Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 177 (2018) 780–815

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Random intertemporal choice ☆

Jay Lu [a,*], Kota Saito [b]

[a] University of California, Los Angeles, United States
[b] California Institute of Technology, United States

Received 11 April 2017; final version received 27 May 2018; accepted 15 August 2018
Available online 17 August 2018

Abstract

We provide a theory of random intertemporal choice. Agents exhibit stochastic choice over consumption due to preference shocks to discounting attitudes. We first demonstrate how the distribution of these preference shocks can be uniquely identified from random choice data. We then provide axiomatic characterizations of some common random discounting models, including exponential and quasi-hyperbolic discounting. In particular, we show how testing for exponential discounting under stochastic choice involves checking for both a stochastic version of stationarity and a novel axiom characterizing decreasing impatience.

JEL classification: D8; D9

Keywords: Stochastic choice; Intertemporal choice; Discounting; Stationarity

1. Introduction

In many economic situations, it is useful to model intertemporal choices, i.e. decisions involving tradeoffs between earlier or later consumption, as stochastic or random. For instance, in typical models of random utility used in discrete choice estimation, this randomness is driven by

☆ We want to thank David Ahn, Jose Apesteguia, Miguel Ballester, Yoram Halevy, Yoichiro Higashi, Vijay Krishna, Tomasz Strzalecki, Charlie Sprenger and participants at D-Day at Duke, LA Theory Bash, UCSD, Barcelona GSE Stochastic Choice Workshop, D-TEA and the 16th SAET Conference at IMPA for their helpful comments.
* Corresponding author.
E-mail addresses: jay@econ.ucla.edu (J. Lu), saito@caltech.edu (K. Saito).

https://doi.org/10.1016/j.jet.2018.08.005

unobserved heterogeneity where the econometrician is not privy to all the various determinants of discounting attitudes.[1] Even when considering the behavior of a single individual, intertemporal choices can still be stochastic.[2] Decisions involving tradeoffs at different points in time are heavily influenced by visceral factors which are often of an uncertain nature even from the perspective of the decision-maker.[3]

In addition to being descriptively more accurate, a model of random intertemporal choice would also be useful for welfare analysis. An agent whose discounting is random but his utility is deterministic may behave as if his discounting is deterministic but his utility is random.

However, the welfare analysis of an agent whose discounting is random would naturally be different from that of an agent with deterministic discounting. Given all these issues, any careful analysis and interpretation of behavioral patterns in intertemporal choice would require a probabilistic, or random model of discounting.

In this paper, we provide a theoretical framework to study random intertemporal choice. We model random discounting as a distribution of preference shocks to discounting attitudes. Importantly, we focus on random discounting as the sole source of stochastic choice. This allows us to precisely characterize the relationship between random discounting and stochastic choice data.[4] In applications such as demand estimation where the relevant variable of economic interest is probabilistic choice, this is a useful and important exercise. Moreover, since random discounting is modeled as preference shocks on discounting attitudes, our theory yields robust comparative statics as demonstrated in recent work by Apesteguia and Ballester (2018).

Our model is flexible enough to allow for random discounting to be interpreted in two ways. In the first interpretation, we consider stochastic choice as the aggregated choice frequencies made by agents in a group. Aggregated choices are random due to unobserved heterogeneity in the population from the perspective of an outside observer such as an econometrician. This is the case in most applications of discrete choice estimation or in typical intertemporal choice experiments.

In the second interpretation, we consider stochastic choice as probabilistic choice from a single agent due to individual shocks to discounting attitudes. For instance, the agent is asked to choose from a menu of consumption streams repeatedly over a short interval of time. Under this interpretation, we can obtain stochastic data from experiments such as in Tversky (1969), Camerer (1989), Ballinger and Wilcox (1997), and more recently Regenwetter et al. (2011) and Agranov and Ortoleva (2017).[5] In this case, final payoffs are randomized across the agent's choices, so the agent considers each choice problem independently of the others.[6] Random choice is then obtained from the frequency of the agent's repeated choices.

