Federal Reserve Bank of Minneapolis
Research Department

Smooth Nonexpected Utility without State Independence*

Hengjie Ai

Working Paper 637

August 2005

ABSTRACT

We propose a notion of smoothness of nonexpected utility functions, which extends the variational analysis of nonexpected utility functions to more general settings. In particular, our theory applies to state dependent utilities, as well as the multiple prior expected utility model, both of which are not possible in previous literatures. Other nonexpected utility models are shown to satisfy smoothness under more general conditions than the Fréchet and Gateaux differentiability used in the literature. We give more general characterizations of monotonicity and risk aversion without assuming state independence of utility function.

*Ai, University of Minnesota and Federal Reserve Bank of Minneapolis. I thank Michele Boldrin for his advice and continuous encouragement. I thank David Levine for helpful comments. All errors are mine. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1. Introduction

Machina (1982) initiated the "local expected utility analysis" and showed some of the fundamental concepts of expected utility theory can be used for non-expected utility analysis. This approach has been extended by many authors to analyze broader classes of nonexpected utility functions, for example Allen (1987), Chew et al (1987), Machina (1989), Wang (1993), Chew and Mao (1995), among others. However, some of the most important classes of nonexpected utility functions cannot be analyzed by this approach. For example, none of the above literature deals with the multiple prior expected utility (MPEU) model (Gilboa and Schmeidler (1989)), which is probably the most widely used nonexpected utility model in applied work. In addition, all of the above literatures study derivatives of nonexpected utility functions with respect to distribution functions, therefore rule out state dependent utilities. The purpose of this paper is to develop a notion of smoothness of nonexpected utilities that applies to broader classes of nonexpected utility functions; in particular, to fill the gap of state dependent utilities and MPEU. Our approach generalizes the notion of Gateaux and Fréchet differentiability, and allows for calculus of variations analysis of nonexpected utility functions under more general conditions. The notion of smoothness we propose is applicable to MPEU, as well as to the class of state dependent utilities. Furthermore, our approach also overcomes the difficulty in the Fréchet differentiability approach that the domain of the utility function has to be a set of uniformly bounded random variables.

We achieve this generality by extending the existing theory along two dimensions. First, we study utility functions defined on set of random variables, in stead of on distribution functions. Therefore our approach does not assume probabilistic sophistication (Machina and Schmeidler (1992)), and allows for state dependent utilities as well. Many notions associated with probabilistically sophisticated utility functions, such as monotonicity and risk aversion can be generalized to the state dependent case (for example, Werner (2004)). Our formulation allows variational analysis of these properties in a more general setting. A technical advantage associated with this is that we allow utility functions to be defined on set of random variables with unbounded support, while the literature on Fréchet differentiability with respect to distribution functions restricts the domain of the utility function to be a set of random variables with bounded support. Next, we generalize the notion of Gateaux differentiability and Fréchet differentiability to a weaker smoothness condition. Many utility functions, for example, MPEU and the rank dependent expected utility (RDEU) model (Quiggin (1994), Quiggin and Wakker (1994)) that are not even Gateaux differentiable with respect to random variables satisfy our smoothness condition. We show how calculus of variations analysis can be applied to utility functions that are smooth.

We analyze differential properties of the MPEU. We prove the smoothness of MPEU without concavity assumptions. As a byproduct, our results imply that multiple prior expected utilities are Gateaux differentiable on a dense G_δ set of the L_p space of random variables. We characterize the set of subdifferentials of MPEU.

Our theory builds on the "local expected utility" analysis literature. Machina (1982, 1989) introduced the variational approach to nonexpected utility analysis and laid a theoretical foundation for linking local behaviors of nonexpected utility functions to its global properties. Allen (1987) examined the relation between smooth preference and smooth local utility representation and provided conditions under which smooth local utility exists. The notion of differentiability used by Machina (1982, 1989) is L_1 Fréchet differentiability with respect to distribution functions. This turns out to be a strong requirement and is not satisfied in many models used in practice. Efforts are made by many authors to extend Machina's original analysis to more general settings. Chew et al (1987) showed that RDEU, although not Fréchet differentiable in Machina's original formulation, is Gateaux differentiable, and Machina's theory can be extended to RDEU. Chew and Mao (1995) extended Machina's characterization of risk aversion of Fréchet differentiable nonexpected utilities to Gateaux differentiable utilities. Wang (1993) extended Machina's analysis to L_p Fréchet differentiable utility functionals. These above literature all focused on the derivatives of nonexpected utility functions with respect to distribution functions and thus not applicable to state-dependent utility functions. Carlier and Dana (2003) characterized the Gateaux subdifferential of RDEU with respect to random variables under concavity assumptions. Our purpose is to provide a general theory of differentiability of nonexpected utility functions with respect to random variables. Our approach also allows us to characterize the subdifferential of RDEU without assuming concavity.

The paper is organized as follows. Section two lays out some mathematical preliminaries. We define a notion of smoothness of nonexpected utility functions, and compare our notion with the commonly used Gateaux differentiability and Fréchet differentiability in the literature. Section three gives three examples of nonexpected utility models to illustrate the relation of our notion of smoothness and other concept of differentiability in the literature. We show that our notion can be applied to a wider class of nonexpected utility functions. Section four shows how to apply our notion of smoothness to the variational analysis of nonexpected utilities. We relates differential properties of utility functions to some fundamental concepts of preference defined on set of random variables, such as probabilistic sophistication, monotonicity and risk aversion. The last section concludes.

2. Preliminaries and Definition of Smoothness

We study utility functions defined on set of random variables. Let $\Omega = [0,1]$, let $\mathcal{F}$ be the σ field of Lebesgue measurable sets on Ω, and P be the Lebesgue measure [1]. We consider the space of real valued random variables defined on $(\Omega, \mathcal{F}, P)$ endowed with the L_p norm, denoted L_p for $1 \leq p < \infty$. We consider utility function $V : L_p \rightarrow R$ without assuming expected utility representation. We identify elements in L_p as the equivalence class of random variables that are equal P almost surely. By adopting this convention we are assuming the utility function V represents a preference that is indifferent between random variables that differ only on sets of measure 0. With this convention L_p is a complete metric space, and the dual space of L_p is L_q, with $\frac{1}{p} + \frac{1}{q} = 1$. For any $X \in L_p$, we use $F_X(\cdot)$ to denote the distribution function of X, that is $F_X : R \rightarrow [0,1]$, $\forall x \in R$, $F_X(x) = P(\{\omega : X(\omega) \leq x\})$.

The commonly used notion of smoothness is Gateaux differentiability and Fréchet differentiability. We first recall their definitions. Let S be a vector space, and V a real valued function defined on an open set $O \subseteq S$.

Definition 1. [2]*: V is said to be Gateaux differentiable at $x \in O$ if there exists a unique linear functional $DV(X)$ such that $\forall Y \in S$,*

$$DV(X)(Y) = \lim_{\alpha \to 0} \frac{1}{\alpha}[V(X + \alpha Y) - V(X)] \tag{1}$$

In this case, $DV(X)$ is call the Gateaux derivative of V at X. V is said to be Gateaux

[1] This assumption is made for simplicity of exposition. Since any standard Borel space with nonatomic probability measure is isomorphic to the unit interval with Lebesgue measure (Kechris (1995), theorem 17.41, page 116), we are essentially assuming Ω is standard Borel and P is nonatomic.

[2] *The term of Gateaux differentiability were used by many authors with slightly different meanings. Some do not require $DV(X)$ to be linear; some do not require the linear functional $DV(X)$ to be the same for all $Y \in S$.*

differentiable if it is Gateaux differentiable at X for all $X \in O$.

A stronger notion of differentiability used in the literature is Fréchet differentiability:

Definition 2. *Let S be a normed vector space, and V a real valued function defined on an open set $O \subseteq S$. Then V is said to be Fréchet differentiable at $X \in O$, if there exists a continuous linear functional $DV(X)$ such that $\forall Y \in S$,*

$$\lim_{||Y|| \to 0} \frac{||V(X+Y) - V(X) - DV(X)(Y)||}{||Y||} = 0 \tag{2}$$

In this case, the continuous linear functional $DV(X)$ is called the Fréchet derivative of V at X.

V is said to be continuously Fréchet differentiable at X if it is Fréchet differentiable at X and $||X_n - X|| \to 0$ implies $||DV(X_n) - DV(X)|| \to 0$.

V is said to be (continuously) Fréchet differentiable if it is (continuously) Fréchet differentiable at all $X \in O$.

Fréchet differentiability is strictly stronger than Gateaux differentiability, and requires a notion of norm on S. Machina (1982)'s original analysis assumes Fréchet differentiability. Chew et al (1987) and Chew and Mao (1995) showed Machina's approach can be generalized to the Gateaux differentiability case.

In the above mentioned local expected utility analysis literature, the domain of the utility function S is taken to be some set of distribution functions. This formulation assumes that the preference is state independent, i.e. probability sophisticated, in the language of Machina and Schmeidler (1992). However, when states of nature are payoff relevant, agent's preferences over uncertain outcomes are intrinsically state dependent. We study utility functions defined on the set of random variables thus allowing for state dependent utility functions. Our purpose is to develop a proper notion of smoothness that is strong enough to allow for calculus-of-variations type of analysis, yet weak enough to incorporate most of existing models of nonexpected utility.

To allow for a broader class of nonexpected utility to be considered, we generalize the notion of subdifferential from the convex analysis literature (for example, Phelps (1993)), and define sub-Gateaux differential as follows.

Definition 3. *Let $O \subseteq L_p$ be open. $V : O \rightarrow R$ is said to be sub-Gateaux differentiable at $X \in O$ if there exists a set of continuous linear functionals $\Gamma(X)$ such that $\forall Y \in L_p$, $\forall l \in \Gamma(X)$,*

$$g'_L(0) \leq l(Y) \leq g'_H(0) \tag{3}$$

where

$$g(t) = V(X + tY) - V(X) \tag{4}$$

and where $g'_L(0) = \min\{g'(0^+), g'(0^-)\}$ and $g'_H(0) = \max\{g'(0^+), g'(0^-)\}$. In this case, $\Gamma(X)$ is called the set of sub(-Gateaux) differential of V at X.

V is called sub-Gateaux differentiable if it is sub-Gateaux differentiable at all $X \in O$. In this case, the mapping $X \rightarrow \Gamma(X)$ is called the sub(-Gateaux) differential correspondence.

In the above definition, we require that the subdifferential be continuous linear functionals. By Riesz's representation theorem, elements in $\Gamma(X)$ can be represented by vectors in L_q, with $\frac{1}{p} + \frac{1}{q} = 1$. The usual notion of Gateaux differentiability does not require the

Gateaux derivative to be a continuous linear functional. It does not even require a notion of norm to be defined on the domain of the utility function. However, the notion of Gateaux differentiability is usually too weak in this respect. $DV(X) \in L_q$ is usually neccessary for linking properties of the Gateaux derivative of utility functions to its global properties. We therefore build this requirement into the definition of sub-Gateaux differentiability. On the other hand, this requirement is almost innocuous in the sense that for all the nonexpected utility models we consider in this paper, conditions that gurantee Gateaux differentiability also implies $DV(X) \in L_q$ (See section 3 for examples). It is also clear that given $DV(X) \in L_q$ Gateaux differentiability implies sub-Gateaux differentiability, and the sub-Gateaux correspondence Γ is single valued. Finally, we remark that it is straightforward to verify that the subdifferential correspondence Γ is convex and closed (in the weak* topology[3]) valued.

If Γ is the sub-Gateaux differential correspondence of V, we use the notation $\gamma \in \Gamma$ to denote that γ is a selection from Γ, i.e. $\forall X \in L_p$, $\gamma(X) \in \Gamma(X)$. We will use l to denote both a generic linear functional on L_p and its representation in L_q, the meaning of which will often be clear from the context. Note if V is concave, then our notion subdifferentiability coincides with that for concave functions. However, our approach in this paper does not rely on any concavity or concexity assumptions.

