# A Theory of Risk Aversion without the Independence Axiom[1]

Preliminary and Incomplete

January 04, 2004

Hengjie Ai

hai@econ.umn.edu

**Abstract:** I study preferences defined on the set of real valued random variables as a model of economic behavior under uncertainty. It is well-known that under the Independence Axiom, the utility functional has an expected utility representation. However, the Independence Aiom is often found contradictory to empirical evidences. The purpose of this paper is to study risk averse utility functionals without assuming the Independence Axiom.

The first part of the paper studies the relation between convexity of preference and risk aversion. If the utility functional does not have an expected utility representation, then the equivalence between risk aversion and convexity of preference breaks down. This paper shows that under appropriate continuity conditions, risk aversion is equivalent to a weaker condition, which I call "equal-distribution convexity", that is a preference is risk averse iff convex combinations of random variables with the same distribution are preferred to the random variables themselves.

Differential properties of risk averse utility functionals are studied. A representation theorem for the form of the Frechet derivative of continuously differentiable utility functionals is given. Characterization of monotonicity and risk aversion in terms of the Frechet derivative of utility functionals are given. Neccessary and sufficient condition of risk aversion in terms of the Frechet derivative of utility functionals is provided. I also provide a criteria of comparing individual's attitude toward risk by the properties of the Frechet differential of the utility functions. This criteria, when applied to expected utility, reduces to the usual Arrow-Platt measure of absolute risk aversion. I also studied the relation between the notion of differentiability developed here and that in Machina (1982).

[1] I thank my adviser Jan Werner for his guidance. I thank William Neilson, Andy Mclennan, Marcel K. Richter, Fanchang Huang, Yuzhe Zhang participants of Midwest Economic Theory Meeting 2004, and participants of Minnesota Micro/Finance workshop for their helpful comments. All errors are mine.

# A Theory of Risk Aversion without the Independence Axiom

The purpose of this paper is to study risk aversion without assuming the independence axiom, consequently the utility function may not have an expected utility representation. I restrict my attention to preference orders that satify a state-independence condition, that is random variables of the same distribution are indifferent to each other. The approach taken in this paper differs from most of the literature in that I focus on preference orders and utility functions that defined on space of random variables, in stead of probability distributions. The advantage of doing so includes the following: First, I give a simple characterization of strongly risk aversion, which is stated as a property of preference orderings defined on set of random variables. I will also give simple characterizations of risk averse utility functions in terms of its derivatives with respect to random variables. These characterizations, however, are not possible when preference is defined on set of probability distributions. Second, the differential properties of utility function that is derived in this paper is more convenient for many application purposes. For example in studying consumer's portfolio choice problem in the financial market, One needs the derivative of utility functions with respect to random variables to derive the first order condition of consumer's portfolio choice problem. Previous literatures have studied differential properties of utility functions defined on space of distribution functions, for example, Machina ([12]), Chew, Karni and Safra ([4]). As will be shown in the last part of the paper, their differentiability conditions can be used to derive the derivative of utility functions with repect to random variables as well. However, In order to use Machina ([12])'s notion of derivative, one needs assume that the random variable has a compact support, while the results of this paper does not rely on this assumption. Therefore, for purpose of application in asset pricing problem, the results of this paper allows one to keep the important tools developed in those asset pricing models where securiy prices follow diffusion processes, which is clearly unbounded.The independence Axiom, on which the expected utility theory relies, is often found contradictory to empirical evidence. For a comprehensive survey on this, see [25]. Many weakenings and generalizations of the independence Axiom have been proposed as an alternative theoretical foundation of study of economic behavior under unceratainty, for example, the weighted utility theory [4], the anticipated utility theory [16],[17], the dual expected utility theory [30], among many others.

This paper studies properties of risk averse utility functionals[2] without assuming the Independence Axiom, with and without differentiability assumption of the utility functional. Preferences are assumed to satisfy the "state independence" assumption, that is, two random variables are indifferent to each other

[2]Hereafter called strong risk aversion, to distinguish it from the following weaker notion of risk aversion: A preference relation is risk averse if any random variable is less prefered to its expectation.

if they have the same distribution. Under this assumption, preference over set of random variables can be modelled either by utility functions defined on set of random varibles, denoted $V(X)$, or as utility function defined on probability distribution functions, denoted $U(F)$ , where $V$ and $U$ are related by $U(F) = V(X)$ if the distribution function of $X$ is $F$. Most of the literature took the second approach, i.e. studying the properties of the function $U$.

This paper takes a different point of view, i.e. studying utility functions defined on set of random variables directly. The most important reason for this is that in many applications, it is the properties of utility function as a function of random variable that is of direct interest. To see this point, consider the a model consumer's portfolio choice problem:

$$\begin{array}{rl} \max & V(c) \\ \text{s.t.} & \sum_{j=1}^{J} p_j h_j \leq m \\ c_s = & \sum_{j=1}^{J} h_j x_{j,s} \quad \forall s \end{array}$$

Consumer's preference over a random consumption $c$ is represented by utility function $V(\cdot)$. $\{x_{j,s}\}_{j=1,2,\cdots J, s=1,2,\cdots S}$ is the pay-off matrix of the $J$ securities in the market. The first order condition of the above problem is

$$\lambda p_j = \sum_s \frac{\partial V}{\partial c_s}(c) \cdot x_{j,s} \tag{1}$$

, which is the key equation underlying many asset pricing models. Note in order to write down the first order condition, one need to know the derivative of $V$ with respect to the random variable $c$. It may well be argued that the properties of the uitlity function $U$ can be used to derive the property of $V$, thus equation (1) can be derived by using the properties of utlity function $U$. This approach is taken by Machina (1982). Machina (1982) studies differential properies of smooth uitlity functionals defined on set of probability distribution functions. Machina (1982) showed that although linearity is lost if one abandon the independence axiom, smooth nonlinear (in the probability distribution) utility functionals can be locally approximated by a linear functional. Under appropriate assumptions, the Frechet derivative of a smooth utlity functions, as a linear functional that maps probability distribution functions to the real line, has an expected utility representation. That is to say, although the utility functional may not have an expected utility representation, it can be locally approximated by expected utility functions. Machina's results can be applied to study asset pricing implications of non-expected utility functions. However, this approach does not always work. In particular, in establishing the existence of the locally expected utility, one need the set of probability distribution under consideration to have a compact support to invoke the Riesz theorem. As will be shown, the approach developed here does not rely on this assumption. Considering the importance of asset pricing models that rely on asset prices driven by diffusion processes, which is clearly unbounded, one implication of my result is that applying non-expected utility analysis to asset pricing theory does not neccessarily means one has to abandon the powerful tool that is developed in continuous time asset pricing models. In the last part of the paper, I study the relationship between the differentiability of $U$ as a function of probability distribution function, and the differentiability of $V$ as a function of random variables. The general conclusion is the some the two notions of differentiability translate into each other, some time not. For applications in which the differentiability of $V$ is of direct interest, it is desirable to study the properties of $V$ function directly.

A second reason to study utility functions defined on set of random variables is that this alternative view leads to a deeper understanding of the notion of risk aversion. Although the notion of *monotonicity* and *risk aversion* of preference is defined as properties of preference over set of probability distribution functions, it can be defined as properties of preference over set of random variables as well, without any reference to the distributional properties of the random variables. This second view leads to a deeper understanding of the relation between risk aversion and convexity of preference.

# 1 Introduction

This paper studies preferences on the set of random variables that can be represented by some differentiable utility functional. Fomally, let $(\Omega, \mathcal{F}, P)$ be a nonatomic probability space. We further assume $\Omega$ is a Polish space (complete metric space with a countable dense subset), and let $\mathcal{F}$ be the Borel $\sigma$ field. Therefore $(\Omega, \mathcal{F}, P)$ is a standard Borel space[3]. Let $([0,1], \mathcal{B}, m)$ denote the unit interval equipped with the Borel $\sigma$ field, and the Lebesgue measure. Since any standard Borel space is isomorphic to $([0,1], \mathcal{B}, m)$[4], without loss of generality, we will let $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}, m)$ whenever this specification simplifies exposition. Preferences are defined on $L^1(\Omega, \mathcal{F}, P)$, the vector space of $(R, \mathcal{B})-$valued integrable random variables equipped with the $L^1$ norm, where $\mathcal{B}$ is the Borel $\sigma$ field on $R$. For any $X \in L^1$, let $F_X(\cdot)$ be the distribution function of $X$, that is $F_X : R \to [0,1]$, $\forall x \in R$, $F_X(x) = P(\{\omega : X(\omega) \leq x\})$. The following stochastic order relations can be defined on $L^1$:

**Definition 1 (First Order Stochastic Dominance)** *Let $X, Y \in L^1$,then the distribution of $X$ first order stochastically dominates that of $Y$, denoted $X \succsim_{FSD} Y$, if $\forall x \in R$, $F_X(x) \leq F_Y(x)$. If in addition, $F_X(\widehat{x}) < F_Y(\widehat{x})$ holds for some $\widehat{x} \in R$, then $X$ is said to strictly first order stochastic dominate $Y$, denote $X \succ_{FSD} Y$.*

**Definition 2 (Second Order Stochastic Dominance)** *Let $X, Y \in L^1$, then the distribution of $X$ second order stochastically dominates that of $Y$, denoted $X \succsim_{SSD} Y$, if $EX = EY$, and $\int_{-\infty}^{x} F_X(t)dt \leq \int_{-\infty}^{x} F_Y(t)dt \ \ \forall x \in R$.*

[3] See [9], Page 74

[4] That is, $\exists T : (\Omega, \mathcal{F}, P) \to ([0,1], \mathcal{B}, m)$ such that 1) $T$ is one-to-one and onto; 2) Both $T$ and $T^{-1}$ are measurable fuctions. 3) $T$ is measure preserving. This result is essentially due to [21], see also the Isomorphism Theorem in [9](Theorem 17.41, Page 116).

Let's note the following facts about stochastic order relations:

**Lemma 3** *1) $\forall X, Y \in L^1$, $X \succsim_{FSD} Y$ iff $\exists \tilde{X}, \tilde{Y} \in L^1$,such that $X \approx_d \tilde{X}$ and $Y \approx_d \tilde{Y}$[5], and $\tilde{X} \geq \tilde{Y}$ a.s.*

*2) $\forall X, Y \in L^1$, $X \succsim_{SSD} Y$ iff $\exists \tilde{X}, \tilde{Y} \in L^1$,such that $X \approx_d \tilde{X}$ and $Y \approx_d \tilde{Y}$, and $E[\tilde{Y}|\tilde{X}] = \tilde{X}$ a.s., and the conditional distribution $Q_{Y|X=x}$ is stochastically increasing in $x$.*

In the above lemma, the existence of the random varibles $\tilde{X}, \tilde{Y}$ defined on some probability space, is well known. We show in Appendix I that such random variables could be constructed on $L^1$. It also follows from the lemma that if $X \succ_{FSD} Y$, then $\tilde{X} > \tilde{Y}$ holds on a set of positive measure. The notion of simple mean preserving spread (MPS) (Roschild and Stiglitz [22]) is very useful in studying second order stochastic dominance. Let us repeat the key theorems for later reference:

**Definition 4 (Single Crossing Property)** *Let $X, Y \in L^1$, let $F_X$, $F_Y$ be the distribution functions of $X$ and $Y$, respectively. Then $F_X$ single cross $Y$ from below if $\exists \overline{x} \in R$ such that $\forall x \leq \overline{x}$, $F_X(x) \leq F_Y(x)$, and $\forall x \geq \overline{x}$, $F_X(x) \geq F_Y(x)$.*

**Definition 5 (Differ by a simple MPS)** *Let $X, Y \in L^1$, let $F_X$, $F_Y$ be the distribution functions of $X$ and $Y$, respectively. Then $F_Y$ is said to differ from $F_X$ by a simple MPS if $F_X$ single cross $Y$ from below, and in additon $EX = EY$.*

**Proposition 6** [6]*Let $X, Y \in L^1$, then $X \succsim_{SSD} Y$ iff $\exists \{X_n\}_{n=1}^{\infty}$ such that $X \approx_d X_1$, and for each $n = 1, 2, \cdots$, the distribution of $X_{n+1}$ differ from that of $X_n$ by a simple MPS, and $X_n \Longrightarrow Y$.*

We are interested in studying preference ordering on $L^1$, Formally, let $\succcurlyeq$ be a complete preorder on $L^1$, and let $\succ$ and $\sim$ be the assymmetric part and the symmetric part of it, respectively. We will make frequent use of the following assumptions:

*A*1 State Independence: $\forall X, Y \in L^1$, $X \approx_d Y$ implies $X \sim Y$

*A*2 Monotonicity: $\forall X, Y \in L^1$, $X \geq Y$ *a.s.* implies $X \succcurlyeq Y$.

*A*2′ Strict Monotonicity: $\forall X, Y \in L^1$, $X \geq Y$ *a.s.*, and $X > Y$ on a set of strictly positive measure implies $X \succ Y$.

*A*3 Strong Risk Aversion: $\forall X, Y \in L^1$, $E[Y|X] = X$ *a.s.* implies $X \succcurlyeq Y$.

*A*3′ Weak Risk Aversion: $\forall X \in L^1$, $E(X) \succcurlyeq X$.

*A*4 Convexity: If $X \succcurlyeq Y$, then $\lambda X + (1-\lambda)Y \succcurlyeq Y$, $\forall \lambda \in [0,1]$.