In both interpretations, we interpret stochastic choice as arising from *ex-ante* choices with commitment. By *ex-ante*, we mean that we observe choices before any consumption is realized. Indeed, this is the case in the experiments mentioned above. For example, in

1 For further discussions on random utility and discrete choice estimation, see McFadden (2001) and Train (2009).

2 For some recent evidence, see Short Experiments 2 of Agranov and Ortoleva (2017).

3 Frederick et al. (2002) provides a detailed discussion on such visceral influences on intertemporal choice.

4 In Section 6, we generalize our model to allow for taste shocks as well.

5 See the introduction of Agranov and Ortoleva (2017) for more experiments on stochastic data.

6 Assuming the agent is an expected utility maximizer, this payment procedure means that he considers each choice problem separately.

Agranov and Ortoleva (2017), each subject is presented with the same choice problem repeatedly and his choices are elicited before any payment.[7]

Our main contributions are twofold. First, we show that the distribution of random discounting can be uniquely identified from random choice. In other words, an outside observer such as an econometrician can recover the entire distribution of discount attitudes given sufficient stochastic choice data. Second, we provide axiomatic characterizations of our model including random exponential and quasi-hyperbolic discounting as special cases. As a result, we extend the characterizations of classic models of intertemporal choice to the domain of random choice with novel implications.

In particular, our characterization of random exponential discounting sheds new light on one of the benchmark properties of rational choice, stationarity. Originally proposed by Koopmans (1960), the classic stationarity axiom states that choices are not affected when all consumptions are delayed by the same amount of time and it is a defining property of exponential discounting. We propose a stochastic version of the stationarity axiom which we call *Stochastic Stationarity*. It states that choice *probabilities* are not affected when all consumptions are delayed by the same amount of time.

Under stochastic choice, the relationship between Stochastic Stationarity and exponential discounting is weaker; one can find a random choice model that satisfies Stochastic Stationarity and all the standard properties but is not exponential discounting.[8] In fact, testing for random exponential discounting involves checking not only for Stochastic Stationarity but a new axiom which we call Decreasing Impatience. By itself, Decreasing Impatience exactly characterizes decreasing discount ratios.[9] Testing for quasi-hyperbolic discounting involves checking for a weaker version of Stochastic Stationarity and Decreasing Impatience. All these relationships are novel and unique to random intertemporal choice.

In general, our main model is a random utility maximization model with discounted utilities. A *discount function* D is a decreasing function over time that has value 1 initially and satisfies the tail condition $\sum_{s>t} D(s) \to 0$ as $t \to \infty$. Random choice is characterized by a distribution μ on the set $\mathcal{D}$ of discount functions and a fixed taste utility u. More precisely, the probability that an infinite-period consumption stream $f = (f(0), f(1), \dots)$ is chosen from a menu F of consumptions streams is the probability that f is ranked higher than every other consumption streams in F. In other words, if we let $\rho_F(f)$ denote this probability, then

$$\rho_F(f) = \mu\left\{ D \in \mathcal{D} \;\middle|\; \sum_t D(t)\, u(f(t)) \geq \sum_t D(t)\, u(g(t)) \text{ for all } g \in F \right\}.$$

We call this a *random discounting* model. One could interpret each realization of a discount function as corresponding to a particular agent in a population (as in most applied models of random utility) or to a particular realization of a preference shock to discounting for an individual.

[7] On the other hand, by *ex-post*, we mean that we observe choices only after all consumptions are realized. It may be difficult to take this ex-post interpretation literally as consumption steams in our paper have infinite length. However, with a slight modification of our axioms for finite-period consumption streams, we can interpret our model as ex-post as well. We consider infinite streams as it is a standard domain in the literature of intertemporal choice and allows for easy axiomatic comparisons.

[8] See Proposition 2 for an explicit example.

[9] If we let $D(t)$ denote the discount factor at time t, then the discount ratio at t is given by $D(t)/D(t+1)$.

Theorem 1 shows that under a random discounting model, the distribution of discounting functions can be uniquely identified from random choice. Moreover, this identification can be achieved using only binary choice data. This extends related uniqueness results of random utility representations to our setup with infinite-period consumption streams.

Theorem 2 provides an axiomatic characterization of our main model. We introduce three new axioms: Initial Determinism, Time Monotonicity, and Impatience. Initial Determinism requires choice to be deterministic when all consumption streams differ only at time 0. Time Monotonicity requires consumption streams that dominate at every time period to be chosen for sure. Impatience requires that when a menu consists of early and delayed consumption streams, the early streams are chosen for sure. We show that these three axioms along with the standard axioms for random utility representations fully characterize the random discounting model.