To motivate our definition of smoothness, note that the following relation links local properties of the function V and its global property on the domain $O \subseteq S$ and is essential

[3]Note the dual space of L_p is L_q. The weak* topology on L_q is the one induced by weak* convergence defined by : $X_n \to X$ weak* if $\forall Y \in L_p$,

$$\int YX_n dP \to \int YX dP$$

in any of the differential analysis of nonexpected utility literature. Fix $X, Y \in O$, define $g : [0,1] \to R$ as in (4), then

$$V(X+Y) - V(X) = g(1) - g(0) = \int_0^1 g'(t)dt \tag{5}$$

$$= \int_0^1 DV(X+tY)Y dt \tag{6}$$

In general, Gateaux differentiability is not enough to guarantee the operation in (5) and (6), not even Fréchet differentiability. A sufficient condition is V being continuously Fréchet differentiable (See Tapia (1971), proposition 4.2, page 71. See also the discussion in Wang (1993)). However, Fréchet differentiability is often a too strong restriction for many of the nonexpected utility models (for example, Chew et al (1987)). We observe that the necessary and sufficient condition for operation in (5) is g being an absolute continuous function on $[0,1]$. Given absolute continuity of g, a sufficient condition for operation in (6) is V being sub-Gateaux differentiable. We thus propose the following definition of smoothness. Consider a utility function $V : L_p \to R$, where $1 \leq p < \infty$.

DEFINITION 4. *V is said to be* smooth *if*

1) $\forall X, Y \in L_p$, the function $V(X+tY)$ is absolutely continuous in t on $[0,1]$.

2) V is sub-Gateaux differentiable on L_p.

If a utility function $V : L_p \to R$ is smooth, then $\forall X, Y \in L_p$, (5) is true by the absolute continuity of g. Also, by definition of sub-Gateaux differentiability, whenever $g'(t)$ exists, we have

$$\int_\Omega \gamma(X+tY)Y dP = g'(t) \tag{7}$$

where $\gamma \in \Gamma$ is an arbitrary selection of Γ. Since $g'(t)$ exists a.s. on $[0, 1]$,

$$V(X+Y) - V(X) = \int_0^1 \int \gamma(X+tY) Y dP dt \tag{8}$$

Our definition of smoothness thus guarantees the validity of the operations in (5)-(6).

It maybe easily verified that neither Gateaux differentiability not Fréchet differentiability implies smoothness, while continuous Frechet differentiability is a sufficient condition for smoothness. On the other hand, many functions that are not even Gateaux differentiable are smooth. We argue that the smoothness condition is an appropriate notion for variational analysis of nonexpected utility functions, in the sense that it allows most of the nonexpected utility models to be "differentiable", while at the same time validate the operations in (5)-(6).

We will also make frequent use of the following two conditions on smooth nonexpected utility functions.

Condition 1. There exists a dense subset of L_p, denoted $\mathcal{D}$, such that Γ is single-valued on $\mathcal{D}$.

Condition 2. The subdifferential correspondence Γ is L_p to weak*[4] upper hemi-continuous and (weak*) compact valued.

To see why we need condition 1 and condition 2, consider functions defined on the real line. Suppose we want to make a differential characterization of monotonicity, that is, we want to make statements like:" $f : R \to R$ is nondecreasing if and only f' is nonnegative. " The "if" part is true if f is absolutely continuous. In our context, the analogue of absolute continuity

[4]See footnote 3 for the definition of weak* topology on L_q.

is the smoothness condition defined above. The "only if" part of the above statement is not true if we only assumes absolute continuity. The reason is that f' can be negative on a set of measure 0, yet f is still increasing. However, if we impose stronger differentiability conditions on f, for example, continuous differentiability, then the above statement is true. The generalization of continuous differentiability in infinite dimensional space is continuous Frechet differentiability. In fact, it is straightforward to verify that continuously Frechet differentiability implies condition 1 and condition 2. Our condition 1 and 2 play exactly the role of continuous differentiability in the above example, but is much weaker than continuous Frechet differentiability. The examples in the next section show that our smoothness condition and condition 1 and 2 are satisfied by most of the nonexpected utility models, many of which are not even Gateaux differentiable.

3. Examples

In this section, we verify the smoothness of several important nonexpected utility models that appear in the literature, and characterize their subdifferential correspondences. We do not intend to cover all examples of nonexpected utility models; in stead we emphasize two points: our approach can be applied to broader classes of utility functions, and it allows the utility functions to be defined on larger domains. In the first example, we study the relationship between our notion of smoothness and Machina (1982) and Wang (1993)'s notion of Fréchet differentiability. We show that, under mild conditions, the class of models that can be analyzed in Machina (1982, 1989) and Wang (1993) can also be analyzed with our approach. However, our approach allows utility functions to be defined on larger domains. The second example deals with expected utility and the weighted utility (Chew (1983, 1989), Dekel (1986)). These two models are all differentiable in Machina's approach. The purpose of this example is to illustrate the form of the derivatives of nonexpected utility functions with respect to random variables, and compare that with derivatives with respect to distribution functions. We use these models as an example to illustrate why our approach allows utility to be defined on a larger domain, in particular, allow for unbounded random variables. In the Third example, we study MPEU. MPEU serve as a good example to illustrate the strength of our theory. In general, it does not satisfy probability sophistication, and is not differentiable with respect to distribution functions even if it satisfies probabilistic sophistication. However, we show that it is smooth and satisfies condition 1 and 2; consequently, all of our theorems in section four would apply to this model. Other comparative statics results in Machina (1982, 1989) and Wang (1993), with proper modifications, also applies to MPEU. MPEU models also serve as an example in which the subdifferential correspondence is not single valued. We give characterizations of its subdifferential correspondence. We establish an important property of the subdifferential correspondence, i.e. it is single-valued on a dense G_δ set of L_p.

A. Example 1: L_p Fréchet differentibility and Smoothness.

This section studies the relation between Fréchet differentiability with respect to distribution functions and our notion of smoothness. We show utility functions that satisfy L_p differentiability also satisfies our notion of smoothness under mild conditions, yet we allow utility functions to be defined on larger domains. One of the significance of the local expected utility analysis approach is that it allows one to derive first order conditions of consumer's optimization problem and do comparative statics analysis. Machina (1982, 1989) laid a theoretical foundation for this approach. Machina's notion of differentiability is Fréchet differentiability with respect to distribution functions. Wang (1993) extended this approach to L_p Fréchet differentiability and allowed for a larger class of utility functions. Their results require that the domain of the utility function is a set of random variables with a common compact support, which allows one to define a notion of norm on the set of distribution functions. For more discussion on the choice of norm and topological structure on the set of distribution functions in this context, see Allen (1987) and Wang (1993). It is not clear, however, whether a proper notion of norm could be chosen on the space of distribution functions with unbounded support, so that Machina and Wang's result generalize to this case. Our approach however, indicates that if one instead considers differentiability with respect to random variables, this difficulty can be easily overcome.

To links their notion of differentiability to our approach, let's first recall Machina (1982)'s notion of Fréchet differentiability. Let $DF([-M, M])$ denote the set of distribution functions with support $[-M, M]$. The conclusion in Machina (1982) is that if $U(\cdot)$ is a Fréchet differentiable utility function of distribution functions, then $\forall F \in DF([-M, M])$,

$$U'(F)(G - F) = \int u_F(x) d\left[G(x) - F(x)\right] \tag{9}$$

where u_F is the local expected utility function. Machina uses L_1 norm on the set of distribution functions, while Wang (1993) generalize to L_p norm for $1 \leq p < \infty$. It is important to distinguish their use of L_p norm and ours. Machina and Wang use L_p norm on the space of distribution functions, i.e. the L_p space on $([-M, M], \mathcal{B}, Leb)$, where Leb denotes the

Lebesgue measure. The norm is defined by: for distribution functions F and G,

$$||F - G||_p = \left\{ \int_{-M}^{M} |F(t) - G(t)|^p dt \right\}^{\frac{1}{p}} \tag{10}$$

The L_p norm in our framework is defined on the space of random variables. To be precise,$\forall X, Y \in L_p(\Omega, F, P)$

$$||X - Y||_p = \left\{ \int_{\Omega} |X(\omega) - Y(\omega)|^p dP \right\}^{\frac{1}{p}} \tag{11}$$

Note distribution functions does not vanish at infinity, therefore Machina (1982, 1989) and Wang (1993) need require the support of the distributions to be compact to prevent $||F||_p = \infty$. Our formulation, however, does not need this requirement. We can represent the same preference represented by U by a utility function defined on the set of random variables. Let $RV([-M, M]) = \{X \in L^1 : \forall\omega \in [0, 1], -M \le X(\omega) \le M\}$, Define $V : RV([-M, M]) \to R$ through the relation

$$\forall X \in RV([-M, M]), \quad V(X) = U(F_X) \tag{12}$$

The following proposition establishes the link between the two notions of differentiability.

Proposition 1. *Suppose U is L_1 Fréchet differentiable, and $\forall F \in DF([-M, M])$, $u_F : R \to R$ is Lipschitz continuous, then V is smooth and Gateaux differentiable. The Gateaux differential of V at $X \in RV([-M, M])$ is given by: $\forall Y \in RV([-M, M])$,*

$$V'(X)(Y) = \int u'_F(X) \cdot Y dP \tag{13}$$

If further, U is L_1 continuously Fréchet differentiable, then V is smooth and satisfies *condition 1 and 2.*

(Proof in appendix.)

The above theorem shows that continuous L_1 Fechet differentiability of U (with respect to distribution functions) and Lipschitz continuity of the local expected utility function imply smoothness of V and condition 1 and 2. Although Machina (1982)'s original analysis does not impose continuous Fréchet differentiability explicitly, this is in fact needed[5]. The above theorem therefore implies the class of nonexpected utility functions that are differentiable in Machina's framework is also smooth and satisfies condition 1 and 2, provided that the local utility functions are Lipschitz continuous. The later condition is a rather innocuous one as long as one is interested in deriving first order conditions of consumer's optimization problem and performing comparative static analysis, because differentiability of local expected utility is needed anyway.

Wang (1993) extended Machina (1982, 1989)'s analysis and proved Machina's result holds for L_p Fréchet differentiable utility functions as long as the path is L_p-smooth. The extension can be explained by the following: if $p > 1$, then the set of utility functions that are L_p Fréchet differentiable are strictly larger than the the set of L_1 Fréchet differentiable utility functions. This is clear from equation (2). L_p differentiability requires the limit in (2) to exist for a smaller set of $Y's$, since convergence in L_p is a more stringent requirement than convergence in L_1. It is also clear that a weaker notion of differentiability is obtained at the expense of a stronger notion of norm on the space of distribution functions. This means there are less smooth paths in Wang's formulation than in Machina's formulation.

[5]See footnote 9 in Wang ([25]) for discussion of this point.

This is the reason why the results in Wang (1993) holds only along L_p-smooth paths. For the same reason, we are not able to show that all L_p Fréchet differentiable functions are smooth. However, they are smooth along "smooth paths". The following corollary formalizes the statement. Let first recall Wang (1993)'s definition of smooth paths. Let $F(\cdot, \alpha)$ be a local path through F on $D([-M, M])$, and let W be the associated neighborhood of 0. Then $F(\cdot, \alpha)$ is called a weakly smooth path at F if $F_\alpha(\cdot, \alpha)$ exists and is bounded on $[-M, M] \times W$. If in addition, the mapping $\alpha \to F_\alpha(\cdot, \alpha)$ is continuous in the L_p norm, then the path $F(\cdot, \alpha)$ is called L_p-smooth. The following corollary establishes Gateaux differentiability of V along weakly smooth paths:

Corollary 1. Suppose U is L_p Fréchet differentiable, and $\forall F \in D([-M, M])$, u_F is Lipschitz continuous. Suppose at $X, Y \in L_p$, the path $F_{X+\alpha Y}$ is weakly smooth, then V is Gateaux differentiable in the direction of Y.

Although we are not able to prove that all L_p Fréchet differentiable utility functions are smooth in all directions, the models studied in Wang (1993) can all be analyzed using our notion of smoothness. This is not surprising, since the results obtained in Wang only needs differentiability along smooth paths.

B. Example 2: Expected Utility and weighted Utility

In this section, we provide sufficient conditions under which the expected utility and the weighted utility are smooth. It is not our purpose to give a minimum set of conditions under which they are smooth, neither do we intend to provide a comprehensive study of the differentiability of all nonexpected utility models. Our purpose is to use these models as an example to illustrate the properties of derivatives of utility functions with respect to random variables. Since both models satisfy probabilistic sophistication, they also serve as an example for illustrating the results in example 1.

