RISK AND TRADEOFFS

Lara Buchak, UC Berkeley
Dec 2011

APPENDIX: REPRESENTATION THEOREM

For readability, I will repeat all of the technical information from section 5 here, but will omit intuitive explanations of the axioms. MS abbreviates Machina and Schmeidler (1992), and KW abbreviates Kobberling and Wakker (2003).

Spaces, Relations and Notation:

State space S = {…, s, …}, a set of states.
Event space EE = the set of all subsets of S.
Outcome space X = {…, x, …}, a set of outcomes.
Act space A = {…, f(.), g(.), …} = the set of all finite-valued functions from S to X, where $f^{-1}(x) \in EE$.
For any act $f \in A$, there is some partition of EE into $\{E_1, \ldots E_n\}$ and some finite set $Y \subseteq X$ such that $f \in Y^n$.
Strictly speaking, f takes states as inputs. However, I will sometimes write f(E), where $E \in EE$, if f agrees on $s \in E$. Thus, f(E) = f(s) for all $s \in E$.
For any fixed partition of events $M = \{E_1, \ldots, E_n\}$, define $A_M = \{f \in A \mid (\forall E_i \in M)(\exists x \in X)(\forall s \in E_i)(f(s) = x)\}$, the set of acts on M.

≥ is a two-place relation over the act space.[1]
~ and > are defined in the usual way: f ~ g iff f ≥ g and g ≥ f; and f > g if f ≥ g and ¬(g ≥ f).

For all $x \in X$, fx denotes the constant function f(i) = x. Sometimes I will just write x to denote this function, so x ≤ y means fx ≤ fy.
For all $x \in X$, $E \in EE$, $f \in A$, $x_E f$ denotes the function that agrees with f on all states not contained in E, and yields x on any state contained in E. That is, $x_E f(s) = \{x \text{ if } s \in E; f(s) \text{ if } s \notin E\}$. Likewise, for two disjoint events E_1 and E_2, $x_{E1}y_{E2}f$ is the function that agrees with f on all states not contained in E_1 and E_2, and yields x on E_1 and y on E_2: $x_{E1}y_{E2}f(s) = \{x \text{ if } s \in E_1, y \text{ if } s \in E_2, f(s) \text{ if } \neg(s \in E_1 \cup E_2)\}$

We say that an event E is "null" on $F \subseteq A$ iff $(\forall x_E \in F)(\forall f \in F)(x_E f \sim f)$. (MS 749)
If I say that an event is (non-)null without stating a particular F, it should be read as (non-)null on the entire act space A.

We say that outcomes $x^1, x^2, \ldots$ form a *standard sequence* on $F \subseteq A$ if there exists an act $f \in F$, events $E_i \neq E_j$ that are non-null on F, and outcomes y, z with ¬(y ~ z) such that

$$(x^{k+1})_{Ei}(y)_{Ej}f \sim (x^k)_{Ei}(z)_{Ej}f$$

with all acts contained in F (KW, pg. 398).
A standard sequence is bounded if there exist outcomes v and w such that $\forall i(v \geq x^i \geq w)$

We say that $f,g \in A$ are comonotonic if $(\neg\exists s_1,s_2 \in S)(f(s_1) > f(s_2) \;\&\; g(s_1) < g(s_2))$. $C \subseteq A$ is a comoncone if every pair $f, g \in C$ is comonotonic.
Alternatively, take any fixed partition of events $M = \{E_1, \ldots, E_n\}$. A permutation ρ from {1, …, n} to {1, …, n} is a *rank-ordering* permutation of f if $f(E_{\rho(1)}) \geq \ldots \geq f(E_{\rho(n)})$. So a comoncone is a subset C of A_M

[1] Note: I use the "≥" symbol for both the preference relation between acts and the greater-than-or-equal-to relation between real numbers; but it should be clear from context which I am using.

that is rank-ordered by a given permutation: C = {f ∈ A_M | f($E_{\rho(1)}$) ≥ … ≥ f($E_{\rho(n)}$)} for some ρ. For each fixed partition of events of size n, there are n! comoncones. (Definition due to KW, pg. 400).

For each partition M, we define the relation ~*(F) for F ⊆ A_M and outcomes x, y, z, w ∈ X as follows:

xy ~*(F) zw

iff ∃f,g∈ F and ∃E∈ EE such that E is non-null on F and $x_E f \sim y_E g$ and $z_E f \sim w_E g$, where all four acts are contained in F. (Definition due to KW, pg. 396-397)

We write xy ~*(C) zw if there exists a comonocone F ⊆ A_M such that xy ~*(F) zw. (KW, pg. 400)

p: EE → [0, 1] is a *finitely additive probability function* iff:

∀E∈ EE(0 ≤ p(E) ≤ 1)
p(∅) = 0
p(S) = 1
p($E_1 \cup E_2 \cup \ldots \cup E_n$) = Σ$E_n$ for any finite sequence of disjoint events

Let the set of finite-outcome probability distributions over X be denoted by P(X) = {(x_1, p_1; …; x_m, p_m) | m ≥ 1, $\sum_{i=1}^{m} p_i = 1$, x_i∈ X, $p_i \geq 0$}) (Definition due to MS, pg. 753.)