We then focus on the most popular model of intertemporal choice, exponential discounting. In *random exponential discounting*, for each discount function D in the support of μ, there exists a $\delta \in (0, 1)$ such that

$$D(t) = \delta^t.$$

Theorem 3 shows that by adding two new axioms, Stochastic Stationarity and Decreasing Impatience, we can characterize random exponential discounting.

While Stochastic Stationarity is the random analog of Koopman's stationarity axiom, Decreasing Impatience is new and requires that when faced with a consumption stream and two appropriately delayed streams, either the earliest or the latest stream will be chosen for sure. Given a model of random discounting, it is equivalent to decreasing discount ratios, i.e. $D(t)/D(t+1) \geq D(t+1)/D(t+2)$ for all t almost surely. This characterization is of interest as similar deterministic versions have been studied by several papers in the literature, e.g. Halevy (2008), Saito (2011), Chakraborty and Halevy (2017) and Saito (2017).[10] While the classic stationarity axiom along with the standard axioms is sufficient for exponential discounting under deterministic choice, the role of Decreasing Impatience in characterizing random exponential discounting is a novel feature unique to random choice.

Theorem 4 shows that by weakening Stochastic Stationarity, we obtain a model of random quasi-hyperbolic discounting. In *random quasi-hyperbolic discounting*, for each D in the support of μ, there exist $\beta \in [0, 1]$ and $\delta \in (0, 1)$ such that

$$D(t) = \beta\delta^t.$$

Weak Stochastic Stationarity requires Stochastic Stationarity to hold only when comparing consumption streams that have been delayed by at least one period. Analogous to the deterministic quasi-hyperbolic discounting model, random quasi-hyperbolic discounting allows for violations of Stochastic Stationarity when comparing immediate to future consumptions.

Proposition 3 provides comparative statics for our model of random discounting. It provides a behavioral characterization of when the distribution of discount ratios under one model of random discounting first-order stochastically dominates that of another. As in the identification results, such comparisons can be made based on binary choices data only.

Finally, we consider two extensions of our model. In the first extension, we generalize our model by also allowing for unobserved shocks to the utility function. We show that the joint

[10] Those papers study a slightly stronger property where $D(t)/D(t+1) > D(t+1)/D(t+2)$ for all t, which they call *Diminishing Impatience*.

distribution of discounting and utility shocks can be uniquely identified from stochastic choice data. For instance, if we interpret random choice as reflecting repeated choices of an individual, then we can detect when two agents exhibit the same randomness in discounting attitudes but one agent's utility is more random than that of the other. This provides a measurement that captures the volatility of utility shocks independent of discounting attitudes.

In the second extension, we address the issue of dynamic inconsistency. In our main model, we assume that all choices are collected in a small interval of time. Since tests of dynamic inconsistency in the literature typically involve comparing choices from two periods far apart in time, the static nature of our baseline model does not lend itself to addressing the issue of dynamic inconsistency. In Section 6.2, we consider a richer primitive consisting of dynamic random choice data: $\{\rho^t\}_{t\in T}$, where ρ^t is the random choice collected at time period t. Using this richer data, we can then address the issue of dynamic inconsistency. In particular, we axiomatize a dynamic model of random exponential discounting. This model is characterized by the addition of one new axiom, Stochastic Dynamic Consistency, which says that the random choices are the same over time (i.e. $\rho^t = \rho^s$ for all s and t). When we study ρ^t with some given period t, we consider only consumption streams starting after period t. Hence, we interpret ρ^t as ex-ante choices at period t.

The rest of the paper is organized as follows. We first discuss the related literature below. Section 2 then introduces our model and provides the main identification result. Section 3 discusses the axioms for our random general discounting representation. Section 4 provides the characterization results for the special cases of random exponential and quasi-hyperbolic discounting. Section 5 discusses comparative statics and finally, Section 6 considers the two extensions.

1.1. Related literature

There are many recent papers that study the choice-theoretic foundations of random utility. On the theoretical side, the closest papers to ours are Gul and Pesendorfer (2006) and Lu (2016). While they do not study intertemporal choice, we provide a generalization of their results in a larger domain of choice. This extension is necessary in order for us to deal with stochastic choice over the standard domain for intertemporal preferences, i.e., the set of infinite-period consumption streams. Furthermore, our axiomatic characterizations for the random exponential and quasi-hyperbolic discounting models are new and address issues unique to intertemporal choice.