*A*4′ Equal-distribution Convexity: See definition below.

[5]Here and after $\approx_d$ is used to denote that two random variables have the same distribution.
[6]See ([22]), see also ([14]), theorem 1.5.19.

$A5$ Continuity: If $X_n \succcurlyeq Y_n$ $\forall n$, and $X_n \to X$ in $L^1$, $Y_n \to Y$ in $L^1$, then $X \succcurlyeq Y$.

**Definition 7 (Equal-distribution Convexity)** *Let $X \in L^1$, let $E_X = \{Y \in L^1 : Y \approx_d X\}$, then $\succcurlyeq$ is said to be equal-distribution convex with respect to $E_X$ if $\forall Y_1, Y_2 \in E_X$, $\forall \lambda \in [0,1]$, $\lambda Y_1 + (1-\lambda) Y_2 \succcurlyeq X$. $\succcurlyeq$ is said to satisfy the equal-distribution convexity property $(A4')$ if it is equal-distribution convex with respect to $E_X$ for every $X \in L^1$.*

Note under the state independence assumption $(A1)$, $\forall Y \in E_X$, $Y \sim X$, equal-distribution convexity means convex combinations of elements in $E_X$ are weakly preferred to $X$. Therefore under state independence $(A1)$, convexity implies equal distribution convexity. Note by lemma 3, under $A1$, the above definition of monotonicity $(A2)$ is equivalent to "$X \succsim_{FSD} Y$ implies $X \succcurlyeq Y$", and strong risk aversion $(A3)$ is equivalent to $X \succsim_{SSD} Y$ implies $X \succcurlyeq Y$".

The rest of the paper is organized as follows, in the second part, we study the relationship between strong strong risk aversion $(A3)$ and convexity $(A4)$. We argue that under $(A1)$, and $(A5)$, strong risk aversion is equivalent to equal-distribution convexity $(A4')$. It then follows immediately that under $(A1)$, and $(A5)$, convexity implies strong risk aversion. We give an example to show that the converse is not true. The rest of the paper focus on the differential properties of the utility function representation. The third part gives neccessary and sufficient conditions of the state independence property$(A1)$, and shows that under condition $(A1)$, the Frechet derivative of the utility function can be represented by a real valued Borel measurable function. We maintain the assumptions $(A1)$ and explore the properties of the this representation function. Part four provides neccessary and sufficient conditons of monotonicity $(A2)$ and strong risk aversion$(A3)$. We also present a second characterization of strong risk aversion that directly relates to the equal-distribution convexity property. Part five compares individual's attitudes toward risk. Part six compares the notion of differentiability developed here and that in Machina (1982), part seven concludes.

# 2 Strong Risk Aversion and Convexity

It is well known that if $\succcurlyeq$ satisfies the Independence Axiom, then strong risk aversion is equivalent to the cocavity of the von-Neuman Morgenstern utility function. However, what if the Independence Axiom is not assumed? The rest of this section is devoted to this question. We argue that convexity is strictly stronger than strong risk aversion, in fact strong risk aversion can be viewed as "equal-distribution convexity" as we defined above$(A4')$. This also gives a equivalent characterization of strong risk aversion.

We will first need a conditional version of the strong law of large numbers:

**Lemma 8 (Conditional Strong Law of Large Numbers)** *Let $X_1, X_2, \cdots$ be a sequence of random variables that is conditionally i.i.d. give $Y$, where*

*$Y \in L^1$, and $X_i \in L^1$ for each $i$. Suppose also $E[X_1|Y] = Y$ a.s.. Then $\frac{1}{n}\sum_{i=1}^{n} X_i \to Y$ a.s. and in $L^1$.*

**Proof.** See Appendix II. ■

Given any $X \in L^1$, the following lemma provides a neccessary and sufficient condition for $Y \succsim_{SSD} X$ to hold:

**Lemma 9** [7]*$\forall X \in L^1$, the set of distributions that second order stochastically dominate the distribution of $X$ is precisely the set of distributions that can be obtained through convex combinations of random variables that have the same distribution with $X$, or as $L^1$ limit of it*[8].

**Proof.** First, suppose $Y \succsim_{SSD} X$, we need to show the distibution of $Y$ could be obtained as the distribution of convex combinations of random variables that has the same distribution with $X$. Since only the distribution of random variables matter, by lemma 3, we can assume without loss of generality $E[X|Y] = 0$ *a.s.*. Let $R_0$ be the distribution of $Y$, and for each $x \in R$, let $Q_{X|Y=x}(\cdot)$ be the conditional distribution of $X$ given $Y = x$. Then one can construct a sequence of random variables $\widetilde{X}_1, \widetilde{X}_2, \cdots$ , and a random variable $\widetilde{Y}$ on $L^1$ such that the distribution of $\widetilde{Y}$ is $R_0$, $\widetilde{X}_i$'s are conditional i.i.d. given $\widetilde{Y}$, and the conditional distribution of $\widetilde{X}_i$ given $\widetilde{Y}$ is $Q_{X|Y}$. [9] Consider $\frac{1}{n}\sum_{i=1}^{n} \widetilde{X}_i$, since $\widetilde{X}_i$'s all have the same distribution with $\widetilde{X}$, for each $n$, $\frac{1}{n}\sum_{i=1}^{n} \widetilde{X}_i$ is a convex combination of random variables that have the same distribution with $X$. Note $\frac{1}{n}\sum_{i=1}^{n} \widetilde{X}_i \to E[\widetilde{X}_1|\widetilde{Y}] = \widetilde{Y}$ *a.s.* and in $L^1$ by lemma 8, as needed.

Next, need to show the reverse inclusion. Take any $X_1, X_2$, such that $X_1 \approx_d X_2 \approx_d X$, need to show $\forall \lambda \in [0,1]$, $\lambda X_1 + (1-\lambda)X_2 \succsim_{SSD} X$. It is enough to show for any concave function $u$ such that the expectations of $u\ (\lambda X_1 + (1-\lambda)X_2)$ and $u\ (X)$ both exist, then $\int u\ (\lambda X_1 + (1-\lambda)X_2)dP \geq \int u\ (X)dP$. This is

[7] A similiar result in [11](theorem 15.6, page 137) implies that $\forall X \in L^p$, the set of random variables that second order stochastic dominate $X$ is the $\sigma(L^p, L^q)-$closed convex hull of all random variables $Y$ satisfying $Y \approx_d X$. The probabilistic proof we give here, however, is completely different from that in [11]. The reader will also see that Luxemberg's theorem is not enough to prove proposition 10, in fact it needs a stronger continuity assumption than $A6$, namely, continuity in the $\sigma(L^p, L^q)$ topology.

[8] The proof can be easily adapted to show that the statement is also true if one replace $L^1$ limit in the lemma with *a.s.* limit.

[9] Let $R_0$ be the distribution of $Y$, and for each $x \in R$, let $R_1(x, \cdot) = (Q_{X|Y=x}(\cdot))^\infty$ be a conditional probability distribution. Then it follows from lemma 1 on page 430 in Fristedt and Gray (1997), that there exist random variables $\widehat{X}$, and $\widehat{Y}$ defined on some probability space, such that the distribution of $\widehat{Y}$ is $R_0$, the conditional distribution of $\widehat{X}$ given $\sigma(\widehat{Y})$ is $R_1$. Note $\widehat{X} = (\widehat{X}_1, \widehat{X}_2, \cdots)$ is a sequence of conditional i.i.d. random variables. Note also since $(\widehat{Y}, \widehat{X})$ takes values in $(R \times R^\infty, \mathcal{B})$, which is a standard Borel space, it follows from **??** in Appendix I that $\exists$ random variables $(\widetilde{Y}, \widetilde{X})$ define on $(\Omega, F, P)$, such that $(\widetilde{Y}, \widetilde{X})$ and $(\widehat{Y}, \widehat{X})$ have the same distribution. Therefore $\widetilde{X} = (\widetilde{X}_1, \widetilde{X}_2, \cdots)$ is the sequence of random variables that is needed.

immediate, since

$$
\begin{aligned}
& \int u(\lambda X_1 + (1-\lambda)X_2)dP \\
\geq \quad & \lambda \int u(X_1)dP + (1-\lambda)\int u(X_2)dP = \int u(X)dP
\end{aligned}
$$

Therefore it follows that any finite convex combination of random variables with the same distribution second order stochastically dominate $X$. Now suppose $Y$ is the $L^1$ limit of a sequence of random variables $X_n$, where for each $n$, $X_n$ is a convex combination of random variables that have the same distribution with $X$. Need to show $Y \succsim_{SSD} X$. Since $X_n \to Y$ in $L^1$, we have $EY = \lim_{n\to\infty} E(X_n) = E(X)$. Therefore only need to show $\int_{-\infty}^{x} F_Y(t)dt \leq \int_{-\infty}^{x} F_X(t)dt$ for all $x \in R$, where $F_Y$ and $F_X$ are the distribution functions of $Y$ and $X$, respectively. For each $n$, let $F_n$ be the distribution function of $Y_n$. Note convergence in $L^1$ implies convergence in distribution, we have $F_n(t) \to F_Y(t)$, for all continuity ponits of $F_Y$. Note also there can be at most countably many ponits of discontinuities, therefore $F_n \to F_Y$ $a.s.$. By Fatau's lemma,$\forall x \in R$,

$$
\int_{-\infty}^{x} F_Y(t)dt \leq \lim_{n\to\infty}[\int_{-\infty}^{x} F_n(t)dt] \leq \int_{-\infty}^{x} F_X(t)dt
$$

the second inequality is true because $\forall n$, $Y_n \succsim_{SSD} X$. This gives the desired result. ■

**Proposition 10** *Let $\succcurlyeq$satisfies $A1$ and $A5$ then $\succcurlyeq$ is strongly risk averse ($A3$) if and if it satisfies the equal-distribution convex property($A4'$).*

**Proof.** First strongly risk averse implies equal-distribution convexity ($A4'$) follows directly from the above lemma. To see the reverse implication, note if $\succcurlyeq$ satisfies the equal-distribution convexity assumption, then take any $X \in L^1$, if $Y \succsim_{SSD} X$, then in the prove of the above lemma, we showed $\exists \widetilde{Y} \in L^1$, $\widetilde{Y} \approx_d Y$, such that $\widetilde{Y}$ could be obtained as the $L^1$ limit of a sequence of random variables that are convex combinations of elements in $E_X$, by the continuity assumption ($A6$), we have $Y \succcurlyeq X$. ■

Note if $V$ is the utility function that represents $\succcurlyeq$, then the equal-distribution convexity property of $\succcurlyeq$ implies that $V$ resembles the property of concave functions on the set $E_X$ for every $X \in L^1$, namely

$$
\forall X_1, X_2 \in E_X,\ V(\lambda X_1 + (1-\lambda)X_2) \geq \lambda V(X_1) + (1-\lambda)V(X_2) \qquad (2)
$$

In fact, the above proposition implies that (2) provides an equivalent definition of strong risk aversion under assumptions ($A1$), and ($A5$). The reader will see a differential version of condition (2) in part four of the paper.

It follows immediately from the above proposition that under $A1$, and $A5$, convexity of $\succcurlyeq$ implies strong risk aversion. Hence we obtain the following Corollary:

**Corollary 11** *Suppose $\succcurlyeq$satisfies A1, and A5, then convexity (A4) implies strong risk aversion (A3).*

The following example shows that the converse of the above corollary is not neccessarily true.

**Example 12** *Let $(\Omega, F, P) = ([0,1], \mathcal{B}, m)$. Let $\succcurlyeq$ be a preference order on $\mathbf{L}^1$ that is represented by $V(X) = -E[X^+] \cdot E[X^-]$, where $x^+ = \max\{0, x\}$, and $x^- = \max\{0, -x\}$. Note $V$ satisfies A1, A6 and A3. To verify A3, suppose $X \succsim_{SSD} Y$, without loss of generality assume $E[Y|X] = X \quad a.s.$. Then*

$$V(Y) = -E[Y^+] \cdot E[Y^-] = -E[E(Y^+|X)] \cdot E[E(Y^-|X)] \tag{3}$$

*Note $x^+$ and $x^-$ are both convex. Therefore $E[Y^+|X] \geq (E[Y|X])^+ = X^+$, similarly, $E[Y^-|X] \geq X^-$ by conditional Jensen's inequality. Therefore one has (3)$\leq -E[X^+] \cdot E[X^-] = V(X)$, i.e. $\succcurlyeq$ is strongly risk averse.*

*However, $\succcurlyeq$ is not convex, to see this, enough to show $V$ fails to be quasiconcave. Consider*

$$X(\omega) = \begin{cases} 3 & \text{if } \omega \in [0, 1/2] \\ 1 & \text{if } \omega \in (1/2, 1] \end{cases} \qquad Y(\omega) = \begin{cases} -1 & \text{if } \omega \in [0, 1/2] \\ -3 & \text{if } \omega \in (1/2, 1] \end{cases}$$

*Then it is straightforward to verify $V(X) = V(Y) = 0$, yet $V(\frac{1}{2}X + \frac{1}{2}Y) = -1 < \min\{V(X), V(Y)\}$, i.e. $V$ is not quasiconcave.*

The choice of topology in this setting deserves some comment. In studying preference defined on set of probability distribution, it seems always desirable to use the weak convergence topology, and require the preference to be continuous in this topology. For example, Machina ([12]) studies probability distributions defined on a compact set $[0, M]$. Machina ([12]) worked with $L^1$ norm on the set of distribution functions on $[0, M]$, the topology of which is equivalent to the weak topology. However, as argued in Allen ([1]), sometimes it is desirable to embed the consumption set in a Hilbert space especially when one is concerned about using smooth utility functions to obtain smooth demand functions. However, Allen ([1]) pointed out the general difficulty in embedding a general class of probability measures into a Hilbert manifold for the topology of weak convergence of probability measures. The topology we use here is stronger than the weak convergence topology, therefore the continuity assumption here we use is strictly weaker than requiring continuity in the weak topology. Allen ([1]) believed it is desirable to give up the weak topology in favor of the Hilbert manifold structure. Here we give another reason for giving up the weak topology, when one is interested in risk aversion. That is the strongly risk averse property $(A3)$ is closed in the topology we use here, but not closed in the weak topology. In general, one cannot require a preference to be strictly monotone $(A2')$, strongly risk averse$(A3)$ and continuous in the weak topology at the same time. Consider the following example: Let $P(X = -1) = 1$, and $\forall n$, let $P(Y_n = 0) = 1 - \frac{1}{n}$ and $P(Y_n = -n) = \frac{1}{n}$. Then $X \succsim_{SSD} Y_n$ for all $n$. Therefore $X \succcurlyeq Y_n$ for all $n$.

