kihlstrom and Mirman [9], Levy and Levy [13], Paroush [16], Shah [19]); the results are surveyed in Shah [19]. We should mention here that, with the exception of the last cited article, all the results are restricted to Euclidean spaces. The equivalence results in Shah [19] hold for the class of locally convex topological vector spaces, which (arguably) includes most, if not all, vector outcome spaces occurring in economic models. In order to demonstrate the inclusiveness and usefulness of such settings, it is shown that members of a very important class of processes commonly used in economic models, namely Wiener process and various processes derived from it, are representable by lotteries on such spaces.

2 Conventions

2.1 General conventions

Relation $\subset$ denotes weak inclusion. A subset of a topological space is given the subspace topology. A product of topological spaces is given the product topology. The real line $\Re$ is given the Euclidean topology.

A topological space E is given the Borel σ-algebra $\mathcal{B}(E)$ and $\Delta(E)$ is the set of countably-additive probability measures (henceforth, lotteries) on E.

Given a set T, let $\Re^T$ be the real vector space of all functions $f : T \to \Re$ with the usual pointwise definitions of vector addition and scalar multiplication. For $x, y \in \Re^T$, we say $x = y$ if $x(t) = y(t)$ for every $t \in T$.

Given $x, y \in \Re$, let $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$. Given $D \subset \Re$ and $f : D \to \Re$, we say that f is increasing if $x > y$ implies $f(x) > f(y)$ for $x, y \in D$.

For a function f, we denote the first (resp. second) derivative by Df (resp. $D^2 f$). Subsets of $\Re^n$ with Lebesgue measure zero are called negligible. If a property holds everywhere on a set other than a negligible subset, then we say that it holds almost everywhere (henceforth, a.e.) on the set.

2.2 Hilbert spaces

Consider a real vector space X with scalar product $\langle ., . \rangle : X^2 \to \Re$. The scalar product yields the norm on X, given by $\|x\| = \langle x, x \rangle^{1/2}$, and the distance function on X, given by $d(x,y) = \|x-y\|$. $(X, \langle .,. \rangle)$ is said to be a (real) **Hilbert space** if (X,d) is a complete metric space.

$B \subset X$ is called a **Hilbert basis** of $(X, \langle .,. \rangle)$ if it is orthonormal and X is the topological closure of the span of B. A Hilbert space $(X, \langle .,. \rangle)$ with $X \neq \{0\}$ has a Hilbert basis (Lang [12], Chapter V, Corollary 1.7).

Consider a Hilbert space $(X, \langle .,. \rangle)$. Let X^d be the vector space of all continuous real-valued linear functionals on X. X^d is isomorphic to X (Lang [12], Chapter V, Theorem 2.1). $X^* \subset X^d$ is said to be **total** if $x^*(x) = 0$ for all $x^* \in X^*$ implies $x = 0$. If $x \in X \setminus \{0\}$, then there exists $x^* \in X^d$ such that $x^*(x) \neq 0$ (Dunford and Schwartz [6], Corollary V.2.13). So, X^d is an example of a total set of continuous linear functionals on X.

The following class of vector spaces will be of particular interest in many results: a Hilbert space $(X, \langle .,. \rangle)$ is called a **reproducing kernel Hilbert (henceforth, RKH) space** if $X \subset \Re^T$ for a set T and the evaluation functional $e(t,.) : X \to \Re$, given by $e(t,x) = x(t)$, is continuous for every $t \in T$. For our purposes, the key property of an RKH space $(X, \langle .,. \rangle)$ is the following: as $e(t,.)$ is a continuous linear functional for $t \in T$, the Riesz representation theorem (Dunford and Schwarz [6], Theorem IV.4.5) implies the existence of a unique $k_t \in X$, called the reproducing kernel for t, such that $\langle k_t, . \rangle = e(t,.)$ on X.[2]

3 Model

Our results will concern the following setting for outcome spaces.

Definition 3.1 $((X, \langle .,. \rangle), X^*, \geq)$ *is called a* ***Hilbert (resp. RKH) setting*** *if* $(X, \langle .,. \rangle)$ *is a Hilbert (resp. RKH) space with a Hilbert basis* B*,* X^* *is a total set of continuous real-valued linear functionals on* X*, and for* $x, y \in X$*,* $x \geq y$ *if and only if* $\langle x-y, b \rangle \geq 0$ *for every* $b \in B$*. Given* $x, y \in X$*, we say* $x > y$ *if* $x \geq y$ *and* $x \neq y$*.*

A familiar RKH setting is specified by $X = \Re^n$, $\langle x,y \rangle = \sum_{i=1}^n x_i y_i$, $X^* = \{\langle e_i, . \rangle \mid i = 1, \ldots, n\}$ where $B = \{e_1, \ldots, e_n\}$ is the standard basis for $\Re^n$, and $x \geq y$ if and only if $x_i = \langle e_i, x \rangle \geq \langle e_i, y \rangle = y_i$ for $i = 1, \ldots, n$.

Definition 3.2 O *is called a* ***Hilbert (resp. RKH) outcome space*** *if it is a nonempty convex set that is open in* $(X, \langle .,. \rangle)$ *for a Hilbert (resp. RKH) setting* $((X, \langle .,. \rangle), X^*, \geq)$*.*

For a Hilbert outcome space O, we may identify $\Delta(O)$ with $\{\mu \in \Delta(X) \mid \mu(O) = 1\}$ as Lemma 4.20 in Aliprantis and Border [1] implies $\mathcal{B}(O) \subset \mathcal{B}(X)$.

[2] See Berlinet and Thomas-Agnan [3] for an exposition of RKH spaces. See Appendix A for examples of RKH spaces and examples of Hilbert spaces that are not RKH spaces.

Definition 3.3 *Given a Hilbert outcome space O, the set of **admissible lotteries** $\Delta(O)_0$ consists of lotteries $\mu \in \Delta(O)$ with a mean outcome $m_\mu \in O$ satisfying $x^*(m_\mu) = \int_O \mu(dz)\, x^*(z)$ for every $x^* \in X^*$.*[3]

The definition of X^* implies that $\mu \in \Delta(O)_0$ has a unique mean $m_\mu \in O$.

Given a Hilbert outcome space O, a utility $u : O \to \Re$ is said to be **risk averse** if u is measurable and $u(m_\mu) \geq U(\mu, u) := \int_O \mu(dy)\, u(y)$ for every $\mu \in \Delta(O)_0$, where the abstract Lebesgue integral $\int_O \mu(dy)\, u(y)$ is the expected utility from μ and u. A utility $u : O \to \Re$ is said to be **increasing** if $x, y \in O$ and $x > y$ implies $u(x) > u(y)$.

Definition 3.4 *Given a Hilbert outcome space O, the set of **admissible utilities** $\mathcal{U}$ consists of risk averse, increasing and twice differentiable utilities $u : O \to \Re$.*

We say that $u, v \in \mathcal{U}$ are **ordinally congruent** if, for all $x, y \in O$, $u(x) \geq u(y)$ if and only if $v(x) \geq v(y)$.

For $u, v \in \mathcal{U}$, let $u \equiv v$ if $u = a + bv$ where $a, b \in \Re$ and $b > 0$. Then, $\mathcal{U}/\equiv$ is the quotient set generated by the equivalence relation $\equiv$ and $[u]$ denotes the equivalence class containing u.

The derivative of $u \in \mathcal{U}$ is a mapping $Du : O \to X$ and the second derivative of u is a mapping $D^2u : O \to L(X, X)$, where $L(X, X)$ is the space of continuous linear maps from X to X.

4 Equivalence result

We prepare for our main definition with the following observation.

Lemma 4.1 *If O is a Hilbert outcome space and $u \in \mathcal{U}$, then $Du(.) > 0$ and $\|Du(.)\| > 0$ on O.*

Using this fact, we define the **generalized Arrow-Pratt coefficient (GAP coefficient)** of $u \in \mathcal{U}$ at $x \in O$ by

$$a_u(x) = \frac{-D^2u(x)Du(x)}{\|Du(x)\|^2} \tag{1}$$

If $X = \Re$, then a_u reduces to the AP coefficient. Moreover, a_u is invariant across utilities equivalent to u: if $b, c \in \Re$ with $c > 0$, then $a_u = a_{b+cu}$. So, a_u is determined purely by the cardinal preference underlying u. As $D^2u(x) \in L(X, X)$ and $Du(x) \in X$, we have $a_u(x) \in X$. So, $a_u : O \to X$ is the **GAP function** generated by $u \in \mathcal{U}$.

[3]Outcome m_μ is the weak integral of the identity function on O *à la* Pettis [17]. For example, in the Euclidean setting, the mean m_μ of $\mu \in \Delta(O)_0$ must satisfy $\langle e_i, m_\mu \rangle = \int_O \mu(dz)\, \langle e_i, z \rangle$ for $i = 1, \ldots, n$, which amounts to computing it component-by-component.

Consider a Hilbert outcome space O and $u \in \mathcal{U}$. The **acceptance set** of u at $x \in O$ is $A(x, u) = \{\mu \in \Delta(O)_0 \mid u(x) \leq U(\mu, u)\}$ (Yaari [20]). The set of **risk premia** of u at $\mu \in \Delta(O)_0$ is $\pi(\mu, u) = \{\pi \in X \mid u(m_\mu - \pi) = U(\mu, u)\}$ (Paroush [16]).

With this preparation, we now define various alternative criteria that may be used to compare risk aversion. Our aim is to show their equivalence.

Definition 4.2 *Given a Hilbert outcome space O and $u, v \in \mathcal{U}$, we say*

(a) $u \succeq_G v$ if u and v are ordinally congruent and $a_u \geq a_v$ on O,

(b) $u \succeq_C v$ if there exists an increasing concave function $f : v(O) \to \Re$ such that $u = f \circ v$,

(c) $u \succeq_A v$ if $A(x, u) \subset A(x, v)$ for every $x \in O$, and

(d) $u \succeq_\pi v$ if, for every $\mu \in \Delta(O)_0$, there do not exist $\pi_u \in \pi(\mu, u)$ and $\pi_v \in \pi(\mu, v)$ such that $\pi_v > \pi_u$.

The auxiliary ordinal congruence condition in the definition of $\succeq_G$ requires some discussion.

We first show that the comparability of GAP functions *per se* does not imply that the underlying utilities are ordinally congruent. Let $X = \Re^2$ and $O = \Re^2_{++}$. Consider utilities $u : O \to \Re$ and $v : O \to \Re$ given by

$$u(x, y) = x + e^{-3}y \qquad \text{and} \qquad v(x, y) = \ln xy \tag{2}$$

It is easy to verify that $u, v \in \mathcal{U}$ and

$$a_u(x, y) = (0, 0) \leq \left(\frac{y^2}{x(x^2 + y^2)}, \frac{x^2}{y(x^2 + y^2)} \right) = a_v(x, y)$$

for every $(x, y) \in O$, but u and v are not ordinally congruent as

$$u(e, e) = e + e^{-2} > 2 = u(1, e^3) \quad \text{and} \quad v(e, e) = 2 < 3 = v(1, e^3) \tag{3}$$

In the real outcomes setting, i.e., $O \subset X = \Re$, *all the utilities in* $\mathcal{U}$ are ordinally congruent as they are increasing. Since the auxiliary condition is trivially satisfied, it can be dropped in this case.

