Journal of Mathematical Economics 73 (2017) 142–148

Contents lists available at ScienceDirect

Journal of Mathematical Economics

journal homepage: www.elsevier.com/locate/jmateco

Probabilistic intertemporal choice

Pavlo R. Blavatskyy

Montpellier Business School, Montpellier Research in Management, 2300 Avenue des Moulins, 34185, Montpellier Cedex 4, France

ARTICLE INFO

Article history:
Received 27 April 2017
Received in revised form 2 October 2017
Accepted 5 October 2017
Available online 20 October 2017

Keywords:
Intertemporal choice
Time preference
Additively separable utility
Probabilistic choice
Fechner model of random errors
Luce's choice model

ABSTRACT

Probabilistic intertemporal choice involves situations when a decision maker does not choose the same stream of intertemporal outcomes when presented with the same decision problem repeatedly; or when a decision maker makes non-repeated choice decisions that are inherently inconsistent (*i.e.* they cannot be represented by any rational time preferences); or when an aggregated choice pattern of several decision makers is contradictory. This paper presents behavioural characterization (axiomatization) of an additively separable utility (that includes discounted utility, quasi-hyperbolic discounting, generalized hyperbolic discounting and liminal discounting as special cases) embedded into Fechner model of random errors (also known as strong utility) and Luce's choice model (also known as strict utility). Such probabilistic extensions of classical utility representations of time preferences are consistent with some behavioural patterns that challenge the descriptive validity of the original (deterministic) theories (*e.g.* some instances of the common difference effect).

1. Introduction

Intertemporal choice involves outcomes that are received at different moments in time. Classical models of intertemporal choice, such as discounted utility theory[1] (Samuelson, 1937), are typically deterministic. Yet, empirical data often require a model of probabilistic intertemporal choice. For example, such necessity arises when a decision maker does not choose the same alternative when presented with the same decision problem on several occasions; or when a decision maker reveals inconsistent time preferences across different decision problems; or when an aggregate choice pattern of several decision makers is contradictory.

Historically, the first approach to modelling probabilistic choice over time is to embed a deterministic theory of intertemporal choice into a random utility/random preference model. In this paradigm, a decision maker is assumed to have several preference relations, one of which is drawn at random when a choice decision is made (*e.g.* Falmagne, 1985; Loomes and Sugden, 1995). The most popular random preference approach to intertemporal choice is a model with a randomly distributed discount factor *e.g.*, Coller and Williams (1999, p. 115, section 4.2), Warner and Pleeter (2001, p. 38, Section III.A), Harrison et al. (2002, p. 1611, section III.A).

Unfortunately, the random preference approach may lead to violations of weak stochastic transitivity that are rarely documented in empirical data (*e.g.*, Rieskamp et al., 2006). This problem is similar to the Condorcet paradox in social choice. It can be illustrated with the following three choice alternatives:

(A) to receive \$74 today;
(B) to receive \$100 tomorrow;
(C) to receive \$37 today and to receive \$64 the day after tomorrow.

Consider a decision maker who maximizes discounted utility with a linear utility function and a random daily discount factor that is equally likely to take one of three possible values {1/2, 3/4, 1}. This decision maker chooses B over A with probability 2/3 (when the realized value of the discount factor happens to be 3/4 and 1). This decision maker also chooses C over B with probability 2/3 (when discount factor happens to be 1/2 and 1). Yet, paradoxically, the same decision maker also chooses A over C with probability 2/3 (when discount factor happens to be 1/2 and 3/4).[2]

Another approach to generating probabilistic choice over time in empirical applications (*e.g.*, for estimating time preferences) is to embed a deterministic theory of intertemporal choice into a Fechner (1860) model of random errors (*e.g.*, Chabris et al., 2008, p. 248; Ida and Goto, 2009, p. 1174, formula 1; Tanaka et al., 2010, p. 567, equation 1; Olivier et al., 2013. section 2.1, p. 617) or Luce (1959) choice model (*e.g.*, Andersen et al., 2008, p. 599, equation 9; Meier and Sprenger, 2015, p. 276, equation 1). This paper presents behavioural characterization (axiomatization) of such models. In

E-mail address: p.blavatskyy@montpellier-bs.com.

1 Also known as constant or exponential discounting.

2 Jackson and Yariv (2015, p. 164) give a similar example in the context of collective decision making over time when three individuals with different discount factors vote by the majority rule. Jackson and Yariv (2015, proposition 2) prove that collective intertemporal choice by majority voting is intransitive whenever the largest group of individuals with identical discount factors is smaller than the majority.

https://doi.org/10.1016/j.jmateco.2017.10.002
0304-4068/

particular, we consider a general class of an additively separable utility that includes discounted utility (Samuelson, 1937), quasi-hyperbolic discounting (Phelps and Pollak, 1968), generalized hyperbolic discounting (Loewenstein and Prelec, 1992) and liminal discounting (Pan et al., 2013).

The remainder of the paper is organized as follows. Mathematical notation is introduced in Section 2. Behavioural assumptions (axioms) required for embedding an additively separable utility into Fechner (1860) model of random errors (also known as strong utility) are presented in Section 3. Section 4 presents examples where probabilistic extensions of classical utility representations of time preferences are consistent with some behavioural patterns that challenge the descriptive validity of the original (deterministic) theories (*e.g.* some instances of the common difference effect). Behavioural assumptions (axioms) required for embedding an additively separable utility into Luce (1959) choice model (also known as strict utility) are presented in Section 5. Section 6 concludes.

2. Mathematical notation

Let T denote a nonempty finite set that is called *time*. An element $t \in T$ is called a *moment of time*. Let X denote a connected set. An element $x \in X$ is called an *outcome* or *payoff*. A *stream* of intertemporal outcomes $f : T \to X$ is a function from T to X. The set of all streams is denoted by F. A constant stream that yields one outcome $x \in X$ at all moments of time, is denoted by $\boldsymbol{x} \in F$.