Yet $Y_n \Rightarrow Y$, with $P(Y=0)=1$. And $Y \succ_{FSD} X$. If $\succcurlyeq$ satisfies $A2'$, we have $Y \succ X$. But by continuity in the weak topology, we should have $X \succcurlyeq Y$, which is a contradiction.

In the last part, we showed that without the expected utility assumption, risk aversion is equivalent to equal-distribution convex, which when $(A1)$ is assumed, in general, is strictly weaker than convexity. For comparison and completeness, we also point out here that if the preference relation has an expected utility representation, then strong risk aversion does imply convexity, as is summerized in the following proposition.

**Proposition 13** *Suppose $\succcurlyeq$ satisfies $A1$, and $V$ represents $\succcurlyeq$. Suppose also $V$ has an expected utility representation, i.e. $\exists u : R \longrightarrow R$ such that $V(X) = \int_\Omega u(X(\omega))dP(\omega)$ Then the followings are equivalent:*

*1) $\succcurlyeq$ is convex.*

*2) $\succcurlyeq$ is equal-distribution convex.*

*3) $\succcurlyeq$ is strongly risk averse.*

*4) $\succcurlyeq$ is weakly risk averse.*

**Proof.** Note 1) obviously implies 2), and 2) implies 3) by proposition 10. 3) implies 4) is straight forward. It is enough to show that 4) implies 1).

Consider the expected utility representation $V(X) = \int_\Omega u(X)dP$. Note 4) implies the $u$ is concave, this implies $V$ is concave since

$$\begin{aligned} V(\lambda X + (1-\lambda)Y) &= \int u(\lambda X + (1-\lambda)Y)dP \\ &\geq \int [\lambda u(X) + (1-\lambda)u(Y)]dP \\ &= \lambda V(X) + (1-\lambda)V(Y) \end{aligned} \tag{4}$$

Therefore the proposition is proved. ■

# 3 Representation Function of the Frechet Derivative

>From this section on, we assume that $\succcurlyeq$ can be represented by a utility function $V$ that is continuously Frechet differentiable, and study the differential properties of $V$. This section is devoted to investigate the form of the Frechet derivative of $V$. $\forall X \in L^1$, let $DV(X)$ denote the Frechet derivative of $V$ evaluated at $X$, then $DV(X)$ is a continuous linear functional on $L^1 = L^1(\Omega, \mathcal{F}, P)$. Then by Riesz representation theorem, $DV(X)$ has a representation in $L^\infty(\Omega, \mathcal{F}, P)$. If we assume the state indepence condition$(A1)$, then it is possible to obtain a sharper characterization of the form of $DV(X)$. The rest of this section is devoted to this question and we also give a neccessary and sufficient condition for state independence $(A1)$ under continuous differentiability assumptions.

**Lemma 14** *Suppose $V$ represents $\succcurlyeq$ that is defined on $L^1$. Suppose that $\succcurlyeq$ satisfies condition (A1) and (A2) and $V$ is continuously Frechet differentiable. Then $\forall X \in L^1$, the Rieze representation of $DV(X)$, as an element of $L^\infty(\Omega, \mathcal{F}, P)$, is $\overline{\sigma(X)}$ measurable, where $\overline{\sigma(X)}$ denote the $P-$completion of the $\sigma$ field generated by $X$.*

**Proof.** In the proof, without any confusion, I use $DV(X)$ to denote both the continuous linear functional, and the representation of it. By the above lemma, we have the following identity:$\forall h \in L^1$, $DV(X)(h) = \int_\Omega DV(X)(\omega) \cdot h(\omega) dP$. It is the measurability part that needs to be shown.

First, assume $X$ is a simple function, i.e. $X = \sum_{i=1}^n c_i I_{C_i}$, where $I$ is the indicator function, $c_i \in R$, $C_i \in \mathcal{F}$ $\forall i = 1, 2, \cdots n$, and $\{C_i\}_{i=1}^n$ is a partition of $\Omega$. To show $DV(X)$ is measurable with respect to $\overline{\sigma(X)}$ , enough to show $DV(X)$ is constant $a.s.$ on $C_i$ for each $i$ such that $P(C_i) > 0$.

Let us first prove the following result: Fix $1 \le j \le n$, let $T_j$ be any measure preserving (m.p. hereafter) transformation on $C_j$ (that is, $T_j : (C_j, \mathcal{F}_j, P) \to (C_j, \mathcal{F}_j, P)$ is measure preserving. where $\mathcal{F}_j = \mathcal{F}|_{C_j}$ for $j = 1, 2, \cdots, n$). Define $T : (\Omega, \mathcal{F}, P) \to (\Omega, \mathcal{F}, P)$ such that

$$T(\omega) = \begin{cases} \omega & if \quad \omega \in C_j^C \\ T_j(\omega) & if \quad \omega \in C_j \end{cases} \tag{5}$$

then $DV(X) = DV(X \circ T)$.

Note $T$ defined above is m.p. .Note also the fact that $V$ is Frechet differentiable implies the Gateaux derivative exists and they are equal. Therefore $\forall h \in L^1$,

$$\begin{aligned} DV(X \circ T)(h) &= \lim_{\alpha \to 0} \frac{1}{\alpha} \{V(X \circ T + \alpha h) - V(X \circ T)\} \\ &= \lim_{\alpha \to 0} \frac{1}{\alpha} \{V(X + \alpha h) - V(X)\} = DV(X)(h) \end{aligned} \tag{6}$$

The second equality above needs some justification. $T$ is m.p. implies $X$ and $X \circ T$ has the same distribution, therefore $V(X \circ T) = V(X)$. One can also prove that $X \circ T + \alpha h$ and $X + \alpha h$ have the same distribution, because $\forall r \in R$,

$$\begin{aligned} P(\{\omega &: X \circ T(\omega) + \alpha h(\omega) < r\}) = P(\cup_{i=0}^n \{\omega : X \circ T(\omega) = c_i ; c_i + \alpha h(\omega) < r\}) \\ &= P(\cup_{i=0}^n [\{\omega : X \circ T(\omega) = c_i\} \cap \{\omega : c_i + \alpha h(\omega) < r\}]) \\ &= P(\cup_{i=0}^n [\{\omega : X(\omega) = c_i\} \cap \{\omega : c_i + \alpha h(\omega) < r\}]) \\ &= P(\{\omega : X(\omega) + \alpha h(\omega) < r\}) \end{aligned}$$

The next-to-last line is true since $\{\omega : X(\omega) = c_i\} = \{\omega : X \circ T(\omega) = c_i\}$ for all $i = 1, 2, \cdots n$ by the definition of $T$.

Now let us prove that $DV(X)$ is constant $a.s.$ on $C_i$ $\forall i$ such that $P(C_i) > 0$. Suppose this is not true, then $\exists$ $C_i$ with $P(C_i) = \delta > 0$ and $r \in R$ such that $\varepsilon = P(\{\omega \in C_i : DV(X)(\omega) > r\}) < \delta$ for some $0 < \varepsilon < \delta$. Let's assume $\varepsilon \le \delta - \varepsilon$ (The case where $\varepsilon > \delta - \varepsilon$ can be proved analogously). Let

$B = \{\omega \in C_i : DV(X)(\omega) > r\}$, and let us define the m.p. transformation $T_i : C_i \to C_i$ such that $T_i^{-1}(B) \subseteq C_i \backslash B$[10]. Let the transformation $T$ be defined as in (5). Note $T$ is m.p. and by the result in the last paragraph, $DV(X \circ T) = DV(X)$ *a.s.*. This implies

$$\forall h \in L^1, \quad \int_\Omega DV(X \circ T)(\omega) \cdot h(\omega) dP = \int_\Omega DV(X)(\omega) \cdot h(\omega) dP \qquad (7)$$

Particularly, let $h = I_{T^{-1}(B)}$, then the above implies $\int DV(X \circ T)(\omega) \cdot h(\omega) dP = \int_{T^{-1}(B)} DV(X)(\omega) dP$. Note however, $T^{-1}(B) \subseteq C_i \backslash B$ implies $DV(X)(\omega) \leq r$ on $T^{-1}(B)$. Hence $\int_{T^{-1}(B)} DV(X)(\omega) dP \leq rP(T^{-1}(B)) = r\varepsilon$. On the other hand,

$$\begin{aligned}
\int DV(X)(\omega) \cdot h(\omega) dP &= \lim_{\alpha \to 0} \frac{1}{\alpha} \{V(X \circ T + \alpha h) - V(X \circ T)\} \\
&= \lim_{\alpha \to 0} \frac{1}{\alpha} \{V(X \circ T + \alpha I_{T^{-1}(B)}) - V(X \circ T)\} \\
&= \lim_{\alpha \to 0} \frac{1}{\alpha} \{V(X \circ T + \alpha I_B \circ T) - V(X \circ T)\} \\
&= \lim_{\alpha \to 0} \frac{1}{\alpha} \{V(X + \alpha I_B) - V(X)\} \\
&= \int DV(X)(\omega) \cdot I_B(\omega) dP \qquad (8)
\end{aligned}$$

But $DV(X)(\omega) > r$ on $B$ implies that

$$\int DV(X \circ T)(\omega) \cdot h(\omega) dP = \int DV(X)(\omega) \cdot I_B(\omega) dP > \varepsilon r$$

which is a contradiction to (7).

Finally, suppose $X$ is any measurable function, define $X_n$ in the following way: for $k, n \geq 1$, let

$$\begin{aligned}
A^+_{k,n} &= \{\omega : (k-1)/2^n \leq X^+(\omega) < k/2^n\}, \;\; B^+_{k,n} = \{\omega : X^+(\omega) \geq n\} \\
A^-_{k,n} &= \{\omega : (k-1)/2^n \leq X^-(\omega) < k/2^n\}, \;\; B^-_{k,n} = \{\omega : X^-(\omega) \geq n\} \\
X_n &= n \cdot I_{B^+_n} + \frac{1}{2^n} \sum_{k=1}^{n \cdot 2^n} (k-1) I_{A^+_{k,n}} - n \cdot I_{B^-_n} - \frac{1}{2^n} \sum_{k=1}^{n \cdot 2^n} (k-1) I_{A^-_{k,n}}
\end{aligned}$$

[10] To see such measure preserving transformation exists, let's take any Borel subset $A$ of $C_i \backslash B$ with $P(A) = \varepsilon$. Let $\mathcal{B}(B)$ and $\mathcal{B}(A)$ denote the Borel subsets of $B$ and $A$, respectively. Note $(A, \mathcal{B}(A), P|_A)$ and $(B, \mathcal{B}(B), P|_B)$ are standard Borel spaces (see Corollary 13.4 on page 82 in [9]) and $P|_A$ and $P|_B$ are nonatomic measures. Therefore $(A, \mathcal{B}(A), P|_A)$ and $(B, \mathcal{B}(B), P|_B)$ are isomorphic probability spaces and there exist an Borel isomorphism $T^* : (A, \mathcal{B}(A), P|_A) \to (B, \mathcal{B}(B), P|_B)$ that is measure preserving. Define $T : C_i \to C_i$ such that

$$T(\omega) = \begin{cases} T^*(\omega) & \text{if } \omega \in A \\ T^{*-1}(\omega) & \text{if } \omega \in B \\ \omega & \text{if } \omega \in C \backslash (A \cup B) \end{cases}$$

It is then straightforward to verify that $T$ is a measure preserving transformation and $T^{-1}(B) = A \subseteq C_i \backslash B$.