Now consider a general Hilbert outcome space O and utilities $u, v \in \mathcal{U}$. Examples such as u and v in equations (2) and (3) show that utilities in $\mathcal{U}$ may not be ordinally congruent. So, unlike in the real outcomes case, assuming that utilities are increasing does not ensure ordinal congruence. However, ordinal congruence is necessary for well-founded comparisons of risk aversion, as the next result shows.[4]

[4] Indeed, this result holds in far more general settings and for much larger sets of admissible utilities (Shah [19], Lemma 4.4).

Theorem 4.3 *Consider a Hilbert outcome space O and $u, v \in \mathcal{U}$.*

(A) If $u \succeq_C v$ (resp. $u \succeq_A v$), then u and v are ordinally congruent.

(B) Suppose, for all $x, y \in O$, there exists $z \in O$ such that $z \geq x$ and $z \geq y$. If $u \succeq_\pi v$, then u and v are ordinally congruent.

Given these facts, it is clear that, if we wish to define a criterion that (a) compares risk aversion by comparing GAP functions, and (b) is equivalent to well-founded criteria embodied in relations $\succeq_C$, $\succeq_A$ and $\succeq_\pi$, then such a definition must require ordinal congruence as an auxiliary condition. This has been done in the definition of $\succeq_G$.

We note the following facts as preparation for our main result.

Lemma 4.4 *If O is a Hilbert outcome space and $u, v \in \mathcal{U}$ are ordinally congruent, then*

(A) there is a unique function $f : v(O) \to \Re$ such that $u = f \circ v$; moreover, f is increasing and twice differentiable, and

(B) $Df > 0$ and $a_v - a_u = DvD^2f(v)/Df(v)$.

We now have our equivalence result.

Theorem 4.5 *If O is a Hilbert outcome space, then $\succeq_G = \succeq_C = \succeq_A = \succeq_\pi$ on $\mathcal{U}$.*

5 Duality of processes and lotteries

This section displays two dualities between processes and lotteries. For an outcome space that is an RKH path-space, these results serve to identify the risks embodied in lotteries on the outcomes with those embodied in processes with that path-space. These results provide substantive motivation for extending the theory of AP coefficients beyond Euclidean spaces.

Fix a **time domain** $T \subset \Re$. A vector subspace of $\Re^T$ is called a **path-space**. A **random process** is a collection $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$, where $(\Omega, \mathcal{F}, P)$ is a probability space, T is the time-domain and $y : T \times \Omega \to \Re$ is such that $y(t, .)$ is measurable for every $t \in T$, i.e., each $y(t, .)$ is a random variable. Each $\omega \in \Omega$ yields the **sample path** $\hat{y}(\omega)(.) = y(., \omega) \in \Re^T$.

Definition 5.1 *Consider a Hilbert outcome space $O \subset \Re^T$. If a process $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$ is such that $\hat{y}(\Omega) \subset O$, $\hat{y} : \Omega \to O$ is measurable and $P \circ \hat{y}^{-1} \in \Delta(O)_0$, then $\mathcal{Y}$ is said to be* ***representable*** *with respect to O.*

This connects processes to the model in Section 3 as the sample paths of a representable process become admissible outcomes and the distribution generated by it on the sample paths becomes an admissible lottery.

The possibility of representing a process by a lottery on its path-space is important because these representations are **decision-theoretically equivalent**: given O, an admissible utility $u : O \to \Re$ and a representable process $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$ with $\hat{y}(\Omega) \subset O$, the change-of-variable formula implies

$$\int_{\Omega} P(d\omega)\, u \circ \hat{y}(\omega) = \int_{O} P \circ \hat{y}^{-1}(dy)\, u(y) \tag{4}$$

The measurability of $\hat{y}$, which is required for defining $P \circ \hat{y}^{-1}$, is not inherent in the definition of a process and must be verified for each process.

Given a process $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$, let $m_{\mathcal{Y}}(.) = \int_{\Omega} P(d\omega)\, y(., \omega)$. If $m_{\mathcal{Y}} \in \Re^T$, then $m_{\mathcal{Y}}$ is called the **mean function** of $\mathcal{Y}$.

Theorem 5.2 *Consider an RKH outcome space $O \subset \Re^T$.*

(A) If process $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$ is such that $\hat{y}(\Omega) \subset O$ and $\hat{y}$ is measurable, then $m_{P \circ \hat{y}^{-1}} \in O$ if and only if $m_{\mathcal{Y}} \in O$. In either case, $m_{P \circ \hat{y}^{-1}} = m_{\mathcal{Y}}$.

(B) For every $P \in \Delta(O)$, the process $\mathcal{E}(P) = (O, \mathcal{B}(O), P; T, e)$ is such that $\hat{e}(O) = O$, $\hat{e}$ is measurable and $P \circ \hat{e}^{-1} = P$, where $e : T \times O \to \Re$ is the evaluation functional given by $e(t, x) = x(t)$.

Part (A) is useful for verifying whether a process $\mathcal{Y}$ is representable as it is often easier to verify $m_{\mathcal{Y}} \in O$ than to verify $m_{P \circ \hat{y}^{-1}} \in O$. Part (B) implies that the "coordinate processes" generated by lotteries in $\Delta(O)_0$ are representable.

Consider an RKH outcome space $O \subset \Re^T$. Let $\mathcal{P}(O)$ be the set of processes that are representable with respect to O. We say that processes $\mathcal{Y}_1 = (\Omega_1, \mathcal{F}_1, P_1; T, y_1)$ and $\mathcal{Y}_2 = (\Omega_2, \mathcal{F}_2, P_2; T, y_2)$ in $\mathcal{P}(O)$ are equivalent, denoted by $\mathcal{Y}_1 \sim \mathcal{Y}_2$, if $P_1 \circ \hat{y}_1^{-1} = P_2 \circ \hat{y}_2^{-1}$. Therefore, all processes belonging to an equivalence class in the quotient space $\mathcal{P}(O)/\!\sim$ generate the same distribution in $\Delta(O)_0$. Consequently, we have the function $\Psi : \mathcal{P}(O)/\!\sim\, \to \Delta(O)_0$, given by $\Psi([\mathcal{Y}]) = P \circ \hat{y}^{-1}$, where $[\mathcal{Y}] \in \mathcal{P}(O)/\!\sim$ is the equivalence class containing $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y) \in \mathcal{P}(O)$. Ψ is injective by the definition of $\sim$ and surjective by (B). So, our first duality result is that Ψ is a bijection between $\mathcal{P}(O)/\!\sim$ and $\Delta(O)_0$.

Next, we set the stage for a sharper duality result. Given an RKH path-space $(X, \langle ., . \rangle)$ with $X \subset \Re^T$, the evaluation $e(t, .) : X \to \Re$ is a continuous linear functional for every $t \in T$. Also, $e(t, x) = 0$ for every $t \in T$ if and only if $x = 0$. Therefore, we may set $X^* = \{e(t, .) \mid t \in T\}$.

So, consider an RKH setting $((X, \langle ., . \rangle), X^*, \geq)$ with $X \subset \Re^T$ and $X^* = \{e(t, .) \mid t \in T\}$. Let $\Delta(X)_2$ be the set of lotteries $\mu \in \Delta(X)$ such that $\int_X \mu(dx)\, \|x\|^2 < \infty$. Let $\mathcal{P}(X)_2$ be the set of processes $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$ such that $\hat{y}(\Omega) \subset X$, $\hat{y}$ is measurable and $P \circ \hat{y}^{-1} \in \Delta(X)_2$.

A process $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$ is called a **second-order process** if $\int_{\Omega} P(d\omega)\, y(t, \omega)^2 < \infty$ for every $t \in T$.

Theorem 5.3 *Given an RKH setting* $((X, \langle ., . \rangle), X^*, \geq)$,

(A) $\mathcal{P}(X)_2 \subset \mathcal{P}(X)$ *and every* $\mathcal{Y} \in \mathcal{P}(X)_2$ *is a second-order process, and*

(B) $\mathcal{E}(P) = (X, \mathcal{B}(X), P; T, e) \in \mathcal{P}(X)_2$ *and* $P \circ \hat{e}^{-1} = P$ *for every* $P \in \Delta(X)_2$.

Part (A) means that $\mathcal{P}(X)_2$ consists of second-order processes that are representable with respect to X by lotteries in $\Delta(X)_2$. Part (B) means that every lottery in $\Delta(X)_2$ is the distribution of some process in $\mathcal{P}(X)_2$.

6 Estimating the GAP coefficient

Given an RKH setting $((X, \langle ., . \rangle), X^*, \geq)$ with $X \subset \Re^T$ and $T \subset \Re$, consider the RKH outcome space $O = X$.[5]

Consider a utility $u \in \mathcal{U}$ and an observer who does not know u. Therefore, the observer cannot compute the GAP function $a_u : X \to X$ *ex ante*. If outcome $x \in X$ occurs, then the observer's data-set $D = \{(t_j, y_j) \in T \times \Re \mid j = 1, \ldots, n\}$ is interpreted to be a subset of the graph of the realized GAP coefficient $a_u(x) : T \to \Re$, i.e., $y_j = a_u(x)(t_j)$ for $j = 1, \ldots, n$. Since only one outcome occurs, there cannot be empirical data regarding $a_u(y)$ for $y \neq x$.

The observer's problem is to use data D to estimate $a_u(x)$ by some $b \in X$. The choice of b may be guided by two considerations.

On the one hand, high in-sample goodness-of-fit is desirable for explaining D. Goodness-of-fit is modeled by a loss function $\Gamma : (T \times \Re)^n \times \Re^n \to \Re_+$, where $\Gamma(D; b(t_1), \ldots, b(t_n))$ is the loss entailed by estimate $b \in X$, e.g., $\Gamma(D; b(t_1), \ldots, b(t_n)) = \sum_{j=1}^n (y_j - b(t_j))^2$.

On the other hand, it is desirable that out-of-sample predictions $b(t)$ depend on t in a regular fashion, i.e., small variations in t should not cause large variations in the prediction $b(t)$. In order to model regularity, let $(X, \langle ., . \rangle)$ be the RKH space $(A, \langle ., . \rangle_A)$ as *per* Theorem A.2. So, $\|b\|_X^2 = |b(0)|^2 + \sum_{i=0}^k \|D^{i+1} b\|_{L^2}^2$ where $k \in \mathcal{N} \cup \{0\}$. If we set $k = 0$, then $\|b\|_X^2 = |b(0)|^2 + \|Db\|_{L^2}^2$. For any intercept $b(0)$, $\|Db\|_{L^2}^2 \approx 0$ amounts to b being almost a constant function. Being a constant function is the strongest notion of the regularity of b. Less stringent forms of regularity, modeled by larger values of k, have analogous interpretations. Using $\|b\|_X^2$ as a measure of the irregularity of b, the loss caused by this irregularity is modeled by $\Lambda(\|b\|_X^2)$, where $\Lambda : \Re_+ \to \Re$ is increasing.