A decision maker chooses between streams of intertemporal outcomes in a probabilistic manner. The primitive of choice is a binary choice probability function $P : F \times F \to [0, 1]$. In particular, notation $P(f, g) \in [0, 1]$ denotes the probability that a decision maker chooses stream $f \in F$ over stream $g \in F$ in a direct binary choice.

A traditional binary preference relation is a special case of a binary choice probability function: a decision maker strictly prefers stream f over stream g when $P(f, g) = 1$; a decision maker is indifferent between streams f and g when $P(f, g) \in (0, 1)$; and a decision maker strictly prefers stream g over stream f when $P(f, g) = 0$. However, axiomatic characterization of the models of probabilistic intertemporal choice can benefit from the representation theorems for the following auxiliary binary preference relation. A decision maker strictly prefers stream f over stream g when $P(f, g) > 0.5$; a decision maker is indifferent between streams f and g when $P(f, g) = 0.5$; and a decision maker strictly prefers stream g over stream f when $P(f, g) < 0.5$.

For a compact notation, let notation $x_t f$ denote a stream of intertemporal outcomes that results from stream f by replacing $f(t)$ with outcome $x \in X$ for some moment $t \in T$. A moment of time $t \in T$ is called *null* (or inessential) if $P(x_t f, y_t f) = 0.5$ for all $x, y \in X$ and $f \in F$. In this case, a decision maker does not really care which outcome is received at moment t and chooses at random (with probabilities 50%–50%) between any two streams that yield the same outcomes in all moments of time but moment t. For example, a moment of time, which will occur in 100 years, is null for many decision makers who may not care about outcomes received in such a distant future. If there exist two outcomes $x, y \in X$ such that $P(x_t f, y_t f) \neq 0.5$, then a moment of time t is called *nonnull* (or essential).

Any model of intertemporal choice is degenerate (static) when there is only one nonnull moment of time. In this case, for the existence of a continuous utility function we need an additional assumption that X is a separable set (*i.e.*, X contains a countable subset whose closure is X), *cf.* Debreu (1954, Theorem I, p. 162).

Let $\boldsymbol{x} \in F$ denote a stream that yields outcome $x \in X$ at every moment $t \in T$. Finally, let $t_0 \in T$ denote the present moment of time.

3. Axiomatization of an additively separable utility embedded into Fechner model

We assume that a binary choice probability function $P : F \times F \to [0, 1]$ satisfies the following two standard axioms.

Axiom 1 (*Completeness*)**.** $P(f, g) + P(g, f) = 1$ for all $f, g \in F$.

Axiom 2 (*Weak Stochastic Transitivity*)**.** If $P(f, g) \geq 0.5$ and $P(g, h) \geq 0.5$ then $P(f, h) \geq 0.5$ for all $f, g, h \in F$.

Let us consider an auxiliary binary preference relation $f \succcurlyeq g$ if and only if $P(f, g) \geq 0.5$. This preference relation must be rational (due to Axioms 1–2). It can be represented by a real-valued utility function $U : F \longrightarrow \mathbb{R}$ when $f \succcurlyeq g$ if and only if $U(f) \geq U(g)$, for all $f, g \in F$. Next, we add two additional behavioural assumptions (Axioms 3 and 4), which insure that auxiliary binary preference relation $\succcurlyeq$ admits a (continuous and additively separable) real-valued utility representation.

Axiom 3 (*Probabilistic Continuity*)**.** The sets $\{f \in F : P(f, g) \geq 0.5\}$ and $\{f \in F : P(g, f) \geq 0.5\}$ are closed for all $g \in F$.

Axiom 3 (formulated for a binary relation $\succcurlyeq$) is used in a connected topology approach. Alternatively, Axiom 3 can be replaced with two implications of continuity that are known as solvability and the Archimedean axiom. This is known as an algebraic approach (*e.g.*, Wakker, 1988; Köbberling and Wakker, 2003, p. 398).

Axiom 4 (*Probabilistic Cardinal Independence*)**.** If $P(x_t f, y_t g) \geq 0.5$, $P(x_t g, z_t f) \geq 0.5$ and $P(y_s h, x_s k) \geq 0.5$ then $P(x_s h, z_s k) \geq 0.5$ for all $x, y, z \in X$; $f, g, h, k \in F$; any nonnull moment of time $t \in T$ and any moment of time $s \in T$.

Axiom 4 (formulated for a binary relation $\succcurlyeq$) is known as cardinal independence (*e.g.*, Blavatskyy, 2013) or standard sequence invariance (e.g., Krantz et al., 1971, Section 6.11.2). It is a weaker version of tradeoff consistency (Wakker, 1984, 1989) or Reidemeister closure condition in geometry (Blaschke and Bol, 1938).

Intuitively, Axiom 4 can be interpreted as follows. A decision maker is more likely to choose stream f over stream g when $f(t) = x$ and $g(t) = y$. However, this decision maker becomes more likely to choose g over f when we replace outcome x with outcome z in stream f and outcome y—with outcome x in stream g. The same decision maker is also more likely to choose stream h over stream k when $h(s) = y$ and $k(s) = x$. Now, if we replace outcome x with outcome z in stream k and outcome y—with outcome x in stream h, this should make h even more desirable (in the absence of any intertemporal substitution or complementarity effects). Thus, we should expect our decision maker to continue to choose h with a higher probability than k. Axiom 4 imposes such consistency for all moments of time. Effectively, Axiom 4 rules out any intertemporal substitution or complementarity effects.

Proposition 1. *A binary choice probability function $P : F \times F \to [0, 1]$ satisfies Axioms 1, 2, 3 and 4 if and only if there exist a discount function $D : T \longrightarrow [0, 1]$ and a continuous utility function $u : X \longrightarrow \mathbb{R}$ such that an auxiliary binary preference relation $\succcurlyeq$ can be represented by an additively separable utility function (1).*[3]

$$U(f) = \sum_{t \in T} D(t)\, u \circ f(t) \tag{1}$$

[3] In other words, $P(f, g) \geq 0.5$ implies $U(f) \geq U(g)$ and vice versa.