We say that a probability distribution P = (x_1, p_1; …; x_m, p_m) *stochastically dominates* Q = (y_1, q_1; …; y_m, q_m) with respect to the order ≥ if for all z∈ X,

$$\sum_{\{i|x_i \leq z\}} p_i \leq \sum_{\{j|y_j \leq z\}} q_j$$

p strictly stochastically dominates q if the above holds with strict inequality for some z*∈ X. (Definition due to MS, pg. 754)

r: [0, 1] → [0, 1] is a *risk function* iff:

r(0) = 0
r(1) = 1
a ≤ b => r(a) ≤ r(b)
a < b => r(a) < r(b)

Axioms:

Each axiom corresponds to an axiom in either Machina and Schmeidler (1993), pg. 749-750, 761, or Kobberling and Wakker (2003), pages noted in the parentheses.

A1. Ordering (MS P1): ≥ is complete, reflexive, and transitive.

(A1) implies that ≥ is a weak ordering. Note that since the act space is the set of *all* finite-valued functions, completeness implies that the agent must satisfy what is sometimes called the Rectangular Field Assumption: he must have preferences over all possible functions from states to outcomes.

A2. Nondegeneracy (MS P5): There exist outcomes x and y such that fx > fy.

A3. State-wise dominance: If f(s) ≥ g(s) for all s∈ S, then f ≥ g. If f(s) ≥ g(s) for all s∈ S and f(s) > g(s) for all s∈ E⊆S, where E is non-null on A, then f > g.

A3 + A1 implies event-wise monotonicity (MS P3):

A3'. Event-wise monotonicity (MS P3): For all outcomes x and y, events E that are non-null on A, and acts f(.),
$x_E f \geq y_E f$ iff $x \geq y$ (that is, if $fx \geq fy$)
(So, given A1, $x_E f > y_E f$ iff $x > y$.)

Proof: (<=) Assume $x \geq y$. Then since $x_E f(s) = f(s) = y_E f(s)$ for all $s \notin E$ and $x_E f(s) = x \geq y = y_E f(s)$ for all $s \in E$, we have, by A3, $x_E f \geq y_E f$.
(=>) Assume $\neg(x \geq y)$. Then, by A1, $y > x$. Since $y_E f(s) = x_E f(s)$ for all $s \notin E$ and $y_E f(s) = y > x = x_E f(s)$ for all $s \in E$, we have, by A3, $y_E f > x_E f$. So $\neg(x_E f \geq y_E f)$.

A3 implies weak (finite) monotonicity, KW 396:

A3". Weak (finite) monotonicity (KW 396). For any fixed partition of events $E_1, \ldots, E_n$ and acts $f(E_1, \ldots, E_n)$ on those events (acts that yield a single outcome for all states in those events, i.e. acts f such that $\forall i \exists x (\forall s \in E_i)(f(s) = x)$),
if $f(E_j) \geq g(E_j)$ for all j, then $f \geq g$.

Proof: If $f(E_j) \geq g(E_j)$ for all j, then $f(s) \geq g(s)$ for all $s \in S$, so $f \geq g$.

A4. Continuity ((i)KW 398, Solvability, and (ii)MS P6 Small Event Continuity):
(i) For any fixed partition of events $E_1, \ldots, E_n$, and for all acts $f(E_1, \ldots, E_n)$, $g(E_1, \ldots, E_n)$ on those events, outcomes x, y, and events E_i with $x_E f > g > y_E f$, there exists an "intermediate" outcome z such that $z_E f \sim g$.
(ii) For all acts $f > g$ and outcome x, there exists a finite partition of events $\{E_1, \ldots, E_n\}$ such that for all i, $f > x_{Ei} g$ and $x_{Ei} f > g$

A5. Comonotonic Archimedean Axiom (KW 398, 400):
For each comoncone F, every bounded standard sequence on F is finite.

A6. Comonotonic Tradeoff Consistency (KW 397, 400): Improving an outcome in any ~*(C) relationship breaks that relationship. In other words, xy ~*(C) zw and y' > y entails ¬(xy' ~*(C) zw).

A7. Strong Comparative Probability (MS P4*): For all pairs of disjoint events E_1 and E_2, outcomes $x' > x$ and $y' > y$, and acts $g, h \in A$,

$$x'_{E1} x_{E2} g \geq x_{E1} x'_{E2} g \quad \Rightarrow \quad y'_{E1} y_{E2} h \geq y_{E1} y'_{E2} h$$

Theorem (Representation): If a relation $\geq$ satisfies (A1) through (A8), then there exist:

(i) a unique finitely additive, non-atomic probability function p: EE -> [0, 1]
(ii) a unique risk function r: [0, 1] -> [0, 1] and
(iii) a utility function unique up to linear transformation such that

for any $f \in A$, each of which we can write as $\{O_1, F_1; \ldots; O_n, F_n\}$ where for all i, $f(s) = O_i$ for all $s \in F_i \subseteq S$ and $fO_1 \leq fO_2 \leq \ldots \leq fO_n$,

REU(f) =

$$u(O_1) + r(\sum_{i=2}^{n} p(F_i))(u(O_2) - u(O_1)) + r(\sum_{i=3}^{n} p(F_i))(u(O_3) - u(O_2)) + ... + r(p(F_n))(u(O_n) - u(O_{n-1}))$$

represents the preference relation $\leq$ on A.