These criteria raise the issue of how they are to be employed. The observer may use them sequentially. One way is to create a set of estimates $b \in X$, say $X' \subset X$, with acceptably low goodness-of-fit loss $\Gamma(D; b(t_1), \ldots, b(t_n))$, possibly with 0 loss if X is rich enough, and then use the regularity loss $\Lambda(\|b\|_X^2)$ to choose from X'. Alternatively, the observer may restrict attention to a class of estimates $X' \subset X$ with acceptable

[5] The discussion of this section can be generalized to situations with $T \subset \Re^m$.

regularity and minimize the goodness-of-fit loss $\Gamma(D; b(t_1), \ldots, b(t_n))$ with respect to $b \in X'$.[6]

The third way is to use the criteria simultaneously and arrive at an endogenous optimal trade-off. Suppose the observer chooses $b \in X$ to minimize

$$\Gamma(D; b(t_1), \ldots, b(t_n)) + \Lambda(\|b\|_X^2) \tag{5}$$

Remarkably, even if X is infinite dimensional, problem (5) is reduced to a finite dimensional problem by the kernel representation theorem (Kimeldorf and Wahba [10]), which exploits the RKH space structure of $(X, \langle ., . \rangle)$.

Theorem 6.1 *Consider an RKH space $(X, \langle ., . \rangle)$ with $X \subset \Re^T$. Given $D = \{(t_j, y_j) \in T \times \Re \mid j = 1, \ldots, n\}$, let X_n be the subspace of X spanned by the reproducing kernels $\{k_{t_j} \in X \mid j = 1, \ldots, n\}$. For every $a \in X$, there exists $b \in X_n$ such that $\Gamma(D; b(t_1), \ldots, b(t_n)) + \Lambda(\|b\|_X^2) \leq \Gamma(D; a(t_1), \ldots, a(t_n)) + \Lambda(\|a\|_X^2)$.*

The proof shows that, for every $a \in X$, there exists $b \in X_n$ such that a and b have the same goodness-of-fit, i.e., $\Gamma(D; b(t_1), \ldots, b(t_n)) = \Gamma(D; a(t_1), \ldots, a(t_n))$, but b is more regular than a, i.e., $\|b\|_X^2 \leq \|a\|_X^2$.

Thus, the search for $b \in X$ that minimizes (5) may be restricted to X_n. As every $b \in X_n$ is a linear combination of the kernels $\{k_{t_1}, \ldots, k_{t_n}\}$, minimizing (5) amounts to choosing $\beta_1, \ldots, \beta_n \in \Re$ to minimize

$$\Gamma\left(D; \sum_{j=1}^n \beta_j k_{t_j}(t_1), \ldots, \sum_{j=1}^n \beta_j k_{t_j}(t_n)\right) + \Lambda\left(\left\| \sum_{j=1}^n \beta_j k_{t_j} \right\|_X^2\right) \tag{6}$$

In (6), the only unknowns are the choice variables $\beta_1, \ldots, \beta_n \in \Re$. Given the optimal choice $\beta_1^*, \ldots, \beta_n^*$, the best estimate of $a_u(x)$ is $\sum_{j=1}^n \beta_j^* k_{t_j}$.

7 Portfolio choice

Consider the problem of investing wealth $w \in \Re$ in a (risky) stock and a (risk-free) bond. It is well-known that, if the returns on these assets are real numbers, then a more risk averse investor will invest a smaller portion of w in the stock. We generalize this result to assets whose returns belong to a Hilbert path-space.[7]

Let $X \subset \Re^T$ be a Hilbert path-space with time-domain T. Let $X^* = X^d$. For simplicity of exposition, let the outcome space be $O = X$.

Each monetary unit invested in the bond (resp. stock) yields the fixed (resp. random) dividend path $\beta \in O$ (resp. $\beta + y \in O$), where y is a sample

[6] Linear regression theory uses the latter approach by looking for a best fit in the class of affine estimates.

[7] The result holds for any Hilbert space, but our narrative emphasizes path-spaces as their elements are interpretable as sample paths of processes.

path of a process with distribution $\mu \in \Delta(O)_0$. The investor is said to hold portfolio α if $\alpha \in \Re$ is invested in the stock and $w - \alpha$ is invested in the bond. Portfolio α's random dividend path is $w\beta + \alpha y$.

Let E be an open subset of $\Re$ such that $\Re_+ \subset E$. Given a utility $u \in \mathcal{U}$, define $\hat{u} : E \times O \to \Re$ by $\hat{u}(\alpha, y) = u(w\beta + \alpha y) - u(w\beta)$. The expected utility from portfolio α is $\int_O \mu(dy)\, u(w\beta + \alpha y) = \int_O \mu(dy)\, \hat{u}(\alpha, y) + u(w\beta)$.

We now define the class of utilities used in our result.

Definition 7.1 *Let $\mathcal{U}^*$ be the set of utilities $u \in \mathcal{U}$ satisfying*

(a) $\int_O \mu(dy)\, \hat{u}(\alpha, y) \in \Re$ for every $\alpha \in E$,

(b) there is a function $g : O \to \Re$ such that $\int_O \mu(dy)\, g(y) \in \Re$ and $|D_1\hat{u}(\alpha, .)| < g(.)$ for every $\alpha \in E$, and

(c) for every $(\alpha, y) \in E \times O$ such that $\alpha > 0$, $D_1\hat{u}(\alpha, y) > 0$ implies $\hat{u}(\alpha, y) > 0$ and $D_1\hat{u}(\alpha, y) < 0$ implies $\hat{u}(\alpha, y) < 0$.

Conditions (a) and (b) are used to ensure that $D \int_O \mu(dy)\, \hat{u}(., y) = \int_O \mu(dy)\, D_1\hat{u}(., y)$ for $u \in \mathcal{U}^*$ (Lang [12], Chapter VIII, Lemma 2.2).

Condition (c), which may be called directional monotonicity, has economic content. Consider $\alpha > 0$ and $y \in O \setminus \{0\}$. If the portfolio changes from 0 to α, then the perturbation of the dividend path in the direction y is αy. The resulting variation in utility is $\hat{u}(\alpha, y)$. Since u is increasing, condition (c) holds trivially if $y > 0$ or $y < 0$: if $y > 0$ (resp. $y < 0$), then $\hat{u}(\alpha, y) > 0$ (resp. $\hat{u}(\alpha, y) < 0$) and $D_1\hat{u}(\alpha, y) > 0$ (resp. $D_1\hat{u}(\alpha, y) < 0$). However, suppose neither $y > 0$ nor $y < 0$.[8] Condition (c) determines the sign of $\hat{u}(\alpha, y)$ in such cases: if the investor prefers to invest more (resp. less) than α in the stock, then $\hat{u}(\alpha, y)$ must be positive (resp. negative), i.e., portfolio α is superior (resp. inferior) to portfolio 0.

Given $u \in \mathcal{U}^*$, the investor's problem is to choose a portfolio α_u that maximizes $\int_O \mu(dy)\, \hat{u}(., y)$ over $\Re_+$. The restriction $\alpha \geq 0$ means that the investor cannot short sell the stock, but may buy any amount of it by selling the bond. Given our assumptions, there exists $\lambda \in \Re$ such that

$$\int_O \mu(dy)\, D_1\hat{u}(\alpha_u, y) + \lambda = 0 \qquad \lambda, \alpha_u \geq 0 \qquad \lambda\alpha_u = 0 \tag{7}$$

Theorem 7.2 *Consider w, X, X^*, O, β and μ as specified above and $u, v \in \mathcal{U}^*$. If*

(a) $\langle Dv(w\beta), m_\mu \rangle > 0$, and

(b) u and v are ordinally congruent and $a_u > a_v$ on O,

then $\alpha_u, \alpha_v > 0$. Furthermore, if

(c) $\mu(\{y \in O \mid D_1\hat{v}(\alpha_v, y) \neq 0\}) > 0$,

then $\alpha_u < \alpha_v$.

[8] The incompleteness of $\geq$ on X allows this possibility. It cannot occur if $X = \Re$.

The hypotheses of this result are generalizations of the conditions underlying its scalar version. In the scalar setting, assumption (a) is equivalent to the condition $m_\mu > 0$; otherwise, a risk averse investor will not invest in the stock. Assumption (b), *via* Theorem 4.5, implies that u is strictly more risk averse than v.

Assumption (c) is a non-degeneracy condition. It is well-specified because the continuity of Dv and the scalar product imply the continuity of $D_1\hat{v}(\alpha_v, .) = \langle Dv(w\beta + \alpha_v .), . \rangle$, and so, $\{y \in O \mid D_1\hat{v}(\alpha_v, y) \neq 0\}$ is open, and therefore Borel measurable. Given the directional monotonicity of v, $\{y \in O \mid D_1\hat{v}(\alpha_v, y) \neq 0\} \subset \{y \in O \mid \hat{v}(\alpha_v, y) \neq 0\} = \{y \in O \mid v(w\beta + \alpha_v y) \neq v(w\beta)\}$. So, the utilities from portfolios α_v and 0 are different for a non-negligible set of dividend paths.

8 Duality of utility and GAP functions

In this section, we study the duality between utility functions and Arrow-Pratt functions. In order to guide and motivate the theory for the vector outcomes setting, we first consider the familiar duality in the setting $X = \Re$.

8.1 The real outcomes case

Let $O = X = \Re$. We define two sets of admissible utilities.

Definition 8.1 *$\hat{\mathcal{U}}^1$ is the set of twice differentiable functions $u : O \to \Re$ with $Du > 0$ and $D^2u \leq 0$.*

$\mathcal{U}^1 \subset \hat{\mathcal{U}}^1$ consists of functions u such that $\lim_{|x|\uparrow\infty} |u(x)| = \infty$ and $|Du(.)| \in (\alpha, \beta)$ for some $\beta > \alpha > 0$.

The **AP coefficient** of $u \in \hat{\mathcal{U}}^1$ at $x \in O$ is

$$\chi u(x) = -D \ln Du(x) \geq 0$$

The resulting mapping $\chi u : O \to \Re_+$ is called the **AP function** generated by u. Next, we define a set of functions whose members will be rationalized as AP functions generated by members of $\hat{\mathcal{U}}^1$.

Definition 8.2 *$\hat{\mathcal{R}}^1$ is the set of functions $a : O \to \Re_+$ such that $a = Df$ for a differentiable function $f : O \to \Re$.*

$\mathcal{R}^1 \subset \hat{\mathcal{R}}^1$ consists of functions a such that f is bounded.

It is easy to check that, if $u \in \hat{\mathcal{U}}^1$ (resp. $u \in \mathcal{U}^1$), then $\chi u \in \hat{\mathcal{R}}^1$ (resp. $\chi u \in \mathcal{R}^1$). Since χ is invariant on $[u]$, we have the mappings $\chi : \hat{\mathcal{U}}^1/\!\equiv\, \to \hat{\mathcal{R}}^1$ and $\chi \circ \iota : \mathcal{U}^1/\!\equiv\, \to \mathcal{R}^1$, where $\iota : \mathcal{U}^1 \to \hat{\mathcal{U}}^1$ is the inclusion mapping given by $\iota(u) = u$.

Very elementary arguments yield two dualities.