Discount function $D(.)$ is unique under conventional normalization $D(t_0) = 1$, except for a trivial case when all moments of time are null. Utility function $u(.)$ is unique up to a positive affine transformation if at least two moments of time are nonnull.

The proof follows immediately from proposition 1 in Blavatskyy (2013).

Special cases of additively separable utility (1) include discounted utility or constant (exponential) discounting model of Samuelson (1937), quasi-hyperbolic discounting (Phelps and Pollak, 1968), generalized hyperbolic discounting (Loewenstein and Prelec, 1992) and liminal discounting (Pan et al., 2013).

Next, we assume that any two streams of intertemporal outcomes can be swapped (interchanged) in a binary choice problem (without affecting binary choice probabilities) whenever a decision maker is "probabilistically indifferent" between these two streams (*i.e.* when the decision maker chooses with probabilities 50%–50% between these two streams). This property is known as "interchangeability" in the literature (*e.g.*, axiom 5 in Blavatskyy, 2008, p. 1051) and it is implied by strong stochastic transitivity and completeness (*cf.* Davidson and Marschak, 1959).

Axiom 5 (*Interchangeability*). $P(f, g) = P(f, h)$ for all $f, g, h \in F$ such that $P(g, h) = 0.5$.

According to Axiom 5, binary choice probability $P(f, g)$ does not change if we replace stream f with another stream k such that $P(f, k) = 0.5$ and it also does not change if we replace stream g with another stream h such that $P(g, h) = 0.5$. In other words, binary choice probability $P(f, g)$ can be written as a function $\varphi(U(f), U(g))$, where $\varphi : \mathbb{R} \times \mathbb{R} \longrightarrow [0, 1]$ and $U : F \longrightarrow \mathbb{R}$ is a real-valued utility function that represents auxiliary binary preference relation $\succcurlyeq$. Proposition 1 together with Axiom 5 effectively implies that a binary choice probability function, which satisfies Axioms 1–5, depends only on the additively separable utility of the two streams. Several classical approaches to probabilistic choice generate binary choice probability functions that satisfy this condition. In particular, Fechner (1860) argued that the probability with which a decision maker chooses one choice alternative over another is a function of utility differences between these two choice alternatives. Such a model of probabilistic intertemporal choice (also known as strong utility) requires the following behavioural assumption.

Axiom 6 (*Outcome Substitution*). $P(\boldsymbol{x}, \boldsymbol{y}) = P(\boldsymbol{z}, \boldsymbol{w})$ for any four outcomes $x, y, z, w \in X$ such that $P(x_t f, y_t g) = P(z_t f, w_t g) = 0.5$ for some nonnull moment of time $t \in T$ and some $f, g \in F$.

Intuitively, Axiom 6 can be interpreted as follows. The probability that a decision maker chooses stream $x_t f$ over stream $y_t g$ does not change when an outcome x is replaced with outcome z and an outcome y is replaced with outcome w.[4] If such a simultaneous replacement of outcomes in one moment of time does not affect behaviour, we could also expect that such a replacement in all moments of time also does not change choice probabilities. Axiom 6 shares some similarity with axiom 8 in Blavatskyy (2012a), in the context of choice under uncertainty, and axiom 5 in Blavatskyy (2014), in the context of choice under risk. However, these two papers use the mixture operation (correspondingly on the set of acts and the set of lotteries) that is not feasible in our current context of intertemporal choice.

Proposition 2. *A binary choice probability function* $P : F \times F \to [0, 1]$ *satisfies Axioms 1–6 if and only if* $P(f, g) = \varphi(U(f) - U(g))$, *where* $\varphi : \mathbb{R} \longrightarrow [0, 1]$ *is a function satisfying* $\varphi(v) + \varphi(-v) = 1$ *for all* $v \in \mathbb{R}$ *as well as* $\varphi(v) > 0.5$ *for all* $v > 0.5$ *and utility function* $U(.)$ *takes an additively separable form (1). Uniqueness results are the same as in Proposition 1.*

Proof is presented in Appendix.

Intuitively, Axioms 5 and 6 are used in the proof of Proposition 2 as follows. Given Axiom 5, any stream f can be substituted with a "probabilistically equivalent" constant stream $\boldsymbol{x}$ such that $P(f, \boldsymbol{x}) = 0.5$. Thus, in order to construct a representation for all streams, we only need to construct a representation for constant streams. Axiom 6 accomplishes this job: a binary choice probability for constant streams depends only on the utility difference between the outcomes of these constant streams.[5]

Proposition 2 shows that Axioms 1–6 are necessary and sufficient conditions for an additively separable utility (1) embedded into a Fechner model of probabilistic choice. This model has an intuitive interpretation as an econometric model of discrete choice with homoscedastic random errors. A decision maker, who maximizes (deterministic) additively separable utility, chooses stream f over stream g when $U(f) - U(g) > 0$. If the revealed choices of this decision maker are affected by random errors or imprecision or noise, then stream f is chosen over stream g when $U(f) - U(g) > \varepsilon$, where ε denotes a random variable with zero mean that is independently and identically distributed across all pairs of streams f and g. Thus, if function $\varphi : \mathbb{R} \longrightarrow [0, 1]$ denotes a cumulative distribution function of this random variable ε, stream f is chosen over stream g with probability $P(f, g) = \varphi(U(f) - U(g))$. Moreover, if random variable ε is symmetrically distributed around its mean (zero) then function $\varphi(.)$ satisfies the restriction $\varphi(v) + \varphi(-v) = 1$. Popular parametric forms of function $\varphi(.)$ include cumulative distribution functions of the normal and logistic distribution, which corresponds to a (homoscedastic) probit and logit model of discrete choice.