We first consider the expected utility function, denoted V_E, where

$$V_E(X) = \int u(X)dP \tag{14}$$

PROPOSITION 2. *Suppose u is bounded and Lipschitz continuous, then V_E is smooth, Gateaux differentiable and satisfies condition 1 and 2. The Gateaux derivative of V is given by:* $\forall Y \in L_p$,

$$DV_E(X)(Y) = \int u'(X)Y dP \tag{15}$$

Proof. Since expected utility is a special case of weighted utility (with $w \equiv 1$), the result here can be obtained as a special case of proposition 3.

From the above theorem it clear that the subdifferential correspondence of expected utility is single valued. The Gateaux differential of expected utility takes the form:

$$\Gamma(X) = u'(X) \in L_q \tag{16}$$

Note the Gateaux differential of expected utility is not constant. This is in contrast with Machina's formulation. The expected utility defines a linear funtional on the set of distribution functions; therefore its derivative with respect to distribution functions is a constant. However, expected utility is not linear in random variables unless the von Neumann-Morgenstern utility function is linear. Therefore its derivative with respect to random variables is not constant. Note even the expected utility is not Fréchet differentiable with respect to distribution functions unless the set of random variable on which the utility is defined have a common compact support. However, it is smooth and satisfies condition 1 and 2. Here we assume that the von Neumann-Morgenstern utility function u is bounded for simplicity. It could be replaced by more general integrability conditions, in which case the bounded convergence theorems in the proof can be replaced by the dominated convergence theorem. We next consider the weighted utility. The weighted utility is defined as:

$$V_B(X) = \frac{\int w(X)u(X)dP}{\int w(X)dP} \tag{17}$$

for some $w : R \to R$, such that $w > 0$, and $u : R \to R$. The following proposition gives sufficient conditions under which V_B is smooth.

PROPOSITION 3. *Suppose w and u are both bounded and Lipschitz continuous, assume also, w is bounded away from 0, then V_B is smooth, Gateaux differentiable and satisfy condition 1 and 2. The Gateaux differential of V_B is given by:*

$$\Gamma(X) = \frac{[w'(X)u(X) + w(X)u'(X)]}{\int w(X)dP} - \frac{\int w(X)u(X)dP}{[\int w(X)dP]^2} w'(X) \tag{18}$$

(Proof in appendix)

C. Example 3: MPEU and RDEU

In this section, we study the smoothness of MPEU. The MPEU model formalizes the idea of Knighten uncertainty and is of fundamental importance in nonexpected utility analysis. It is perhaps the most widely used nonexpected utility model in practice. For a recent survey on applications of MPEU models in asset pricing theory, game theory, contract theory and others, see Luo and Ma (1999). However, none of the previous local utility analysis literature addresses MPEU. We show that under fairly general conditions (no more stringent than the conditions under which expected utility is smooth), MPEU is *smooth*, and satisfies condition 1 and 2. We also characterize the sub-Gateaux derivative of MPEU. We show it is Gateaux differentiable on a dense G_δ[6] set of L_p.

Consider the following utility function defined on L_p:

$$V_M(X) = \inf_{\pi \in \Pi} \int u(X) d\pi \tag{19}$$

where Π is a convex set of probability measures on Ω. We assume $\forall \pi \in \Pi$, π is absolute continuous with respect to P, so that the set of probability measures Π can also be represented by densities, denote

$$M = \{\frac{d\pi}{dP} : \pi \in \Pi\} \tag{20}$$

Then V_M can be written as:

$$V_M(X) = \inf_{\phi \in M} \int \phi u(X) dP \tag{21}$$

The set of densities M is saided to be bounded if $\exists K > 0$, such that $\forall \phi \in M$, $\phi \leq K$; closed if it is closed under almost sure limit. Closedness and boundedness of M implies the inf in

[6]Recall a set is G_δ if it can be represented as intersection of countably many open sets.

(21) is always achieved[7], we can write:

$$V_M(X) = \min_{\phi \in M} \int \phi u(X) dP \tag{22}$$

Proposition 4. *Suppose u is bounded and Lipschitz continuous, and M is convex, closed and bounded, then the MPEU V_M is smooth, and satisfies condition 1 and 2. The subdifferential correspondence of V_M is given by:*

$$\Gamma(X) = \{\phi u'(X) : \phi \in \Psi(X)\} \tag{23}$$

where

$$\Psi(X) = \arg\min_{\phi \in M} \int \phi u(X) dP \tag{24}$$

(Proof in Appendix).

Corollary 2. Under the conditions of proposition 4, V_M is Gateaux differentiable on a dense G_δ set of L_p.

Proof. See lemma 6 in appendix.

Note MPEU cannot be Gateaux differential on L_p unless M is a singleton, in which case it reduces to the expected utility. The above proposition shows it is subGateaux differentiable and smooth under fairly general conditions. If u is concave, so that V is a concave function of random variables, results from convex analysis can be applied to show the existence of

[7] Closedness and boundedness of M implies it is compact in some properly chosen topology, thus the inf is always achieved. For proof of this, see the proof of lemma 4 in appendix B.

subdifferentials. However, our results do not rely on any concavity assumption. Finally, we consider the rank-dependent expected utility (RDEU) model (Quiggin (1982), Quiggin and Wakker (1994)) defined by:

$$V_R(X) = \int_R u(x) dg \circ F_X(x)$$

where $g : [0, 1] \to [0, 1]$ is increasing, concave, continuous and

$$g(0) = 1 - g(1) = 0$$

and F_X denote the distribution function of X. Since RDEU can be viewed as a special case of MPEU where the set of prior is the core of some convex distortion of the probability measure P (Schmeidler (1986)), RDEU is smooth and satisfies condition 1 and 2 by proposition 3. Chew et al (1987) showed that RDEU is not Fréchet differentiable in Machina (1982)'s sense, but it is Gateaux differentiable, therefore local expected utility analysis is still possible. Motivated by the fact that Fréchet differentiability yields strong results and allows for study of larger class of comparative static analysis, Wang (1993) introduced the notion of L_p Fréchet differentiability (with respect to distribution functions), and recovered the Fréchet differentiability of RDEU under a different notion of norm. Carlier and Dana (2003) argued that it is often more useful to know the derivative with respect to random variables in applications, and they characterize the set of Gateaux (sub)differentials of RDEU. Their results are based on the assumption that u is concave. The above proposition establishes subGateaux differentiability of RDEU without concavity. The following proposition characterizes the subdifferential correspondence of RDEU:

PROPOSITION 5. *Suppose u is strictly increasing, bounded and Lipschitz continuous, g' is bounded a.s on $[0,1]$, then V is smooth and satisfies condition 1 and 2. The subdifferential correspondence of V is given by:*

$$\Gamma(X) = co\left\{g'(\sigma) \cdot u'(X) : \sigma \text{ is measure preserving, and } X = X^* \circ \sigma\right\} \tag{25}$$

for all $X \in L_p$. Moreover, Γ is single-valued on a dense G_δ set of L_p.

(Proof in appendix)

In equation (25), *co* means convex hull. It is easy to verify that $F_X(X)$ is a measure preserving transformation that satisfies $X = X^* \circ F_X(X)$. See also Corollary 2 of Carlier and Dana (2003). Therefore

$$g'(F_X(X)) \cdot u'(X) \in \Gamma(X) \tag{26}$$

and $\Gamma(X) = g'(F_X(X)) \cdot u'(X)$ whenever $\Gamma(X)$ is a singleton. We also note $g'(F_X(X)) \cdot u'(X)$ is a measurable function of X.

There are certainly other nonexpected utility models that we do not discuss here, for example, the quadratic utility and the implicit weighted utility. However, they are all smooth under fairly general conditions. The analysis of these cases is very similar to the analysis we did for weighted utility above. We thus do not repeat here.

4. Smooth Nonexpected Utility Functions

In this section, we link global properties of nonexpected utility functions to its local properties, we show that the variational analysis of nonexpected utilities can be done under the smoothness condition and condition 1 and 2. No attempt is made to exhaust all applications of the variational analysis approach. Interested reader are referred to Chew and Mao (1995), Machina (1982, 1989), and Wang (1993), among others. We just remark here that although their theorems are stated for utility functions that satisfies probabilistic sophistication, as far as comparative statics analysis are concerned, they can all be reformulated in our settings as well.

The purpose of this section is to extend the differential characterizations of monotonicity, risk aversion of nonexpected utility in previous literature on variational analysis of nonexpected utilities to the class of smooth utility functions, and demonstrate why our notion of smoothness and condition 1 and 2 proposed in section 2 are sufficient for variational analysis of nonexpected utility functions. We give characterizations of probabilitistic sophistication (i.e. state independence), monotonicity, and risk aversion of the nonexpected utility functions in terms of properties of its Gateaux derivative. Utility functions are defined as functions of random variables. This formulation allows for state dependence. The notion of monotonicity and risk aversion Werner (2004) for state independent utilities can be generalized to the state dependent case. Our theorems are formulated to allow for this generality.

We consider smooth utility functions $V : L_p \to R$. Let $\Gamma : L_p \to L_q$ denote the subdifferential correspondence of V. We first reformulate Machina and Schmeidler (1992)'s definition of probabilistic sophistication in our context: A utility function $V : L_p \to R$ is probabilistically sophisticated if $\forall X, Y \in L_p$, X and Y have the same distribution implies $V(X) = V(Y)$.

The notion of probabilistic sophistication is proposed by Machina and Schmeidler (1992). Their purpose is to derive subjective probability without assuming decision maker's preference over lotteries conforms to the expected utility hypothesis. Previous literatures on local expected utility analysis deal with utility function defined on set of distribution functions, therefore assumes probabilistic sophistication automatically. On the other hand, some important non-expected utility models does not satisfy the probabilistic sophistication condition, for example, MPEU. It is therefore important to give a differential characterization of probabilistic sophistication in this more general setting.

Recall that for any smooth utility function $V : L_p \to R$, the sub-Gateaux derivative of V, is a correspondence from L_p to L_q. If $\gamma \in \Gamma$ is a selection of Γ, then $\forall X \in L_p$, $\gamma(X) \in L_q$. Probability sophistication implies more structure on Γ, as is stated in the following lemma:

Lemma 1. Suppose V is smooth, if V is probabilistically sophisticated, then $\forall \gamma \in \Gamma$, $E[\gamma(X)|X] \in \Gamma(X)$.

In general, $\forall X \in L_p$, $\gamma(X)$ may or may not be measurable with respect to (the completion of $\sigma-$field generated by) X. The above lemma implies that under the probability sophistication condition, $\Gamma(X)$ is closed with under taking conditional expectations with respect to X. Since $E[\gamma(X)|X]$ is measurable with respect to X, we immediately conclude that under probability sophistication, there is at least some $\gamma \in \Gamma$, such that $\gamma(X)$ is X measurable for all $X \in L_p$. This is summarized in the following corollary:

Corollary 3. If a smooth utility function V satisfies probabilistic sophistication, then $\exists \gamma \in \Gamma$ such that $\forall X \in L_p$, $\gamma(X)$ is $\overline{\sigma(X)}$ measurable, where $\overline{\sigma(X)}$ denote the $P-$completion of the σ field generated by X.

In the examples of section three, L_p-Fréchet differentiable utility functions, expected utility function, and weighted utility are all probability sophisticated and have a single valued subdifferential correspondence. In this case, the above lemma implies that the representation of the Gateaux differential at $X \in L_p$ is $\overline{\sigma(X)}$ measurable, and thus can be represented as a measurable function of X. It is easy to verify that their Gateaux differentials all satisfy this condition. See equation (13), (16), (18). RDEU also satisfies probabilistic sophistication, but the subdifferential correspondence is not single valued, however, the lemma implies there exist a selection of the subdifferential correspondence that satisfies the measurability condition. This selection is given in (26). There is no guarantee however, that every selection of Γ satisfy the measurability condition. In the case of RDEU, it is easy to construct elements in $\Gamma(X)$ that does not when $\Gamma(X)$ is not a singleton. MPEU, on the other hand, does not satisfy probability sophistication in general.