Theorem 8.3 *$\chi : \hat{\mathcal{U}}^1/\!\equiv\, \to \hat{\mathcal{R}}^1$ and $\chi \circ \iota : \mathcal{U}^1/\!\equiv\, \to \mathcal{R}^1$ are bijections.*

8.2 The vector outcomes case

We now consider the duality problem with $O = X = \Re^n$. Given $u \in \mathcal{U}$, let $\Gamma u : O \to \Re^n$ be the GAP function. For a duality result, Theorem 8.3 suggests the program of defining a set $\mathcal{U}^n \subset \mathcal{U}$ of admissible utilities and a set $\mathcal{R}^n$ of admissible GAP functions $a : O \to \Re^n$ such that Γ is a bijection between these sets. Surjectivity of Γ means that, for every $a \in \mathcal{R}^n$, the system of n partial differential equations (henceforth, PDEs) $\Gamma u = a$ has a solution $u \in \mathcal{U}^n$. Injectivity of Γ requires that the system $\Gamma u = a$ cannot have multiple solutions in $\mathcal{U}^n$ for a given $a \in \mathcal{R}^n$. This program is unimplementable as we are unaware of results that guarantee the existence and uniqueness of solutions for such systems.

However, observe that $\Gamma u(.) = -D \ln \|Du(.)\|$. Using this representation of a GAP function, a plausible modified strategy for deriving a duality is to specify a set $\mathcal{F}$ of functions $f : O \to \Re$ such that:

1. $\mathcal{R}^n = \{Df : O \to \Re^n \mid f \in \mathcal{F}\}$, i.e., the admissible GAP functions are the gradients of functions in $\mathcal{F}$.

2. For every $f \in \mathcal{F}$, there is a unique admissible utility $u \in \mathcal{U}^n$ such that

$$\|Du(.)\| = e^{-f(.)} \tag{8}$$

3. For every $u \in \mathcal{U}^n$, there is some $f \in \mathcal{F}$ such that (8) holds.[9]

We observe a number of implications of these properties.

First, if $u, v \in \mathcal{U}^n$ are equivalent, denoted by $u \equiv v$, then $\Gamma u = \Gamma v$. So, we may specify the domain of Γ as the equivalence classes in the quotient set $\mathcal{U}^n/\equiv$.

Given $u \in \mathcal{U}^n$, property (3) implies the existence of $f \in \mathcal{F}$ such that $\Gamma u = Df$ and property (1) implies $Df \in \mathcal{R}^n$. So, $\Gamma : \mathcal{U}^n/\equiv \to \mathcal{R}^n$.

Given $a \in \mathcal{R}^n$, property (1) implies the existence of $f \in \mathcal{F}$ such that $a = Df$, and property (2) implies the existence of $u \in \mathcal{U}^n$ such that $\Gamma u = Df = a$, i.e., Γ is surjective.

Suppose $u, v \in \mathcal{U}^n$ and $\Gamma u = \Gamma v$. Property (3) implies the existence of $f, g \in \mathcal{F}$ such that $Df = \Gamma u = \Gamma v = Dg$. Using the mean value theorem, $f = g - c$ for some $c \in \Re$. Therefore, $\|Du(.)\| = e^{-f(.)} = e^c e^{-g(.)} = \|De^c v(.)\|$. By property (2), $u = e^c v$. So, $u \equiv v$ and Γ is injective.

Property (2) requires, for every $f \in \mathcal{F}$, the unique solvability of the eikonal PDE (8) in the set $\mathcal{U}^n$. Unfortunately, it is well-known that (8) is

[9]This strategy reduces the system of n PDEs to a single eikonal PDE (8). Such Hamilton-Jacobi PDEs characterize value functions in control theory (Cannarsa and Sinestrari [4]) and are widely studied in physics, especially in the area of geometric optics (Luneburg [14]). To the best of our knowledge, this is their first sighting in economics.

not generally solvable in the classical sense of it being satisfied at *all points* by a twice differentiable function.

So, the notion of solving (8) has to be relaxed and the sets $\mathcal{U}^n$, $\mathcal{R}^n$ and $\mathcal{F}$ have to accommodate this change. A number of weaker notions of solving (8) are available, but given our particular problem, the notion of a generalized solution, i.e., one that satisfies (8) at *almost all points*, is adequate and closest to the classical notion of a solution. An immediate consequence is that properties (1)-(3) cannot be required to hold everywhere and the best we can hope for is that they hold almost everywhere.

Furthermore, it is essential that $\mathcal{F}$ and $\mathcal{U}^n$ satisfy the properties required for the unique solvability of (8) in terms of generalized solutions. The first change in this respect is that $\mathcal{U}^n$ must admit not just concave functions but also the more general semiconcave functions. Theorem 8.6 shows that a preference underlying a semiconcave utility is "arbitrarily close on compact sets" to a risk averse preference. The second change is that we cannot insist on increasing utilities as (8) simply cannot fix the sign of Du: if u solves (8), then so does $-u$.

We now implement these changes. First, we modify the definition of a GAP function. Given a utility $u : O \to \Re$, let O_u be the set of outcomes $x \in O$ where u is twice differentiable and $\|Du(x)\| > 0$. Let

$$\Gamma u(x) = \begin{cases} -D \ln \|Du(x)\|, & x \in O_u \\ 0, & x \in O \setminus O_u \end{cases} \tag{9}$$

$\Gamma u(x)$ is the GAP coefficient of u at $x \in O_u$. We call $\Gamma u : O \to \Re^n$ the **generalized GAP function** generated by u.

As preparation for the definition of $\mathcal{F}$, we say that $g \in \mathcal{C}^{2,\alpha}$ for $g : O \to \Re$ if its partial derivatives up to second order are α-Hölder continuous for some $\alpha \in (0, 1)$.[10]

Definition 8.4 *$\mathcal{F}$ is the set of functions $f : O \to \Re$ with $e^{-f} \in \mathcal{C}^{2,\alpha}$ and $|f(.)| < k$ for some $k \in \Re$.*

As preparation for the definition of $\mathcal{U}^n$, we define semiconcave functions and note their properties. A function $u : O \to \Re$ is said to be **semiconcave** (with linear modulus) if there exists $C \geq 0$ such that the function $x \mapsto u(x) - C\|x\|^2/2$ is concave on O; u is called semiconvex if $-u$ is semiconcave. Clearly, if u is concave, then it is semiconcave. Also, if u is semiconcave, then every $v \in [u]$ is semiconcave, which makes it meaningful to say "preference $[u]$ is semiconcave" if u is semiconcave. We note two regularity properties of semiconcave functions.

Remark 8.5 *Consider a semiconcave utility $u : O \to \Re$.*

[10] Hölder continuity is a ubiquitous and essential notion of regularity in the study of PDEs; see Ziemer [22] for the definition.

(A) A generalization of Alexandrov's theorem (Cannarsa and Sinestrari [4], Theorem 2.3.1) implies that u is twice differentiable a.e. Thus, if $\|Du(.)\| > 0$ a.e., then the first case of (9) determines Γu a.e.

(B) u is locally Lipschitzian (Cannarsa and Sinestrari [4], Theorem 2.1.7).

A semiconcave function can be non-concave as long as the non-concavity can be compensated by the concavity of $x \mapsto -C\|x\|^2/2$ for some $C > 0$. For instance, consider the functions

$$u(x) = \begin{cases} x, & x < 0 \\ x^2/2, & x \geq 0 \end{cases} \qquad \text{and} \qquad u(x) - x^2/2 = \begin{cases} x - x^2/2, & x < 0 \\ 0, & x \geq 0 \end{cases}$$

Although u is not concave, $x \mapsto u(x) - x^2/2$ is concave. So, u is semiconcave.

We say that *preferences $[u]$ and $[v]$ are arbitrarily close on compact sets*, denoted by $[u] \approx [v]$, if for every compact set $O' \subset O$ and every $\epsilon > 0$, there exists $u' \in [u]$ and $v' \in [v]$ such that $\sup\{|u'(x) - v'(x)| \mid x \in O'\} < \epsilon$. The next result provides an economic interpretation of semiconcave preferences.

Theorem 8.6 *If $[u]$ is a semiconcave preference, then there is a risk averse preference $[v]$ such that $[v] \approx [u]$.*

The set of admissible utilities satisfying the generalized property (3) is as follows.

Definition 8.7 *$\mathcal{U}^n$ is the set of semiconcave functions $u : O \to \Re$ satisfying $\lim_{\|x\| \to \infty} u(x) = \infty$ and $\|Du(.)\| = e^{-f(.)}$ a.e. for some $f \in \mathcal{F}$.*

If $u \in \mathcal{U}^n$, then $\|Du(.)\| > 0$ a.e. and Remark 8.5(A) implies that u is twice differentiable a.e. Thus, $O \setminus O_u$ is negligible and the first case in (9) generates Γu a.e.

The set of admissible GAP functions satisfying the generalized property (1) is as follows.

Definition 8.8 *$\mathcal{R}^n$ is the set of functions $a : O \to \Re^n$ satisfying $a = Df$ a.e. for some $f \in \mathcal{F}$. We say that $a, b \in \mathcal{R}^n$ are equivalent, denoted by $a \simeq b$, if $a = b$ a.e.*

Finally, given $f \in \mathcal{F}$, a **generalized solution** (Kružkov [11]) of (8) is a semiconcave function $u : O \to \Re$ such that $\|Du(.)\| = e^{-f(.)}$ a.e. and $\lim_{\|x\| \to \infty} u(x) = \infty$. Consequently, given $f \in \mathcal{F}$, a generalized solution of (8) is some $u \in \mathcal{U}^n$. Note the existence and uniqueness properties of generalized solutions.

Lemma 8.9 *Consider $f \in \mathcal{F}$.*

(A) If $u : O \to \Re$ and $v : O \to \Re$ are generalized solutions of (8), then $u = v$.

(B) (8) has a generalized solution $u : O \to \Re$.

We now have our duality result.

Theorem 8.10 Γ *maps* $\mathcal{U}^n/\equiv$ *to* $\mathcal{R}^n$ *such that, for every* $a \in \mathcal{R}^n$*, there is a unique* $[u] \in \mathcal{U}^n/\equiv$ *with* $\Gamma[u] \simeq a$.

Using the results of Kružkov [11], an analogous result can be derived if O has a boundary $\partial O \neq \emptyset$. Although we do not present this result here, we should mention that the main complication introduced by a boundary is that the uniqueness result analogous to Lemma 8.9 requires the supplementary restriction that u should match some prescribed boundary data $h : \partial O \to \Re$ on ∂O. Consequently, the duality is not between $\mathcal{U}^n$ and $\mathcal{R}^n$, but between $\mathcal{U}^n$ and $\mathcal{R}^n \times \mathcal{H}$, where $\mathcal{H}$ is an appropriate set of boundary data $h : \partial O \to \Re$.

9 Conclusions

We have defined a vector-valued GAP coefficient generated by a risk averse, increasing and twice differentiable utility defined on an outcome space contained in a Hilbert space. It specializes to the AP coefficient in the real outcomes setting. It also meets our desiderata for a satisfactory generalization of the AP coefficient to vector outcomes. Specifically, it is preference-determined, the GAP function compares risk aversion in the same way as all the decision-theoretically compelling criteria for doing so, and finally, Hilbert outcome spaces allow the representation of a rich class of random processes as admissible risks in the form of lotteries on the outcomes.