4. Examples

A probabilistic extension of classical utility representations of time preferences is consistent with some behavioural patterns that challenge the descriptive validity of the original (deterministic) theories. For example, Thaler (1981, p. 202) argued that an individual could choose one apple today over two apples tomorrow and, at the same time, he or she could choose two apples in one year plus one day over one apple in one year. Discounted utility theory (Samuelson, 1937) cannot rationalize such a choice pattern. Yet, discounted utility embedded into Fechner model of probabilistic choice could rationalize some instances of such switching behaviour.

The probability that a decision maker chooses one apple today over two apples tomorrow is given by $p = \varphi(u(1) - \delta u(2))$, where $\delta \in (0, 1)$ denotes a daily discount factor. The probability that a decision maker chooses two apples in one year plus one day over one apple in one year is $q = \varphi\left(\delta^{366} u(2) - \delta^{365} u(1)\right)$. Note that utility difference $\delta^{366} u(2) - \delta^{365} u(1)$ is closer to zero than utility difference $u(1) - \delta u(2)$. Hence, the second binary choice probability q is closer to 0.5 than the first binary choice probability p.

This implies that a decision maker can choose one apple today over two apples tomorrow with a probability very close to one; but the same decision maker chooses two apples in one year plus one day over one apple in one year with a probability close to (but less than) one half. In other words, we can observe many choice patterns where a decision maker prefers one apple today but prefers two apples in one year plus one day. On the other hand, we

[4] In the tradition of Fechner (1860), one could argue that the "strength of preference" for x over y is the same as the "strength of preference" for z over w.

[5] If $P(x_t f, y_t g) = P(z_t f, w_t g) = 0.5$ then the utility difference between outcomes x and y is the same as that between z and w due to Proposition 1.

can observe only few instances of a decision maker choosing two apples tomorrow and choosing one apple in one year. Discounted utility embedded into Fechner model of probabilistic choice can rationalize a statistically significant asymmetry between these two choice patterns. However, the model cannot rationalize a switching modal pattern, when a decision maker chooses one apple today over two apples tomorrow with a probability greater than 0.5 and, at the same time, he or she chooses two apples in one year plus one day over one apple in one year also with a probability greater than 0.5. Similarly, discounted utility embedded into Fechner model can rationalize some instances of the common difference effect (Loewenstein and Prelec, 1992, section II.1, p. 574; Shane et al., 2002, p. 361).[6]

Rubinstein (2003, section 3.1, p. 1211) reports an experiment where subjects prefer receiving \$607.07 on June 17th, 2005 instead of receiving \$467 on June 17th, 2004. At the same time, the same subjects prefer receiving \$467 on June 16th, 2005 instead of receiving \$467.39 on June 17th, 2005. Rubinstein (2003, p. 1212) showed that such a switching choice pattern falsifies discounted utility (Samuelson, 1937), quasi-hyperbolic discounting (Phelps and Pollak, 1968) and generalized hyperbolic discounting (Loewenstein and Prelec, 1992). Yet, these theories embedded into a Fechner model can rationalize some instances of such switching behaviour.

Specifically, an individual chooses a stream that yields \$607.07 on June 17th, 2005 over a stream that yields \$467 on June 17th, 2004 with probability $\varphi(\Delta_1)$, where

$$\Delta_1 = D(17.06.05)\, u(\$\,607.07) - D(17.06.04)\, u(\$\,467).$$

The same decision maker chooses a stream that yields \$467 on June 16th, 2005 over a stream that yields \$467.39 on June 17th, 2005 with probability $\varphi(\Delta_2)$, where

$$\Delta_2 = D(16.06.05)\, u(\$\,467) – D(17.06.05)\, u(\$\,467.39).$$

Under discounted utility, quasi-hyperbolic discounting and generalized hyperbolic discounting (with a conventional non-increasing marginal utility of money) both utility differences Δ_1 and Δ_2 cannot be simultaneously positive (*cf.* (Rubinstein, 2003, p. 1212). Yet, if utility difference $\Delta_1 > 0$ is larger in absolute value than utility difference $\Delta_2 < 0$[7] then binary choice probability $\varphi(\Delta_2)$ is closer to 0.5 than binary choice probability $\varphi(\Delta_1)$. In such a case, a decision maker is more likely to choose \$607.07 on June 17th, 2005 in the first decision problem and \$467 on June 16th, 2005—in the second decision problem rather than to choose \$467 on June 17th, 2004 in the first decision problem and \$467.39 on June 17th, 2005—in the second decision problem. Therefore, discounted utility, quasi-hyperbolic discounting or generalized hyperbolic discounting theory embedded into Fechner model of probabilistic choice can rationalize a statistically significant asymmetry between these two choice patterns. However, these theories cannot rationalize a switching modal choice pattern (when the majority of subjects prefer receiving \$607.07 on 17.06.2005 in the first decision problem and \$467 on 16.06.2005—in the second decision problem).

5. Axiomatization of an additively separable utility embedded into Luce model

We already established that a binary choice probability function satisfying Axioms 1–5 depends only on the additively separable utility of the two streams (due to Proposition 1 and Axiom 5). So far, we considered a special case of such a binary choice probability function that depends on the difference in additively separable utilities of the two streams (which is known as Fechner (1860) model of random errors or strong utility). Alternatively, we can consider another special case when a binary choice probability function depends on the ratio of additively separable utilities of the two streams (which is known as Luce 1959, choice model or strict utility). Specifically, instead of Axiom 6 we make an alternative behavioural assumption presented as Axiom 7.

Axiom 7 (*Odds Ratio Independence*)**.** The odds ratio $\left[\frac{P(f,g)}{P(g,f)}\right] / \left[\frac{P(f,h)}{P(h,f)}\right]$ is independent of f for any streams $f, g, h \in F$.