That is, there exist a unique probability function p, a unique risk function r, and a utility function unique up to linear transformation such that for all outcomes $x_1, \ldots, x_n, y_1, \ldots, y_m \in X$ and events $E_1, \ldots, E_n, G_1, \ldots, G_m \in$ EE,

$\{x_1, E_1; \ldots; x_n, E_n\} \geq \{y_1, G_1; \ldots; y_m, G_m\}$ iff

$$u(O_1) + r(\sum_{i=2}^{n} p(F_i))(u(O_2) - u(O_1)) + r(\sum_{i=3}^{n} p(F_i))(u(O_3) - u(O_2)) + ... + r(p(F_n))(u(O_n) - u(O_{n-1}))$$

$\geq$

$$u(Q_1) + r(\sum_{i=2}^{m} p(H_i))(u(Q_2) - u(Q_1)) + r(\sum_{i=3}^{m} p(H_i))(u(Q_3) - u(Q_2)) + ... + r(p(H_m))(u(Q_m) - u(Q_{m-1}))$$

for some rank-ordering permutations Δ and Ψ such that $x_{\Delta(1)} \leq \ldots \leq x_{\Delta(n)}$, $O_i = x_{\Delta(i)}$, and $F_i = E_{\Delta(i)}$, and $y_{\Psi(1)} \leq \ldots \leq y_{\Psi(n)}$, $Q_i = y_{\Psi(i)}$, and $H_i = G_{\Psi(i)}$.

If there are three such functions so that REU(f) represents the preference relation, we say that REU holds.

Furthermore, in the presence of (A2) and (A4i), if REU holds with a continuous r-function, then the remaining axioms are satisfied. Therefore, if we assume non-degeneracy (A2) and solvability (A4i), we have:

(A1), (A3), (A4ii), (A5), (A6), (A7) are sufficient conditions for REU.
(A1), (A3), (A4ii), (A5), (A6), (A7) are necessary conditions for REU with continuous r-function.

Proof:

Assume we have a preference ordering that satisfies (A1) through (A8).

Part one. Derive a Weighting Function of States (following Kobberling and Wakker (2003))

Koberling and Wakker (Theorem 8, pg 399-400) show that if the relation ≥ over A satisfies (A1), (A3"), (A4i), (A5), and (A6), then for each fixed partition of events M = $\{E_1, \ldots, E_n\}$ and acts $f(E_1, \ldots, E_n)$ on those events which form a subset $A_M \subseteq A$, there exists:

(i) A *weighting function* W: 2^M -> [0, 1] such that W(∅) = 0, W(M) = 1, and A ⊆ B => W(A) ≤ W(B)
(ii) A *utility function* u: X -> ℜ
such that preferences can be represented by:

CEU: $f \mapsto \sum_{j=1}^{n} \pi_{\Phi(j)} U\big(f(E_{\Phi(j)})\big)$

Where the "decision weights" $\pi_{\Phi j}$ are nonnegative and sum to one. Furthermore, for some rank-ordering permutation Φ on {1, …, n} such that $f(E_{\Phi(1)}) \geq \ldots \geq f(E_{\Phi(n)})$, we have:
$\pi_{\Phi(j)} = W(E_{\Phi(1)}, \ldots, E_{\Phi(j)}) - W(E_{\Phi(1)}, \ldots, E_{\Phi(j-1)})$ for all j (and $\pi_{\Phi(1)} = W(E_{\Phi(1)})$).

Furthermore:

Case (i). In the trivial case where at least one comoncone has only null events, all comoncones only have null events, the preference relation is trivial, utility is constant, and the weighting function can be chosen arbitrarily.

However, (A2) rules out the case in which the preference relation is trivial, so this case (which is possible on Kobberling and Wakker's axioms) is ruled out by our axioms.

Case (ii). Kobberling and Wakker consider the degenerate case in which there is exactly one non-null event for each comoncone, and they add an additional restriction (that there is a countable order-dense subset of X) to determine a unique utility function in this case. This case corresponds, for example, to the preferences of a maximinimizer or a maximaximizer: the only event that will matter in each comoncone will be the event with the worst outcome, or the event with the best outcome. However, this case turns out to be ruled out by our other axioms: ruled out by the axioms that are necessary to deriving a probability function; I will postpone this proof to Part Two.

Case (iii). In the nondegenerate case where at least one comoncone has two or more non-null events, utility is unique up to linear transformation, and the weighting function is uniquely determined.
Since this is the only case not ruled out by the other axioms, it is the only case we are concerned with.

In section 5.1 (pg. 402), Kobberling and Wakker extend the result to arbitrary state spaces (i.e. our S), with finite valued functions from S to X (i.e. our act space A). As noted above, every act can be denoted as a subset of some A^M. In case (iii), we can again derive a unique weighting function defined over all events in EE and a utility function unique up to unit and location.