Note $X_n$'s are simple functions and are all measurable with respect to $\overline{\sigma(X)}$. Note also$\forall n$, $|X_n| \leq |X|$, therefore, $\{|X_n|\}_{n=1}^{\infty}$ are dominated by $|X|$. It then follows that $\{|X_n|\}_{n=1}^{\infty}$ is a uniformly integrable sequence, therefore, the fact that $X_n \rightarrow X\, a.s$ implies $X_n \rightarrow X$ in $L^1$. Hence by the assumption of continuously Frechet differentiability, this implies $DV(X_n) \rightarrow DV(X)$ in $L^{\infty}$ norm (Note the operator norm: $\|DV(X_n) - DV(X)\|$ coincides with the $L^{\infty}$ norm: $\|DV(X_n) - DV(X)\|_{\infty}$). Since $X_n$'s are all simple, $DV(X_n)$ is $\overline{\sigma(X_n)}$ measurable for all $n$, therefore $DV(X_n) \in L^{\infty}(\Omega, \overline{\sigma(X)}, P)$ for all $n$. Also note $L^{\infty}(\Omega, \overline{\sigma(X)}, P)$ is a complete metric space (if we identify functions with the equivalence class modulo the relation:" $a.s.$ equal"), therefore one has $DV(X) \in L^{\infty}(\Omega, \overline{\sigma(X)}, P)$. This completes the proof. ■

In view of the above lemma, if $V(X)$ is Frechet differentiable, then the representation of $DV(X)$, viewed as an element of $L^{\infty}(\Omega, \mathcal{F}, P)$, is a measurable function of $X$, it follows that there exist a measurable function $R_X$: $(R, \mathcal{B}) \longrightarrow (R, \mathcal{B})$, such that $DV(X)(\omega) = R_X \circ X(\omega)$ $a.s.$ This motivate the following definition:

**Definition 15 (Representation Function of the Frechet Derivative)** *Let $V : L^1 \longrightarrow R$ be a continuously Frechet differentiable utility functional that satisfies the state independence condition $(A1)$. $\forall X \in L^1$, let $R_X : (R, \mathcal{B}) \longrightarrow (R, \mathcal{B})$ be the measurable function such that $DV(X)(\omega) = R_X(X(\omega))$ a.s.. Then $R_X$ is called the representation function of $DV(X)$.*

Note $R_X$ is not neccessarily unique; in particular, it can take any value outside the range of $X$. However, $R_X$ is unique on the set $X(\Omega)$ except on a set of $P$ measure 0, where $X(\Omega)$ is the range of the random variable $X$, that is $R_X$ is unique $Q_X$ a.s., where $Q_X$ denote the distribution of $X$. The above lemma shows that for preferences that satisfies the state independence assumption, and that is represented by a continuously Frechet differentiable utility function, the Frechet derivative can be represented by a function that is defined on the real line. This brings us a lot of convenience, because in stead of studying the properties of functions that is defined on the probability space $(\Omega, \mathcal{F}, P)$, we now only need to study the properties of functions defined on the real line to characterize the derivatives of the utility function $V$. Functions of a real variable is a lot easier to work with, in fact, as we will see shortly, positivity and monotonicity of the representation function corresponds to assumptions $(A2)$ and $(A3)$, respectively. Before we give the equivalent characterizations of these assumptions, let's first observe the following propety of the representation function, which is very useful, namely, random variables that differ by a measure preserving transformation have the same representation function.

**Lemma 16** *Let $V : L^1 \longrightarrow R$ be continuous Frechet differentiable utility function which represents a preference order that satisfies condition $(A1)$. Let $T : (\Omega, \mathcal{F}, P) \rightarrow (\Omega, \mathcal{F}, P)$ be a measure preserving transformation. Let $R_X$, $R_{X \circ T}$ be representation function of $X$ and $X \circ T$, respectively, then $R_X = R_{X \circ T}$ $Q_X$ a.s.*

**Proof.** Condition $(A2)$ implies that $\forall h \in L^1$

$$\begin{aligned} DV(X \circ T)(h \circ T) &= \lim_{\alpha \to 0} \frac{1}{\alpha}[V(X \circ T + \alpha h \circ T) - V(X \circ T)] \\ &= \lim_{\alpha \to 0} \frac{1}{\alpha}[V(X + \alpha h) - V(X)] = DV(X)(h) \end{aligned}$$

That is, $\forall h \in L^1$,

$$\begin{aligned} \int_\Omega [R_X(X(\omega))] \cdot [h(\omega)] dP &= \int_\Omega [R_{X \circ T}(X \circ T(\omega))] \cdot [h \circ T(\omega)] dP \\ &= \int_\Omega [R_{X \circ T}(X(\omega))] \cdot [h(\omega)] d\mu \end{aligned} \tag{9}$$

where $\mu$ is a probability measure on $(\Omega, \mathcal{F}, P)$ defined by $\forall A \in \mathcal{F}$, $\mu(A) = P(\{\omega : T(\omega) \in A\})$. Note that fact that $T$ is measure preserving implies $P = \mu$. Let $h$ be the set of indicator functions, then (9) implies $\forall F \in \mathcal{F}$,

$$\int_F [R_X(X(\omega))] dP = \int_F [R_{X \circ T}(X(\omega))] dP$$

It follows that $R_X = R_{X \circ T}$ $a.s.$ on $X(\Omega)$. (For example let $F = \{\omega : R_X > R_X\}$, then above implies $P(F) = 0$. Similarly, $P(\{\omega : R_X > R_X\}) = 0$.) ■

As is in the case of utility functions defined on real numbers, integration of the derivative of the utility functions gives back the value of the utility function up to a constant, and this relates the properties of the derivatives of the utility function and that of the utility function itself. In this case, we will frequently use the following relation to calculate $V(Y) - V(X)$ through the Frechet derivative of $V$: given $X, Y \in L^1$, for $t \in [0,1]$, one can define $f(t) = V(tY + (1-t)X)$, then[11]:

$$\begin{aligned} V(Y) - V(X) &= f(1) - f(0) = \int_0^1 f'(t) dt \\ &= \int_0^1 DV(tY + (1-t)X)(Y - X) dt \end{aligned}$$

Now we are ready to give a neccessary and sufficient conditions for the state independence condition$(A2)$ in terms of the Frechet derivative of the utility functional $V$:

**Proposition 17** ***Proposition 2.2*** *Suppose $V$ is continuously Frechet differentiable. Then the following are equivalent:*

1)$V$ satisfies state independence $(A1)$.

2)$\forall X, Y \in L^1$, if $X$ and $Y$ has the same distribution, then $R_X = R_Y$ $Q_X - a.s.$

[11] For an excellent treatment of the integration theory of nonlinear operators, see [28]

**Proof.** First show 1) $\Rightarrow$ 2), the existence of the representation function is implied by lemma 14. We need to show $X, Y \in L^1$, and $X \approx_d Y$ implies $R_X = R_Y$. Define $X^* = \inf\{x \in R : F(x) \geq \omega\}$, where $F$ is the common distribution function of $X$ and $Y$. Then $X, Y$, and $X^*$ have the same distribution, and there exist measure preserving transformations $T_X : (\Omega, \mathcal{F}, P) \to (\Omega, \mathcal{F}, P)$, and $T_Y : (\Omega, \mathcal{F}, P) \to (\Omega, \mathcal{F}, P)$, such that $X = X^* \circ T_X$, and $Y = X^* \circ T_Y$[12]. Therefore by lemma 16,$R_X = R_{X^*} = R_Y$.

Next, suppose 2) is true, take any $X, Y \in L^1$, and $X \approx_d Y$, need to show $V(X) = V(Y)$. Define $X^*$ as in the preceeding paragraph, it is enough to show $V(X) = V(X^*)$. Let $\theta$ denote the random variable that equals 0 $a.s.$. Note $X = X^* \circ T$ for some measure preserving transformation $T$ by Ryff's theorem. Then we have:

$$\begin{aligned} V(X) &= V(X^* \circ T) = V(\theta) + \int_0^1 DV(t \cdot X^* \circ T)(X^* \circ T)dt \\ &= V(\theta) + \int_0^1 [\int R_{t \cdot X^* \circ T}(t \cdot X^* \circ T)(X^* \circ T)dP]dt \\ &= V(\theta) + \int_0^1 [\int R_{t \cdot X^*}(t \cdot X^* \circ T)(X^* \circ T)dP]dt \\ &= V(\theta) + \int_0^1 [\int R_{t \cdot X^*}(t \cdot X^*)(X^*)dP]dt \\ &= V(X^*) \end{aligned}$$

where the third to last line uses the fact that $\forall t \in [0,1]$, $t \cdot X^* \circ T$ and $t \cdot X^*$ have the same distribution, and the second to last equality is true because $T$ is measure preserving. ■

The above proposition says under continuous Frechet differentiability conditions, state independence ($A1$) is equivalent to the condition that random variables with the same distribution have the same representation function. This implies that under $A1$, to characterize the property of the Frechet derivative of $V$, one can without loss of generality assume that the random variable $X$ is weaky increasing on $[0,1]$. This simplification is possible since for any real valued random variable $X$, there exists a random $\widetilde{X}$ variable defined on $[0,1]$, such that $X$ and $\widetilde{X}$ have the same distribution and $\widetilde{X}$ is weakly increasing (One can take for example, $X = \inf\{x \in R : F_X(x) \geq \omega\}$, where $F_X$ is the distribution function of $X$). By the above proposition, if $X$ and $\widetilde{X}$ have the same distribution, we can take $R_X = R_{\widetilde{X}}$, then the properties of $R_{\widetilde{X}}$ can be easily translated into properties of $R_X$.

[12] This is sometimes called Ryff's Theorem, see [23], see also Theorem 7.5, page 82 in [2]. [2] also contains an example showing that $X \approx_d Y$ does not imply $X = Y \circ T$ for some measure preserving transformation $T$. See example 7.7 on Page 89 in [2].

# 4 Characterization of Monotonicity and Risk-aversion

In this section, we will assume $\succcurlyeq$ satisfies state independence $(A1)$, and is represented by a continuously Frechet differentiable utility function $V : L^1 \rightarrow R$. Neccessary and sufficient conditions for $\succcurlyeq$ satisfying $(A2)$ and $(A3)$ will be provided. Roughly, under the above assumptions, Monotonicity $(A2)$ is equivalent to the nonnegativity of the representation function $R_X$, and strong risk aversion is equivalent to the representation function $R_X$ being nonincreasing. lemma 18, lemma 22, lemma 24, and lemma 25 study the properties of almost surely nonnegative and almost surely nonincreasing functions defined on $X(\Omega)$. Proof of these lemmas are relegated to Appendix III. The main results of this section are proposition 20,proposition 26, and proposition 27.

**Lemma 18** *Let $V : L^1 \longrightarrow R$ be continuously Frechet differentiable. $\forall X \in L^1$, let $R_X$ be the representation function of $DV(X)$. Suppose $\exists A, B \in \mathcal{F}$, and real numbers $c, d > 0$, such that for some $X \in L^1$,*

$$c \int_A R_X(X(\omega))dP > d \int_B R_X(X(\omega))dP$$

*then $\exists \delta > 0$, such that $\|Y - X\| < \delta$ implies*

$$c \int_A R_Y(Y(\omega))dP > d \int_B R_Y(Y(\omega))dP$$

***Proof.*** *See Appendix III.* ■

**Corollary 19** *Let $V : L^1 \longrightarrow R$ be continuously Frechet differentiable. Suppose for some $A \in \mathcal{F}$, and $X \in L^1$, $\int_A R_X(X(\omega))dP < 0$, then $\exists \delta > 0$, such that $\|Y - X\| < \delta$ implies $\int_A R_Y(Y(\omega))dP < 0$.*

**Proof.** Take $c = 0$, and $d = 1$ in lemma 18. ■

The following propositions give equivalent conditions to monotonicity $(A3)$ in terms of the properties of the Frechet derivative of the utility function.

**Proposition 20** *Suppose also $V : L^1 \rightarrow R$ represents $\succcurlyeq$ and is continuously Frechet differentiable then $\succcurlyeq$ satisfies monotonicity $(A2)$ iff $\forall X \in L^1, R_X \geq 0$ $Q_X$ a.s..*

**Proof.** First suppose $R_X \geq 0$ $Q_X$ *a.s.*, need to show $\forall X, Y \in L^1$, $Y \succsim_{FSD} X$ implies $V(Y) \geq V(X)$. Because of condition $(A2)$ and lemma 3, we can assume $Y \geq X$ *a.s.*. $\forall t \in [0, 1]$, let $Y^t = tY + (1-t)X$, and $R_t = R_{tY+(1-t)X}$ . Then,

$$\begin{aligned} V(Y) - V(X) &= \int_X^Y DV(\xi)(Y - X)d\xi = \int_0^1 DV(Y^t)(Y - X)dt \\ &= \int_0^1 [\int_\Omega R_t(Y^t(\omega)) \cdot (Y(\omega) - X(\omega))dP]dt \end{aligned}$$

$R_t \geq 0$ $a.s.$ and $Y - X \geq 0$ $a.s.$ implies that $V(Y) \geq V(X)$, as needed.