Using a different primitive, Higashi et al. (2009) also provide a model of random discounting which includes random exponential discounting as a special case. In their model, choice data consists of a preference relation over menus reflecting an agent's anticipation of future uncertainty in discount rates. In contrast, our primitive consists of random choice. More recently, Higashi et al. (2016) propose a behavioral definition of comparative impatience using their model.

Pennesi (2015) studies an intertemporal version of the famous Luce model of stochastic choice. As in Luce's model, the probability that an agent chooses a consumption stream is its weighed average utility in the menu of consumption streams. In his baseline model, each utility is evaluated according to exponential discounting but he also provides a generalization which accounts for quasi-hyperbolic discounting as well.

More recently, Apesteguia and Ballester (2018) analyze the robustness of certain random utility models of intertemporal and risky choice. They show the possibility of a fundamental problem in comparative statics that arises in the standard application of random utility models. As mentioned before, since our random discounting model belongs to the class of what they call random parameter models, we are free of their criticisms.[11] In a more recent paper, Apesteguia et al. (2017) study a case of random parameter models where the parameters can be ordered according to the single-crossing property.

2. Model

2.1. Primitives and notation

We consider agents choosing an infinite-period stream of risky payoffs, i.e. lotteries. Let time be denoted by $T := \{0, 1, 2, \dots\}$, that is, the set of all nonnegative integers. Let X be some finite set of payoffs. We model consumption at each time period as a risky payoff, that is a lottery in ΔX. Thus, a *consumption stream* corresponds to a sequence of lotteries in the space $(\Delta X)^T$. We let H denote the set of all possible consumption streams endowed with the product topology.[12] For any $p \in \Delta X$, we sometimes abuse notation and also let p denote the *constant* consumption stream that yields p in every period.

The use of risky payoffs to model consumption allows for a straightforward characterization of our model but in general, any mixture space would work as well. For any $p, q \in \Delta X$ and $a \in [0, 1]$, we let $ap + (1 - a)q$ denote the lottery (i.e. an element of ΔX) that yields $x \in X$ with probability $ap(x) + (1 - a)q(x)$. Each consumption stream $f \in H$ yields a lottery $f(t)$ at every time period t. For any $f, g \in H$ and $a \in [0, 1]$, we let $af + (1 - a)g$ denote the consumption stream (i.e. an element of H) that yields the lottery $af(t) + (1 - a)g(t)$ at every period $t \in T$.[13]

Agents choose a stream from a *menu*, that is, a finite set of consumption streams. Let $\mathcal{K}$ be the set of all menus of consumption streams endowed with the Hausdorff metric. Given any menu $F \in \mathcal{K}$, we let $\text{ext}F$ denote the extreme points of F.

Choice data in our model is a *random choice rule (RCR)* that specifies a choice distribution over consumption streams for every menu $F \in \mathcal{K}$. Let ΔH be the set of all measures over consumption streams and endow it with the topology of weak convergence. Formally, a RCR is a function $\rho : \mathcal{K} \to \Delta H$ such that $\rho_F(F) = 1$. We use the notation $\rho_F(f)$ to denote the probability that consumption stream f will be chosen in the menu F. For binary menus $F = \{f, g\}$, we use the condensed notation $\rho(f, g)$ to denote $\rho_F(f)$.

Following Lu (2016), we model indifferences by relaxing the restriction that all choice probabilities have to be fully specified. This is analogous to how under classic deterministic choice, if the agent is indifferent between two streams, then the model is silent about which stream the agent will choose. This allows the modeler to be agnostic about data that is orthogonal to the parameters of interest. For example, if two consumption streams are tied, then the stochastic choice does not specify individual choice probabilities for either stream. We model this as non-

[11] For further details, see our comparative statics results in Section 5.

[12] The product topology corresponds to point-wise convergence in that $f_k \to f$ if $f_k(t) \to f(t)$ for all $t \in T$. The corresponding metric can be defined as $d(f, g) := \sum_t \frac{1}{2^t} \frac{\|f(t)-g(t)\|}{1+\|f(t)-g(t)\|}$, where $\|\cdot\|$ is the Euclidean norm in ΔX. On a technical note, we could have alternatively used uniform convergence but this would have resulted in a continuity axiom that would be too weak. Of course, if T is finite, then both notions of convergence agree.