We can reorganize the terms of Kobberling's and Wakker's equation:

$$\text{CEU(f)} = \sum_{j=i}^{n} \pi_{\Phi(j)} U\left(f\left(E_{\Phi(j)}\right)\right) = \sum_{j=1}^{n} \left(W\left(E_{\Phi(1)},\ldots,E_{\Phi(j)}\right) - W\left(E_{\Phi(1)},\ldots,E_{\Phi(j-1)}\right)\right) U\left(f\left(E_{\Phi(j)}\right)\right)$$

$$= \quad U\left(f\left(E_{\Phi(n)}\right)\right) + \sum_{j=1}^{n-1} \left(W\left(E_{\Phi(1)},\ldots,E_{\Phi(j)}\right)\right)\left(U\left(f\left(E_{\Phi(j)}\right)\right) - U\left(f\left(E_{\Phi(j+1)}\right)\right)\right)$$

where, again, Φ orders the states from best to worst.

This is equivalent to

$$\text{CEU(f)} = U\left(f\left(E_{\Delta(1)}\right)\right) + \sum_{j=2}^{n} \left(W\left(E_{\Delta(j)},\ldots,E_{\Delta(n)}\right)\right)\left(U\left(f\left(E_{\Delta(j)}\right)\right) - U\left(f\left(E_{\Delta(j-1)}\right)\right)\right)$$

where Δ is a rank-ordering permutation on {1, …, n} that orders states from *worst* to *best*.

Part Two. Derive a Probability Function of States (following Machina and Schmeidler (1993))

Machina and Schmeidler (Theorem 2, pg. 766) show that if the ≥ relation over A satisfies (A1), (A2), (A3’), (A4ii), and (A7) then:

There exists a unique, finitely additive, non-atomic probability measure p: EE -> [0, 1] and a non-constant, mixture continuous preference functional $V(P) = V(x_1, p_1; \ldots; x_n, p_n)$ on the set of finite-outcome probability distributions such that V(.) exhibits monotonicity with respect to stochastic dominance (that is, V(P) ≥ V(Q) whenever P stochastically dominates Q and V(P) > V(Q) whenever P strictly stochastically dominates Q), such that the relation ≥ can be represented by the preference functional:

$У(f(.)) = V(x_1, p(f^{-1}(x_1)); \ldots x_n, p(f^{-1}(x_n)))$, where $\{x_1, \ldots, x_n\}$ is the outcome set of the act f(.).

That is, for an act $f = \{x_1, E_1; \ldots; x_n, E_n\}$, $У(f(.)) = V(x_1, p(E_1)); \ldots x_n, p(E_n)))$

This implies that an agent will be indifferent between two distributions of outcomes over states that give rise to the same probability distribution over outcomes (see also Theorem 1, MS, pg. 765). That is, if $p(f^{-1}(x_i)) = p(g^{-1}(x_i))$ for all i, then f ~ g.

By the claim that p is non-atomic, we mean that for any event E with p(E) > 0 and any $c \in (0,1)$, there exists some event $E^* \subset E$ such that $p(E^*) = c \cdot p(E)$. (MS, pg. 751, footnote 12).

We are now in a position to say why case (ii), above, is ruled out:

Recall that case (ii) is the degenerate case in which there is exactly one non-null event for each comoncone. That there is one non-null event for each comoncone implies that there is exactly one non-null event on A_M for each partition M of events. For assume there are at least two non-null events on A_M: E_1 and E_2. By (A2), ∃x∃y(fx > fy). Now consider the gamble $y_{E1}y_{E2}fy$ (= fy) and the gamble $x_{E1}y_{E2}fy$ (=$x_{E1}fy$). Since E_1 is non-null, by (A3’) we will have $x_{E1}y_{E2}fy > y_{E1}y_{E2}fy$. Now consider the gamble $x_{E1}x_{E2}fy$. Since E_2 is non-null, by (A3’) we will have $x_{E1}x_{E2}fy > x_{E1}y_{E2}fy$. But $y_{E1}y_{E2}fy$, $x_{E1}y_{E2}fy$, and $x_{E1}x_{E2}fy$ are comonotonic: they are part of any comonocone whose rank-ordering permutation orders the events from worst to best starting with any ordering of the events in ($\neg E_1 \cap \neg E_2$), followed by E_2, followed by E_1. So there are two non-null events on that comoncone; this is a contradiction, so there is only one non-null event on A_M.

Thus, for any fixed partition of events M = {E_1, …, E_n}, there is exactly one non-null event on A_M (if there were no non-null events, we would have case (i), which, as stated above, is ruled out). Call the event E. Since the event is non-null, p(E) > 0.[2] Therefore, since p is non-atomic, there exists some event $E^* \subset E$ such that p(E*) = (0.5)p(E). By definition of a probability function, p(E-E*) = (0.5)p(E). By (A3), $x_{E^*}f > y_{E^*}f$ and $x_{E\text{-}E^*}f > y_{E\text{-}E^*}f$ for all f. Take the partition M* = {E_1, …, E*, E-E*, …, E_n}. Both E* and E-E* are non-null on A_M, so we have a contradiction.

So case (ii) is impossible given our axioms.

Part Three. Derive a Risk Function

So far, since case (iii) is the only remaining case that is consistent with our axioms, we know that we will be able to find a utility function u: X -> $\Re$ unique up to linear transformation, a unique weighting function W: EE -> [0, 1], and a unique probability function p such that:

(I) ≤ is representable by:

$$\text{CEU(f)} = U\left(f\left(E_{\Delta(1)}\right)\right) + \sum_{j=2}^{n}\left(W\left(E_{\Delta(j)},\ldots,E_{\Delta(n)}\right)\right)\left(U\left(f\left(E_{\Delta(j)}\right)\right) - U\left(f\left(E_{\Delta(j-1)}\right)\right)\right)$$

where Δ is a rank-ordering permutation on {1, …, n} that orders states from worst to best.