Under the conditions of lemma 1, if γ is a selection of Γ such that $\gamma(X)$ is $\overline{\sigma(X)}$ measurable, then $\gamma(X)$ can be represented as a measurable function of X, i.e. $\gamma(X) = \rho \circ X$ for some $\rho : R \to R$. We define such function ρ as the representation function of $\gamma(X)$:

Definition 5. *Let $V : L_p \longrightarrow R$ be a smooth utility function. Suppose for some $X \in L_p$, $l \in \Gamma(X)$, $l(\omega) = \rho \circ X(\omega)$ a.s. for some measurable function $\rho : R \to R$, then ρ is called the representation function of l.*

Suppose there exists a family of measurable functions $\varrho = \{\rho_X : X \in L_p, \rho_X : R \to R\}$ and a selection of Γ, $\gamma \in \Gamma$, such that $\forall X \in L_p$, $\gamma(X)(\omega) = \rho_X \circ X(\omega)$ a.s., then ϱ is called a system of representation functions of Γ.

If $\gamma(X) \in \Gamma(X)$ is $\overline{\sigma(X)}$ measurable, the representation function ρ_X exists. However, ρ_X may not be unique. In particular, it can take arbitrary values outside the range of X. However, it can be easily verified ρ_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 ρ_X is unique Q_X a.s., where Q_X denote the distribution of X. We will identify ρ_X as the equivalent class of $Q_X - a.s.$ equal functions. Under this convention, there is a unique representation function system for the sub-Gateaux differential correspondence of any smooth utility function.

For each $X \in L_p$, the representation function maybe different. We use the notation ρ_X to emphasize the dependence of ρ on X. For example, the representation function of weighted utility at $X \in L_p$ is

$$\frac{[w'(\cdot)u(\cdot) + w(\cdot)u'(\cdot)]}{\int w(X)dP} - \frac{\int w(X)u(X)dP}{[\int w(X)dP]^2}w'(\cdot)$$

and the system of representation functions is given by:

$$\left\{\frac{[w'(\cdot)u(\cdot) + w(\cdot)u'(\cdot)]}{\int w(X)dP} - \frac{\int w(X)u(X)dP}{[\int w(X)dP]^2}w'(\cdot) : X \in L_p\right\}$$

Similiarly, the form of representation functions of the Gateaux derivative of expected utility, L_p-Fréchet differentiable utility and RDEU are given in (16), (13) and (26), respectively. Expected utility is special in the sense that the representation function does not depend on the distribution of X. For L_p-Fréchet differentiable utility, weighted utility, and RDEU, the representation function ρ_X depends nontrivially on X. However, they depends on X only through its distribution.

The next proposition gives a characterization of probability sophistication for smooth utility functions. It turns out that for smooth utility functions that satisfy condition 1 and 2, probability sophistication is equivalent to existence of a representation function system, and representation functions being invariant to random variables with the same distribution. As we have noted above, expected utility, L_p-Fréchet differentiable utility, weighted utility, and RDEU all satisfy this condition. MPEU, on the other hand, does not satisfy this condition in general.

PROPOSITION 6. *Let $V : L_p \to R$ be smooth. If the system of representation functions ϱ satisfies: $\forall X, Y \in L_p$, X and Y have the same distribution implies $\rho_X = \rho_Y \ Q_X - a.s.$, then V is probability sophisticated. Suppose in addition V satisfies condition 1 and 2, then the converse of the above theorem also holds.*

(Proof in appendix)

If a system of representation functions exists for V, properties of the subdifferential γ can be stated in terms of the properties of the representation function ρ. The theorems in Machina (1982, 1989), Wang (1993) and Chew et al (1987) can viewed as linking global properties of utility function V to the properties of the representation function ρ. In fact, it may be easily checked that for L_p-Fréchet differentiable utilities, the representation function and the local expected function $u_F(\cdot)$ is related through: $u'_F(t) = \rho(t), \quad \forall t.$

Our theorems are concerned with properties of the subdifferential correspondence and do not rely on probabilistic sophistication therefore the existence of representation functions. However, whenever representation functions exist, our theorems can be formulated in terms of properties of the representation functions as well. We give characterizations of monotonicity, and risk aversion of utility functions in terms of properties of the sub-differential correspondence. Standard notions of monotonicity and risk aversion for nonexpected utilities are generalized to allow for state-dependent utilities as well. Our definition of risk aversion follows Werner (2004). Under probabilistic sophistication, we also give characterization of monotonicity and risk aversion in terms of properties of the representation functions.

Definition 6. *A utility function* $V : L_p \to R$ *is monotone if* $\forall X, Y \in L_p$, $V(X+Y) \geq V(X)$ *whenever* $Y \geq 0$ *a.s..*

Under probabilistic sophistication, the standard notion of monotonicity is the following: V is monotone if $V(X) \geq V(Y)$ whenever X first order stochastic dominate Y. Since X first order stochastic dominate Y if and only $\exists$ random variables $\widetilde{X}$,$\widetilde{Y}$ such that $\widetilde{X}$ and X have the same distribution, and $\widetilde{Y}$ and Y have the same distribution, and $\widetilde{X} \geq \widetilde{Y}$ a.s.. It follows immediately that our definition of monotonicity is equivalent to the standard definition under probabilistic sophistication. The following proposition gives a characterization of monotonicity without assuming probabilistic sophistication:

Proposition 7. *Suppose* $V : L_p \to R$ *is smooth. If* $\exists \gamma \in \Gamma$, *such that* $\forall X \in L_p$, $\gamma(X) \geq 0$ *a.s., then* V *is monotone. Suppose in addition,* V *satisfies condition 1 and 2, then the converse is also true.*

(Proof in appendix)

If V is probability sophisticated, the above theorem can be stated in terms of the representation function:

Corollary 4. Suppose $V : L_p \to R$ is smooth and satisfies probabilistic sophistication. Let ϱ be the system of representation functions for Γ. If $\forall \rho \in \varrho$, $\rho \geq 0$, then V is monotone. Suppose in addition, V satisfies condition 1 and 2, the converse is also true.

Proof. Follows directly from proposition 5.

It is clear from example 1, if the utility function is continuously L_1 Fréchet differentiable in the sense of Machina (1982), and the local expected utility is Lipschitz, then it is also smooth and satisfies condition 1 and 2. In this case, the representation is the (almost sure) first order derivative of the local expected utility function. Machina (1982) characterize monotonicity as the local expected utility being nondecreasing. Our proposition is thus a generalization of Machina's theorem.

We next provide equivalent characterizations of risk aversion. The notion of risk aversion can also be generalized to utility functions that do not satisfy the probabilistic sophistication condition. Our definition of risk aversion for state dependent utility functions follows Werner (2004):

Definition 7. *A utility function $V : L_p \to R$ is averse to mean independent risk if $\forall X, Z \in L_p$, such that $E(Z|X) = 0$ a.s., $V(X + \lambda Z) \geq V(X + Z)$ for all $0 \leq \lambda \leq 1$.*

If V satisfies probabilistic sophistication, the above definition is equivalent to the Rothchild-Stiglitz definition of risk aversion Rothchild and Stiglitz (1970), i.e., V is risk averse if $V(X) \geq V(Y)$ whenever X second order stochastic dominate Y. For further discussions of the notion of aversion to mean-independent risk, see Werner (2004). The following theorem characterizes mean-independent risk aversion in terms of properties of the subdifferential correspondence.

Proposition 8. *Suppose $V : L_p \to R$ is smooth. If $\exists \gamma \in \Gamma$, such that one of the following two conditions hold, then V is averse to mean-independent risk. Suppose in addition, V*

satisfies condition 1 and 2, then the converse is also true.

1) $\forall X \in L_p$, $\exists$ *a null set* $N \in \mathcal{F}$, *such that* $\forall \omega, \omega' \in \Omega$,

$$[X(\omega) - X(\omega')][\gamma(X)(\omega) - \gamma(X)(\omega')] \leq 0 \tag{27}$$

2) $\forall X \in L_p$, $\forall$ *sub* σ *field* $\mathcal{G} \subseteq \mathcal{F}$,

$$\int \gamma(X)[X - E(X|\mathcal{G})]dP \leq 0 \tag{28}$$

Condition (27) is sometimes called the "negative comonotone" condition that also appears, for example in Machina (1982) and Chew and Mao (1995). Thus the above proposition is a generalization of those characterizations of risk aversion to the state dependent case, and to a weaker notion of differentiability. Under probabilistic sophistication, mean-independent risk aversion reduces to the usual Rothchild-Stiglitz notion of risk aversion. We give additional characterizations of risk aversion for this case.

Corollary 5. Suppose $V : L_p \to R$ is smooth and satisfies probabilistic sophistication. Let ϱ be the system of representation functions, and let $\gamma \in \Gamma$ be generated by ϱ, i.e. $\forall X \in L_p$, $\gamma(X) = \rho_X(X)$. V is Rothchild-Stiglitz risk averse if one of the following two conditions hold. If in addition, V satisfies condition 1 and 2, then the converse is also true.

1)$\forall X \in L_p$, $\forall \rho_X \in \varrho$, ρ_X is nonincreasing.

2)$\forall X, Y \in L_p$, X and Y have the same distribution implies

$$\int [\gamma(X) - \gamma(Y)](X - Y)dP \leq 0 \tag{29}$$

Proof. Condition 1) is just a restatement of (27) in terms of the representation functions. Proof of condition 2) is in appendix.

The first condition says Rothchild-Stiglitz risk aversion is equivalent to there existing a system of nonincreasing representation functions. In the expected utility case, this reduces to the first order derivative of von Neumann-Morgenstern utility function being a decreasing function. Condition (29) is new. It says the subdifferential of V is a negatively monotone operator on the set of random variables that has the same distribution. Recall a function is concave if and only if the sub-Gateaux derivative is a negatively monotone operator (See Phelps (1993)). Note also, a probabilistic sophisticated utility function $V : L_p \to R$ is Rothchild-Stiglitz risk averse if and only if is quasiconcave on the set of random variables that has the same distribution (Ai (2004)). Condition (29) is thus a differential analogue of the above result.

The above theorems extended the variational analysis in Chew and Mao (1995), Machina (1982, 1989), and Wang (1993) to smooth nonexpected utility functions. Many nonexpected utility functions that are neither Gateaux differentiable nor concave satisfies our smoothness condition. In fact, as was shown in section 2, most of the nonexpected utility models in the existing literature satisfy smoothness under fairly general conditions, including MPEU, which is not differentiable under any notion of differentiability used in this literature. Our approach also allow for state dependent utility functions. We do not attempt to reformulate all the theorems appeared in this literature in our setting, however, this can be done in an analogous fashion as we did here.

5. Conclusion

We propose a notion of smoothness of nonexpected utility functions. Our notion of smoothness allows the type of variational analysis of nonexpected utility models proposed by Machina (1982) to be performed under very general conditions. In particular, our formulation allows for state dependent utilities, as well as the MPEU model. Other nonexpected utility models, are shown to satisfy our smoothness condition under more general conditions than Fréchet differentiability and Gateaux differentiability used in the literature. We analyze the properties of the subdifferential correspondence of nonexpected utility functions, and show how the type of variational arguement in Machina (1982, 1989), Wang (1993), and Chew et al (1987) among others, can be applied to smooth utility functions. We also give characterizations of monotonicity and risk aversion without assuming probabilistic sophistication. We give a careful analysis of the subdifferential correspondence of MPEU. We establish subdifferentiability of MPEU without concavity assumption. We also show that MPEU is Gateaux differential on a dense subset of L_p.

References

[1] Ai, H. (2004): Risk Aversion without Independence Axiom, working paper.

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

[3] Carlier,G and R.A. Dana. (2003), Core of Convex Distortions of Probability, *Journal of Economic Theory* 113, 199-222.

[4] Chew, S.H. (1983), A Generalization of the Quasilinear mean with applications to the

measurement of income inequality and decision theory resolving the Allais paradox, *Econometrica* 51,, 1065-1092.

[5] Chew, S.H. (1989), Axiomatic Utility Theories with the Betweenness Property, *Annals of Operations Research*. 19, 273-298.

[6] Chew, S.H., E. Karni and Z. Safra. (1987), Risk Aversion in the Theory of Expected Utility with Rank Dependent Probabilities. *Journal of Economic Theory*, 42, 370-381.

[7] Chew, S.H. and M.H. Mao (1995), A Shur Concave Characterization of Risk Aversion for Non-expected Utility Preferences, *Journal of Economic Theory*, 67, 402-435.

[8] Dekel, E. (1986), An Axiomatic Characterization of Preference under Uncertainty: Weakening the Independence Axiom, *Journal of Economic Theory* 40, 304-318.