[13] While in our primitive, consumption lotteries are independent across time, our results would still hold if we adopted a primitive that allowed for temporal correlations of lotteries and assumed that agents are indifferent to randomization. We could accommodate preference for randomization by adopting a more general model such as a random intertemporal version of Saito (2015).

measurability and let ρ denote the corresponding outer measure without loss of generality.[14] With this interpretation, we have $\rho(f, g) = \rho(g, f) = 1$ whenever two streams f and g are tied. Define $\mathcal{K}_0 \subset \mathcal{K}$ as the subset of menus that contain no indifferences.

As mentioned in the introduction, we interpret the RCR ρ as corresponding to ex-ante choices that are observable in experiments for example. However, ρ_F must be defined for all menus F. We admit that it would be difficult to observe choices for all menus, although the richness of the domain allows for unique identification (see Theorems 1 and 6) and has been assumed in the recent literature on random choice.

2.2. *Random discounting representations*

We now describe our main model. Agents evaluate consumption streams using discounted utilities. Discounting attitudes are modeled using a *discount function* $D : T \to [0, 1]$ that is decreasing and satisfies $D(0) = 1$ and the tail convergence condition $\sum_{s>t} D(s) \to 0$ as $t \to \infty$. The tail condition ensures that consumption at time infinity is irrelevant. Let $\mathcal{D}$ be the set of all discount functions. We are now ready to formally define our main model.

Definition. ρ is said to have a *Random Discounting Representation* if there exists a probability measure μ on $\mathcal{D}$ and a vN-M function u on ΔX such that for all $F \in \mathcal{K}$ and $f \in F$

$$\rho_F(f) = \mu \left\{ D \in \mathcal{D} \;\middle|\; \sum_{t \in T} D(t)\, u(f(t)) \geq \sum_{t \in T} D(t)\, u(g(t)) \text{ for all } g \in F \right\}.$$

A Random Discounting Representation is a random utility model where the utilities are discounted utilities. Here, choice is stochastic due to preference shocks that hit discount functions directly. The probability that one stream is chosen over another is exactly the probability that one stream has a higher discounted utility than another. Note that for simplicity, our model assumes that the vN-M utility u is deterministic. In other words, the only source of stochastic choice in this model is random discounting. In Section 6.1, we consider a generalization where the utility u is random as well, in which case, we obtain a model that is characterized by a joint distribution over both discount functions and vN-M utilities.

As in standard applications of random utility, one could interpret the RCR as reflecting the proportion of agents in a population who choose one stream over another. Here, random choice is motivated by unobserved heterogeneity on the part of the econometrician. Note that the model does not impose that each agent in the population must have the same discount function across all menus. It even allows for random discounting at the individual level provided that discount functions are drawn from the same distribution μ.[15]

Alternatively, one could use our model to describe the random choice of a single agent choosing from the same set of consumption streams. For example, this is the case in typical random choice experiments where the agent is required to make repetitive choices from the same choice

[14] Formally, stochastic choice naturally includes a σ-algebra $\mathcal{H}$ on H. Given any menu F, the corresponding choice distribution ρ_F is a measure on the σ-algebra generated by $\mathcal{H} \cup \{F\}$. Without loss of generality, we let ρ denote the outer measure with respect to this σ-algebra. See Lu (2016) for details.

[15] There is some experimental evidence in support of this. In a large field experiment conducted over two years, Meier and Sprenger (2015) elicited time preferences using incentivized choice experiments. Despite changes in discounting at the individual level, they found that the aggregate distributions of discount factors to be unchanged over the two years.

set and one of his choices is randomly selected for payment at the end of the experiment. Note that this payment procedure does not affect the agent's choice if the agent is an expected utility maximizer, which is the case in our model. In this case, the richness of our model can even accommodate learning by the agent. For example, suppose $D(t) = \mathbb{E}[\delta(1) \cdots \delta(t) | \delta(1)]$ where $\delta(t)$ is a discount factor for the consumption at time $t \in T$. Notice that at the initial period, only a realization of $\delta(1)$ is known to the agent and the discounting for period t consumption is the product $\delta(1)\delta(2) \cdots \delta(t)$ of all the discount factors up to t. Since the agent does not know $\delta(2), \ldots, \delta(t)$, he considers the conditional expectation of the product given the realization of $\delta(1)$. This describes an agent who updates his belief about future discount factors based on the realization of his current discount factor.