(II) If two acts f and g give rise to the same probability distribution over outcomes, i.e. if $p(f^{-1}(x)) = p(g^{-1}(x))$ for all $x \in X$, then f ~ g, so, by (I), CEU(f) = CEU(g).

(III) For all $c \in [0,1]$, $\exists N \in EE$ s.t. p(N) = c. This holds because p(S) = 1, p(∅) = 0, and since p is non-atomic, $\forall c \in (0,1)$ there exists $E^* \subset S$ such that $p(E^*) = c \cdot p(S) = c$.

Now for each $c \in [0,1]$, define r(c) = W(A) for any A such that p(A) = c.
We can show that this is well-defined, since any two events that have the same probability must have the same weight: assume there exist A and B such that p(A) = p(B) but W(A) > W(B), and take outcomes x > y. CEU(fx) > CEU(fy), so u(x) > u(y). $CEU(x_A y_{\neg A}) = u(y) + W(A)[u(x) - u(y)]$, and $CEU(x_B y_{\neg B}) = u(y) + W(B)[u(x) - u(y)]$, so $CEU(x_A y_{\neg A}) > CEU(x_B y_{\neg B})$, and thus $x_A y_{\neg A}$ is preferred to $x_B y_{\neg B}$ because CEU(f) represents ≥. But $У(x_A y_{\neg A}) = У(x_B y_{\neg B})$ because the two acts give rise to the same probability distribution over outcomes, so $x_A y_{\neg A} \sim x_B y_{\neg B}$ because У represents ≥. Since this is a contradiction, p(A) = p(B) => W(A) = W(B).
We also know that r is unique, since W and p are unique.
So for all $A \in EE$, r(p(A)) = W(A).

Furthermore, r has all the properties of a risk function:

---

[2] Note that by MS Theorem 1, p represents the relationship in (A7), so that $p(E_1) > p(E_2)$ whenever E_1 and E_2 satisfy $x'_{E1}x_{E2}g \geq x_{E1}x'_{E2}g$ for all x' > x and all $g \in A$; and note further that, by (A3), E_1 = E and E_2 = ∅ satisfy this relationship.

(1) r(0) = 0, because p(∅) = 0 (by definition of a probability function), and W(∅) = 0 (by the properties of the weighting function, given in part one).
(2) r(1) = 1, because p(S) = 1 (by definition of a probability function), and W(S) = 1 (by the properties of the weighting function).
(3) r is non-decreasing: Assume $a \leq b$. Pick B such that p(B) = b. Now pick $A \subseteq B$ such that p(A) = a (we know such an A exists because p is non-atomic). Therefore, by the properties of the weighting function, $W(A) \leq W(B)$. So $r(A) \leq r(B)$.
(4) r is strictly increasing: Assume a < b, and pick $A \subset B$ such that p(A) = a and p(B) = b. If r(a) = r(b), then for outcomes x > y, we will have $CEU(x_A fy) = CEU(x_B fy)$, and so $x_A fy \sim x_B fy$. But by stochastic dominance, $V(x_B fy) > V(x_A fy)$, so $x_B fy$ is preferred to $x_A fy$. Since this is a contradiction, ¬(r(a) = r(b)). And since $r(A) \leq r(B)$, r(A) < r(B).

Therefore, for a function $f = \{x_1, E_1; \ldots; x_n, E_n\}$, we have:

$$U\left(f\left(E_{\Delta(1)}\right)\right)+\sum_{j=2}^{n}\left(W\left(E_{\Delta(j)},\ldots,E_{\Delta(n)}\right)\right)\left(U\left(f\left(E_{\Delta(j)}\right)\right)-U\left(f\left(E_{\Delta(j-1)}\right)\right)\right) =$$

$$U\left(x_{\Delta(1)}\right)+\sum_{j=2}^{n} r\left(p\left(E_{\Delta(j)},\ldots,E_{\Delta(n)}\right)\right)\left(U\left(x_{\Delta(j)}\right)-U\left(x_{\Delta(j-1)}\right)\right)$$

Note that $p(E_{\Delta(j)},\ldots, E_{\Delta(n)}) = \sum_{i=j}^{n} p\left(E_{\Delta(i)}\right)$ by finite additivity.

Now for all members of the act space, finite valued functions $f = \{x_1, E_1; \ldots; x_n, E_n\} \in A$, where $f(s) = x_i$ for all $s \in E_i$ let REU(f)=

$$u(x_{\Delta(1)})+r(\sum_{i=2}^{n} p(E_{\Delta(i)}))(u(x_{\Delta(2)})-u(x_{\Delta(1)}))+r(\sum_{i=3}^{n} p(E_{\Delta(i)}))(u(x_{\Delta(3)})-u(x_{\Delta(2)}))+\ldots+r(p(E_{\Delta(n)}))(u(x_{\Delta(n)})-u(x_{\Delta(n-1)}))$$

for some rank-ordering permutation Δ such that $x_{\Delta(1)} \leq \ldots \leq x_{\Delta(n)}$.