Next, Suppose $\succcurlyeq$ is monotone, yet $\exists X \in L^1$ such that $R_X(X) < 0$ on $A$, where $A \in \mathcal{F}$ and $P(A) > 0$. By the previous lemma, $\exists \delta > 0$, such that $\forall Y \in L^1$, $\|Y - X\| < \delta$ implies $\int_A R_Y(Y(\omega))dP < 0$. Define $Y : (\Omega, \mathcal{F}, P) \to (R, \mathcal{B})$ by

$$Y(\omega) = \begin{cases} X(\omega) & \text{if } \omega \in A^C \\ X(\omega) + \frac{\delta}{2} & \text{if } \omega \in A \end{cases}$$

Let $Y^t$ and $D_t$ be as defined in the first part of the proof, then $\forall t \in [0, 1]$

$$\begin{aligned} \|Y^t - X\| &= \int |Y^t - X| dP = \\ \int_A |Y^t - X| dP &= \frac{1}{2}\delta t[P(A)] < \delta \end{aligned}$$

Therefore, $\forall t \in [0, 1]$, $\int_A R_t(Y^t(\omega))dP < 0$ implies

$$\begin{aligned} V(Y) - V(X) &= \int_0^1 \{\int_\Omega [R_t(Y^t(\omega))] \cdot [Y(\omega) - X(\omega)]dP\}dt \\ &= \int_0^1 \{\int_A [R_t(Y^t(\omega))] \cdot \frac{\delta}{2} dP\}dt < 0 \end{aligned}$$

Yet $Y \succsim_{FSD} X$ by construction, a contradiction. ■

**Corollary 21** *In the setting of the above theorem, suppose $\forall X \in C, R_X > 0$ $Q_X$ a.s., then $\succcurlyeq$ satisfies A2′.*

Next, let's assume that $\succcurlyeq$ is monotone, and investigate the differential properties of $V$ that satisfies the strong risk aversion condition ($A3$). To prepare for the proof of the main theorem, i.e. proposition 26 and proposition 27, lemma 22 gives a simple way of constructing mean preserving spreads of a random variable, lemma 24 and lemma 25 characterize almost surely weakly decreasing functions:

**Lemma 22** *Let $X \in L^1$, suppose $\exists A, B \in \mathcal{F}$ such that $P(A)$, $P(B) > 0$, and $\forall \omega \in A$, $\forall \omega' \in B$, for $a > 0$, $b > 0$, define $Y : (\Omega, F, P) \to (R, B)$ such that*

$$Y(\omega) = \begin{cases} X(\omega) & \text{if } \omega \in (A \cup B)^C \\ X(\omega) + b & \text{if } \omega \in B \\ X(\omega) - a & \text{if } \omega \in A \end{cases}$$

*Then the distribution function of $X$ cross that of $Y$ from below.*

**Proof.** See Appendix III. ■

**Corollary 23** *In the above setting, if $P(A) = P(B)$, and $a = b$, then the distribution function of $Y$ differs from that of $X$ by a simple MPS.*

**Lemma 24** *Take $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}, m)$, let $f : L^1 \times (R, \mathcal{B}) \to (R, \mathcal{B})$ be such that $\forall X \in L^1$, $f_X$ is a measurable function. Let $X \in L^1$, then $f_X$ is almost surely weakly decreasing iff for any two disjoint nontrivial intervals $A$, $B$, $\exists$ a null set $N$, such that $\forall \omega \in A \backslash N$, $\forall\ \omega' \in B \backslash N$,*

$$[X(\omega) - X(\omega')] \cdot [f_X(X(\omega)) - f_X(X(\omega'))] \le 0. \tag{10}$$

**Proof.** : See Appendix III. ■

**Lemma 25** *Take Take $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}, m)$, let $f : L^1 \times (R, \mathcal{B}) \to (R, \mathcal{B})$ be such that $\forall X \in L^1$, $f_X$ is a measurable function. Let $X \in L^1$, and suppose $X$ is weakly increasing on $[0,1]$, then $f_X$ is not almost surely weakly decreasing on $[0,1]$ iff $\exists$ sets of positive measure $A, B \in \mathcal{F}$, such that*

$$\forall \omega \in A,\ \forall \omega' \in B,\ X(\omega) < X(\omega') \tag{11}$$

*and*

$$\exists \alpha \in R, \text{ such that } \forall \omega \in A,\ \forall \omega' \in B,\ f_X(X(\omega)) \le \alpha < f_X(X(\omega')) \tag{12}$$

**Proof.** See Appendix III. ■

Note in the above lemma, one can take sets $A$, $B$ to be of the same measure. Now we are ready to chracterize strong risk aversion in terms of the Frechet derivative of the utility functional.

**Proposition 26** *Suppose $V$ represents $\succcurlyeq$ and is continuously Frechet differentiable. Then $\succcurlyeq$ is risk averse iff $\forall X \in L^1$, $R_X$ weakly decreasing $Q_X$ a.s.. That is, $\forall X \in L^1$, $(x - x') \cdot [R_X(x) - R_X(x')] \le 0$ for every $x, x' \in R$, $Q_X - a.s.$*

**Proof.** First suppose $R_X$ is weakly decreasing for all $X \in C$. Let $X \succsim_{SSD} Y$, need to show $V(X) \ge V(Y)$. By A2 and lemma 3, we can assume $Y = X + Z$ and $E[Z|X] = 0\ a.s.$. Then

$$\begin{aligned}
V(Y) - V(X) &= \int_X^Y DV(\xi)(Y - X)d\xi = \int_0^1 DV(X + tZ)(Y - X)dt \\
&= \int_0^1 [\int_\Omega R_t(X(\omega) + tZ(\omega)) \cdot Z(\omega)dP]dt^{13} \\
&= \int_0^1 E[E[R_t(X + tZ) \cdot Z|\sigma(X)]]dt
\end{aligned} \tag{13}$$

Let $F_X$ be the distribution function of $X$, and let $F_{Z|X=x}$ be the d.f. of $Z$ conditional on $X = x$, then the above is written as:

$$= \int_0^1 [\int_R \int_R R_t(x + tz)zdF_{Z|X=x}(z)dF_X(x)]dt \tag{14}$$

Note $R_t \ge 0$ by $(A3)$ and Proposition 20. Therefore $\forall t \in [0.1], \forall x \in R$, $R_t(x + tz)z \le R_t(x)z$. (There are two cases, if $z \ge 0$, $R_t(x + tz) \le R_t(x)$, since $R_t$ is decreasing. If $z < 0$, $R_t(x+tz) \geq R_t(x)$ and $R_t(x+tz)z \leq R_t(x)z$, therefore the result holds as well). Therefore:

$$\begin{aligned}(14) \quad &\leq \int_0^1 [\int_R \int_R R_t(x) z dF_{Z|X=x}(z) dF_X(x)]dt \\ &= \int_0^1 [\int_R R_t(x) \int_R z dF_{Z|X=x}(z) dF_X(x)]dt = 0\end{aligned}$$

The last equality is true since $\forall x \in R$, $E[Z|X=x] = 0$.

To see the reverse implication, let us assume $\succcurlyeq$ is risk averse, yet $\exists X \in L^1$, such that $R_X$ does not satisfy the almost sure decreasing property. Assume, without loss of generality, $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}, m)$, and $X$ is weakly increasing on $[0,1]$. By lemma 25, $\exists A, B \in \mathcal{F}$, such that $P(A) = P(B) = m > 0$, and $\forall \omega \in A$, $\forall \omega' \in B$, $X(\omega) \leq X(\omega')$ and $R_X(X(\omega)) < R_X(X(\omega'))$. Note this implies

$$\int_A R_X(X(\omega))dP < \int_B R_X(X(\omega))dP$$

By lemma 18, $\exists \delta > 0$, such that $\|Y - X\| < \delta$ implies $\int_A R_Y(Y(\omega))dP < \int_B R_Y(Y(\omega))dP$.

Now define:

$$Y(\omega) = \begin{cases} X(\omega) & \text{if } \omega \in (A \cup B)^C \\ X(\omega) + \delta & \text{if } \omega \in B \\ X(\omega) - \delta & \text{if } \omega \in A \end{cases}$$

and let $Y^t = tY + (1-t)X$, then $\forall t \in [0,1]$,

$$\|Y^t - X\| = \int (Y^t - X)dP = t\delta\sqrt{2m} < \delta$$

Therefore, $\forall t \in [0,1]$,

$$\int_A R_t(Y^t(\omega))dP < \int_B R_t(Y^t(\omega))dP$$

This implies

$$\begin{aligned} V(Y) - V(X) &= \int_0^1 [\int_\Omega R_t(Y^t(\omega)) \cdot (Y(\omega) - X(\omega))dP]dt \\ &= \delta \cdot \int_0^1 \{\int_B R_t(Y^t)dP - \int_A R_t(Y^t)dP\}dt > 0 \end{aligned}$$

However, corollary 23 implies $Y$ differs from $X$ by a simple MPS, this contrdicts $\succcurlyeq$ being risk averse. ■

We showed that strong risk aversion $(A3)$ can be defined as equal-distribution convexity. The following proposition gives an alternative characterization of strong risk aversion$(A3)$ using Fechet derivatives, which can be viewed as a differential version of the equal-distribution convexity property$(A4')$.

**Proposition 27** *Suppose $V$ represents $\succcurlyeq$ and is continuously Frechet differentiable. Then $\succcurlyeq$ is risk averse iff*

$$\forall X, Y \in L^1,\ X \approx_d Y \text{ implies } [DV(X) - DV(Y)](X - Y) \leq 0 \tag{15}$$

.

**Proof.** First suppose $\succcurlyeq$ is strongly risk averse, let $X, Y \in L^1$, and $X \approx_d Y$, then by proposition 17, $R_X = R_Y$, therefore,

$$\begin{aligned}
& [DV(X) - DV(Y)](X - Y) \\
= & \int [R_X(X(\omega)) - R_X(Y(\omega))] \cdot [X(\omega) - Y(\omega)] dP \leq 0
\end{aligned} \tag{16}$$

The last inequality is true because by proposition 26,

$$[R_X(X(\omega)) - R_X(Y(\omega))] \cdot [X(\omega) - Y(\omega)] \leq 0 a.s.$$

To see the reverse implication, suppose condition (15) is true, yet $\succcurlyeq$ is not risk averse, therefore $\exists X \in L^1$, such that $R_X$ does not satisfy the weakly decreasing property, by lemma 25, $\exists A, B \in \Omega$, with $P(A) = P(B) = m > 0$, and $\exists \alpha \in R$ such that $\forall \omega \in A, \forall \omega' \in B$, $X(\omega) < X(\omega')$ , ans $R_X(X(\omega)) \leq \alpha < R_X(X(\omega'))$. Using the same construction as in footnote 10, one can define a measure preserving transformation $T : (\Omega, F, P) \to (\Omega, F, P)$ such that $\forall \omega \in A$, $T(\omega) \in B$, $\forall \omega' \in B$, $T(\omega') \in A$, and $T(\omega) = \omega$ if $\omega \in (A \cup B)^C$. Consider

$$\begin{aligned}
& [DV(X) - DV(X \circ T)](X - X \circ T) \\
= & \int [R_X(X(\omega)) - R_X(X \circ T(\omega))] \cdot [X(\omega) - X \circ T(\omega)] dP \\
= & \int_A [R_X(X(\omega)) - R_X(X \circ T(\omega))] \cdot [X(\omega) - X \circ T(\omega)] dP \\
& + \int_B [R_X(X(\omega)) - R_X(X \circ T(\omega))] \cdot [X(\omega) - X \circ T(\omega)] dP \\
> & \ 0
\end{aligned}$$

The last inequality is true since on $A$, $X(\omega) < X \circ T(\omega)$ and $R_X(X(\omega)) < R_X(X \circ T(\omega))$, and on $B$, $X(\omega) > X \circ T(\omega)$ and $R_X(X(\omega)) > R_X(X \circ T(\omega))$, this gives the desired contradiction. ■

The statement of the above proposition immediately remind one of a characterizing property of differentiable concave functions, namely $V : C \to R$ is concave iff $\forall X, Y \in L^1$, $[DV(X) - DV(Y)](X - Y) \leq 0$. Note instead of requiring equation (15) holds for all $X, Y \in L^1$, strong risk aversion only requires it holds for $X$, $Y$ such that $X$ and $Y$ have the same distribution. Corollary 11 established that under $(A1)$ and $(A5)$, quasiconcavity of the utility functional implies strong risk aversion. This can also be easily derived from proposition 27 if $V$ is continuously Frechet differentiable.

**Example 28 (Quasiconcavity Implies Strong Risk-aversion)** *Suppose the preference order* $\succcurlyeq$ *satisfies A1, and is represented by a continuously Frechet differentiable utility function* $V$[14]*. If* $V$ *is qusiconcave, then it is strongly risk averse. To see this note continuously Frechet differentiable quasiconcave functions satisfy the following condition*[15]*:*

$$\forall X, Y \in L^1,\ V(Y) \geq V(X) \text{ implies that } DV(X)(Y - X) \geq 0$$

*Therefore if* $X \approx_d Y$*, by (A2),* $V(X) = V(Y)$*. If* $V$ *is quasiconcave, then* $DV(X)(Y - X) \geq 0$*, and* $DV(Y)(X - Y) \geq 0$*, which implies* $[DV(X) - DV(Y)](X - Y) \leq 0$*.*

# 5 Comparison of Attitudes toward Risk

**Definition 29 (Differ by a Simple Compensated Spread)** *Let* $X, Y \in L^1$*, let* $F_X$ *and* $F_Y$ *be the distribution function of* $X$ *and* $Y$*, respectively, and let* $\succcurlyeq$ *be a preference order on* $L^1$ *that satisfies* $A1 - A3$*. Then* $Y$ *is said to differ from* $X$ *by a simple compensated spread with respect to* $\succcurlyeq$*, if* $X \sim Y$*, and* $\exists \overline{x} \in R$ *such that* $\forall x \leq \overline{x}$*,* $F_X(x) \leq F_Y(x)$*, and* $\forall x \geq \overline{x}$*,* $F_X(x) \geq F_Y(x)$*.*

**Definition 30 (More Risk-averse than (Machina 1982))** *Let* $\succcurlyeq_1$ *and* $\succcurlyeq_2$ *be two preference orders on* $L^1$*, then* $\succcurlyeq_1$ *is said to be more risk-averse than* $\succcurlyeq_2$ *if* $\forall X,\ Y \in L^1$*,* $Y$ *differs from* $X$ *by a simple compensated spread with respect to* $\succcurlyeq_2$ *implies* $X \succcurlyeq_1 Y$*.*

**Proposition 31** *Let* $\succcurlyeq_1$ *and* $\succcurlyeq_2$ *be two preference orders on* $L^1$ *that satisfies A1, A2, and A3, suppose also,* $\succcurlyeq_1$ *and* $\succcurlyeq_2$ *are represented by continuously Frechet differentiable utility functions* $V_1$*, and* $V_2$*, respectively. For each* $X \in L^1$*, let* $R_{1X}$ *and* $R_{2X}$ *be the representation function of* $DV_1(X)$ *and* $DV_2(X)$ *respectively. Suppose further* $\forall X \in L^1$*,* $R_{2X} > 0$ *a.s. on* $X(\Omega)$*. Then* $\succcurlyeq_1$ *is more risk-averse than* $\succcurlyeq_2$ *if and only if* $\forall X \in L^1$*,* $\dfrac{R_{1X}}{R_{2X}}$ *is weakly decreasing* $Q_X$ *a.s.. That is,* $\forall X \in L^1$*,* $[x - x'] \cdot [\dfrac{R_{1X}}{R_{2X}}(x) - \dfrac{R_{1X}}{R_{2X}}(x')] \leq 0$ *for every* $x, x'$ $Q_X$ *a.s..*

**Proof.** First let us assume $\dfrac{R_{1X}}{R_{2X}}$ is weakly decreasing $Q_X$ *a.s.*. $\forall X \in L^1$, Suppose $Y$ differ from $X$ by a simple compensated spread from the point of view of $\succcurlyeq_2$, need to show $X \succcurlyeq_1 Y$.