We also provide a systematic method for estimating the GAP coefficient from data when outcomes belong to a reproducing kernel Hilbert space.

We use the GAP function to model the effect of risk aversion on the optimal allocation of wealth between a stock and a bond whose returns are generated by processes with sample paths in a Hilbert path-space. Generalizing the result obtained in the real returns setting, we show that greater risk aversion lowers the investment in the stock.

Finally, we derive a duality between utility functions and GAP functions on Euclidean spaces, which generalizes the familiar duality between utility functions and AP functions on the real line.

A Appendix: Examples of RKH spaces

We present three examples of RKH path-spaces. Their respective time domains are finite, denumerable and uncountable. The third setting is of particular interest as it is used for estimating GAP coefficients in Section 6. We also characterize second-order processes with these path-spaces.

First, if $T = \{1, \ldots, n\}$, then $\Re^T$ with the scalar product $\langle x, y\rangle = \sum_{t\in T} x_t y_t$ is a Hilbert space. As the Euclidean topology generated by $\langle ., .\rangle$

coincides with the product topology and $e(t,.)$ is a projection mapping, it is continuous. Thus, $(\Re^T, \langle .,. \rangle)$ is an RKH space.

Second, set $T = \mathcal{N}$ and let l^2 be the set of sequences $x \in \Re^T$ such that $\sum_{t\in T} x_t^2 < \infty$. Define the scalar product of $x, y \in l^2$ by $\langle x, y\rangle_{l^2} = \sum_{t\in T} x_t y_t$.

Theorem A.1 *$(l^2, \langle .,. \rangle_{l^2})$ is an RKH space.*

Proof. It is well-known that $(l^2, \langle .,. \rangle_{l^2})$ is a Hilbert space. Consider $t \in T$, $x \in l^2$ and $\epsilon > 0$. If $y \in l^2$ and $\|x - y\|_{l^2} < \epsilon$, then $|x_t - y_t|^2 \leq \sum_{s\in T}(x_s - y_s)^2 = \|x - y\|_{l^2}^2 < \epsilon^2$. Therefore, $|e(t, x) - e(t, y)| = |x_t - y_t| < \epsilon$ and $e(t, .)$ is continuous. ■

Third, fix the measure space $(T, \mathcal{B}(T), \mathrm{Leb})$ with $T = [0, 1]$ and Lebesgue measure Leb. Measurable functions $x, y \in \Re^T$ are said to be equivalent if $x = y$ Leb a.e. Let L^2 be the resulting collection of equivalence classes of measurable functions $x \in \Re^T$ such that $\int_T dt\, |x(t)|^2 < \infty$. $(L^2, \langle .,. \rangle_{L^2})$ is a Hilbert space given the scalar product $\langle x, y\rangle_{L^2} := \int_T dt\, x(t)y(t)$ (Yosida [21], Section I.9, Proposition 2).

Let $H^0 = L^2$. For $k \in \mathcal{N}$, let H^k be the collection of $x \in L^2$ with distributional derivatives $D^i x \in L^2$ for $i = 1, \ldots, k$. $(H^k, \langle .,. \rangle_{H^k})$ is a Sobolev space (see Ziemer [22]) that is also a Hilbert space given the scalar product $\langle x, y\rangle_{H^k} := \sum_{i=0}^k \langle D^i x, D^i y\rangle_{L^2}$ with $D^0 x := x$ (Yosida [21], Section I.9, Proposition 5). Clearly, $\|x\|_{H^k}^2 = \sum_{i=0}^k \|D^i x\|_{L^2}^2 \geq \|D^j x\|_{L^2}^2$ for $j = 0, \ldots, k$.

As L^2 and H^k consist of equivalence classes of functions, their evaluations $e(t, .)$ are not well-defined. **Therefore, L^2 and H^k are not RKH spaces**.

Let A be the vector subspace of $\Re^T$ consisting of absolutely continuous $x \in \Re^T$ with distributional derivative $Dx \in H^k$. Let $\langle x, y\rangle_A := x(0)y(0) + \langle Dx, Dy\rangle_{H^k}$ for $x, y \in A$. Let $\|x\|_A^2 := \langle x, x\rangle_A = |x(0)|^2 + \|Dx\|_{H^k}^2$ for $x \in A$.

Theorem A.2 *$(A, \langle .,. \rangle_A)$ is an RKH space.*

Proof. Let $x \in A$ and $t \in T$. As $Dx \in H^k \subset L^2$ and $1_{[0,t]} \in L^2$, we have $|\langle Dx, 1_{[0,t]}\rangle_{L^2}| \leq \|Dx\|_{L^2}\|1_{[0,t]}\|_{L^2} = t^{1/2}\|Dx\|_{L^2} \leq \|Dx\|_{L^2} \leq \|Dx\|_{H^k}$. As $\int_{[0,t]} ds\, Dx(s) = \langle Dx, 1_{[0,t]}\rangle_{L^2}$, we have $|x(0) + \int_{[0,t]} ds\, Dx(s)| \leq |x(0)| + |\langle Dx, 1_{[0,t]}\rangle_{L^2}| \leq |x(0)| + \|Dx\|_{H^k}$. It follows that

$$|e(t, x)| = |x(t)| = \left| x(0) + \int_{[0,t]} ds\, Dx(s) \right| \leq |x(0)| + \|Dx\|_{H^k} \tag{10}$$

If $\|x\|_A = 0$, then $|x(0)| = 0 = \|Dx\|_{H^k}$, and (10) implies $x = 0$. As the other properties of scalar products and norms are easily verified, $\langle .,. \rangle_A$ is a scalar product and $\|.\|_A$ is a norm.

Suppose $\|x\|_A < \epsilon$. It follows that $|x(0)| < \epsilon$, $\|Dx\|_{H^k} < \epsilon$, and (10) implies $|e(t, x)| < 2\epsilon$. Therefore, $e(t, .) : A \to \Re$ is continuous. Thus, if $(A, \langle .,. \rangle_A)$ is complete, then it is an RKH space.

To verify completeness, consider a Cauchy sequence (x_n) in $(A, \langle ., . \rangle_A)$. We show the existence of $x \in A$ such that $\lim_n \|x_n - x\|_A = 0$.

As (x_n) is a Cauchy sequence, for every $\epsilon > 0$, there exists $N \in \mathcal{N}$ such that $m, n > N$ implies $\|Dx_m - Dx_n\|_{H^k} \leq \|x_m - x_n\|_A < \epsilon$. Thus, (Dx_n) is a Cauchy sequence in $(H^k, \langle ., . \rangle_{H^k})$. As $(H^k, \langle ., . \rangle_{H^k})$ is complete, there exists $g \in H^k$ such that $\lim_n \|Dx_n - g\|_{H^k} = 0$.

Consider $t \in T$ and $\epsilon > 0$. As (x_n) is a Cauchy sequence, there exists $N \in \mathcal{N}$ such that $m, n > N$ implies $\|x_m - x_n\|_A < \epsilon$. Using (10), $|x_m(t) - x_n(t)| = |e(t, x_m - x_n)| < 2\epsilon$. Thus, $(x_n(t))$ is a Cauchy sequence in $\Re$. As $\Re$ is complete, we can define $x \in \Re^T$ pointwise by $x(.) = \lim_n x_n(.)$.

We verify that $x \in A$. Applying Jensen's inequality, $\|Dx_n - g\|_{H^k}^2 \geq \|Dx_n - g\|_{L^2}^2 = \int_T ds\, |Dx_n(s) - g(s)|^2 \geq [\int_T ds\, |Dx_n(s) - g(s)|]^2$, and for every $t \in T$, $\int_{[0,t]} ds\, |Dx_n(s) - g(s)| \geq |\int_{[0,t]} ds\, [Dx_n(s) - g(s)]|$. Therefore, $\|Dx_n - g\|_{H^k} \geq \int_T ds\, |Dx_n(s) - g(s)| \geq \int_{[0,t]} ds\, |Dx_n(s) - g(s)| \geq |\int_{[0,t]} ds\, [Dx_n(s) - g(s)]| \geq 0$ for every $t \in T$. As $\lim_n \|Dx_n - g\|_{H^k} = 0$, we have $\lim_n \int_{[0,t]} ds\, Dx_n(s) = \int_{[0,t]} ds\, g(s)$ for every $t \in T$. It follows that $x(t) = \lim_n x_n(t) = \lim_n [x_n(0) + \int_{[0,t]} ds\, Dx_n(s)] = x(0) + \int_{[0,t]} ds\, g(s)$ for every $t \in T$. Therefore, x is absolutely continuous, $Dx = g \in H^k$ and $x \in A$.

Moreover, $\lim_n \|x_n - x\|_A^2 = \lim_n [x_n(0) - x(0)]^2 + \lim_n \|Dx_n - Dx\|_{H^k}^2 = \lim_n \|Dx_n - g\|_{H^k}^2 = 0$, as required. ■

We now characterize second-order processes that have $(\Re^T, \langle ., . \rangle)$ and $(l^2, \langle ., . \rangle_{l^2})$ as path-spaces. The characterizations involve the **covariance kernel function** $K_\mathcal{Y} \in \Re^{T \times T}$ of a process $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$, defined by $K_\mathcal{Y}(s,t) = \int_\Omega P(d\omega)\, (y(s,\omega) - m_\mathcal{Y}(s))(y(t,\omega) - m_\mathcal{Y}(t))$.

Theorem A.3 *If $\mathcal{Y} \in \mathcal{P}(l^2, \langle ., . \rangle_{l^2})$, then $\mathcal{Y} \in \mathcal{P}(l^2, \langle ., . \rangle_{l^2})_2$ if and only if $\operatorname{tr} K_\mathcal{Y} < \infty$. An analogous characterization holds for the Euclidean path-space $(\Re^T, \langle ., . \rangle)$.*

Proof. By Theorem 5.2, $m_\mathcal{Y} \in l^2$. Therefore, $\int_\Omega P(d\omega)\, y(t,\omega)^2 = m_\mathcal{Y}(t)^2 + K_\mathcal{Y}(t,t)$ for $t \in \mathcal{N}$. Applying the monotone convergence theorem (Dunford and Schwartz [6], Corollary III.6.17), we have $\int_{l^2} P \circ \hat{y}^{-1}(dx)\, \|x\|_{l^2}^2 = \int_\Omega P(d\omega)\, \|\hat{y}(\omega)\|_{l^2}^2 = \int_\Omega P(d\omega) \sum_{t \in \mathcal{N}} y(t,\omega)^2 = \sum_{t \in \mathcal{N}} \int_\Omega P(d\omega)\, y(t,\omega)^2 = \|m_\mathcal{Y}\|_{l^2}^2 + \operatorname{tr} K_\mathcal{Y}$. The result follows as $\|m_\mathcal{Y}\|_{l^2} < \infty$. ■

In the same spirit, we characterize a second-order process with the RKH path-space $(A, \langle ., . \rangle_A)$ with $k = 0$, i.e., $\langle x, y \rangle_A := x(0)y(0) + \langle Dx, Dy \rangle_{L^2}$.