Therefore, for any $f \in A$, each of which we can write as $\{x_1, E_1; \ldots; x_n, E_n\}$ where $f(s) = x_i$ for all $s \in E_i$ and again as $\{O_1, F_1; \ldots; O_n, F_n\}$ where for some rank-ordering permutation Δ such that $x_{\Delta(1)} \leq \ldots \leq x_{\Delta(n)}$ we have $O_i = x_{\Delta(i)}$ and $F_i = E_{\Delta(i)}$, there exists

(i) a unique finitely additive, non-atomic probability function p: EE -> [0, 1]
(ii) a unique risk function r: [0, 1] -> [0, 1] and
(iii) a utility function unique up to linear transformation such that

REU(f) =

$$u(O_1)+r(\sum_{i=2}^{n} p(F_i))(u(O_2)-u(O_1))+r(\sum_{i=3}^{n} p(F_i))(u(O_3)-u(O_2))+\ldots+r(p(F_n))(u(O_n)-u(O_{n-1}))$$

represents the preference relation ≤ on A.

Part Four. In the presence of (A2) and (A4i), REU with a continuous r-function implies the remaining axioms.

Again, REU holds if there is:

(i) a unique finitely additive, non-atomic probability function p: EE -> [0, 1]

(ii) a unique risk function r: [0, 1] -> [0, 1] and

(iii) a utility function unique up to linear transformation such that

for any $f \in A$, each of which we can write as $\{O_1, F_1; \ldots; O_n, F_n\}$ where for all i, $f(s) = O_i$ for all $s \in F_i \subseteq S$ and $fO_1 \leq fO_2 \leq \ldots \leq fO_n$, REU(f) =

$$u(O_1) + r(\sum_{i=2}^{n} p(F_i))(u(O_2) - u(O_1)) + r(\sum_{i=3}^{n} p(F_i))(u(O_3) - u(O_2)) + \ldots + r(p(F_n))(u(O_n) - u(O_{n-1}))$$

represents the preference relation $\leq$ on A.

(i) KW's axioms: (A3"), (A5), and (A6).

If REU holds, then CEU holds; let W(A) = r(p(A)).
KW's Theorem 8 (pg. 400) states that if (A4i) holds, and if CEU holds, then (A3"), (A5), and (A6) hold.

(ii) MS's axioms: (A1), (A2), (A3'), (A4ii), and (A7).

MS's Theorem 2 (pg. 766) states that if there exists a unique, finitely additive, non-atomic probability measure p: EE -> [0, 1] and a non-constant, mixture continuous preference functional $V(P) = V(x_1, p_1; \ldots; x_n, p_n)$ on the set of finite-outcome probability distributions such that V(.) exhibits monotonicity with respect to stochastic dominance (that is, $V(P) \geq V(Q)$ whenever P stochastically dominates Q and $V(P) > V(Q)$ whenever P strictly stochastically dominates Q), such that the relation $\geq$ can be represented by the preference functional $У(f(.)) = V(x_1, p(f^{-1}(x_1)); \ldots x_n, p(f^{-1}(x_n)))$, then (A1), (A2), (A3'), (A4ii), and (A7) hold.

We know that p is unique, finitely additive, non-atomic, and by (A2), we know that REU(f) is non-constant. We must show that (1) REU(f) is mixture continuous and (2) REU(f) $\geq$ REU(g) whenever f stochastically dominates g and REU(f) > REU(g) whenever f strictly stochastically dominates g:

(1) Since we have a probability function of states, we can identify an act f with its probability distribution $f^* \in P(X)$ such that all acts identified with that probability distribution have the same REU; we can write REU(f*). As defined on pg. 755 of MS, REU(f) is mixture continuous if, for any probability distributions f*, g*, and h* in P(X), the sets $\{\lambda \in [0,1] \mid REU(\lambda f^* + (1-\lambda)g^*) \geq REU(h^*)\}$ and $\{\lambda \in [0,1] \mid REU(\lambda f^* + (1-\lambda)g^*) \leq REU(h^*)\}$ are both closed.

Consider the set of outcomes $\{x_1, \ldots, x_n\}$ such that x_i is an outcome of f* or an outcome of g*, and $fx_1 \leq fx_2 \leq \ldots \leq fx_n$. Then we have:

$f^* = \{p_1, x_1; \ldots; p_n, x_n\}$ for some $\{p_1, \ldots, p_n\}$
$g^* = \{q_1, x_1; \ldots; q_n, x_n\}$ for some $\{q_1, \ldots, q_n\}$
$\lambda f^* + (1-\lambda)g^* = \{(\lambda p_1 + (1-\lambda)q_1, x_1; \ldots; \lambda p_n + (1-\lambda)q_n\}$

For each i, $\lambda p_i + (1-\lambda)q_i$ is continuous in λ. So for each k, $\sum_{i=k}^{n}(\lambda p_i + (1-\lambda)q_i)$ is continuous in λ. Since r is continuous, $r(\sum_{i=k}^{n}(\lambda p_i + (1-\lambda)q_i))$ is continuous in λ. So ,

$r(\sum_{i=k}^{n}(\lambda p_i + (1-\lambda)q_i))[u(x_k) - u(x_{k-1})]$ is continuous in λ. So the sum of these terms from k = 2 to n is continuous in λ. So REU(f) is continuous in λ. Therefore, $\{\lambda \in [0,1] \mid REU(\lambda f^* + (1-\lambda)g^*) \geq m\}$ and $\{\lambda \in [0,1] \mid REU(\lambda f^* + (1-\lambda)g^*) \leq m\}$ are closed for all m, in particular for m = REU(h*), so REU(f) is mixture continuous.