Intuitively, Axiom 7 can be illustrated on the following example. Consider an investor who chooses among three intertemporal streams: holding cash (f), investment in bonds (g), and investment in stocks (h). We compare the relative chance that this investor decides to hold cash instead of investing into bonds with the relative chance that this investor decides to hold cash instead of investing into stocks. According to Axiom 6, this odds ratio depends only on the characteristics of bonds and stocks and it does not depend on the characteristics of the third intertemporal stream (holding cash).

A weaker version of axiom 7 was used in Blavatskyy (2011, axiom 6, p. 544) and Blavatskyy (2012b, axiom 6) in the context of choice under risk. These papers require the odds ratio to be independent of an irrelevant alternative only if the least upper bound of the two lotteries with this irrelevant alternative (in terms of the first-order stochastic dominance) is the same. Axiom 7 always requires the odds ratio to be independent of an irrelevant alternative (whether the least upper bounds in terms of the first-order temporal dominance are the same or not).

Proposition 3. *A binary choice probability function* $P : F \times F \to [0, 1]$ *satisfies Axioms 1–5 and 7 if and only if* $P(f, g) = \nu \circ U(f) / [\nu \circ U(f) + \nu \circ U(g)]$, *where* $\nu : \mathbb{R} \longrightarrow \mathbb{R}$ *is a continuous function that is either decreasing in the negative range or increasing in the positive range and utility function* $U : F \longrightarrow \mathbb{R}$ *takes an additively separable form (1). Uniqueness results are the same as in Proposition 1.*

Proof is presented in the Appendix.

Proposition 3 shows that Axioms 1–5 and 7 are necessary and sufficient conditions for embedding an additively separable utility (1) into Luce's choice model (strict utility). In a special case when $\nu(.)$ is an exponential function, this model becomes equivalent to an additively separable utility embedded into Fechner model with function $\varphi(.)$ being the cumulative distribution functions of the logistic distribution.

6. Conclusion

This paper considers choice between streams of intertemporal payoffs when a decision maker chooses in a probabilistic manner. First of all, we consider a model where the probability that stream f is chosen over stream g depends on the difference in utility between f and g, which is known as Fechner (1860) model of random errors or a strong utility. Second, we consider a model where this probability depends on the ratio of utilities of f and g, which is known as Luce (1959) choice model or strict utility.

The main contribution of the paper is to provide behavioural restrictions on a binary choice probability function that are necessary and sufficient for the two well-known models of probabilistic intertemporal choice. Such restrictions are known for (subjective) expected utility theory embedded into Fechner (1860) or Luce (1959) model in the context of choice under risk (uncertainty) *cf.* Blavatskyy (2008, 2012a). However, axiomatic characterization of

[6] This empirical observation is similar to the observation that expected utility theory embedded into Fechner model can rationalize some instances of the common ratio effect.

[7] For example, such a case is possible when all discount factors are nearly identical and utility function $u(.)$ is linear.

probabilistic choice under risk/ uncertainty typically exploits the mixture operation on the set of lotteries/acts that does not have an intuitive counterpart in the context of intertemporal choice. This paper provides the required behavioural assumptions (most notably Axioms 6 and 7) that allow constructing a probabilistic extension of an additively separable utility (characterized by standard Axioms 1–4) for choice over time.

An additively separable utility that we consider in this paper is unique up to a positive affine transformation (except for the static case when there is only one nonnull moment of time). Thus, for example, multiplication of utility function by a positive constant should not affect binary choice probabilities. In general, this does not hold in the first (Fechner) model, which may be viewed as its disadvantage. Similarly, addition of a positive constant to utility function should not affect binary choice probabilities. In general, this does not hold in the second (Luce's choice) model with a power function $\nu(.)$, which may be viewed as its disadvantage.

Acknowledgements

I am grateful to the Co-Editor Simon Grant as well as two anonymous referees for their helpful comments and suggestions.

Appendix

Proof of Proposition 2. First, we prove that binary choice probability function $P(f,g) = \varphi(U(f) - U(g))$ satisfies Axioms 1–6.

For all $f, g \in F$ we have $P(f,g) + P(g,f) = \varphi(U(f) - U(g)) + \varphi(U(g) - U(f)) = 1$ with the last equality due to the property $\varphi(v) + \varphi(-v) = 1$ of function $\varphi : \mathbb{R} \longrightarrow [0,1]$. Thus, Axiom 1 is satisfied.

For all $f, g \in F$ such that $P(f,g) \geq 0.5$ we must have $U(f) \geq U(g)$ (otherwise, if $U(g) - U(f) > 0$, then $\varphi(U(g) - U(f)) > 0.5$ due to the property $\varphi(v) > 0.5$ for all $v > 0.5$ of function $\varphi : \mathbb{R} \longrightarrow [0,1]$; and $P(g,f) > 0.5$, which together with $P(f,g) \geq 0.5$ contradicts Axiom 1). Similarly, for $g, h \in F$ such that $P(g,h) \geq 0.5$ we must have $U(g) \geq U(h)$. If $U(f) \geq U(g)$ and $U(g) \geq U(h)$ then $U(f) - U(h) \geq 0$ and $\varphi(U(f) - U(h)) \geq 0.5$ due to the property $\varphi(v) > 0.5$ for all $v > 0.5$. Hence, we must have $P(f,h) \geq 0.5$ and Axiom 2 is satisfied.

The sets $\{f \in F : P(f,g) \geq 0.5\}$ and $\{f \in F : P(g,f) \geq 0.5\}$ are equivalent to the sets $\{f \in F : U(f) \geq U(g)\}$ and $\{f \in F : U(g) \geq U(f)\}$ correspondingly (as a consequence of property $\varphi(v) > 0.5$ for all $v > 0.5$). The latter two sets are closed for a continuous function $U(.)$. Thus, Axiom 3 is satisfied.