We call the probability measure μ a *discount distribution*. Two natural special cases of the representation are the following.

Definition. A discount distribution μ is

(1) *exponential* if and only if μ-a.s. for each $t \in T$

$$D(t) = \delta^t$$

for some $\delta \in (0, 1)$;

(2) *quasi-hyperbolic* if and only if μ-a.s. for each $t > 0$

$$D(t) = \beta\delta^t$$

for some $\delta \in (0, 1)$ and $\beta \in [0, 1]$.

We say that a discount distribution is *regular* if the random utilities of two consumption streams are either always or never equal. In other words, ties either never occur or occur almost surely.[16] Regular discount distributions are dense in the set of all discount distributions. For example, if μ is quasi-hyperbolic where the distribution on $(\beta, \delta) \in [0, 1]^2$ is diffuse, then μ is regular.[17] They are a relaxation of the standard restriction in traditional random utility models where utilities are never equal and allows us to allow for indifferences. Going forward, we only consider regular discount distributions in Random Discounting Representations.

If ρ has a Random Discounting Representation, we say that it is represented by some (μ, u) where μ is regular. Our first result below shows that the discount distribution can be uniquely identified by only looking at binary choices over consumption streams.

Theorem 1. *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then the following are equivalent.*

(1) $\rho(f, g) = \tau(f, g)$ *for all* $f, g \in H$
(2) $\rho = \tau$
(3) $(\nu, v) = (\mu, \alpha u + \beta)$ *for some* $\alpha > 0$ *and* $\beta \in \mathbb{R}$.

Proof. See Appendix A.1. □

[16] Formally, this means that for all $z \in [0, 1]^T$, $D \cdot z = 0$ occurs with μ-measure zero or one.

[17] On the other hand, if the distribution of (β, δ) has multiple mass points, then it may not be regular.

Here, we provide a brief sketch of the proof for why binary choices are sufficient for identification. By condition (1) on binary choices, for any finite subset J of T, we can pin down the distributions of $\sum_{t\in J} D(t)z(t)$ for any $z \in \mathbb{R}^J$. By the Cramer–Wold Theorem, this implies that the two distributions ν and μ must have the same marginal distribution on $(D(t))_{t\in J}$. Since this is true for any finite $J \subset T$, it follows from Kolmogorov's Extension theorem that $\mu = \nu$.

3. Characterizing random discounting

We now provide an axiomatic characterization of our general discounting model. The first five axioms are known in the literature for their role in characterizing random expected utility. We will present them with limited discussion and focus on the new axioms that are novel for our model of random intertemporal choice.

The first axiom, Monotonicity, is a standard condition necessary for any random utility model. It states that the probability that a stream is chosen from a menu does not increase if we enlarge the menu.

Axiom *(Monotonicity). For any $F, G \in \mathcal{K}$, if $G \subset F$, then $\rho_G(f) \geq \rho_F(f)$.*

The next two axioms are direct consequences of the fact that the utilities in our random utility model are linear in consumption streams. Linearity is the random choice analog of the standard independence axiom.

Axiom *(Linearity). For any $F \in \mathcal{K}$, $g \in H$, and $a \in (0,1)$,*

$$\rho_F(f) = \rho_{aF+(1-a)g}(af + (1-a)g).$$

The next axiom, Extremeness, is from Gul and Pesendorfer (2006) and states that an agent can restrict himself to extreme options of a menu without loss of generality. This follows from the fact that the utilities in our model are linear. For instance, given two consumption streams and a third that is an interior mixture of the two, an agent will either choose the first or the second but never the third.

Axiom *(Extremeness). For any $F \in \mathcal{K}$, $\rho_F(\text{ext}F) = 1$.*

The following Continuity axiom is standard given the topologies we defined previously. Recall that $\mathcal{K}_0$ is the set of menus without indifferences.

Axiom *(Continuity). $\rho : \mathcal{K}_0 \to \Delta H$ is continuous.*

Finally, to avoid degenerate cases, we assume the following nondegeneracy axiom. This rules out the case where all consumption streams are tied and the agent is indifferent between all consumption streams.