(2) Take two acts f and g, such that f stochastically dominates g. Consider the set of outcomes $\{x_1, \ldots, x_n\}$ such that x_i is an outcome of f or an outcome of g, and $fx_1 \leq fx_2 \leq \ldots \leq fx_n$. Then we have:

$f^* = \{p_1, x_1; \ldots; p_n, x_n\}$ for some $\{p_1, \ldots, p_n\}$

$g^* = \{q_1, x_1; \ldots; q_n, x_n\}$ for some $\{q_1, \ldots, q_n\}$

Since f stochastically dominates g, for all k, $\sum_{\{i|x_i \leq x_k\}} p_i \leq \sum_{\{i|x_i \leq x_k\}} q_i$

Since $\sum_{\{i\}} p_i = \sum_{i} q_i = 1$, we have, for all k, $\sum_{\{i|x_i > x_k\}} p_i \geq \sum_{\{i|x_i > x_k\}} q_i$.

Thus, since r is strictly increasing, we have $r(\sum_{\{i|x_i > x_k\}} p_i) \geq r(\sum_{\{i|x_i > x_k\}} q_i)$ for all k.

Since $u(x_k)$ is non-decreasing,

$r(\sum_{\{i|x_i > x_k\}} p_i)(u(x_{k+1}) - u(x_k)) \geq r(\sum_{\{i|x_i > x_k\}} q_i)(u(x_{k+1}) - u(x_k))$ for all k.

For each k, the above terms will either be equal to the k^{th} term in the REU equation for REU(f) and REU(g), respectively, or else in the k^{th} terms of REU(f) or REU(g), the "r" coefficient will be different than the above but the difference $u(x_{k+1}) - u(x_k)$ will be zero.

So REU(f) ≥ REU(g)

If f strictly stochastically dominates g, then $\sum_{\{i|x_i \leq x_m\}} p_i < \sum_{\{i|x_i \leq x_m\}} q_i$ for some m.

So, by the same chain of reasoning, $r(\sum_{\{i|x_i > x_m\}} p_i) > r(\sum_{\{i|x_i > x_m\}} q_i)$ for some m.

Take the least m for which the strict inequality holds. Then, for the least k such that x_k is preferred to x_m (we know there is some such k, otherwise both sums would be 1), the difference $u(x_k) - u(x_{k-1})$ will be non-zero, so we have:

$r(\sum_{\{i|x_i > x_m\}} p_i)(u(x_k) - u(x_{k-1})) > r(\sum_{\{i|x_i > x_m\}} q_i)(u(x_k) - u(x_{k-1}))$

This will be the k^{th} term in the REU equation for REU(f) and REU(g), respectively; and for all other terms, the weak inequality will hold, i.e. the j^{th} term of REU(f) will be greater than or equal to the j^{th} term of REU(g).

Therefore, by MS Theorem 2, (A1), (A2), (A3'), (A4ii), and (A7) hold.

(iii) (A3). We want to show that if f(s) ≥ g(s) for all s∈S, then f ≥ g. If f(s) ≥ g(s) for all s∈S and f(s) > g(s) for all s∈E⊆S, where E is some non-null event on A, then f > g.

> Assume f(s) ≥ g(s) for all s∈S. Then, for all x, {s | g(s) > x} ⊆ {s | f(s) > x}, since if g(s) > x, then f(s) > x. Consequently, for all x, p({s | g(s) > x}) ≤ p({s | f(s) > x}), by definition of a probability function. That is, f stochastically dominates g. To put it in directly in terms of our definition of stochastic dominance: take a partition of M = {E_1, …, E_n} of S such that ∀i∃x_i∃y_i(∀s∈E_i)(f(s) = x_i & g(s) = y_i). Then {s | g(s) ≥ x} and {s | f(s) ≥ x} are subsets of that partition, and $\sum_{\{i|y_i>y\}} p(E_i) \leq \sum_{\{i|x_j>x\}} p(E_j)$. So $\sum_{\{i|y_i\leq y\}} p(E_i) \geq \sum_{\{i|x_j\leq x\}} p(E_j)$. That is, f stochastically dominates g. So, by (2ii) in part four, REU(f) ≥ REU(g). So f is weakly preferred to g.
>
> Now assume f(s) ≥ g(s) for all s∈S and f(s) > g(s) for all s∈E⊆S, where E is some non-null event of A. Take the set {x | (∃s∈E)(g(s) = x)}. We know this set is non-empty. We also know this set is finite, because the outcomes space of g is finite; so we can pick the greatest member of the set, x*. Now we want to show that {r | f(r) ≤ x*} ⊂ {r | g(r) ≤ x*}. Consider some element t∈{r | f(r) ≤ x*}. Since f(t) ≤ x*, we have g(t) ≤ x*, so {r | f(r) ≤ x} ⊆ {r | g(r) ≤ x}. Now consider some element t*∈E such that g(t*) = x*. We have t*∈{r | g(r) ≤ x*}, but we have f(t*) > x*, so we have ¬(t*∈{r | f(r) ≤ x*}). Therefore, {r | f(r) ≤ x*} ⊂ {r | g(r) ≤ x*}. Consequently, f strictly stochastically dominates g. Again by (2ii) in part four, REU(f) > REU(g). So f is weakly preferred to g.
>
> Therefore, (A3) holds.