Theorem A.4 *If $\mathcal{Z} = (\Omega, \mathcal{F}, P; T, z)$ is a process such that $\hat{z}(\Omega) \subset L^2$, z is measurable, $\hat{z}$ is measurable and $m_\mathcal{Z} \in L^2$, then*

(A) $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y)$ is a process with $y(t,\omega) = \int_{[0,t]} ds\, z(s,\omega)$ for $(t,\omega) \in T \times \Omega$, $\hat{y}(\Omega) \subset A$ and $\hat{y}$ is measurable.

(B) Moreover, $\mathcal{Y} \in \mathcal{P}(A, \langle ., . \rangle_A)_2$ if and only if $\int_T dt\, K_\mathcal{Z}(t,t) < \infty$.

Proof. (A) For $\omega \in \Omega$, as $z(.,\omega) = \hat{z}(\omega)(.) \in L^2$, it is measurable. As $|y(t,\omega)| = |\int_{[0,t]} ds\, z(s,\omega)| = \langle \hat{z}(\omega), 1_{[0,t]} \rangle_{L^2} \leq \|\hat{z}(\omega)\|_{L^2} \|1_{[0,t]}\|_{L^2} = t^{1/2} \|\hat{z}(\omega)\|_{L^2} < \infty$ for $(t,\omega) \in T \times \Omega$, $y : T \times \Omega \to \Re$ is well-defined. Since $z = z^+ - z^-$, where $z^+(.) = z(.) \vee 0 \geq 0$ and $z^-(.) = -(z(.) \wedge 0) \geq 0$ are measurable functions, Tonelli's theorem (Dunford and Schwartz [6], Theorem III.11.14) implies that $y(t,.) = \int_{[0,t]} ds\, z^+(s,.) - \int_{[0,t]} ds\, z^-(s,.)$ is measurable for every $t \in T$. So, $\mathcal{Y}$ is well-defined.

For every $\omega \in \Omega$, $\hat{y}(\omega)$ is absolutely continuous and its distributional derivative is $\hat{z}(\omega) \in L^2 = H^0$. So, $\hat{y}(\Omega) \subset A$.

The linear mapping $F : L^2 \to A$, defined by $F(g)(t) = \int_{[0,t]} ds\, g(s)$, is continuous at $0 \in L^2$ as $\|F(g)\|_A = \|g\|_{L^2}$ for $g \in L^2$. So, F is continuous and therefore measurable. As $\hat{z}$ is measurable, so is $\hat{y}(.) = F \circ \hat{z}(.)$.

(B) As z is measurable, Tonelli's theorem (Dunford and Schwartz [6], Theorem III.11.14) implies that $\int_A P \circ \hat{y}^{-1}(dx)\, \|x\|_A^2 = \int_\Omega P(d\omega)\, \|\hat{y}(\omega)\|_A^2 = \int_\Omega P(d\omega)\, \|\hat{z}(\omega)\|_{L^2}^2 = \int_\Omega P(d\omega) \int_T dt\, z(t,\omega)^2 = \int_T dt \int_\Omega P(d\omega)\, z(t,\omega)^2$. As $\int_\Omega P(d\omega)\, z(t,\omega)^2 = \int_\Omega P(d\omega)\, [z(t,\omega) - m_{\mathcal{Z}}(t) + m_{\mathcal{Z}}(t)]^2 = K_{\mathcal{Z}}(t,t) + m_{\mathcal{Z}}(t)^2$, we have $\int_A P \circ \hat{y}^{-1}(dx)\, \|x\|_A^2 = \int_T dt\, [K_{\mathcal{Z}}(t,t) + m_{\mathcal{Z}}(t)^2] = \int_T dt\, K_{\mathcal{Z}}(t,t) + \|m_{\mathcal{Z}}\|_{L^2}^2$. The result follows as $\|m_{\mathcal{Z}}\|_{L^2} < \infty$. ■

B Appendix: Proofs

Proof of Lemma 4.1 Let $x \in O$ and $b \in B$, where B is a Hilbert basis of $(X, \langle .,. \rangle)$. As O is open in X, there exists $s > 0$ such that $x + sb \in O$. Since O is convex, $x + tsb = (1-t)x + t(x + sb) \in O$ for every $t \in (0,1)$. As $b > 0$, we have $sb > 0$, and as u is increasing, $u(x + sb) > u(x)$. As u is risk averse, it is concave. Since B is orthonormal, $\|b\| = 1$. So, for every $t \in (0,1)$, $t[u(x + sb) - u(x)] \leq u(x + tsb) - u(x) = t\langle Du(x), sb \rangle + tsr(tsb)$, where $\lim_{\|\tau\| \downarrow 0} r(\tau) = 0$. Dividing by t and letting $t \downarrow 0$, we have $\langle Du(x), sb \rangle \geq u(x + sb) - u(x) > 0$. As $s > 0$, we have $\langle Du(x), b \rangle > 0$. As this holds for every $b \in B$, we have $Du(x) > 0$. ■

Proof of Theorem 4.3 We start with
Observation 1: *If $y, z \in O$, $f, g \in \mathcal{U}$, $f(z) < f(y)$ and $g(z) \geq g(y)$, then there exists $x \in O$ such that $f(x) < f(y)$ and $g(x) > g(y)$.*
Proof. Let $b \in B$. As B is a Hilbert base, $b > 0$. As O is open, there exists $N \in \mathcal{N}$ such that $z + b/n \in O$ for every $n > N$. Otherwise, there is an increasing sequence $(k_n) \subset \mathcal{N}$ such that $\lim_n k_n = \infty$ and $z + b/k_n \in X \setminus O$ for every n. This implies $z = \lim_n (z + b/k_n) \in X \setminus O$ as $X \setminus O$ is closed in X, which is a contradiction.

There exists $n_0 > N$ such that $f(z + b/n_0) < f(y)$; otherwise, $f(z + b/n) \geq f(y)$ for every $n > N$, and by the continuity of f, this means $f(z) = \lim_n f(z + b/n) \geq f(y)$, a contradiction. As g is increasing, $g(z + b/n_0) > g(z) \geq g(y)$. Set $x = z + b/n_0$. •

(A) Suppose $u \succeq_C v$. Then, there exists an increasing function $f : v(O) \to \Re$ such that $u = f \circ v$. So, for $x, y \in O$, $u(x) - u(y) = f \circ v(x) - f \circ v(y) \geq 0$ if and only if $v(x) \geq v(y)$.

Now we prove the result for $\succeq_A$. Suppose u and v are not ordinally congruent. Then, there exist $y, z \in O$ such that, either $u(z) \geq u(y)$ and $v(z) < v(y)$, or $u(z) < u(y)$ and $v(z) \geq v(y)$. We show that $\neg u \succeq_A v$.

In the first case, as $\delta_z \in A(y, u) \setminus A(y, v)$, we have $\neg u \succeq_A v$. In the second case, Observation 1 implies the existence of $x \in O$ such that $u(x) < u(y)$ and $v(x) > v(y)$. So, $\delta_y \in A(x, u) \setminus A(x, v)$. Therefore, $\neg u \succeq_A v$.

(B) We start with

Observation 2: *If $x, y \in O$ and $u, v \in \mathcal{U}$ are such that $u(x) > u(y)$ and $v(x) < v(y)$, then there exist $r_1, r_2 \in X$ such that $r_1 > 0$, $r_2 > 0$, $x + r_2, y + r_1 \in O$, $u(y + r_1) = u(x)$ and $v(x + r_2) = v(y)$.*

Proof. Consider $x, y \in O$ and $u, v \in \mathcal{U}$ such that $u(x) > u(y)$ and $v(x) < v(y)$. Then, there exists $r \in X_+ := \{z \in X \mid z \geq 0\}$ such that $y + r \in O$ and $y + r \geq x$. As u is increasing and $u(x) > u(y)$, we have $r > 0$ and $u(y + r) \geq u(x) > u(y)$. As $y \in O$, we have $O \cap [y + X_+] \neq \emptyset$. As O and $y + X_+$ are convex, so is $O \cap [y + X_+]$. Thus, $O \cap [y + X_+]$ is connected. As u is continuous, $u(O \cap [y + X_+]) \subset \Re$ is an interval. As $u(y), u(y+r) \in u(O \cap [y+X_+])$, we have $u(x) \in u(O \cap [y+X_+])$. Thus, there exists $r_1 \in X_+$ such that $y + r_1 \in O$ and $u(y + r_1) = u(x)$; as $u(y) < u(x)$, we have $r_1 > 0$. By an analogous argument, there exists $r_2 \in X$ such that $r_2 > 0$, $x + r_2 \in O$ and $v(x + r_2) = v(y)$. •

Suppose $u, v \in \mathcal{U}$ are not ordinally congruent. Then, there exist $y, z \in O$ such that, either $u(z) \geq u(y)$ and $v(z) < v(y)$, or $u(z) < u(y)$ and $v(z) \geq v(y)$. We show that $\neg u \succeq_\pi v$.

Consider the first case. Using Observation 1, there exists $x \in O$ such that $u(x) > u(y)$ and $v(x) < v(y)$. Using Observation 2, there exist $r_1, r_2 \in X$ such that $r_1 > 0$, $r_2 > 0$, $x + r_2, y + r_1 \in O$, $u(y + r_1) = u(x)$ and $v(x + r_2) = v(y)$. As u is increasing, $u(x + r_2) > u(x) > u(y)$. Let $\mu = t\delta_y + (1 - t)\delta_{x+r_2}$, where $t \in (0, 1)$ satisfies $tu(y) + (1 - t)u(x + r_2) = u(x)$. Then, $m_\mu = ty + (1 - t)(x + r_2) \in O$ as $y, x + r_2 \in O$ and O is convex. So, $\mu \in \Delta(O)_0$. By construction, $U(\mu, u) = tu(y) + (1 - t)u(x + r_2) = u(x) = u(y + r_1)$. Thus, $m_\mu - y - r_1 \in \pi(\mu, u)$. As $U(\mu, v) = tv(y) + (1 - t)v(x + r_2) = v(y)$, $m_\mu - y \in \pi(\mu, v)$. As $m_\mu - y > m_\mu - y - r_1$, we have $\neg u \succeq_\pi v$.

The arguments for the second case copy the arguments for the first case with appropriate changes. ■

Proof of Lemma 4.4 (A) As u and v are ordinally congruent, u is constant over $v^{-1}(\{r\})$ for every $r \in v(O)$. Therefore, we define $f : v(O) \to \Re$ by $f(.) = u \circ v^{-1}(\{.\})$. Then, $f \circ v(.) = u \circ v^{-1}(\{v(.)\}) = u(.)$.

If $g : v(O) \to \Re$ is such that $u = g \circ v$, then $g(.) = g \circ v \circ v^{-1}(\{.\}) = u \circ v^{-1}(\{.\}) = f \circ v \circ v^{-1}(\{.\}) = f(.)$ on $v(O)$. Thus, f is unique.

Consider $r, s \in v(O)$ such that $r > s$. Then, there exist $x, y \in O$ such that $v(x) = r > s = v(y)$. As u and v are ordinally congruent, we have $f(r) = f \circ v(x) = u(x) > u(y) = f \circ v(y) = f(s)$. So, f is increasing.