Axiom *(Nondegeneracy). $\rho_F(f) < 1$ for some F and some $f \in F$.*

We now introduce three new axioms that are unique to random discounting. In our model, the utility u over consumption is fixed while discounting can be random. If the consumption streams are different only at period 0, then the choices over such consumption streams must be deterministic. This requirement is formalized by the following axiom.

Axiom *(Initial Determinism). For any* $F \in \mathcal{K}$ *and any* $f, g \in F$*, if* $f(t) = g(t)$ *for all* $t > 0$*, then* $\rho_F(\cdot) \in \{0, 1\}$.

We now introduce some useful notation. Given any two consumption streams f and g and time period $t \in T$, define the *spliced consumption stream* ftg such that

$$ftg(s) = \begin{cases} f(s) & \text{if } s < t, \\ g(s-t) & \text{if } s \geq t. \end{cases}$$

Thus, ftg is the consumption stream that is f up to period $t-1$ and then restarts with g from t onwards. In other words,

$$ftg = (f(0), f(1), \ldots, f(t-1), g(0), g(1), \ldots).$$

For any menu $F \in \mathcal{K}$ and any stream $g \in H$ we can also define the *spliced menu*

$$Ftg := \{ftg \in H \mid f \in F\}$$

Note that the sequence of menus $(Ftg)_{t \in T}$ converges to the menu F as $t \to \infty$ under the product topology. By the Continuity axiom, $\rho_{Ftg} \to \rho_F$. In other words, only consumptions in finite time matter.

Given Initial Determinism and the fact that the set of final payoffs is finite, we can pin down preferences using time 0 choice data and find a *worst consumption stream* $w \in H$ where w is a constant consumption stream and for all $f, g \in F$,

$$\rho(f1g, w1g) = 1.$$

Lemma 2 in the Appendix shows that given the standard axioms, the worst consumption stream is well-defined.

For any menu $F \in \mathcal{K}$ and time period $t \in T$, let $F(t)$ denote the menu of constant consumption streams that yield $f(t)$ at each period for all $f \in F$. Formally, $F(t) = \{f(t) \in H \mid f \in F\}$. We can now define our next axiom.

Axiom *(Time Monotonicity). For all* $F \in \mathcal{K}$ *and* $f \in F$*, if* $\rho_{F(t)1w}(f(t)1w) = 1$ *for all* $t \in T$*, then* $\rho_F(f) = 1$.

Time Monotonicity says that if the consumption at every time period of a stream is the best in a menu, then that stream must be chosen for sure. For example, suppose $F = \{f, g\}$ only consists of two consumption streams. Thus, at every period $t \in T$, $F(t)1w = \{(f(t), w, w, \ldots), (g(t), w, w, \ldots)\}$. If $(f(t), w, w, \ldots)$ is chosen over $(g(t), w, w, \ldots)$ for every t for sure, then f must be chosen over g for sure. It is the natural temporal analog of standard monotonicity axioms. Note that given the random expected utility axioms of Gul and Pesendorfer (2006), we can in fact replace the worst outcome w in Time Monotonicity with any other fixed consumption and our result would still hold.

Finally, we define *delayed consumptions*. For any $f \in H$ and $t \in T$, let $f^t := wtf$. Hence, f^t is a consumption stream that consists of f delayed by t and with w at the beginning. In other words,

$$f^t = (\underbrace{w, \ldots, w}_{t}, f(0), f(1), f(2), \ldots).$$

For example, $f^0 = f$ and $f^1 = (w, f(0), f(1), \ldots)$. Impatience below states that earlier streams are always chosen over delayed ones.

Axiom *(Impatience). For any $f \in H$ and $t \in T$, $\rho\left(f, f^t\right) = 1$.*

We are now ready to state our general representation theorem.

Theorem 2. *ρ has a Random Discounting Representation if and only if it satisfies Monotonicity, Linearity, Extremeness, Continuity, Nondegeneracy, Initial Determinism, Time Monotonicity and Impatience.*

Proof. See Appendix A.2. □

The proof of Theorem 2 consists of two main steps. First, we use the standard arguments to obtain a representation for menus that consist of streams that yield non-worst consumptions only in a finite number of time periods. The second step consists of using Kolmogorov's Extension theorem along with our Continuity to obtain the representation for all menus.