[9] Gilboa, I. and D. Schmeidler (1989): Maxmin Nonexpected Utility with Nonunique Prior, *Journal of Mathematical Economics*.

[10] Lou, X. and C. Ma (1999) Recent Advancements in the Theory of Choice under Knighten Uncertainty and Their Applications in Economics, working paper.

[11] Kechris, A.S. (1995), "*Classical Descriptive Set Theory*", Springer.

[12] Lunberger, D.G. (1969), "*Optimization by Vector Space Methods*", New York: John Wiley and Sons, Inc.

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

[14] Machina, M.J. (1989), Comparative Statics and Nonexpected Utility Analysis, *Journal of Economic Theory* 47, 393-405.

[15] Machina, M.J. and D. Schmeidler (1992): A More Robust Definition of Subjective Expected Utility. *Econometrica.*

[16] Muller, A. and D. Stoyan. (2002), "*Comparison Methods for Stochastic Models and Risks*", John Wiley and Sons, Inc.

[17] Phelps, R.R (1993), "*Convex Functions, Monotone Operators and Differentiability*", Springer-Verlag.

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

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

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

[21] Rudin, W. (1974), *Real and Complex Analysis*, McGraw-Hill, second edition.

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

[23] Schmeidler, D. (1986), Integral representation without additivity, *Proc. AMS* 97 (2), 255-261.

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

[25] Wang, T (1993). L_p-Fréchet Differentiable Preference and Local Utility Analysis, *Journal of Economic Theory*. 61, 139-159.

[26] Werner, J.(2004), Risk and Risk Aversion When States of Nature Matter, working paper.

Appendix

A1. Proof of proposition 1 and 3.

In the proof of the propositions, we will make frequent use of the following lemma:

LEMMA 2. Suppose $u : R \to R$ is Lipschitz continuous, then $\forall X, Y \in L_p$,

$$\lim_{t\to 0} \frac{1}{t} \int [u(X+tY) - u(X)]dP = \int u'(X)YdP \tag{A1}$$

Proof. First note Lipschitz continuity imply absolute continuity of u, we have u' exists almost surely, and $\forall x, y \in R$,

$$u(y) - u(x) = \int_x^y u'(t)dt \tag{A2}$$

Note also, u' is bounded by the Lipschitz constant K. The left hand side of (A1) can be written as

$$\begin{aligned} & \lim_{t\to 0} \frac{1}{t} \int \int_0^t u'(X+\theta Y)YdtdP \\ = & \lim_{t\to 0} \frac{1}{t} \int_0^t \int u'(X+\theta Y)YdPdt \\ = & \int u'(X)YdP \end{aligned}$$

The first equality is Fubini's theorem. This is valid because

$$\int_0^t \int |u'(X+\theta Y)Y|dPdt \le \int_0^t \int K|Y|dPdt = tKE|Y| < \infty$$

The second equality is because as $\theta \to 0$,

$$\int u'(X+\theta Y)YdP \to \int u'(X)YdP$$

By dominated convergence theorem. Since for every θ, $u'(X+\theta Y)Y$ is dominated by the integrable function $K|Y|$. This proves the lemma.

Proof of Propostion 1:

First, assume U is Fréchet differentiable with respect to distribution functions. We first prove V is Gateaux differentiable with respect to random variables, i.e.

$$\lim_{\alpha\to 0}\frac{1}{\alpha}[V(X+\alpha Y)-V(X)] = \int u'_F(X)\cdot YdP \qquad \text{(A3)}$$

Let F_α denote the distribution function of $X+\alpha Y$, We have

$$\begin{aligned} V(X+\alpha Y)-V(X) &= U(F_\alpha)-U(F) \\ &= U'(F)(F_\alpha - F) + o||F_\alpha - F|| \end{aligned} \qquad \text{(A4)}$$

Note

$$o||F_\alpha - F|| = o|\alpha| \qquad \text{(A5)}$$

To see this, note X and Y are both bounded by M, we have

$$F_\alpha(x) = P(X+\alpha Y\le x)\le P(X-\alpha M\le x) = F(x+\alpha M)$$

similarly, $F_\alpha(x) \geq F(x - \alpha M)$. Also, $F(x-\alpha M) \leq F(x) \leq F(x+\alpha M)$ therefore,

$$
\begin{aligned}
||F_\alpha - F|| &\leq \int [F(x+\alpha M) - F(x-\alpha M)]\, dx \\
&= \int_{-\infty}^{+\infty} \int_{x-\alpha M}^{x+\alpha M} 1 dF(t) dx \\
&= \int_{-\infty}^{\infty} \int_{t-\alpha M}^{t+\alpha M} dx dF(t) \\
&= 2\alpha M \qquad \text{(A6)}
\end{aligned}
$$

This proves (A5).

To see Gateaux differentiability, note

$$
\begin{aligned}
U'(F)(F_\alpha - F) &= \int u_F(x) d(F_\alpha - F) \\
&= \int u_F(X+\alpha Y) dP - \int u_F(X) dP \qquad \text{(A7)}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\lim_{\alpha \to 0} \frac{1}{\alpha}[V(X+\alpha Y) - V(X)] &= \lim_{\alpha \to 0} \frac{1}{\alpha}[\int u_F(X+\alpha Y) - u_F(X) dP] \\
&= \int u_F'(X) Y dP \qquad \text{(A8)}
\end{aligned}
$$

by lemma 2. This establishes the Gateaux differentiability and hence condition 1.

To prove the second part of the theorem, suppose U is continuously differentiable, we need to prove V is smooth and satisfies condition 2. We first prove condition 2. Since the subdifferential correspondence is single valued, this amounts to proving the following:

$X_n \to X$ in L_p implies

$$\int u'_{F_n}(X_n)YdP \to \int u'_F(X)YdP \tag{A9}$$

for every $Y \in RV([-M, M])$, where F_n is the distribution function of X_n, and F is the distribution function of X.

To see (A9) is true, note $X_n \to X$ in L_p implies $F_n \to F$ weakly, i.e. $F_n(t) \to F(t)$ whenever F is continuous at t. Note also, F is an nondecreasing function, hence the set of discontinuity of F is of measure 0. Therefore

$$\int |F_n(t) - F(t)|dt \to 0$$

by bounded convergence, i.e. $F_n \to F$ in L_1. Continuous Fréchet differentiability of U implies $u_{F_n} \to u_F$ pointwise[8]. Note $u_{F_n} \to u_F$ pointwise implies $u'_{F_n} \to u'_F$ a.s. To see this, $u_{F_n} \to u_F$ point wise implies $\forall x, y \in R$

$$\left| \int_x^y u'_{F_n}(t)dt - \int_x^y u'_F(t)dt \right| \to 0$$

Therefore

$$\left| \int_A \left[u'_{F_n}(t) - u'_F(t) \right] dt \right| \to 0$$

On any A with positive measure, i.e. $u'_{F_n} \to u'_F$ a.s.. To prove (A9), note $X_n \to X$ in L_p implies $X_n \to X$ a.s., togehter with $u'_{F_n} \to u'_F$ a.s., (A9) can be obtained by applying the bounded convergence theorem.

[8]Continuous Frechet differentiability of U implies $u_{F_n} \to u_F$ in the $|| \cdot ||_p^*$ norm defined in Wang (1993), in particular, this implies convergence almost surely. Continuity of u_F implies convergence is in fact point wise.

To prove smoothness, we need to show the absolute continuity of $g(t) = V(X+tY)$ on $[0,1]$. We prove absolute continuity by showing $g(t)$ is continuously differentiable on $[0,1]$. By equation (A3), for $t_n \to t$, $t_n \in [0,1]$, all n,

$$\begin{aligned} g'(t) &= \int u'_{F_t}(X+tY) \cdot Y dP \\ g'(t_n) &= \int u'_{F_n}(X+t_nY) \cdot Y dP \end{aligned}$$

where F_t and F_n are the distribution function of $X+tY$ and $X+t_nY$, respectively. As $t_n \to t$, $X+t_nY \to X+tY$ in L_p, by the proof of condition 2, $g'(t_n) \to g'(t)$, as needed.

Proof of the Corollary:

All the above aguments go through except the proof of equation (A5), where the norm $||F_\alpha - F||$ is understood as the L_p norm defined in (10). The path $F_{X+\alpha Y}(\cdot)$ is weakly smooth implies

$$\begin{aligned} \lim_{\alpha \to 0} \frac{||F_\alpha - F||}{\alpha} &\leq \left\{ \int \left[\frac{|F_{X+\alpha Y}(x) - F_X(x)|}{\alpha} \right]^p dx \right\}^{\frac{1}{p}} \\ &\leq C \end{aligned}$$

for some constant C by weak smoothness of the path.

Proof of proposition 3:

We first prove V_B is Gateaux differentiable and $\forall Y \in L_p$

$$\lim_{\alpha\to 0}\frac{1}{\alpha}[V_B(X+\alpha Y)-V_B(X)] = \int \Gamma(X) Y dP \tag{A10}$$

where $\Gamma(X)$ is given by (18). Denote

$$X^t = X + tY \tag{A11}$$

We have:

$$\begin{aligned} & V_B(X+tY)-V_B(X) \\ = \ & \frac{\int w(X^t)u(X^t)dP}{\int w(X^t)dP} - \frac{\int w(X)u(X)dP}{\int w(X)dP} \\ = \ & \frac{\int w(X^t)u(X^t)dP \cdot \int w(X)dP - \int w(X^t)dP \cdot \int w(X)u(X)dP}{\int w(X^t)dP \cdot \int w(X)dP} \end{aligned} \tag{A12}$$

The numerator in (A12) can be written as:

$$\begin{aligned} & \int w(X^t)u(X^t)dP \cdot \int w(X)dP - \int w(X)dP \cdot \int w(X)u(X)dP \\ & + \int w(X)dP \cdot \int w(X)u(X)dP - \int w(X^t)dP \cdot \int w(X)u(X)dP \\ = \ & \int \Big[w(X^t)u(X^t) - w(X)u(X)\Big]\, dP \cdot \int w(X)dP \qquad \text{(A13)} \\ & + \int \Big[w(X) - w(X^t)\Big]\, dP \cdot \int w(X)u(X)dP \qquad \text{(A14)} \end{aligned}$$

Therefore the left hand side of (A10) is:

$$\begin{aligned}
& \lim_{t\to 0}\frac{1}{t}\left\{\frac{\int\left[w(X^t)u(X^t)-w(X)u(X)\right]dP\cdot\int w(X)dP}{\int w(X^t)dP\cdot\int w(X)dP}\right. \\
& \left.-\frac{\int\left[w(X^t)-w(X)\right]dP\cdot\int w(X)u(X)dP}{\int w(X^t)dP\cdot\int w(X)dP}\right\} \\
= & \frac{\int_\Omega\left[w'(X)u(X)+w(X)u'(X)\right]YdP\cdot\int w(X)dP-\int_\Omega w'(X)YdP\cdot\int w(X)u(X)dP}{\left[\int w(X)dP\right]^2} \\
= & \int\left\{\frac{\left[w'(X)u(X)+w(X)u'(X)\right]}{\int w(X)dP}-\frac{\int w(X)u(X)dP}{\left[\int w(X)dP\right]^2}w'(X)\right\}YdP
\end{aligned}$$

The first equality is by lemma 2, note Lipschitz continuity and boundedness of u and w implies uw is Lipschitz. Therefore V_B is Gateaux differentiable, and the Gateaux differential is given by (18). The Gateaux differential is an element of L_p by boundedness of u and w.