Without loss of generality, let's take $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}, m)$, and let $F_X$, $F_Y$ be the distribution function of $X$ and $Y$, respectively. If one define

$$\widetilde{X}(\omega) = \inf\{x : F_X(x) \geq \omega\}, \text{ for every } \omega \in \Omega$$

[14]Note Frechet differentiability implies continuity, therefore, $\succcurlyeq$ also satisfies $A5$.

[15]To see this is true, note $V(Y) \geq V(X)$ implies $V(X + t(Y - X)) \geq V(X)$ for all $t \in [0,1]$, therefore $V'(X)(Y - X) = \lim_{\alpha \to 0} \frac{1}{\alpha}[V(X + \alpha(Y - X) - V(X)] \geq 0$.

and

$$\widetilde{Y}(\omega) = \inf\{x : F_Y(x) \geq \omega\}, \text{ for every } \omega \in \Omega$$

Enough to show $V_1(X) \geq V_1(Y)$. Without any confusion, let write $X, Y$ for $\widetilde{X}$ and $\widetilde{Y}$, respectively. If $X = Y$ $a.s.$, then the conclusion is trivial, therefore let's also assume $X$ differ from $Y$ on a set of positive measure.

Note $F_Y$ differ from $F_X$ by a simple compenstated spread implies that $\exists\overline{\omega}$ such that for $\omega < \overline{\omega}$, $X(\omega) \geq Y(\omega)$, and for $\omega > \overline{\omega}$, $X(\omega) \geq Y(\omega)$. (For example, one can take $\overline{\omega} = F_Y(\overline{x})$. ). For each $t \in [0,1]$, $\eta \in [0,1]$, define $Y(t,\eta) : (\Omega, \mathcal{F}, P) \to (R, \mathcal{B})$ by

$$Y(t,\eta)(\omega) = \begin{cases} tY(\omega) + (1-t)X(\omega) & \text{if } \omega \leq \overline{\omega} \\ tY(\omega) + (1-t)X(\omega) & \text{if } \omega > \overline{\omega} \end{cases}$$

Then $\forall t \in [0,1]$, there exist a unique $\eta^*$ such that $V_2(X) = V_2(Y(t,\eta^*))$. To see this is true, first note since $R_{2X} > 0$, corollary 21 implies $\succcurlyeq_2$ is strictly monotone. Therefore $V_2(Y(t,\eta))$ is continuous in $(t,\eta)$, strictly decreasing in $t$ and strictly increasing in $\eta$[16]. Note also for each $t \in [0,1]$, $X \succsim_{FSD} Y(t,0)$ and $Y(t,1) \succsim_{FSD} Y$, therefore $V_2(Y(t,0)) \leq V_2(X) = V_2(Y) \leq V_2(Y(t,1))$. By the mean value theorem, $\exists\eta^* \in [0,1]$ such that $V_2(X) = V_2(Y(t,\eta^*))$. Uniqueness of $\eta^*$ follows from the strict monotonicity of $V_2(Y(t,\eta))$ with respect to $\eta$.

Therefore one can define $\eta : [0,1] \to [0,1]$ such that for each $t \in [0,1]$, $V_2(Y(t,\eta(t))) = V_2(X)$. $\eta(t)$ such defined is also continuously differentiable. To see this, it is enough to show $\frac{\partial}{\partial t}V_2(Y(t,\eta))$ and $\frac{\partial}{\partial \eta}V_2(Y(t,\eta))$ both exist and is continuous. Using the chain rule,

$$\frac{\partial}{\partial t}V_2(Y(t,\eta)) = DV_2(Y(t,\eta))(\frac{\partial}{\partial t}Y(t,\eta)) \tag{17}$$

to justify (17), need to show $\frac{\partial}{\partial t}Y(t,\eta)$ is a Frechet derivative. Note $\frac{\partial}{\partial t}Y(t,\eta)$ exists in the Gateaux sense:

$$\lim_{\alpha \to 0} \frac{1}{\alpha}[Y(t+\alpha,\eta) - Y(t,\eta)] = Z^- \tag{18}$$

where

$$Z^-(\omega) = \begin{cases} Y(\omega) - X(\omega) & \text{if } \omega \leq \overline{\omega} \\ 0 & \text{otherwise} \end{cases}$$

The limit in (18) is, of course in $L^1$. It is straightforward to verify this is indeed a Frechet derivative, and $Y(t,\eta)$ is continuously differentiable in $t$. Similarly, one

[16] To see this true, note for a fixed $\eta$, $t_1 < t_2$ implies $Y(t_1,\eta) \succ_{FSD} Y(t_2,\eta)$, and for a fixed $t$, $\eta_1 < \eta_2$ implies $Y(t,\eta_1) \prec_{FSD} Y(t,\eta_2)$. To see $V_2(Y(t,\eta))$ is continuous in $t$ and $\eta$. Note $(t_n,\eta_n) \to (t,\eta)$ implies $Y(t_n,\eta_n) \to Y(t,\eta)$ pointwise, note also $Y(t_n,\eta_n)$'s are dominated by the $L^p$ bounded random variable $(X^+ \vee Y^+) - (X^- \vee Y^-)$, therefore $Y(t_n,\eta_n) \to Y(t,\eta)$ in $L^p$. Hence $V_2(Y(t_n,\eta_n)) \to V_2(Y(t,\eta))$ by continuity.

can show $Y(t,\eta)$ is continuously differentiable in $\eta$ as well, and $\dfrac{\partial}{\partial\eta}Y(t,\eta) = Z^+$, where

$$Z^+(\omega) = \begin{cases} 0 & \text{if } \omega \le \varpi \\ Y(\omega) - X(\omega) & \text{otherwise} \end{cases}$$

Therefore, the implicit function theorem (for functions of real numbers) implies $\eta(t)$ is continuously differentiable and

$$\begin{aligned} \eta'(t) &= -\frac{\frac{\partial}{\partial t}V_2(Y(t,\eta(t)))}{\frac{\partial}{\partial \eta}V_2(Y(t,\eta(t)))} = -\frac{DV_2(Y^t)(\frac{\partial}{\partial t}Y(t,\eta(t)))}{DV_2(Y^t)(\frac{\partial}{\partial \eta}Y(t,\eta(t)))} \\ &= -\frac{\int R_{2t}(Y^t(\omega)) \cdot Z^-(\omega)dP}{\int R_{2t}(Y^t(\omega)) \cdot Z^+(\omega)dP} > 0 \end{aligned} \tag{19}$$

where we wrote $Y^t$ for $Y(t,\eta(t))$, and $R_{it}$ for the representation function of $DV_i(Y^t)$, $i = 1, 2$, respectively.

Now consider $f(t) = V_1(Y(t,\eta(t)))$, then $f(t)$ is continuously differentiable, note $f(0) = V(X)$, and $f(1) = V(Y)$. Therefore $V(Y) - V(X) = \int_0^1 f'(t)dt$, where

$$\begin{aligned} f'(t) &= \frac{\partial}{\partial t}V_1(Y(t,\eta(t))) + \frac{\partial}{\partial \eta}V_1(Y(t,\eta(t)))\eta'(t) \\ &= DV_1(Y^t)(Z^-) + DV_1(Y^t)(Z^+)\eta'(t) \\ &= \int_0^{\varpi} R_{1t}(Y^t(\omega)) \cdot Z^-(\omega)dP + \int_{\varpi}^1 R_{1t}(Y^t(\omega)) \cdot Z^+(\omega)dP \cdot \eta'(t) \\ &= \int_0^{\varpi} \frac{R_{1t}(Y^t)}{R_{2t}(Y^t)} R_{2t}(Y^t) \cdot Z^- dP \\ &\quad + \int_{\varpi}^1 \frac{R_{1t}(Y^t)}{R_{2t}(Y^t)} R_{2t}(Y^t) \cdot Z^+ dP \cdot \eta'(t) \end{aligned} \tag{20}$$

Note on $[0,\varpi)$, $\frac{R_{1t}(Y^t)}{R_{2t}(Y^t)}$ is decreasing and $Z^- \le 0$, and on $(\varpi, 1]$, $\frac{R_{1t}(Y^t)}{R_{2t}(Y^t)}$ is decreasing and $Z^+ \ge 0$. Therefore,

$$\begin{aligned} (20) &\le \frac{R_{1t}(Y^t(\varpi))}{R_{2t}(Y^t(\varpi))}\{\int_0^{\varpi} R_{2t}(Y^t(\omega)) \cdot Z^-(\omega)dP \\ &\quad + \int_{\varpi}^1 R_{2t}(Y^t(\omega)) \cdot Z^+(\omega)dP \cdot \eta'(t)\} \\ &= \frac{R_{1t}(Y^t(\varpi))}{R_{2t}(Y^t(\varpi))} \cdot \frac{d}{dt}V_2(Y(t,\eta(t))) = 0 \end{aligned}$$

It then follows $V(Y) \le V(X)$, as needed.

To see the reverse implication, suppose $\succcurlyeq_1$ is more risk averse than $\succcurlyeq_2$, yet for some $X \in C$, $\frac{R_{1X}}{R_{2X}}$ does not satisfy the almost sure decreasing property.

Without loss of generality, assume $X$ is weakly increasing on $[0,1]$. Then by lemma 25, $\exists A, B \in F$, such that $P(A) = PB) = m > 0$ and $\forall \omega \in A$, $\forall \omega' \in B$, $X(\omega) \leq X(\omega')$, and $\exists \alpha \in R$, such that

$$\frac{R_{1X}}{R_{2X}}(X(\omega)) \leq \alpha \leq \frac{R_{1X}}{R_{2X}}(X(\omega')) \tag{21}$$

Define the following random variable:

$$Y(\omega) = \begin{cases} X(\omega) & \text{if } \omega \in (A \cup B)^C \\ X(\omega) + 1 & \text{if } \omega \in B \\ X(\omega) - \gamma & \text{if } \omega \in A \end{cases}$$

where $\gamma$ is chosen such that $V_2(Y) = V_2(X)$. Again, by continuity and strict monotonicity of $V_2$, $\gamma$ exists and is unique, furthermore, $0 \leq \gamma \leq 1$. For $t \in [0,1]$, one can define:

$$Y(t, \eta(t))(\omega) = \begin{cases} X(\omega) & \text{if } \omega \in (A \cup B)^C \\ X(\omega) + t & \text{if } \omega \in B \\ X(\omega) - \eta(t) & \text{if } \omega \in A \end{cases}$$

such that $\forall t \in [0,1]$, $V_2(Y(t, \eta(t))) = V_2(X)$. Lemma 22 then implies for each $t \in [0,1]$, $Y(t, \eta(t))$ differs from $X$ by a simple compensated spread from the point of view of $\succcurlyeq_2$. Note also for $t \in [0,1]$,

$$\begin{aligned} \frac{d}{dt} V_2(Y(t, \eta(t))) &= DV(Y^t)(Y - X) \\ &= \int_B R_{2t}(Y^t(\omega))dP - \gamma \int_A R_{2t}(Y^t(\omega))dP = 0 \end{aligned}$$

where at $t = 0$ the derivative is taken from the right. One has

$$\int_B R_{2X}(X(\omega))dP - \gamma \int_A R_{2X}(X(\omega))dP = 0$$

Therefore

$$\begin{aligned} & \int_B R_{1X}(X(\omega))dP - \gamma \int_A R_{1X}(X(\omega))dP \\ = & \int_B \frac{R_{1X}(X)}{R_{2X}(X)} \cdot R_{2X}(X)dP - \gamma \int_A \frac{R_{1X}(X)}{R_{2X}(X)} \cdot R_{2X}(X)dP \\ > & \alpha \{ \int_B R_{2X}(X(\omega))dP - \gamma \int_A R_{2X}(X(\omega))dP \} = 0 \end{aligned}$$

Now using lemma 18, $\exists \delta > 0$, such that $\|Y - X\| < \delta$ implies $\int_B R_{1Y}(Y)dP - \gamma \int_A R_{1Y}(Y)dP > 0$. Note for $t \leq \frac{\delta}{2}$,

$$\|Y(t, \eta(t)) - X\| = \int |Y^t - X| dP \leq t < \delta$$

Therefore let $\widehat{Y} = Y(\frac{\delta}{2}, \eta(\frac{\delta}{2}))$, then $\widehat{Y}$ differ from $X$ by a simple compensated spread w.r.t. $\succcurlyeq_2$, yet

$$\begin{aligned} V_1(\widehat{Y}) - V_1(X) &= \int_0^{\frac{\delta}{2}} DV_1(Y^t)(Y - X)dt \\ &= \int_0^{\frac{\delta}{2}} \{\int_B R_{1t}(Y^t)dP - \gamma \int_A R_{1t}(Y^t)dP\}dt > 0 \end{aligned}$$

Since $\int_B R_{1t}(Y^t)dP - \gamma \int_A R_{1t}(Y^t)dP > 0$ for all $t \in [0, \frac{\delta}{2}]$, which is a contradiction. ■

# 6 Conclusion

We studied preferences defined on $L^1$. If the utility functional that represents the preference has an expected utility representation, then under $(A1)$ and $(A5)$ strong risk aversion, equal-distribution convexity, and convexity are equivalent. In the nonexpected utility case, the equivalence of equal-distribution convexity and convexity breaks down, however, strong risk aversion is still equivalent to equal-distribution convexity. Since convexity is a very attractive property in many cases, for example, in establishing the existence of general equilibrium. The above arguement has the implication that without the Indepence Axiom, strong risk aversion alone does not provide enough ground for assuming convexity.