If $P(x_t f, y_t g) \geq 0.5$ then $U(x_t f) \geq U(y_t g)$ as a consequence of property $\varphi(v) > 0.5$ for all $v > 0.5$. Similarly, $P(x_t g, z_t f) \geq 0.5$ implies $U(x_t g) \geq U(z_t f)$ and $P(y_s h, x_s k) \geq 0.5$ implies $U(y_s h) \geq U(x_s k)$. If function $U(.)$ has an additively separable form (1) then inequalities $U(x_t f) \geq U(y_t g)$ and $U(x_t g) \geq U(z_t f)$ imply

$$D(t)\left[u(x) - u(y)\right] > \sum_{\substack{r \in T \\ r \neq t}} D(r)\left[u \circ g(r) - u \circ f(r)\right] > D(t)\left[u(z) - u(x)\right]. \tag{2}$$

Hence, we must have $u(x) - u(y) > u(z) - u(x)$. If function $U(.)$ has form (1) then inequality $U(y_s h) \geq U(x_s k)$ implies

$$\sum_{\substack{r \in T \\ r \neq s}} D(r)\left[u \circ h(r) - u \circ k(r)\right] > D(s)\left[u(x) - u(y)\right] > D(s)\left[u(z) - u(x)\right]$$

Rearranging the left-most and the right-most part of this inequality yields $U(x_s h) - U(z_s k) \geq 0$. This implies $\varphi(U(x_s h) - U(z_s k)) \geq 0.5$ due to the property $\varphi(v) > 0.5$ for all $v > 0.5$. Hence, $P(x_s h, z_s k) \geq 0.5$ and Axiom 4 holds.

For all $f, g, h \in F$ such that $P(g,h) = 0.5$ we have $\varphi(U(g) - U(h)) = 0.5$. If $U(g) > U(h)$ then we must have $\varphi(U(g) - U(h)) > 0$, due to the property $\varphi(v) > 0.5$ for all $v > 0.5$, which contradicts $\varphi(U(g) - U(h)) = 0.5$. Similarly, $U(g) < U(h)$ implies $\varphi(U(g) - U(h)) < 0$, which contradicts $\varphi(U(g) - U(h)) = 0.5$. Hence, we must have $U(g) = U(h)$. In such a case, $P(f,g) \equiv \varphi(U(f) - U(g)) = \varphi(U(f) - U(h)) \equiv P(f,h)$. Thus, Axiom 5 is satisfied.

Finally, if $P(x_t f, y_t g) = 0.5$ then $U(x_t f) = U(y_t g)$ and if $P(z_t f, w_t g) = 0.5$ then $U(z_t f) = U(w_t g)$. If function $U(.)$ has an additively separable form (1) then equalities $U(x_t f) = U(y_t g)$ and $U(z_t f) = U(w_t g)$ imply

$$D(t)\left[u(x) - u(y)\right] = \sum_{\substack{r \in T \\ r \neq t}} D(r)\left[u \circ g(r) - u \circ f(r)\right] = D(t)\left[u(z) - u(w)\right].$$

Thus, if moment of time $t \in T$ is nonnull, we must have $u(x) - u(y) = u(z) - u(w)$. This implies $U(\boldsymbol{x}) - U(\boldsymbol{y}) = U(\boldsymbol{z}) - U(\boldsymbol{w})$. Hence, $P(\boldsymbol{x}, \boldsymbol{y}) \equiv \varphi(U(\boldsymbol{x}) - U(\boldsymbol{y})) = \varphi(U(\boldsymbol{z}) - U(\boldsymbol{w})) \equiv P(\boldsymbol{z}, \boldsymbol{w})$. Thus, Axiom 6 is satisfied.

Next, we prove that a binary choice probability function, which satisfies Axioms 1–6, must be $P(f,g) = \varphi(U(f) - U(g))$ with function $U(.)$ taking the additively separable form (1).

Let us consider two arbitrary streams $f, g \in F$. Let $x \in X$ denote an outcome such that $P(f, \boldsymbol{x}) = 0.5$ and let $y \in X$ denote an outcome such that $P(g, \boldsymbol{y}) = 0.5$ (the existence of these two outcomes follows from Axiom 5 and the fact that X is a connected set). Due to Axiom 5 we must have $P(f,g) = P(\boldsymbol{x}, \boldsymbol{y})$. If Axioms 1–4 hold then, by Proposition 1, equality $P(f, \boldsymbol{x}) = 0.5$ implies

$$\sum_{t \in T} D(t)\, u \circ f(t) = u(x)\left[\sum_{t \in T} D(t)\right] \tag{3}$$

and equality $P(g, \boldsymbol{y}) = 0.5$ implies

$$\sum_{t \in T} D(t)\, u \circ g(t) = u(y)\left[\sum_{t \in T} D(t)\right] \tag{4}$$

where $D : T \longrightarrow [0,1]$ denotes a discount function and $u : X \longrightarrow \mathbb{R}$ denotes a continuous utility function (with uniqueness results the same as in Proposition 1).