To prove smoothness, need to show $g(t)=V(X+tY)$ is absolutely continuous. We prove a slightly stronger result, i.e. g is Lipschitz continuous. Take $t,s\in[0,1]$,

$$\begin{aligned}
g(t)-g(s) &= \frac{\int w(X^t)u(X^t)dP}{\int w(X^t)dP}-\frac{\int w(X^s)u(X^s)dP}{\int w(X^s)dP} \\
&= \frac{\int w(X^t)u(X^t)dP\cdot\int w(X^s)dP-\int w(X^s)u(X^s)dP\cdot\int w(X^t)dP}{\int w(X^t)dP\cdot\int w(X^s)dP} \\
&= \frac{\int\left[w(X^t)u(X^t)-w(X^s)u(X^s)\right]dP\cdot\int w(X^s)dP}{\int w(X^t)dP\cdot\int w(X^s)dP} \\
&\quad +\frac{+\int w(X^s)u(X^s)dP\cdot\int\left[w(X^s)-w(X^t)\right]dP}{\int w(X^t)dP\cdot\int w(X^s)dP} \qquad \text{(A15)}
\end{aligned}$$

Note w,u and $w\cdot u$ are all Lipschitz continuous. Let K denote the upperbound on u, w, and the Lipschitz constant of u and w, and wu, We have

$$\left|w(X^t)u(X^t)-w(X^s)u(X^s)\right| \le K|X^t-X^s|$$

$$= \quad K|t-s||Y| \tag{A16}$$

Similarly,

$$\left|w(X^s) - w(X^t)\right| \leq K|t-s||Y|$$

Therefore,

$$\begin{aligned} |g(t)-g(s)| &\leq \frac{K|t-s| \int |Y|dP \cdot K}{\varepsilon} + \frac{K^2 \cdot K|t-s| \int |Y|dP}{\varepsilon} \\ &\leq \bar{K}|t-s| \end{aligned} \tag{A17}$$

for some constant $\bar{K}$, as needed.

Next, we need to verify condition 1 and 2. Condition 1 is trivial since V_B is Gateaux differentiable everywhere. To see condition 2 is true, we need to verify $X_n \to X$ in L_p implies $\gamma(X_n) \to \gamma(X)$ in the weak* topology, i.e. $\forall Y \in L_p$,

$$\int \gamma(X_n) Y dP \to \int \gamma(X) Y dP \tag{A18}$$

Using (18), boundedness of u,w,u', w' implies $\gamma(X_n) \to \gamma(X)$ a.s. and $\gamma(X_n)$ are bounded by some constant K for all n. Hence $\gamma(X_n)Y$ are dominated by $K|Y|$, (A18) is true by dominated convergence.

A2. Proof of proposition 4 and 5.

We establish the smoothness of MPEU through several lemmas.

Lemma 3. $\forall X, Y \in L_p$, define

$$g(t) = V_M(X + tY) \text{ on } t \in [0,1] \tag{A19}$$

then g is absolutely continuous.

Proof. We prove absolute continuity by verifying Lipschitz continuity. We define X^t as in (A11). Define the correspondence $\Psi : L_p \to M$ as in (24), and for $t \in [0,1]$, let $\phi_t \in \Psi(X^t)$. First, take $t, s \in [0,1]$

$$\begin{aligned}
g(t) - g(s) &= \int \phi_t u(X^t) dP - \int \phi_s u(X^s) dP \\
&\geq \int \phi_t \left[u(X^t) - u(X^s) \right] dP \\
&\geq -\int \phi_t \left| u(X^t) - u(X^s) \right| dP \\
&\geq -\int K \cdot K \left| X^t - X^s \right| dP \\
&\geq -\int K^2 |t-s| \cdot |Y| dP \\
&\geq -K^2 |t-s| \int |Y| dP
\end{aligned} \tag{A20}$$

The second line of (A20) is by definition of ϕ_s. The fourth line is by boundedness of M and the Lipschitz continuity of u, where K is both the bound on M, and the Lipschitz constant for u. Similarly, we have:

$$\begin{aligned}
g(t) - g(s) &\leq \int \phi_s \left[u(X^t) - u(X^s) \right] dP \\
&\leq \int \phi_s K |X^t - X^s| dP
\end{aligned}$$

$$\leq \quad K^2|t-s|\int |Y|dP \tag{A21}$$

Combining (A20) and (A21), we have

$$|g(t)-g(s)| \leq K^2|t-s|\int |Y|dP$$

as needed.

Lemma 4. Define the correspondence $\Psi : L_p \to M$ as in (24), then Ψ is L_p to weak* upper hemi-continuous and compact valued, where the weak* topology is generated by the following convergence concept:

$$\phi_n \to \phi \text{ weak }^* \text{ if } \int \phi_n X dP \to \int \phi X dP \text{ for every } X \in L_1$$

Proof. We first prove the function $F(\phi, X) = \int \phi u(X)dP$ is continuous in the weak* $\times$ L_p topology. To see this, Take $(\phi_n, X_n) \to (\phi, X)$, we have:

$$\begin{aligned}
& \left|\int \phi_n u(X_n)dP - \int \phi u(X)dP\right| \\
\leq \; & \left|\int \phi_n u(X_n)dP - \int \phi_n u(X)dP\right| + \left|\int \phi_n u(X)dP - \int \phi u(X)dP\right| \\
\leq \; & \int \phi_n \left|u(X_n) - u(X)\right| dP + \int |\phi_n - \phi| \, |u(X)|dP
\end{aligned} \tag{A22}$$

Note $\phi_n \to \phi$ in weak* implies $\phi_n \to \phi$ a.s.[9], $X_n \to X$ in L_p implies convergence a.s..

[9] To see this, not convergence in weak* implies $\int \phi_n I_A dP \to \int \phi I_A dP$ for every measurable $A \in \mathcal{F}$, where

Therefore the two terms in the last line of (A22) both converge to 0 by bounded convergence.

Next, we prove M is compact in the weak* topology. To see this, enough to show it is a closed subset of some weak* compact set. Closedness is by assumption. Note $\forall \phi \in M$,

$$\frac{|\int \phi X dP|}{\int |X| dP} \leq B$$

Hence M is a subset of $\{l \in dual(L^1) : ||l|| \leq B\}$, the later is a compact set by Alaoglu's theorem (Luengerger (1969), theorem 1, page 128). Since M is compact, and F is continuous, Ψ is u.h.c. and compact valued by Berge's theorem.

LEMMA 5. $\forall X, Y \in L_p$, let g be defined as in (A19), then $\forall t \in [0, 1]$, $g'_+(t)$ and $g'_-(t)$ exists, and

1) $\forall \phi \in \Psi(X + tY)$

$$g'_+(t) \leq \int \phi u'(X + tY) Y dP \leq g'_-(t) \tag{A23}$$

2) $\exists \phi_+, \ \phi_- \in \Psi(X + tY)$ such that

$$g'_+(t) = \int \phi_+ u'(X + tY) Y dP \tag{A24}$$

and

$$g'_-(t) = \int \phi_- u'(X + tY) Y dP \tag{A25}$$

I_A is the indicator function. Take $A = \{\omega : \limsup \phi_n > \phi\}$, $\int \phi_n I_A dP \rightarrow \int \phi I_A dP$ implies $P(A) = 0$. Similarly one can show $P(\liminf \phi_n < \phi) = 0$. This proves convergence almost surely.

Proof. Let X^t be defined as in (A11), we first prove

$$\limsup_{h\to 0^+} \frac{1}{h}[g(t+h)-g(t)] \le \int \phi_t u'(X^t) Y dP \quad \text{for every } \phi_t \in \Psi(X^t) \tag{A26}$$

To see this, $\forall h > 0$,

$$\begin{aligned} \frac{1}{h}[g(t+h)-g(t)] &= \frac{1}{h}\left[\int \phi_{t+h} u(X^{t+h}) dP - \int \phi_t u(X^t) dP\right] \\ &\le \frac{1}{h}\left[\int \phi_t [u(X^t + hY) - u(X^t)] dP\right] \end{aligned} \tag{A27}$$

The second line is by definition of ϕ_{t+h}. By lemma 2 the last line has a limit as $h \to 0^+$, we have

$$\lim_{h\to 0^+} \frac{1}{h}\left[\int_0^h \int \phi_t u'(X^t + \theta Y) Y dP d\theta\right] = \int \phi_t u'(X^t) Y dP$$

Taking limsup on both sides (A27) gives us (A26).

By similar arguement, one can prove

$$\liminf_{h\to 0^-} \frac{1}{h}[g(t+h)-g(t)] \ge \int \phi_t u'(X^t) Y dP \quad \text{for every } \phi_t \in \Psi(X^t) \tag{A28}$$

Next, we will prove:

$$\limsup_{h\to 0^+} \frac{1}{h}[g(t+h)-g(t)] = \liminf_{h\to 0^+} \frac{1}{h}[g(t+h)-g(t)] \tag{A29}$$

and

$$\limsup_{h\to 0^-} \frac{1}{h}[g(t+h)-g(t)] = \liminf_{h\to 0^-} \frac{1}{h}[g(t+h)-g(t)] \tag{A30}$$

To see (A29), note for $h > 0$,

$$\begin{aligned}\frac{1}{h}[g(t+h)-g(t)] &\geq \frac{1}{h}\left[\int \phi_{t+h}[u(X^t+hY)-u(X^t)]dP\right] \\ &= \frac{1}{h}\left[\int_0^h \int \phi_{t+h}u'(X^t+\theta Y)YdPd\theta\right] \qquad \text{(A31)}\end{aligned}$$

by similar arguement as in (A27). Since M is compact, as $h \to 0^+$, $\phi_{t+h} \to \phi_+$ for some $\phi_+ \in M$. Note $h \to 0$ and $Y \in L_p$ implies $X^t + hY \to X^t$ in L_p. Upper hemicontinuity of Ψ then implies $\phi_+ \in \Psi(X^t)$. As $\theta \to 0^+$,

$$\int \phi_{t+h}u'(X^t+\theta Y)YdP \to \int \phi_+ u'(X^t)YdP$$

by dominated convergence. Taking liminf on both sides of (A31), we have

$$\liminf_{h\to 0^+} \frac{1}{h}[g(t+h)-g(t)] \geq \int \phi_+ u'(X^t)YdP \text{ with } \phi_+ \in \Psi(X^t) \qquad \text{(A32)}$$

Compare (A32) with (A26), we get (A29). This implies $g'_+(t)$ exists and the left inequality in (A23) is true. At the same time, we proved $\exists \phi_+ \in \Psi(X^t)$ such that equation (A24) is true. The rest part of theorem can be proved in a similar way.

Combining lemma 3-5, we established V is sub-Gateaux differentiable everywhere and the subdifferential correspondence is given by $\Gamma(X) = \{\phi u'(X) : \phi \in \Psi(X)\}$. The following two lemmas establish condition 1 and 2, respectively:

Lemma 6. V_M is Gateaux differentiable on a dense G_δ set of L_p.

Proof. First note L_p is seperable (Rudin (1974), page 97). Let $\{\xi_n\}_{n=1}^{\infty}$ be a countable dense subset of L_p. Let N be the set of natural numbers, for each $m, n \in N$ define

$$A_{m,n} = \{X \in L_p : \exists \phi, \phi^{'} \in \Psi(X) \text{ such that } \int (\phi - \phi')u'(X)\xi_n dP \geq \frac{1}{m}\}$$

Since $\{\xi_n\}$ is dense, and $\int (\phi - \phi')u'(X)YdP$ is continuous in Y, $\Psi(X)$ is not a singleton if and only if $X \in A_{m,n}$ for some $m, n \in N$, i.e. $X \in \cup_{m=1}^{\infty} \cup_{n=1}^{\infty} A_{m,n}$. We first prove $\forall m, n$, $A_{m,n}$ is closed. To see this, take $\{Z_k\}_{k=1}^{\infty} \subseteq A_{m,n}$, and $Z_k \to \bar{Z}$, need to show $\bar{Z} \in A_{m,n}$. $\forall k$, $Z_k \in A_{m,n}$ implies $\exists \phi_k, \phi_k^{'} \in \Psi(Z_k)$ such that $\int (\phi_k - \phi_k')u'(Z_k)\xi_n dP \geq \frac{1}{m}$. Note M is compact, therefore $\phi_k \to \phi$ and $\phi_k' \to \phi'$ at least along some subsequence. Since Ψ is u.h.c. and compact valued, $\phi, \phi' \in \Psi(\bar{Z})$. Therefore

$$\int (\phi_k - \phi_k')u'(Z_k)\xi_n dP \geq \frac{1}{m}$$

implies

$$\int (\phi - \phi')u'(\bar{Z})\xi_n dP \geq \frac{1}{m}$$

by dominated convergence. This shows $A_{m,n}$ is closed. Therefore $L_p \backslash A_{m,n}$ is open. Next we show $\forall m, n$, $L_p \backslash A_{m,n}$ is dense in L_p. Take any $X \in L_p$, we need to construct $X_k \to X$ in L_p, and $X_k \in L_p \backslash A_{m,n}$ for all k. Define

$$g(t) = V_M(X + t\xi_n) \text{ on } t \in [0, 1]$$

By lemma 1, $g'(t)$ exists a.s. on $[0, 1]$. Pick a sequence of $\{t_k\}$ such that $t_k \to 0$ and $g'(t_k)$

exists, we have, by lemma ,

$$g'(t_k) = \int \phi u'(X + t_k\xi_n)\xi_n dP$$

for all $\phi \in \Psi(X+t_k\xi_n)$, i.e. $X+t_k\xi_n \in L_p\backslash A_{m,n}$ for all m. Note $t_k \to 0$ implies $X+t_k\xi_n \to X$ in L_p. We have $L_p\backslash A_{m,n}$ is dense in L_p. For each m, n, $L_p\backslash A_{m,n}$ is dense and open in L_p, also L_p is a complete metric space, we have $\cap_{m=1} \cap_{n=1} \{L_p\backslash A_{m,n}\} = L_p\backslash\{\cup_{m=1}^{\infty} \cup_{n=1}^{\infty} A_{m,n}\}$ is dense in L_p by Baire's theorem (Rudin (1974), page 102). $\mathcal{D} = L_p\backslash\{\cup_{m=1}^{\infty} \cup_{n=1}^{\infty} A_{m,n}\}$ is the dense subset on which Γ is single valued. This proves condition 1. It is also clear by construction that $\mathcal{D}$ is G_δ (intersection of countably many open sets).