It remains to show that f is twice differentiable. Fix $x \in O$ and $b \in B$.

As O is open, there exists $t > 0$ such that $x - tb, x + tb \in O$. As $t > 0$ and $b > 0$, we have $e := tb > 0$. Let $E = \{y \in O \mid x - e < y < x + e\}$. As O is convex, $x - e + re = (1 - r/2)(x - e) + (x + e)r/2 \in O$ for every $r \in (0, 2)$. As E is convex, it is connected. As v is continuous, $v(E)$ is connected.

Define $w : (0, 2) \to \Re$ by $w(r) = v(x - e + re)$. As $e > 0$ and v is increasing, w is increasing. As v is twice differentiable, so is w. So, $Dw > 0$ and it is easily verified that $w((0, 2)) = v(E) = (v(x - e), v(x + e))$.

It follows that w has the increasing function inverse $h : v(E) \to (0, 2)$. By the inverse function theorem, h is differentiable. So, $Dh(y) = 1/Dw \circ h(y) = \psi \circ Dw \circ h(y)$ for $y \in v(E)$, where $\psi(z) := 1/z$ for $z > 0$. As $Dw > 0$ and ψ, Dw and h are differentiable, h is twice differentiable.

If $r \in v(E)$, then $h(r) \in (0, 2)$. So, $x - e + h(r)e \in O$. Define $\phi : v(E) \to \Re$ by $\phi(r) = u(x - e + h(r)e)$. As u and h are twice differentiable, so is ϕ.

Consider $y \in E$. Then, $v(y) \in v(E)$ and $\phi \circ v(y) = u(x - e + h \circ v(y)e)$. As $w((0, 2)) = v(E)$, there exists $k \in (0, 2)$ such that $v(y) = w(k) = v(x - e + ke)$. Consequently, $\phi \circ v(y) = u(x - e + h \circ w(k)e) = u(x - e + ke) = u(y)$, as u and v are ordinally congruent and $v(y) = v(x - e + ke)$. Thus, f coincides with ϕ on $v(E)$. As ϕ is twice differentiable, so is f.

(B) By the chain rule, $Du(.) = Df(v(.)) \circ Dv(.)$ on O, i.e., $\langle Du(x), y\rangle = Df(v(x))\langle Dv(x), y\rangle = \langle Df(v(x))Dv(x), y\rangle$ for all $x \in O$ and $y \in X$. Thus,

$$Du = Df(v)Dv \tag{11}$$

So, Lemma 4.1 implies $Df > 0$. By an analogous argument,

$$D[Df(v)] = D[Df \circ v] = D^2 f(v)Dv \tag{12}$$

Using the product rule to differentiate (11), and using (12), we have

$$\begin{aligned} D^2u(x)y &= \langle D[Df(v(x))], y\rangle Dv(x) + Df(v(x))D^2v(x)y \\ &= \langle D^2f(v(x))Dv(x), y\rangle Dv(x) + Df(v(x))D^2v(x)y \\ &= D^2f(v(x))\langle Dv(x), y\rangle Dv(x) + Df(v(x))D^2v(x)y \end{aligned}$$

for all $x \in O$ and $y \in X$. By Lemma 4.1, $\|Du(x)\| > 0$ and $\|Dv(x)\| > 0$. Set $y = Du(x)$, divide by $\|Du(x)\|^2$ and use (11) to get the result. ■

Proof of Theorem 4.5 Since $\succeq_C = \succeq_A = \succeq_\pi$ (Shah [19], Theorem 4.5), it suffices to show that $\succeq_G = \succeq_C$.[11] Consider $u, v \in \mathcal{U}$.

Suppose $u \succeq_C v$. Then, there exists an increasing concave function $f : v(O) \to \Re$ such that $u = f \circ v$. As f is increasing, u and v are ordinally

[11] The cited result holds for significantly more general settings than Hilbert settings and for much larger sets of admissible utilities.

congruent. By Lemma 4.4, $Dv > 0$, f is twice differentiable and $Df > 0$. As f is concave, $D^2 f \leq 0$. Lemma 4.4 implies $a_v \leq a_u$ on O. So, $u \succeq_G v$.

Suppose $u \succeq_G v$. Then, u and v are ordinally congruent and $a_v \leq a_u$ on O. By Lemma 4.4, $Du > 0$, $Dv > 0$ and there is an increasing and twice differentiable function $f : v(O) \to \Re$ such that $u = f \circ v$ and $Df > 0$. Lemma 4.4 implies that $D^2 f \leq 0$. Thus, f is concave. So, $u \succeq_C v$. ■

Proof of Theorem 5.2 (A) Suppose $\hat{y}(\Omega) \subset O$ and $\hat{y}$ is measurable. For every $t \in T$, as $e(t,.)$ is continuous, it is measurable. The result follows as $m_{\mathcal{Y}}(.) = \int_\Omega P(d\omega)\, y(.,\omega) = \int_\Omega P(d\omega)\, e(.,\hat{y}(\omega)) = \int_O P \circ \hat{y}^{-1}(dx)\, e(.,x) = e(., m_{P\circ\hat{y}^{-1}}) = m_{P\circ\hat{y}^{-1}}(.)$.

(B) For every $t \in T$, as $e(t,.)$ is continuous, it is measurable. Therefore, $\mathcal{E}(P)$ is a process. The other claims follow as $\hat{e}(x)(.) = e(.,x) = x(.)$. ■

Proof of Theorem 5.3 (A) We first show that $\Delta(X)_2 \subset \Delta(X)_0$. Let $\mu \in \Delta(X)_2$. As $\int_X \mu(dx)\, \|x\| \leq \int_X \mu(dx)\, (\|x\|^2 + 1_X(x)) < \infty$, application of the Jensen and Cauchy-Schwarz inequalities yields $|\int_X \mu(dx)\, \langle h, x\rangle| \leq \int_X \mu(dx)\, |\langle h, x\rangle| \leq \|h\| \int_X \mu(dx)\, \|x\| < \infty$ for every $h \in X$. It follows that $\int_X \mu(dx)\, \langle ., x\rangle$ is a continuous real-valued linear functional on X. The Riesz representation theorem (Dunford and Schwarz [6], Theorem IV.4.5) implies that there is a unique $m_\mu \in X$ such that $e(t, m_\mu) = \langle k_t, m_\mu\rangle = \int_X \mu(dx)\, \langle k_t, x\rangle = \int_X \mu(dx)\, e(t,x)$ for every $t \in T$, where $k_t \in X$ is the reproducing kernel for t. Thus, $\mu \in \Delta(X)_0$.

Next, we show that $\mathcal{P}(X)_2 \subset \mathcal{P}(X)$. Let $\mathcal{Y} = (\Omega, \mathcal{F}, P; T, y) \in \mathcal{P}(X)_2$. Then, $\hat{y}(\Omega) \subset X$, $\hat{y}$ is measurable and $P \circ \hat{y}^{-1} \in \Delta(X)_2 \subset \Delta(X)_0$. Consequently, $\mathcal{Y} \in \mathcal{P}(X)$.

Finally, as $(X, \langle .,. \rangle)$ is an RKH space, $|e(t,h)| = |\langle k_t, h\rangle| \leq \|k_t\| \|h\|$ for all $t \in T$ and $h \in X$. It follows that $|y(t,\omega)| = |e(t, \hat{y}(\omega)| \leq \|k_t\| \|\hat{y}(\omega)\|$, and therefore, $y(t,\omega)^2 \leq \|k_t\|^2 \|\hat{y}(\omega)\|^2$ for all $t \in T$ and $\omega \in \Omega$. Consequently, $\int_\Omega P(d\omega)\, y(t,\omega)^2 \leq \|k_t\|^2 \int_\Omega P(d\omega)\, \|\hat{y}(\omega)\|^2 = \|k_t\|^2 \int_X P \circ \hat{y}^{-1}(dx)\, \|x\|^2 < \infty$ for every $t \in T$. Thus, $\mathcal{Y}$ is a second-order process.

(B) is routinely verified as $\hat{e}$ is the identity mapping on X. ■

Proof of Theorem 6.1 Let $a \in X$. As $X = X_n \oplus X_n^\perp$, where $X_n^\perp$ is the orthogonal complement of X_n, a has a unique representation $a = b + b^\perp$, where $b \in X_n$ and $b^\perp \in X_n^\perp$. As X is an RKH space and $k_{t_j} \in X_n$, $a(t_j) = \langle a, k_{t_j}\rangle_X = \langle b + b^\perp, k_{t_j}\rangle_X = \langle b, k_{t_j}\rangle_X = b(t_j)$ for $j = 1, \ldots, n$. Thus, $\Gamma(D; b(t_1), \ldots, b(t_n)) = \Gamma(D; a(t_1), \ldots, a(t_n))$. Also, $\|a\|_X^2 = \langle b + b^\perp, b + b^\perp\rangle_X = \langle b, b\rangle_X + \langle b^\perp, b^\perp\rangle_X \geq \|b\|_X^2$. Consequently, $\Lambda(\|a\|_X^2) \geq \Lambda(\|b\|_X^2)$. ■

Proof of Theorem 7.2 Suppose assumptions (a) and (b) are satisfied. Lemmas 4.1 and 4.4 imply that $Du(w\beta) > 0$, $Dv(w\beta) > 0$, and there is a unique function $f : v(O) \to \Re$ such that $u = f \circ v$. Moreover, $Df > 0$ and f is twice differentiable.

As $Dv(w\beta) \in X$, we have $\langle Dv(w\beta), .\rangle \in X^*$. If $\alpha_v = 0$, then equation (7) implies $\langle Dv(w\beta), m_\mu\rangle = \int_O \mu(dy)\, \langle Dv(w\beta), y\rangle = \int_O \mu(dy)\, D_1\hat{v}(0,y) \leq 0$, which contradicts assumption (a). So, $\alpha_v > 0$.

Using equations (7) and (11), if $\alpha_u = 0$, then $0 \geq \int_O \mu(dy)\, D_1\hat{u}(0, y) = \int_O \mu(dy)\, \langle Du(w\beta), y\rangle = \langle Du(w\beta), m_\mu\rangle = Df \circ v(w\beta)\langle Dv(w\beta), m_\mu\rangle$. As $Df > 0$, we have $\langle Dv(w\beta), m_\mu\rangle \leq 0$, which contradicts assumption (a). So, $\alpha_u > 0$.

So, equation (7) implies $\int_O \mu(dy)\, D_1\hat{u}(\alpha_u, y) = 0 = \int_O \mu(dy)\, D_1\hat{v}(\alpha_v, y)$.

As $a_u > a_v$, Lemma 4.4 implies $D^2 f < 0$.

Now suppose assumption (c) is also satisfied.