We also study the differential properties of the utility functional. For a differentiable utility functional $V$, we charaterize the Rieze representation of $DV(X)$ under $(A2)$. We then characterize differentiable utility functionals that satisfies monotonicity and strong risk aversion, respectively. We also provide a means for comparing individual's attitude toward risk by differential properties of the Frechet derivative of their utility functionals.

It is worth noting that in a one-period general equilibrium asset pricing model, an economy where the representative consumer have an expected utility functional and an economy with the representative consumer having a nonexpected utility functional are observationally equivalent. The reason is that difference between differentiable expected utility functional and nonexpected utility functional is that when the distribution of the random consumption (denoted $X$) changes, for expected utility functional, the representation function of $DV(X)$ does not change, while for nonexpected utility functional, the representation function of $DV(X)$ does change. However, in the one period general equilibrium model, information on asset prices only yields information on $DV(\cdot)$ evaluated at the equilibrium consumption. Comparison of $DV(X)$ at different $X's$ is not possible. Thus to study the different implications of expected and nonexpected utility functional on asset prices in a general equilibrium model, it is neccessary to move to the multiperiod or infinite horizon setting, and this is left to future research.

## 7 Appendix

### 7.1 Appendix I. Construction of sequence of Random Variables on $(\Omega, \mathcal{F}, P)$

This section shows that given any sequence of random variables, $X = (X_1, X_2, \cdots)$, where for each $i$, $X_i$ takes values in $(R, \mathcal{B})$, then one can contruct at the random variables $\tilde{X} = (\widetilde{X}_1, \widetilde{X}_2, \cdots)$, where $\widetilde{X}$ is defined on $(\Omega, \mathcal{F}, P)$ and $\widetilde{X}$ and $X$ have the same distribution. This result is used in various places in the paper. The following two lemmas are standard and are stated without proof:

**Lemma 32** *Let $\Psi = \bigotimes_{j=1}^{\infty} \Psi_j$ be product space, where for each $j$, $(\Psi_j, \rho_j)$ is a Polish space, with the metric $\rho_j$. Then $\Psi$ is a Polish space with the metric given by $\rho(x, y) = \sum_{j=1}^{\infty} \dfrac{\rho_j(x_j, y_j) \wedge 1}{2^j}$. The topology generated by this metric coincides with the product topology.*

**Lemma 33** *Let $\{(\Psi_j, \mathcal{T}_j)\}_{j=1}^{\infty}$ be a sequence of second countable topological spaces, then*

$$\mathcal{B}(\bigotimes_{j=1}^{\infty} \Psi_j) = \bigotimes_{j=1}^{\infty} \mathcal{B}(\Psi_j)$$

*That is, the Borel $\sigma$ field of the product topology coincides with the product Borel $\sigma$ field.*

Lemma 32 says the product of countably many copies of Polish spaces is Polish. Lemma 33 says for such spaces, the Borel $\sigma$ field of the product topology coincides with the product Borel $\sigma$ field. Note the product Borel $\sigma$ field is the Borel $\sigma$ field generated by the projectionn mappings and $X : (\Omega, \mathcal{F}, P) \to (\Psi, \bigotimes_{j=1}^{\infty} \mathcal{B}(\Psi_j))$ is measurable iff For each $i$, $X_i$ is $\mathcal{B}(\Psi_i)$ measurable. It then follows from lemma 33 that $X : (\Omega, \mathcal{F}, P) \to (\Psi, \mathcal{B}(\Psi))$ is Borel measurable iff for each $i = 1, 2, \cdots$, the coordinate function $X(i)$ is $\mathcal{B}(\Psi_i)$ measurable. Now we ready to prove the main proposition of this section:

**Proposition 34** *Let $Q$ be any probability distribtution on $(R^{\infty}, \mathcal{B}^{\infty})$, where $R^{\infty}$ is the product space of countably many copies of the real line and $\mathcal{B}^{\infty}$ is the product $\sigma$ field. Then there exist a stochastic process $\{X_n\}_{n=1}^{\infty}$, where for each $n$, $X_n : (\Omega, \mathcal{F}, P) \to (R, \mathcal{B})$ is Borel measurable, and the joint distribution of $\{X_n\}_{n=1}^{\infty}$ is $Q$.*

**Proof.** Let $(\Psi, \mathcal{G}, Q) = (R^{\infty}, \mathcal{B}^{\infty}, Q)$ be the probability space. Let $Y : (\Psi, \mathcal{G}, Q) \to (R^{\infty}, \mathcal{B}^{\infty})$ be the identity function, then $Y$ has distribution $Q$.

By lemma 32, $\Psi$ is a Polish space. By lemma 33, $\mathcal{B}^{\infty}$ is the associated Borel $\sigma$ field, therefore it is a standard Borel space. Since $Q$ is a probability measure, it can has at most countably many atoms. Enumerate these atoms by $\psi_1, \psi_2, \cdots$, and let $B = \{\psi_j\}_{j=1}^{\infty}$. Let $\Psi' = \Psi \backslash B$. let $\mathcal{G}' = \mathcal{G}|_{\Psi'}$, then $(\Psi', \mathcal{G}')$

is a standard Borel space as well(See [9], Corollary 13.4, Page 82). Assume for now $Q(\Psi') \neq 0$. Define $Q'$ on $(\Psi', \mathcal{G}')$ such that

$$\forall G \in \mathcal{G}', Q'(G) = \frac{Q(G)}{Q(\Psi')}$$

Take $A_1, A_2, \cdots \in F$ such that $A_i's$ are disjoint, and $P(A_i) = Q(\{\psi_i\})$ for $i = 1, 2, \cdots$. Let $A = \cup_{j=1}^{\infty} A_j$. Let $\Omega' = \Omega \backslash A$, and $\mathcal{F}' = \mathcal{F}|_{\Omega'}$. Similarly, Define $P'$ on $(\Omega', \mathcal{F}')$ such that

$$\forall F \in \mathcal{F}', P'(F) = \frac{Q(F)}{Q(\Omega')}$$

Note $(\Psi', \mathcal{G}', Q')$ and $(\Omega', \mathcal{F}', P')$ are standard Borel spaces, and $Q'$, $P'$ are nonatomic probability measures. Let $T : (\Omega', \mathcal{F}', P') \to (\Psi', \mathcal{G}', Q')$ be the one-to-one and onto, bimeasurable and measure preserving transformation[17]. Define $X : (\Omega, \mathcal{F}, P) \to (R^{\infty}, \mathcal{B}^{\infty})$ by:

$$X(\omega) = \begin{cases} Y \circ T(\omega) & \text{if } \omega \in \Omega' \\ Y(\psi_j) & \text{if } \omega \in A_j, \text{ for } j = 1, 2, \cdots \end{cases}$$

It is straightforward to check that the distribution of $X$ is $Q$. For each $n$, $\forall \omega \in \Omega$, let $X_n(\omega) = X(\omega)(n)$ be the coordinate functions. Then each $X_n : (\Omega, \mathcal{F}, P) \to (R, \mathcal{B})$ is Borel measurable by the comment after 33, and $\{X_n\}_{n=1}^{\infty}$ has the desired distribution. It is easy to see from the above proof that if $Q(\Psi') = 0$, then we do not need to construct the measure preserving transformation $T$, and the rest of the proof goes through without change. ■

Now we briefly sketch the proof of lemma 3. The lemma we have here is exactly the same as Theorem 1.2.4. and Theorem 1.5.20 in [14], except that in addition we claimed that the random variable $\widetilde{X}$ and $\widetilde{Y}$ can be constructed on $L^1$. We refer the reader to [14] for a detailed proof. Here we show that the proof could be adapted to show that $\widetilde{X}$ and $\widetilde{Y}$ can be constructed on $L^1$. Without loss of generality, we take $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}, m)$.

For the first part of the lemma, the construction is straightforward. Suppose $X \succsim_{FSD} Y$, let $F_X$ be the distribution function of $X$, and $F_Y$ be the distribution function of $Y$. Let $\tilde{X}(\omega) = \inf\{x \in R : \quad F_X(x) \geq \omega\}$ for every $\omega \in \Omega$. Similarly, let $\tilde{Y}(\omega) = \inf\{x \in R : \quad F_X(x) \geq \omega\}$. It is straightforward to check that $X \approx_d \tilde{X}$ and $Y \approx_d \tilde{Y}$ and $\tilde{X} \geq \tilde{Y}$.

In the proof of the second part of the lemma, [14] showed that the construction of $\widetilde{X}, \widetilde{Y}$ could be done in the following way: Fix $X$, construct a Markovian martingale sequence of random variables $\{X_n\}_{n=1}^{\infty}$ according to a certain sequence of transition kernels $\{Q_n\}_{n=1}^{\infty}$. Use the martingale convergence theorem to show that the sequence $\{X_n\}_{n=1}^{\infty}$ converge to some random variable $\widetilde{Y}$, finally, show that $\widetilde{Y}$ and $Y$ has the same distribution. In order to show $\widetilde{Y}$ could

[17] Such measure preserving transformation exists because $(\Omega', \mathcal{F}', P')$ and $(\Psi', \mathcal{G}', Q')$ are isomorphic measure spaces. See for example [9], Theorem 17.41, page 116.

be constructed on $([0,1], \mathcal{B}, m)$, and $\widetilde{Y} \in L^1$. Note that the distribution of $X$ and the sequence of transition kernels $\{Q_n\}_{n=1}^{\infty}$ uniquely define a distribution on $(R^{\infty}, \mathcal{B}^{\infty})$ by Komogorov's consistency theorem. Hence the sequence of random variables $\{X_n\}_{n=1}^{\infty}$ could be defined on $(\Omega, \mathcal{F}, P)$. The martingale convergence theorem implies that the limit of $\{X_n\}_{n=1}^{\infty}$ namely, $\widetilde{Y}$ is a random variable on $(\Omega, \mathcal{F}, P)$. $\widetilde{Y}$ is $L^1$ bounded since it has the same distribution with $Y$, therefore, $\widetilde{Y} \in L^1$, as needed.