Lemma 7. Γ is u.h.c. and compact valued.

Proof. This follows directly from the fact that Ψ is u.h.c. and compact valued and the fact that M and u' are both bounded.

Proof of proposition 5:

The form of the Gateaux differential follows from Carlier and Dana (2003)'s characterization of core of g. See theorem 5, and corollary 2 and 3 of Carlier and Dana (2003). The rest follows proposition 3.

A3. Proof of lemma 1 and proposition 6.

Proof of lemma 1:

Let $T : (\Omega, \mathcal{F}, P) \rightarrow (\Omega, \mathcal{F}, P)$ be the measure preserving isomorphism such that $\mathcal{I}_T = \sigma(X)$, where $\mathcal{I}_T$ denote the invariant σ field associated with T, and the equality is interpretated the two $\sigma-$field differ only by sets of measure 0. For existence of such measure preserving transformation see exercise 17.43 in Kechris (1995).

First, if $l \in \Gamma(X)$, then so is $l \circ T$. To see this, note $l \in \Gamma(X)$ implies $\forall Y \in L_p$, either

$$\lim_{\alpha \to 0^+} \frac{1}{\alpha}[V(X+\alpha Y \circ T^{-1}) - V(X)] \leq \int lY \circ T^{-1} dP \leq \lim_{\alpha \to 0^-} \frac{1}{\alpha}[V(X+\alpha Y \circ T^{-1}) - V(X)] \quad \text{(A33)}$$

or

$$\lim_{\alpha \to 0^+} \frac{1}{\alpha}[V(X+\alpha Y \circ T^{-1}) - V(X)] \geq \int lY \circ T^{-1} dP \geq \lim_{\alpha \to 0^-} \frac{1}{\alpha}[V(X+\alpha Y \circ T^{-1}) - V(X)] \quad \text{(A34)}$$

We first assume (A33) is the case, then $\forall \alpha$,

$$\begin{aligned} V(X+\alpha Y) - V(X) &= V(X \circ T + \alpha Y) - V(X \circ T) \\ &= V(X + \alpha Y \circ T^{-1}) - V(X) \end{aligned} \quad \text{(A35)}$$

The first line of (A35) is true since $\mathcal{I}_T = \sigma(X)$ implies $X = X \circ T$ a.s., the second line is because T is a measure preserving isomorphism. Note also

$$\int lY \circ T^{-1} dP = \int l \circ T Y dP$$

Therefore (A33) becomes

$$\lim_{\alpha\to 0^+}\frac{1}{\alpha}[V(X+\alpha Y)-V(X)] \leq \int l\circ T\cdot Y dP \leq \lim_{\alpha\to 0^-}\frac{1}{\alpha}[V(X+\alpha Y)-V(X)] \tag{A36}$$

If in stead, (A34) is true, we have

$$\lim_{\alpha\to 0^+}\frac{1}{\alpha}[V(X+\alpha Y)-V(X)] \leq \int l\circ T\cdot Y dP \leq \lim_{\alpha\to 0^-}\frac{1}{\alpha}[V(X+\alpha Y)-V(X)] \tag{A37}$$

Therefore, $\forall Y \in L_p$, either (A36) or (A37) is true, we have $l\circ T \in \Gamma(X)$, as needed.

Next, since $\Gamma(X)$ is convex and closed, we have $\forall n$,

$$\frac{1}{n}\sum_{j=1}^{n} l\circ T^n \in \Gamma(X)$$

Also, $\forall n$, $l\circ T^n \in L_q \subseteq L_1$, therefore

$$\frac{1}{n}\sum_{j=1}^{n} l\circ T^n \to E[l|\mathcal{I}_T] = E[l|X] \text{ a.s. and in } L_1 \tag{A38}$$

The convergence is due to Birkhauf's ergodic theorm, $E[l|\mathcal{I}_T] = E[l|X]$ bacause $\mathcal{I}_T = \sigma(X)$ by construction. Because $\forall n$, $\frac{1}{n}\sum_{j=1}^{n} l\circ T^n$ are dominated by l (trivially), the convergence in (A38) is also in L_q, which in turn implies weak* convergence, closedness of $\Gamma(X)$ therefore implies $E[l|X] \in \Gamma(X)$.

Proof of proposition 6

First, suppose V is smooth, and $\exists$ a system of representation functions ϱ such that X and Y have the same distribution implies $\rho_X = \rho_Y$, we need to show V is probability sophisticated. First we show if T is a measure preserving transformation on $(\Omega, \mathcal{F}, P)$, then $V(X) = V(X \circ T)$. To see this, let θ denote the r.v. that is 0 a.s. We have:

$$\begin{aligned} V(X \circ T) &= V(\theta) + \int_0^1 \int \rho_{tX \circ T}(tX \circ T) X \circ T dP dt \\ &= V(\theta) + \int_0^1 \int \rho_{tX \circ T}(tX) X dP dt \\ &= V(\theta) + \int_0^1 \int \rho_{tX}(tX) X dP dt \\ &= V(X) \end{aligned} \tag{A39}$$

The second line is because T is measure preserving. The thrid line true since tX and $tX \circ T$ have the same distribution. To prove the proposition, we recall Ryff's theorem: For any random variable Z, one can define the nondecreasing rearrangement of Z by:

$$Z^* = \inf\{x \in R : F_Z(x) \geq \omega\}$$

Ryff's theorem (see Ryff (1970)) states that $\exists$ a measure preserving transformation T such that $Z = Z^* \circ T$. It is clear that if X and Y have the same distribution, then $X^* = Y^*$. Therefore, $X = X^* \circ T_1$, and $Y = X^* \circ T_2$, for some measure preserving transformation T_1 and T_2. By the result in last part, $V(X) = V(X^*) = V(Y)$ as needed.

Next, suppose V is probability sophisticated, lemma 1 implies $\exists \gamma \in \Gamma$ such that $\forall X \in L_p$, $\gamma(X)$ is $\overline{\sigma(X)}$ measurable. For each X nondecreasing, let ρ_X be the measurable function such that $\gamma(X) = \rho_X \circ X$ a.s. Then define $\varrho = \{\rho_{X^*} : X \in L_p\}$. Note if X and Y have the same distribution function, then $X^* = Y^*$, so $\rho_{X^*} = \rho_{Y^*}$. Therefore ϱ is distribution invariant. We next show ϱ is a system of representation functions. Note $\forall X \in L_p$, $X = X^* \circ T$

for some measure preserving transformation T. Note $\rho_{X^*}(X) \in \gamma(X^*)$ by construction, since X^* is nondecreasing. This implies $\forall Y \in L_p$, either

$$\lim_{\alpha \to 0^+} \frac{1}{\alpha}[V(X^* + \alpha Y) - V(X^*)] \leq \int \rho_{X^*}(X^*) Y dP \leq \lim_{\alpha \to 0^-} \frac{1}{\alpha}[V(X^* + \alpha Y) - V(X^*)] \quad \text{(A40)}$$

or

$$\lim_{\alpha \to 0^+} \frac{1}{\alpha}[V(X^* + \alpha Y) - V(X^*)] \geq \int \rho_{X^*}(X^*) Y dP \geq \lim_{\alpha \to 0^-} \frac{1}{\alpha}[V(X^* + \alpha Y) - V(X^*)] \quad \text{(A41)}$$

Suppose (A40) is true, then

$$\lim_{\alpha \to 0^+} \frac{1}{\alpha}[V(X^* \circ T + \alpha Y \circ T) - V(X^* \circ T)] \leq \int \rho_{X^*}(X^* \circ T) Y \circ T dP$$

and

$$\int \rho_{X^*}(X^* \circ T) Y \circ T dP \leq \lim_{\alpha \to 0^-} \frac{1}{\alpha}[V(X^* \circ T + \alpha Y \circ T) - V(X^* \circ T)]$$

i.e.

$$\lim_{\alpha \to 0^+} \frac{1}{\alpha}[V(X + \alpha Y \circ T) - V(X)] \leq \int \rho_{X^*}(X) Y dP \leq \lim_{\alpha \to 0^-} \frac{1}{\alpha}[V(X + \alpha Y \circ T) - V(X)]$$

since T is measure preserving. If instead (A41) is true, we have

$$\lim_{\alpha \to 0^+} \frac{1}{\alpha}[V(X + \alpha Y \circ T) - V(X)] \geq \int \rho_{X^*}(X) Y dP \geq \lim_{\alpha \to 0^-} \frac{1}{\alpha}[V(X + \alpha Y \circ T) - V(X)]$$

Since this holds for all Y, it implies $\rho_{X^*}(X) \in \Gamma(X)$, i.e. ϱ is a representation system, as needed.

A4. Proof proposition 7, 8 and corollary 5.

We first prove the following lemma:

LEMMA 8. Suppose $\Gamma : L_p \to L_q$ is upper hemicontinuous and compact valued. Suppose at $X \in L_p$, $\Gamma(X)$ is single valued, and $\int \Gamma(X) Z dP < 0$ for some $Z \in L_p$. Then $\exists \delta > 0$ and $\gamma \in \Gamma$ such that $||Y - X||_p < \delta$ implies $\int \Gamma(Y) Z dP < 0$.

Proof. Take an arbitrary selection $\gamma \in \Gamma$, upper hemicontinuity of Γ and single valuedness of Γ at X implies if $X_n \to X$ in L_p then $\gamma(X_n) \to \gamma(X)$ in the weak* topology. That is, $\forall Z \in L_p$, $\varepsilon > 0$, $\exists \delta > 0$ such that $||Y - X|| < \delta$ implies $|\int \gamma(Y) Z dP - \int \gamma(X) Z dP| < \varepsilon$, as needed.

Proof of proposition 7: To prove proposition 7, let V be smooth and suppose $\exists \gamma \in \Gamma$, such that $\gamma(X) \geq 0$ for all $x \in L_p$, need to show V is monotone, i.e. $V(X+Y) \geq V(X)$ whenever $Y \geq 0$ a.s.. Define $g(t) = V(X + tY)$ then

$$\begin{aligned} V(X+Y) - V(X) &= \int_0^1 g'(t) dt \\ &= \int_0^1 \int \gamma(X + tY) Y dP dt \geq 0 \end{aligned}$$

the lase inequality is true since $\gamma(X + tY) \geq 0$, and $Y \geq 0$.