Let us fix one arbitrary outcome $x_0 \in X$ and let us select an outcome $z \in X$ such that $P(x_t h, y_t k) = P(z_t h, {x_0}_t k) = 0.5$ for some nonnull moment of time $t \in T$ and some $h, k \in F$. If Axioms 1–4 hold then, by Proposition 1, equality $P(x_t h, y_t k) = 0.5$ implies

$$D(t)\, u(x) + \sum_{\substack{s \in T, \\ s \neq t}} D(s)\, u \circ h(s) = D(t)\, u(y) + \sum_{\substack{s \in T, \\ s \neq t}} D(s)\, u \circ k(s) \tag{5}$$

and equality $P(z_t h, {x_0}_t k) = 0.5$ implies

$$D(t)\, u(z) + \sum_{\substack{s \in T, \\ s \neq t}} D(s)\, u \circ h(s) = D(t)\, u(x_0) + \sum_{\substack{s \in T, \\ s \neq t}} D(s)\, u \circ k(s). \tag{6}$$

Subtracting Eq. (6) from Eq. (5) and dividing by $D(t)$, which cannot be equal to zero because moment of time t is nonnull, we obtain

$$u(z) = u(x_0) + u(x) - u(y). \tag{7}$$

Finally, using Eqs. (3) and (4), we can rewrite Eq. (7) as

$$u(z) = u(x_0) + \frac{\sum_{t\in T} D(t)\, u\circ f(t) - \sum_{t\in T} D(t)\, u\circ g(t)}{\sum_{t\in T} D(t)}.$$

According to Axiom 6, we have $P(\boldsymbol{x}, \boldsymbol{y}) = P(\boldsymbol{z}, \boldsymbol{x}_0)$ so that $P(f, g) = P(\boldsymbol{z}, \boldsymbol{x}_0)$. Since outcome $x_0 \in X$ is fixed, probability $P(f, g)$ depends only on the (utility of) outcome z, which itself if a function of $U(f) - U(g)$. Therefore, binary choice probability $P(f, g)$ must be a function of differences in (continuous and additively separable) utility of streams f and g:

$$P(f, g) = \varphi\left(\sum_{t\in T} D(t)\, u\circ f(t) - \sum_{t\in T} D(t)\, u\circ g(t)\right).$$

If Axiom 1 holds then function $\varphi : \mathbb{R} \longrightarrow [0, 1]$ must satisfy the restriction $\varphi(v) + \varphi(-v) = 1$. According to Proposition 1, if Axioms 1–4 hold then an auxiliary binary preference relation $\succcurlyeq$ can be represented by an additively separable utility function (1) so that $P(f, g) > 0.5$ whenever $U(f) - U(g) > 0$ for all $f, g \in F$. Thus, function $\varphi : \mathbb{R} \longrightarrow [0, 1]$ must satisfy the restriction $\varphi(v) > 0.5$ for all $v > 0.5$. □

Proof of Proposition 3. It is relatively straightforward to show that binary choice probability function $P(f, g) = \nu \circ U(f) / [\nu \circ U(f) + \nu \circ U(g)]$ satisfies Axioms 1–5 and 7. We only prove that a binary choice probability function, which satisfies Axioms 1–5 and 7, must be $P(f, g) = \nu \circ U(f) / [\nu \circ U(f) + \nu \circ U(g)]$ with function $U(.)$ taking the additively separable form (1).

According to Axiom 7, for any streams $f, g, h \in F$ we must have

$$\left[\frac{P(f,g)}{P(g,f)}\right] / \left[\frac{P(f,h)}{P(h,f)}\right] = \vartheta(g,h) \tag{8}$$

where $\vartheta : F \times F \to \mathbb{R}$ is an arbitrary function.

Since Eq. (8) holds for any stream g, it holds also for stream $g = f$. Plugging $g = f$ into Eq. (8) immediately yields

$$\frac{P(h,f)}{P(f,h)} = \vartheta(f,h). \tag{9}$$

Swapping streams f and g in Eq. (8) yields

$$\left[\frac{P(g,f)}{P(f,g)}\right] / \left[\frac{P(g,h)}{P(h,g)}\right] = \vartheta(f,h). \tag{10}$$

Finally, plugging Eq. (9) on the right-hand side of Eq. (10) yields

$$\left[\frac{P(g,f)}{P(f,g)}\right] / \left[\frac{P(g,h)}{P(h,g)}\right] = \frac{P(h,f)}{P(f,h)}. \tag{11}$$

This last Eq. (11) can be rearranged as

$$\frac{P(g,f)}{P(f,g)} = \frac{P(g,h)}{P(h,g)}\frac{P(h,f)}{P(f,h)}. \tag{12}$$

Eq. (12) is a generalized multiplicative Sincov functional equation (*e.g.* Aczél, 1966) for the odds ratio. Eq. (12) is satisfied for any $f, g, h \in F$ if and only if there exists a function $\mu : F \to \mathbb{R}$ such that

$$\frac{P(g,f)}{P(f,g)} = \frac{\mu(g)}{\mu(f)}. \tag{13}$$

If Axiom 1 holds then $P(g, f) = 1 - P(f, g)$ for all $f, g \in F$. Plugging this result into Eq. (13) yields

$$P(f,g) = \frac{\mu(f)}{\mu(f) + \mu(g)} \tag{14}$$

If Axiom 5 holds then function $\mu(.)$ must take the form $\mu(f) = \nu \circ U(f)$ where $\nu : \mathbb{R} \longrightarrow \mathbb{R}$ and $U : F \to \mathbb{R}$ is utility function that represents an auxiliary preference relation $f \succcurlyeq g$ if and only if $P(f, g) \geq 0.5$. Finally, if Axioms 1–4 hold then Proposition 1 implies that

$$P(f,g) = \frac{\nu\left(\sum_{t\in T} D(t)\, u\circ f(t)\right)}{\nu\left(\sum_{t\in T} D(t)\, u\circ f(t)\right) + \nu\left(\sum_{t\in T} D(t)\, u\circ g(t)\right)}$$

where $D : T \longrightarrow [0, 1]$ denotes a discount function and $u : X \longrightarrow \mathbb{R}$ denotes a continuous utility function. Uniqueness results are the same as in Proposition 1.