## 7.2 Appendix II. Conditional Strong Law of Large Numbers

**Proof of 8:**

One can adapt the proof of Strong Law of Large Numbers. A review of Example 6.1 and Exercise 6.1 on page 266-267 in [6] shows that we only need to prove the following version of "Conditional" Hewitt-Savage 0-1 law and apply the reversed martigale theorem in the same way as in the example to conclude the proof:

**Proposition 35 (Conditional Hewitt-Savage 0-1 Law)** *The exchangeable $\sigma$ field of conditionally i.i.d. sequences is conditionally trivial.*

**Proof.** Since the exchangeable $\sigma$ field of an exhangeable sequence is contained in the completion of the tail $\sigma$ field, it is enough to show that the tail $\sigma$ field of conditionally independent sequence is conditionally trivial, i.e. a conditional version of the Komogorov 0-1 law. The proof of the Komogorov 0-1 law can be easily adapted to show that this is indeed the case. ■

## 7.3 Appendix III. Almost surely Nonnegative Functions and Almost Surely Nonincreasing Functions

**Proof of lemma 18:**

Let $\varepsilon = c\int_A R_X(X(\omega))dP - d\int_B R_X(X(\omega))dP > 0$, then

$$\begin{aligned} c\int_A R_Y(Y(\omega))dP &= c\int_A [R_Y(Y) - R_X(X) + R_X(X)]dP \\ &\geq c\int_A R_Y(Y)dP - c\int_A |R_Y(Y) - R_X(X)|dP \end{aligned}$$

Similarly, $d\int_B R_Y(Y(\omega))dP \leq d\int_B R_X(X)dP + d\int_A |R_Y(Y) - R_X(X)|dP$. Therefore,

$$\begin{aligned} & c\int_A R_Y(Y(\omega))dP - d\int_B R_Y(Y(\omega))dP \\ \geq\ & c\int_A R_X(X)dP - d\int_B R_X(X)dP - (c \vee d)\int_\Omega |R_Y \circ Y - R_X \circ X|dP \end{aligned} \quad (22)$$

Note $V$ is continuously differentiable, therefore one can choose $\delta > 0$, such that $\|Y - X\| < \delta$ implies $\|DV(X) - DV(Y)\| < \dfrac{\varepsilon}{2(c \vee d)}$ in the operator norm, i.e., $\|R_X(X) - R_Y(Y)\| < \dfrac{\varepsilon}{2(c \vee d)}$ in the $L^q$ norm. Hence

$$\int_\Omega |R_Y(Y) - R_X(X)| dP = \|R_X(X) - R_Y(Y)\| < \frac{\varepsilon}{2(c \vee d)}$$

Therefore (22)$\geq \varepsilon - \frac{\varepsilon}{2} > 0$, as needed

Proof of lemma 22:
Let $\overline{x} = \sup\{X(\omega) : \omega \in A\}$, then $\forall x \leq \overline{x}$,

$$\begin{aligned} F_Y(x) &= P(\{\omega : Y(\omega) \leq x\}) = P(\{\omega \in B^C : Y(\omega) \leq x\}) \\ &\geq P(\{\omega \in B^C : X(\omega) \leq x\}) = F_X(x) \end{aligned}$$

Similarly, $\forall x \geq \overline{x}$,

$$\begin{aligned} F_Y(x) &= P(\{\omega : Y(\omega) \leq x\}) = P(A \cup \{\omega \in A^C : Y(\omega) \leq x\}) \\ &\leq P(A \cup \{\omega \in A^C : X(\omega) \leq x\}) = F_X(x) \end{aligned}$$

**Proof of lemma 24:**
"Only if" is obvious. To see the "if" part is true, consider a sequence of partitions of the unit interval:

$$P_n = \{[(m-1)2^{-n}, m2^{-n})\}_{m=1}^{2^n} \text{ for } n = 1, 2, \cdots$$

For $n = 1$, by (10), $\exists$ a null set $N_1$ such that on $\Omega \backslash N_1$, if

$$[X(\omega) - X(\omega')] \cdot [f_X(X(\omega)) - f_X(X(\omega'))] > 0 \tag{23}$$

then either $\omega, \omega' \in [0, \frac{1}{2})$, or $\omega, \omega' \in [\frac{1}{2}, 1)$, in any case we must have $|\omega - \omega'| \leq \frac{1}{2}$. Similarly, for $n = 2$, $\exists$ a null set $N_2$ such that on $(N_1 \cup N_2)^C$, if (23) holds, then $|\omega - \omega'| \leq \frac{1}{4}$, $\cdots$ Continuing this way, we can find a sequence of null sets $\{N_i\}_{i=1}^\infty$, such that on $(\cup_{i=1}^\infty N_i)^C$, if (23) holds, then $|\omega - \omega'| \leq 2^{-n}$ for all $n$, i.e. $\omega = \omega'$, which contradicts (23). Note $N = \cup_{i=1}^\infty N_i$ is null as a countable union of null sets. This shows the "only if" part of lemma 24 is true.

**Proof of lemma 25:**
Again, "only if" part is trivial. To prove the "if" part, let's assume $f_X$ is not almost surely weakly decreasing. Note condition (12) implies condition (11), since $f_X(X(\omega)) < f_X(X(\omega'))$ implies $X(\omega) \neq X(\omega')$, the fact that $X$ is weakly increasing implies $X(\omega) < X(\omega')$. Therefore only need to find sets $A$, and $B$ such that condition (12) is satisfied. Let's suppose this is not true and derive a contradiction.

**Proof.** Note by lemma 24, if $f_X$ is not almost surely weakly decreasing on $[0,1]$, then $\exists$ nontrivial disjoint intervals $E, F$,such that

$$\begin{aligned} \forall \text{null set } N, \exists \omega &\in E\backslash N, \omega' \in F\backslash N, \text{such that} \\ [X(\omega) - X(\omega')] \cdot [f_X(X(\omega)) - f_X(X(\omega'))] &> 0 \end{aligned} \tag{24}$$

Without loss of generality, one can assume $E$ is to the left of $F$, i.e. $\forall \omega \in E$, $\forall \omega' \in F$, $\omega < \omega'$.

Let

$$E' = \{\omega \in E : \exists \omega' \in F, \text{ such that } [X(\omega) - X(\omega')] \cdot [f_X(X(\omega)) - f_X(X(\omega'))] > 0\}$$

and let

$$F' = \{\omega' \in F : \exists \omega \in E, \text{ such that } [X(\omega) - X(\omega')] \cdot [f_X(X(\omega)) - f_X(X(\omega'))] > 0\}$$

Let $p = P(E')$, and $q = P(F')$, then condition (24) implies $p, q > 0$. $\forall r \in R$, one can define

$$P_1(r) = P(\{\omega \in E' : f_X(X(\omega)) > r\})$$

and

$$P_2(r) = P(\{\omega \in F' : f_X(X(\omega)) > r\})$$

then $\forall r \in R$, there are six cases:

Case 1: $P_1(r) < p$, $P_2(r) > 0$. This is not possible because if one define $A = \{\omega \in E' : f_X \circ X(\omega) \leq r\}$, and $B = \{\omega \in F' : f_X \circ X(\omega) > r\}$, then $A, B$ are of positive measure and condition (12) is satisfied with $\alpha = r$.

Case 2: $P_1(r) = p$ and $P_2(r) = 0$. This is contradicts condition (24).

Case 3: $0 < P_1(r) < p$, and $P_2(r) = 0$. To derive a contradiction, let's define $\widehat{r} = \inf\{r : P_2(r) = 0\}$ then we have $P_2(\widehat{r}) = 0$. If $P_1(\widehat{r}) = p$, we already have a contradiction since we are back to Case 2. If $0 < P_1(\hat{r}) < p$, there are two subcases: if one can find $\varepsilon > 0$ such that $P_1(\hat{r} - \varepsilon) < p$ and $P_2(\widehat{r} - \varepsilon) > 0$, then we are in case 1. If not, i.e. $\forall \varepsilon > 0$, $P_1(\hat{r} - \varepsilon) = p$, then we have $P(\{\omega \in E' : f_X \circ X(\omega) \geq \hat{r}\}) = p$. Define

$$G = \{\omega \in E' : f_X \circ X(\omega) < \hat{r}\}$$

and

$$H = \{\omega' \in F' : f_X \circ X(\omega') > \widehat{r}\}$$

then both $H$ and $G$ are null sets. Note

$$\forall \omega \in E\backslash G, f_X \circ X(\omega) \geq \hat{r}$$

and

$$\forall \omega' \in F\backslash H, f_X \circ X(\omega) \leq \hat{r}$$

therefore $\forall \omega \in E\backslash G$, $\forall \omega' \in F\backslash H$,

$$[X(\omega) - X(\omega')] \cdot [f_X(X(\omega)) - f_X(X(\omega'))] \leq 0$$

this contradict (24).

Case 4: $P_1(r) = p$, and $0 < P_2(r) < q$. In this case, we define $\widehat{r} = \sup\{r : P_1(r) = p\}$. Then $P_1(\hat{r}) = p$. If $P_2(\hat{r}) = 0$, then we are in case 2. If not, again, there are two subcases: if one can find $\varepsilon > 0$ such that $P_1(\hat{r}+\varepsilon) < p$ and $P_2(\widehat{r}+\varepsilon) > 0$, then we are in case 1. If not, i.e. $\forall\varepsilon > 0$, $P_2(\widehat{r}+\varepsilon) = 0$, we have $P(\{\omega \in F' : f_X \circ X(\omega) \leq \hat{r}\}) = q$. Define

$$G = \{\omega \in E' : f_X \circ X(\omega) \leq \hat{r}\}$$

and

$$H = \{\omega' \in F' : f_X \circ X(\omega') > \widehat{r}\}$$

then both $G$ and $H$ are null sets, we have

$$\forall\omega \in E\backslash G, f_X \circ X(\omega) > \hat{r}$$

and

$$\forall\omega' \in F\backslash H, f_X \circ X(\omega) \leq \hat{r}$$

hence $\forall\omega \in E\backslash G$, $\forall\omega' \in F\backslash H$,

$$[X(\omega) - X(\omega')] \cdot [f_X(X(\omega)) - f_X(X(\omega'))] \leq 0$$

again this contradict (24).

Therefore we are left with only two possible cases, i.e., either

$$\text{Case5: } P_1(r) = p \text{ and } P_2(r) = q$$

or

$$\text{Case 6: } P_1(r) = 0 \text{ and } P_2(r) = 0$$

Let $r^* = \inf\{r \in R : P_1(r) = 0\}$, then one has

$$P(\{\omega \in E' : f_X(X(\omega)) = r^*\}) = p$$

$$P(\{\omega' \in F' : f_X(X(\omega')) = r^*\}) = q$$

But then for almost every $\omega \in E'$, and for almost every $\omega' \in F'$, one has

$$[X(\omega) - X(\omega')] \cdot [f_X(X(\omega)) - f_X(X(\omega'))] = 0$$

which is again a contradiction. ■

# 8 Reference

# References


[1] Allen, Beth(1987): Smooth Preferences and the Approximate Expected Utility Hypothesis. Journal of Economic Theory 41, 340-355 (1987).

[2] Bennett, Colin and Sharpley, Robert (1988), Interpolation of Operators, Academic Press, Inc.

[3] Chong,K.M. and Rice, N.M.(1971), "Equimeasurable Rearrangements of Functions," *Queen's Papers in Pure and Applied Mathematics. 28.*

[4] Chew, S.H., and K.R. MacCrimmon (1979), Alpha-Nu Choice Theory: A Generalization of Expected Utility Theory, Working Paper No.669, University of British Columbia, Faculty of Commerce and Business Administration.

[5] Diamond, P.A. and Stiglitz, J.E.(1974), Increases in Risk and Risk Aversion, Journal of Economic Theory 8, 337-360

[6] Durret, R. (1996), "Probability, Theory and Examples".

[7] Fristedt,B. and Gray,L (1997), "A Modern Approach to Probability Theory"

[8] Halmos, P.R. (1974), "Measure Theory"

[9] Kechris, A.S. (1995), "Classical Descriptive Set Theory"

[10] Lunberger, D.G. (1969), "Optimization by Vector Space Methods"

[11] Luxemberg,W.A.J. (1967) "Rearrangement-invariant Banach Function spaces", *Proceedings of the Symposium in Analysis, Queen's Papers in Pure and Applied Mathematics*. 10., 83-144.

[12] Machina, M.J. (1982), Expected Utility without the Independence Axiom. *Eonometrica,* Vol 50, 2 (1982), 277-324

[13] Marshall,A.W. and Olkin,I. (1979), "Inequalities: Theory of Majorization and Its Applications".

[14] Muller,A. and Stoyan,D. (2002), "Comparison Methods for Stochastic Models and Risks."

[15] Phelps, R.R (1993), "Convex Functions, Monotone Operators and Differentiability".

[16] Quiggin, J.P. (1982) "A Theory of Anticipated Utility", *Journal of Economic Behavior and Organization* 3, 323-343

[17] Quiggin, J.P. and P. Wakker (1994) The Axiomatic Basis of Anticipated Utility: A Clarification, *Journal of Economic Theory* 64, 486-499.

[18] Richter,M.K.(1980), Continuous and Semi-continuous Utility, *International Economic Review*, Vol. 21, No.2, June, 1980.

[19] Armstrong,T.E and Richter,M.K. (1984), The Core-Walras Equivalence, *Journal of Economic Theory*, Vol.33, No.1, June 1984

[20] Roberts, A.W. and Varberg, D.E. (1973), "Convex Functions*"*.

[21] Rohlin,V.A. On the Fundermental Ideas of Measure Theory. *Mat. Sborn.(n.s.)* 25 (1949), 107-150(Russian)—*American Mathematical Society Translatio*n 71 (1952), 1-54.

[22] Rothschild, M. and Stiglitz, J.E. (1970), Increasing Risk: I. A Definition, *Journal of Economic Theory* 2, 225-243.

[23] Ryff,J.V.(1970), "Measure preserving transformations and rearrangements," *Jounal of Mathetical Analysis and Applications*. 118, 315-347.

[24] Schmeidler, D. (1979), A Bibliographical Note on a Theorem of Hardy,, Littlewood and Polya, *Journal of Economic Theory* 20, 125-28.

[25] Seidl, C. (1997)"Experimental and Empirical Research II: Utility under Risk" S.Barbera, P.J. Hammond, and C, Seidl, *Handbook of Utility Theory*.

[26] Stokey, N.L. and Lucas, R.E. with Prescott, E.C. (1989), Recursive Methods in Economic Dynamics.

[27] Szekli,R.(1995), "Stochastic Ordering and Dependence in Applied Probability".

[28] Tapia,R.A. (1971), The Differentiation and Integration of Nonlinear Operators, in *Nonlinear Functional Analysis and Applications*.

[29] Werner, J.(2002), Risk and Risk Aversion When States of Nature Matter.

[30] Yaari, M.E.(1987), The Dual Theory of Choice under Risk, *Econometrica* 55, 95-115.