So, assuming (A2), (A4i), REU implies (A1), (A3), (A4ii), (A5), (A6), and (A7).

Therefore, if we assume non-degeneracy (A2) and solvability (A4i), we have:

(A1), (A3), (A4ii), (A5), (A6), (A7) are sufficient conditions for REU.
(A1), (A3), (A4ii), (A5), (A6), (A7) are necessary conditions for REU with continuous r-function.

APPENDIX B. Tradeoff Equality and Utility Differences.

Claim: For REU maximizers, xy ~*(C)zw holds just in case u(x) – u(y) = u(z) – u(w).

Proof: Assume the agent is an REU maximizer, and xy ~*(C)zw. Then $x_E f \sim y_E g$ and $z_E f \sim w_E g$ for some $x_E f$, $y_E g$, $z_E f$, $w_E g$ that are comonotonic. Therefore, there is some set of states $\{E_1, \ldots, E_n\}$ and some rank-ordering permutation Φ such that $x_E f(E_{\Phi(1)}) \geq \ldots \geq x_E f(E_{\Phi(n)})$, $y_E g(E_{\Phi(1)}) \geq \ldots \geq y_E g(E_{\Phi(n)})$, $z_E f(E_{\Phi(1)}) \geq \ldots \geq z_E f(E_{\Phi(n)})$, $w_E g(E_{\Phi(1)}) \geq \ldots \geq w_E g(E_{\Phi(n)})$ and:

(1) $E_{\Phi(i)} = E$ for some i.

(2) $$\sum_{j=1}^{n}\left(r\left(p\left(E_{\Phi(1)},\ldots,E_{\Phi(j)}\right)\right)-r\left(p\left(E_{\Phi(1)},\ldots,E_{\Phi(j-1)}\right)\right)\right)u\left(x_E f\left(E_{\Phi(j)}\right)\right)=$$
$$\sum_{j=1}^{n}\left(r\left(p\left(E_{\Phi(1)},\ldots,E_{\Phi(j)}\right)\right)-r\left(p\left(E_{\Phi(1)},\ldots,E_{\Phi(j-1)}\right)\right)\right)u\left(y_E g\left(E_{\Phi(j)}\right)\right)$$

(3) $$\sum_{j=1}^{n}\left(r\left(p\left(E_{\Phi(1)},\ldots,E_{\Phi(j)}\right)\right)-r\left(p\left(E_{\Phi(1)},\ldots,E_{\Phi(j-1)}\right)\right)\right)u\left(z_E f\left(E_{\Phi(j)}\right)\right)=$$
$$\sum_{j=1}^{n}\left(r\left(p\left(E_{\Phi(1)},\ldots,E_{\Phi(j)}\right)\right)-r\left(p\left(E_{\Phi(1)},\ldots,E_{\Phi(j-1)}\right)\right)\right)u\left(w_E g\left(E_{\Phi(j)}\right)\right)$$

See part (i) and (iii) of the proof in Appendix A.

Since $x_E f$ and $z_E f$ only differ on $E_{\Phi(i)}$, and $y_E g$ and $w_E g$ only differ on $E_{\Phi(i)}$, subtracting equation (3) from equation (2) yields:

$$\left[r\left(p\left(E_{\Phi(1)},\ldots,E_{\Phi(i)}\right)\right)-r\left(p\left(E_{\Phi(1)},\ldots,E_{\Phi(i-1)}\right)\right)\right][u(x)-u(z)]=\left[r\left(p\left(E_{\Phi(1)},\ldots,E_{\Phi(i)}\right)\right)-r\left(p\left(E_{\Phi(1)},\ldots,E_{\Phi(i-1)}\right)\right)\right][u(y)-u(w)]$$

This simplifies to u(x) – u(z) = u(y) – u(w).

Claim: For EU maximizers, xy~*(A)zw holds just in case u(x) – u(y) = u(z) – u(w).
Proof: Assume the agent is an EU maximizer, and xy ~*(A)zw. Then $x_E f \sim y_E g$ and $z_E f \sim w_E g$ for some $x_E f$, $y_E g$, $z_E f$, $w_E g$ in A. Therefore, there is some set of states $\{E_1, \ldots, E_n\}$ such that:
(1) $E_i = E$ for some i.

(2) $$\sum_{j=1}^{n} p(E_j)u(x_E f(E_j)) = \sum_{j=1}^{n} p(E_j)u(y_E g(E_j))$$

(3) $$\sum_{j=1}^{n} p(E_j)u(z_E f(E_j)) = \sum_{j=1}^{n} p(E_j)u(w_E g(E_j))$$

Since $x_E f$ and $z_E f$ only differ on E_i, and $y_E g$ and $w_E g$ only differ on E_i, subtracting equation (3) from equation (2) yields:
$p(E_i)[u(x) – u(z)] = p(E_i)[u(y) – u(z)]$.
This simplifies to u(x) – u(z) = u(y) – u(w).