According to Proposition 1, if Axioms 1–4 hold then an auxiliary preference relation $\succcurlyeq$ can be represented by an additively separable utility function (1) so that $P(f, g) > 0.5$ whenever $U(f) > U(g)$ for all $f, g \in F$. Thus, function $\nu : \mathbb{R} \longrightarrow \mathbb{R}$ is decreasing in the negative range and increasing in the positive range. By Axiom 3, this function must be continuous. Thus, function $\nu(.)$ must be either decreasing in the negative range or increasing in the positive range (but not both). □

References

Aczél, J., 1966. Lectures on Functional Equations and Their Applications. Academic Press, London.
Andersen, Steffen, Harrison, Glenn W., Lau, Morten I., Elisabet Rutström, E., 2008. Eliciting risk and time preferences. Econometrica 76 (3), 583–618.
Blaschke, W., Bol, G., 1938. Geometrie der Gewebe: Topologische Fragen der Differentialgeometrie. Springer, Berlin.
Blavatskyy, Pavlo, 2008. Stochastic utility theorem. J. Math. Econom. 44 (11), 1049–1056.
Blavatskyy, Pavlo, 2011. A model of probabilistic choice satisfying first-order stochastic dominance. Manage. Sci. 57 (3), 542–548.
Blavatskyy, Pavlo, 2012a. Probabilistic subjective expected utility. J. Math. Econom. 48 (1), 47–50.
Blavatskyy, Pavlo, 2012b. Probabilistic choice and stochastic dominance. Econom. Theory 50 (1), 59–83.
Blavatskyy, Pavlo, 2013. A simple behavioral characterization of subjective expected utility. Oper. Res. 61 (4), 932–940.
Blavatskyy, Pavlo, 2014. Stronger utility. Theory and Decision 76 (2), 265–286.
Chabris, Christopher, Laibson, David, Morris, Carrie, Schuldt, Jonathon, Taubinsky, Dmitry, 2008. Individual laboratory-measured discount rates predict field behavior. J. Risk Uncertain. 37, 237–269.
Coller, Maribeth, Williams, Melonie, 1999. Eliciting individual discount rates. Exp. Econ. 2 (2), 107–127.
Davidson, D., Marschak, J., 1959. Experimental tests of a stochastic decision theory. In: Churchman, C.W., Ratoosh, P. (Eds.), Measurement: Definitions and Theories. Wiley and Sons, New York, NY.
Debreu, G., 1954. Representation of a preference ordering by a numerical function. In: Thrall, R.M., Coombs, C.H., Davis, R.L. (Eds.), Decision Processes. John Wiley and Sons, NY, pp. 159–165.
Falmagne, Jean-Claude, 1985. Elements of Psychophysical Theory. Oxford University Press, New York.
Fechner, G., 1860. Elements of Psychophysics. Holt, Rinehart and Winston, NY.
Harrison, Glenn, Lau, Morten, Williams, Melonie, 2002. Estimating individual discount rates in Denmark: a field experiment. Amer. Econ. Rev. 92 (5), 1606–1617.
Ida, Takanori, Goto, Rei, 2009. Simultaneous measurement of time and risk preferences: stated preference discrete choice modeling analysis depending on smoking behavior. Internat. Econom. Rev. 50 (4), 1169–1182.
Jackson, M.O., Yariv, L., 2015. Collective dynamic choice: The necessity of time inconsistency. Amer. Econ. J. Microeconomics 7 (4), 150–178.
Köbberling, Veronika, Wakker, Peter P., 2003. Preference foundations for nonexpected utility: A generalized and simplified technique. Math. Oper. Res. 28, 395–42.
Krantz, D.H., Luce, R.D., Suppes, P., Tversky, A., 1971. Foundations of Measurement, Vol. 1 (Additive and Polynomial Representations). Academic Press, NY.
Loewenstein, George, Prelec, Drazen, 1992. Anomalies in intertemporal choice: evidence and an interpretation. Quart. J. Econo. 107, 573–597.
Loomes, Graham, Sugden, Robert, 1995. Incorporating a stochastic element into decision theories. Eur. Econ. Rev. 39, 641–648.
Luce, R.D., 1959. Individual Choice Behavior. Wiley, New York.
Meier, S., Sprenger, C., 2015. Temporal stability of time preferences. Rev. Econ. Stat. 97 (2), 273–286.
Olivier, Toubia, Johnson, Eric, Evgeniou, Theodoros, Delquié, Philippe, 2013. Dynamic experiments for estimating preferences: an adaptive method of eliciting time and risk parameters. Manage. Sci. 59 (3), 613–640.

Pan, J., Webb, C., Zank, H., 2013. Discounting the subjective present and future, University of Manchester, Economics discussion paper series EDP-1305.

Phelps, Edmund, Pollak, Robert, 1968. On second-best national saving and game-equilibrium growth. Rev. Econom. Stud. 35, 185–199.

Rieskamp, J., Busemeyer, J., Mellers, B., 2006. Extending the bounds of rationality. J. Econ. Lit. 44, 631–661.

Rubinstein, Ariel, 2003. Economics and psychology? The case of hyperbolic discounting. Internat. Econom. Rev. 44, 1204–1216.

Samuelson, Paul, 1937. A note on measurement of utility. Rev. Econom. Stud. 4, 155–161.

Shane, Frederick, Loewenstein, George, O'Donoghue, Ted, 2002. Time discounting and time preference: A critical review. J. Econ. Lit. 40, 351–401.

Tanaka, T., Camerer, C., Nguyen, Q., 2010. Risk and time preferences: Linking experimental and household survey data from Vietnam. Amer. Econ. Rev. 100 (1), 557–571.

Thaler, Richard H., 1981. Some empirical evidence on dynamic inconsistency. Econom. Lett. 8, 201–207.

Wakker, Peter P., 1984. Cardinal coordinate independence for expected utility. J. Math. Psych. 28, 110–117.

Wakker, Peter P., 1988. The algebraic versus the topological approach to additive representations. J. Math. Psych. 32, 421–435.

Wakker, Peter P., 1989. Additive Representation of Preferences, A New Foundation of Decision Analysis. Kluwer Academic Publishers, Dordrecht, Holland.

Warner, John, Pleeter, Saul, 2001. The personal discount rate: evidence from military downsizing programs. Amer. Econ. Rev. 91 (1), 33–53.