Next assume V also satisfies condition 1 and condition 2. To see the reverse of the theorem is also true, let $\mathcal{D}$ be the dense subset of L_p such that $\Gamma(X)$ is single valued on $\mathcal{D}$. We first prove $\forall X \in \mathcal{D}$, $\Gamma(X) \geq 0$ *a.s.*. Suppose not, then $\exists A \in \mathcal{F}$, $P(A) > 0$ such that

$\int \Gamma(X) I_A dP < -\varepsilon$ for some $\varepsilon > 0$, where I_A denote the indicator function of A. By lemma 8, $\exists \varepsilon > 0$ such that for $0 \leq t \leq \varepsilon$, $\int \gamma(X + tY) I_A dP < 0$. Therefore $\int_0^\varepsilon \int \gamma(X + tI_A) I_A dP dt < 0$, i.e. $V(X + \varepsilon I_A) < V(X)$, contrdicting monotonicity. This is proves $\Gamma(X) \geq 0$ if $X \in \mathcal{D}$. For general X, take $X_n \in \mathcal{D}$, and $X_n \to X$ in L_p, upper hemicontinuity implies at least along a subsequence $\gamma(X_n) \to l$ in weak*, and $l \in \Gamma(X)$. Note for each n, $\gamma(X_n) \geq 0$ $a.s.$, we have $l \geq 0$ $a.s.$, as needed.

To prove proposition 8, we need the following lemmas:

Lemma 9. Condition (27) is not satisfies if and only if $\exists A, B \in \mathcal{F}$ such that $P(A), P(B) > 0$ and $\exists a, b, c, d \in R$ such that $\forall \omega \in A$, $\forall \omega' \in B$,

$$a \leq X(\omega) < b \leq X(\omega') \leq c \tag{A42}$$

and

$$\gamma(\omega) < d \leq \gamma(\omega') \tag{A43}$$

Proof. The proof of this lemma is bit long, but not particularly helpful for understanding the propositions, therefore is omitted here, but availible upon request[10].

Lemma 10. Suppose (27) is true, take any $\mathcal{G} \subseteq \mathcal{F}$, define $S : \Omega^* \to R$ by:

$$\forall \hat{\omega} \in \Omega^*, \quad S(\hat{\omega}) = \sup\{\gamma(\omega) : \omega \in \Omega^*,\ X(\omega) \geq E(X|\mathcal{G})(\hat{\omega})\} \tag{A44}$$

[10] It can be found on the author's webpage: www.econ.umn.edu/~hai/research/appendix.pdf

where $\Omega^* = \Omega \backslash N$, and N is the null set given in (27). Then S is $\mathcal{G}$ measurable, and $\forall \hat{\omega} \in \Omega^*$,

$$\gamma(X)(\hat{\omega})[X(\hat{\omega}) - E(X|\mathcal{G})(\hat{\omega})] \leq S(\hat{\omega})[X(\hat{\omega}) - E(X|\mathcal{G})(\hat{\omega})] \tag{A45}$$

Proof. To see S is $\mathcal{G}$ measurable, define $f : R \to R \cup \{\infty\}$ by

$$f(x) = \sup\{\gamma(\omega) : \omega \in \Omega^*, \ X(\omega) \geq x\}$$

then f is nondecreasing thus measurable. Note $S(\omega) = f(E(X|\mathcal{G})(\omega))$ therefore is $\mathcal{G}$ measurable.

To see (A45) is true, first consider $\hat{\omega}$ such that $X(\hat{\omega}) \geq E(X|\mathcal{G})(\hat{\omega})$. In this case $X(\hat{\omega}) \geq E(X|\mathcal{G})(\hat{\omega})$ implies $S(\hat{\omega}) \geq \gamma(X)(\hat{\omega})$, therefore (A45) is true. Next, if instead $X(\hat{\omega}) < E(X|\mathcal{G})(\hat{\omega})$, then take any ω' such that $X(\omega') \geq E(X|\mathcal{G})(\hat{\omega})$, we have $X(\omega') > X(\hat{\omega})$. Equation (27) implies

$$\gamma(X)(\omega') \leq \gamma(X)(\hat{\omega}) \tag{A46}$$

Therefore $S(\hat{\omega}) \leq \gamma(X)(\hat{\omega})$ by definition of $S(\hat{\omega})$. In this case, equation (A45) is still true since $X(\hat{\omega}) - E(X|\mathcal{G})(\hat{\omega})] < 0$.

LEMMA 11. Suppose V is smooth, let $\gamma \in \Gamma$ be a selection of the subdifferential correspondence, then condition (27) and (28) are equivalent.

Proof. First, (27) implies (28). Suppose (27) is true, if we define $\tilde{S} : \Omega \to R$ such that it agrees with S defined in (A44) on Ω^*, then $\tilde{S}$ is $\mathcal{G}$ measurable and

$$\gamma(X)(\hat{\omega})[X(\hat{\omega}) - E(X|\mathcal{G})(\hat{\omega})] \leq S(\hat{\omega})[X(\hat{\omega}) - E(X|\mathcal{G})(\hat{\omega})] \ a.s.$$

we have

$$\begin{aligned} \int \gamma(X)[X - E(X|\mathcal{G})]dP &\leq \int \tilde{S}[X - E(X|\mathcal{G})]dP \\ &= E\left\{E\{\tilde{S}[X - E(X|\mathcal{G})]|\mathcal{G}\}\right\} \\ &= E\{\tilde{S}E\{[X - E(X|\mathcal{G})]|\mathcal{G}\}\} \\ &= 0 \end{aligned}$$

The second line is law of iterated expectation, and third line is because $\tilde{S}$ is $\mathcal{G}$ measurable.

Next, we prove the reverse direction by showing the following: If (27) is not true, then $\exists \mathcal{G} \subseteq \mathcal{F}$, such that $\int \gamma(X)[X - E(X|\mathcal{G})]dP > 0$. If (27) is not true, by lemma 9, $\exists A, B \in \mathcal{F}$ and $\exists a, b, c, d \in R$ that satisfies condition (A42) and (A43). Since P is nonatomic, we can find $C \subseteq A$, $D \subseteq B$, such that

$$\int_{C \cup D} X dP = b$$

Let $\mathcal{G} = \sigma\{\mathcal{F}|_{(C \cup D)^C}, C \cup D\}$, where $\mathcal{F}|_{(C \cup D)^C}$ denote the restriction of the σ field $\mathcal{F}$ on the complement of $C \cup D$. Then

$$E(X|\mathcal{G})(\omega) = \begin{cases} X(\omega) & \omega \notin C \cup D \\ b & \omega \in C \cup D \end{cases}$$

Hence we have

$$
\begin{aligned}
& \int \gamma(X)[X - E(X|\mathcal{G})]dP \\
= & \int_C \gamma(X)[X - E(X|\mathcal{G})]dP + \int_D \gamma(X)[X - E(X|\mathcal{G})]dP \\
= & \int_C [\gamma(X) - d]\,[X - E(X|\mathcal{G})]dP + \int_D [\gamma(X) - d][X - E(X|\mathcal{G})]dP \\
& + d \int_{C \cup D} [X - E(X|\mathcal{G})]dP \\
= & \int_C [\gamma(X) - d]\,[X - E(X|\mathcal{G})]dP + \int_D [\gamma(X) - d][X - E(X|\mathcal{G})]dP > 0
\end{aligned}
$$

as needed.

Proof of proposition 8:

To prove the proposition, we only need to prove (28). First suppose V is smooth, and (28) is true. Let $E(Z|X) = 0$, and $0 < \lambda < 1$, need to verify $V(X + \lambda Z) \geq V(X + Z)$. Define

$$g(t) = V(X + tZ)$$

then

$$
\begin{aligned}
V(X+Z) - V(X) &= \int_\lambda^1 g'(t)dt \\
&= \int_\lambda^1 \int \gamma(X + tZ) Z dP dt \\
&= \int_\lambda^1 \frac{1}{t} \int \gamma(X + tZ) t Z dP dt \\
&= \int_\lambda^1 \frac{1}{t} \int \gamma(X + tZ)\,[X + tZ - E(X + tZ|X)]\,dP dt \\
&\leq 0 \qquad \text{(A47)}
\end{aligned}
$$

as needed.

Next assume V also satisfies condition 1 and 2, need to show mean independent risk aversion implies (28). First, (28) must hold on $\mathcal{D}$. To see this, suppose the contratrary is true, then

$$\int \gamma(X)[X - E(X|\mathcal{G})]dP > 0$$

for some $X \in \mathcal{D}$, $\mathcal{G} \subseteq \mathcal{F}$. Then by lemma 8, $\exists \varepsilon > 0$, such that for $1 - \varepsilon < t < 1$, $Y = tX + (1-t)E(X|\mathcal{G})$,

$$\int \gamma(Y)[X - E(X|\mathcal{G})]dP > 0$$

Therefore, let

$$g(t) = V(tX + (1-t)E(X|\mathcal{G}))$$

we have:

$$\begin{aligned} & V(X) - V(\varepsilon X + (1-\varepsilon)E(X|\mathcal{G})) \\ = & \int_{1-\varepsilon}^{1} g'(t)dt \\ = & \int_{1-\varepsilon}^{1} \int \gamma(tX + (1-t)E(X|\mathcal{G}))[X - E(X|\mathcal{G})]dPdt \\ > & \ 0 \end{aligned}$$

However, X differ from $E(X|\mathcal{G})$ by a mean independent risk, this contradict risk aversion. For general $X \in L_p$, take $X_n \in D$ all n, $X_n \to X$ in L_p, then $\forall \mathcal{G} \subseteq \mathcal{F}, \forall n$,

$$\int \gamma(X_n)[X_n - E(X_n|\mathcal{G})] \leq 0$$

As $n \to \infty$, $X_n - E(X_n|\mathcal{G}) \to X - E(X|\mathcal{G})$ in L_p, and $\gamma(X_n) \to l \in \Gamma(X)$ in weak*, we have

$$\int l[X - E(X|\mathcal{G})] \leq 0$$

as needed.

Proof of corollary 5:

We only need to prove condition (29). First, suppose V is Rothchild-Stiglitz risk averse. Take $X, Y \in L_p$ such that X and Y have the same distribution. Then by proposition 6, $\rho_X = \rho_Y$, we have

$$\int [\gamma(X) - \gamma(Y)](X - Y)dP = \int [\rho_X(X) - \rho_X(Y)][X - Y]dP \leq 0 \tag{A48}$$

The inequality is true because by first part of the theorem, ρ_X is nonincreasing.

Next, suppose (29) is true, and V is smooth and satisfies condition 1 and 2, need to show V is Rothchild-Stiglitz risk averse. By the first part of the corollary, it is enough to show $\forall X \in L_p$, $\forall \rho_X \in \varrho$, ρ_X is nonincreasing ($Q_X - a.s.$). Suppose this is not true. By lemma 9, $\exists A, B \in \mathcal{F}$ and $a, b, c, d \in R$ such that $P(A) = P(B) > 0$ and $\forall \omega \in A$, $\forall \omega' \in B$,

$$a \leq X(\omega) < b \leq X(\omega') \leq c \tag{A49}$$

and

$$\rho_X(X)(\omega) < d \leq \rho_X(X)(\omega') \tag{A50}$$

(The reason we can choose A, B such that $P(A) = P(B)$ is that the probability space $(\Omega, \mathcal{F}, P)$ is nonatomic.) Let $T : (\Omega, \mathcal{F}, P) \to (\Omega, \mathcal{F}, P)$ be the measure preserving transformation[11] such that $\forall \omega \in A$, $T(\omega) \in B$, $\forall \omega' \in B$, $T(\omega') \in A$, and $T(\omega) = \omega$ if $\omega \in (A \cup B)^C$. Since T is measure preserving, X and $X \circ T$ have the same distribution. Consider;

$$\begin{aligned}
& \int [\gamma(X) - \gamma(Y)](X - Y) dP \\
= & \int [\rho_X(X) - \rho_X(X \circ T)][X - X \circ T] dP \\
= & \int_A [\rho_X(X) - \rho_X(X \circ T)][X - X \circ T] dP \\
& + \int_B [\rho_X(X) - \rho_X(X \circ T)][X - X \circ T] dP \\
> & \; 0
\end{aligned} \tag{A51}$$

The first equality is true because X and $X \circ T$ have the same distribution implies $\rho_X = \rho_{X \circ T}$. The second equality is because on $(A \cup B)^C$ $X = X \circ T$. Note on A, $X(\omega) < X \circ T(\omega)$, and $\rho_X(X)(\omega) < \rho_X(X \circ T)(\omega)$, and on B, $X(\omega) < X \circ T(\omega)$ and $\rho_X(X)(\omega) < \rho_X(X \circ T)(\omega)$ by (A49) and (A50), therefore the strict inequality in (A51) is true. But this contradict (29).

[11] For existence of such measure preserving transformation, see corollary 13.4 on page 82 in Kechris (1995).