Consider $y \in O$ such that $D_1\hat{v}(\alpha_v, y) > 0$. By the directional monotonicity of v, $v(w\beta + \alpha_v y) - v(w\beta) = \hat{v}(\alpha_v, y) > 0$. As $D^2 f < 0$, we have $Df \circ v(w\beta + \alpha_v y) < Df \circ v(w\beta)$, and therefore,

$$Df \circ v(w\beta + \alpha_v y) D_1\hat{v}(\alpha_v, y) < Df \circ v(w\beta) D_1\hat{v}(\alpha_v, y) \tag{13}$$

Consider $y \in O$ such that $D_1\hat{v}(\alpha_v, y) < 0$. By the directional monotonicity of v, $v(w\beta + \alpha_v y) - v(w\beta) = \hat{v}(\alpha_v, y) < 0$. As $D^2 f < 0$, we have $Df \circ v(w\beta + \alpha_v y) > Df \circ v(w\beta)$. Therefore, inequality (13) holds.

So, inequality (13) holds for every $y \in O$ such that $D_1\hat{v}(\alpha_v, y) \neq 0$.

Since $\hat{u}(\alpha, y) = u(w\beta + \alpha y) - u(w\beta) = f \circ v(w\beta + \alpha y) - u(w\beta) = f(\hat{v}(\alpha, y) + v(w\beta)) - u(w\beta)$, we have

$$D_1\hat{u}(\alpha, y) = Df \circ v(w\beta + \alpha y) D_1\hat{v}(\alpha, y) \tag{14}$$

Equation (14), inequality (13) and assumption (c) imply that

$$\begin{aligned}
\int_O \mu(dy)\, D_1\hat{u}(\alpha_v, y) &= \int_O \mu(dy)\, Df \circ v(w\beta + \alpha_v y) D_1\hat{v}(\alpha_v, y) \\
&< Df \circ v(w\beta) \int_O \mu(dy)\, D_1\hat{v}(\alpha_v, y) \\
&= 0
\end{aligned}$$

Consider $y \in X$. As u is concave, $\hat{u}(., y)$ is concave. If $\alpha_u \geq \alpha_v$, then $D_1\hat{u}(\alpha_u, y) \leq D_1\hat{u}(\alpha_v, y)$. It follows that

$$0 = \int_O \mu(dy)\, D_1\hat{u}(\alpha_u, y) \leq \int_O \mu(dy)\, D_1\hat{u}(\alpha_v, y) < 0$$

This contradiction implies $\alpha_u < \alpha_v$. ■

Proof of Theorem 8.3 We show the first result. The second one is a routine verification.

Consider $a \in \hat{\mathcal{R}}^1$. Then, there exists a differentiable function $f : O \to \Re$ with $Df = a \geq 0$. Define $u : O \to \Re$ by $u(x) = \int_0^x dy\, e^{-f(y)}$. It follows that u is twice differentiable. Furthermore, $Du = e^{-f} > 0$ and $D^2 u = -aDu \leq 0$. Thus, $u \in \hat{\mathcal{U}}^1$. As $\chi[u] = -D \ln Du = Df = a$, χ is surjective.

Suppose $u, v \in \hat{\mathcal{U}}^1$ are such that $\chi u = \chi v$. Then, $D \ln Du = D \ln Dv$, which implies $\ln Du = \ln bDv$ for some $b > 0$. Thus, $Du = bDv$. It follows that $u = bv + c$ for some $c \in \Re$. Thus, $u \in [v]$ and χ is injective. ■

Proof of Theorem 8.6 Suppose $[u]$ is a semiconcave preference. Then, u is semiconcave. So, there exists $C \geq 0$ such that $v(.) := u(.) - C\|.\|^2/2$ is concave and $[v]$ is risk averse. If $C = 0$, then $[u]$ itself is risk averse and $[u] \approx [u]$.

Suppose $C > 0$. Consider a compact set $O' \subset O$ and $\epsilon > 0$. If $O' = \emptyset$, then $\sup\{|u(x) - v(x)| \mid x \in O'\} = \sup \emptyset = -\infty < \epsilon$. Suppose $O' \neq \emptyset$. As $\|.\|$ is continuous, $\sup\{\|x\|^2/2 \mid x \in O'\} \in [0, \infty)$. Consequently, there exists $b > 0$ such that $\sup\{|bu(x) - bv(x)| \mid x \in O'\} = bC \sup\{\|x\|^2/2 \mid x \in O'\} < \epsilon$. So, $[v] \approx [u]$. ■

Proof of Lemma 8.9 (A) Consider u and v *ex hypothesi*. Then, $-u$ and $-v$ are semiconvex and satisfy the conditions $\lim_{\|x\| \to \infty} -u(x) = -\infty = \lim_{\|x\| \to \infty} -v(x)$ and $\| - Du(.)\| = e^{-f(.)} = \| - Dv(.)\|$ a.e. Since conditions C1-C6 of Kružkov [11] hold, Theorem 2.1 in Kružkov [11] implies that $u = v$.

(B) As conditions R1-R6 of Kružkov [11] are satisfied, Theorem 5.1 in Kružkov [11] implies the existence of a semiconvex function $v : O \to \Re$ such that $\|Dv(.)\| = e^{-f(.)}$ a.e. and $\lim_{\|x\| \to \infty} v(x) = -\infty$. Then, $u = -v$ is a generalized solution of (8). ■

Proof of Theorem 8.10 Consider $[u] \in \mathcal{U}^n/\equiv$. Since $u \in \mathcal{U}^n$, $\|Du(.)\| = e^{-f(.)}$ a.e. for some $f \in \mathcal{F}$. It follows that $\Gamma u = Df$ a.e. Therefore, $\Gamma u \in \mathcal{R}^n$. Since Γ is invariant on $[u]$, Γ maps $\mathcal{U}^n/\equiv$ to $\mathcal{R}^n$.

Consider $a \in \mathcal{R}^n$. Then, $a \simeq Df$ for some $f \in \mathcal{F}$. Given f, Lemma 8.9 implies that (8) has a generalized solution $u \in \mathcal{U}^n$. So, $\Gamma u \simeq Df$. As $\simeq$ is transitive, $\Gamma u \simeq a$. Since Γ is invariant on $[u]$, $\Gamma[u] \simeq a$.

Consider $a \in \mathcal{R}^n$ and $[u], [v] \in \mathcal{U}^n/\equiv$ such that $\Gamma[u] \simeq a$ and $\Gamma[v] \simeq a$. As $\simeq$ is transitive, $\Gamma[u] \simeq \Gamma[v]$. So, $\Gamma u \simeq \Gamma v$. As $u, v \in \mathcal{U}^n$, there exist $f, g \in \mathcal{F}$ such that $\|Du(.)\| = e^{-f(.)}$ a.e. and $\|Dv(.)\| = e^{-g(.)}$ a.e. It follows that $Df \simeq \Gamma u$ and $\Gamma v \simeq Dg$. As $\Gamma u \simeq \Gamma v$, transitivity of $\simeq$ implies $Df \simeq Dg$.

So, $\mathrm{Leb}(E) = 0$ for $E = \{x \in O \mid D(f-g)(x) \neq 0\}$. Consider $x \in O$. Let $B_{1/n}(x) = \{y \in O \mid \|x - y\| < 1/n\}$ for $n \in \mathcal{N}$. As $\mathrm{Leb}(B_{1/n}(x)) > 0$, there exists $x_n \in B_{1/n}(x)$ such that $D(f - g)(x_n) = 0$; otherwise, $B_{1/n}(x) \subset E$ and $\mathrm{Leb}(E) \geq \mathrm{Leb}(B_{1/n}(x)) > 0$, a contradiction. As $f, g \in \mathcal{C}^{2,\alpha}$, $D(f - g)$ is continuous. By construction, $\lim_n x_n = x$. Therefore, $D(f - g)(x) = D(f - g)(\lim_n x_n) = \lim_n D(f - g)(x_n) = 0$. So, $D(f - g) = 0$ on O.

Fix $y \in O$. Let $(f - g)(y) = c$ and $h(.) = (f - g)(.) - c$ on O. Then, $h(y) = 0$ and $Dh = D(f - g) = 0$ on O. Consider $x \in O$. By the mean value theorem, there exists $t \in (0, 1)$ such that $h(x) - h(y) = \langle Dh(tx + (1-t)y), x - y\rangle$. Therefore, $|h(x)| = |h(x) - h(y)| \leq \|Dh(tx + (1 - t)y)\|\|x - y\| = 0$. It follows that $(f - g)(.) = h(.) + c = c$ on O.

So, $\|De^c u(.)\| = e^c\|Du(.)\| = e^{c-f(.)} = e^{-g(.)}$ a.e. Therefore, given g, $e^c u$ and v are generalized solutions of (8). By Lemma 8.9, $e^c u = v$, i.e., $[v] = [u]$. ■

References

[1] C. D. Aliprantis, K. C. Border: *Infinite Dimensional Analysis: A Hitchhiker's Guide, 3rd ed.*, Springer-Verlag (2006)

[2] K. J. Arrow: *Essays in the Theory of Risk-bearing*, North-Holland Publishing Company (1970)

[3] A. Berlinet, C. Thomas-Agnan: *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer Academic Publishers (2004)

[4] P. Cannarsa, C. Sinestrari: *Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control*, Birkhäuser (2004)

[5] G. T. Duncan: A matrix measure of multivariate local risk aversion, *Econometrica* 45, 895-904 (1977)

[6] N. Dunford, J. Schwartz: *Linear Operators, Part I: General Theory*, Wiley Interscience (1988)

[7] M. F. Hellwig: Risk aversion in the small and in the large when outcomes are multidimensional, *Preprints of the Max Planck Institute for Research on Collective Goods* (2004)

[8] E. Karni: On multivariate risk aversion, *Econometrica* 47, 1391-1401 (1979)

[9] R. E. Kihlstrom, L. J. Mirman: Risk aversion with many commodities, *Journal of Economic Theory* 8, 361-368 (1974)

[10] G. S. Kimeldorf, G. Wahba: Some results on Tchebycheffian spline functions, *Journal of Mathematical Analysis and Applications* 33, 82-95 (1971)

[11] S. Kružkov: Generalized solutions of the Hamilton-Jacobi equations of eikonal type I, *Matematicheskii USSR Sbornik* 27, 406-446 (1975)

[12] S. Lang: *Real and Functional Analysis,* Springer-Verlag (1993)

[13] H. Levy, A. Levy: Arrow-Pratt measures of risk aversion: the multivariate case, *International Economic Review* 32, 891-898 (1991)

[14] R. Luneburg: *Mathematical Theory of Optics*, University of California Press (1965)

[15] A. Mas-Colell, M. D. Whinston, J. R. Green: *Microeconomic Theory*, Oxford University Press (1995)

[16] J. Paroush: Risk premium with many commodities, *Journal of Economic Theory* 11, 283-286 (1975)

[17] B. J. Pettis: On integration in vector spaces, *Transactions of the American Mathematical Society* 44, 277-304 (1938)

[18] J. W. Pratt: Risk aversion in the small and in the large, *Econometrica* 32, 122-136 (1964)

[19] S. A. Shah: Who is more averse to random processes?, available at SSRN: `http://ssrn.com/abstract=2729171` (2016)

[20] M. E. Yaari: Some remarks on measures of risk aversion and on their uses, *Journal of Economic Theory* 1, 315-329 (1969)

[21] K. Yosida: *Functional Analysis, 4th ed.*, Springer (1974)

[22] W. Ziemer: *Weakly Differentiable Functions*, Springer-Verlag (1989)