4. Random exponential and quasi-hyperbolic discounting

In this section, we focus on two of the most popular models of discounting, exponential and quasi-hyperbolic. First, we introduce the stochastic version of the classic stationarity axiom. For any $F \in \mathcal{K}$ and $t \in T$, let $F^t =: \left\{f^t \mid f \in F\right\}$ denote the delayed menu where all streams are delayed by t time periods.

Axiom *(Stochastic Stationarity). For any $f \in H$ and $t \in T$,*

$$\rho_F(f) = \rho_{F^t}\left(f^t\right).$$

This is the stochastic version of the deterministic stationarity axiom as proposed by Koopmans (1960). It is weaker than deterministic stationary in the sense that if choices are deterministic and satisfy Koopmans' stationary, then they must satisfy Stochastic Stationarity. The converse, however, is not true. In fact, there is some empirical evidence that shows how Stochastic Stationarity can be satisfied in aggregated data despite Koopmans' stationarity being frequently violated at the individual level. In a large field experiment conducted over two years, Meier and Sprenger (2015) elicited time preferences using incentivized choice experiments. They found that the aggregate distributions of discount factors and the proportion of present-biased individuals are found to be unchanged over the two years, implying that Stochastic Stationarity is satisfied in their data set.

In order to characterize random exponential discounting, Stochastic Stationarity is insufficient (see example below). First, define a *forward consumption* f^{-1} by $\left(f^{-1}\right)^1 = f$. Note that f^{-1} is well defined if and only if $f(0) = w$.

Axiom *(Decreasing Impatience). For all $f, g, h \in H$, if $f = ag^{-1} + (1-a)w$ and $g = ah^{-1} + (1-a)w$, then*

$$\rho_{\{f,g,h\}}(\{f, h\}) = 1.$$

To understand Decreasing Impatience, note that there are two aspects to intertemporal choices: the level of consumption and the timing of consumption. The condition $g = ah^{-1} + (1-a)\,w$ (or $f = ag^{-1} + (1-a)\,w$) imply that there is a trade-off between these two aspects when an agent is choosing between g and h (or f and g). By choosing g over h (or f over g), the agent consumes earlier but at a lower level of consumption due to the mixing with the worst outcome. Thus, f is the earliest, lowest consumption, h is the latest, highest consumption while g is something in between. If an agent is just impatient enough so that he weakly prefers f to g, then a lower level of impatience one time period later would imply that he weakly prefers h to g. In a menu consists of all three streams, Decreasing Impatience says that he will never choose g (baring ties).[18]

The following result shows that Decreasing Impatience characterizes decreasing discount ratios over time.

Proposition 1. *Let ρ be represented by (μ, u). Then ρ satisfies Decreasing Impatience if and only if for all $t \in T$, μ-a.s. $D(t+2) = 0$ or*

$$\frac{D(t)}{D(t+1)} \geq \frac{D(t+1)}{D(t+2)}.$$

Proof. See Appendix A.3. □

Stochastic Stationarity along with Decreasing Impatience are necessary and sufficient for a random exponential discounting model.

Theorem 3. *Let ρ be represented by (μ, u). Then μ is exponential if and only if ρ satisfies Stochastic Stationarity and Decreasing Impatience.*

Proof. See Appendix A.4. □

In the case when ρ is deterministic, we can obtain a model of exponential discounting from the classical stationarity axiom alone. Thus, the fact that Decreasing Impatience is needed for random exponential discounting is a feature unique to random intertemporal choice.[19] To illustrate this, we provide an example of random discounting that satisfies Stochastic Stationarity but is not exponential.[20] This clarifies the importance of Decreasing Impatience in random intertemporal choice. For each $\omega \in [0, 1]$ define

$$D^{\omega}(t) = \begin{cases} e^{-2n} & \text{if } t = 2n, \\ e^{-2n-\frac{1}{2}-\omega} & \text{if } t = 2n+1. \end{cases}$$

In other words, $D^{\omega} = \left(1, e^{-\frac{1}{2}-\omega}, e^{-2}, e^{-\frac{5}{2}-\omega}, e^{-4}, e^{-\frac{9}{2}-\omega}, \dots\right)$. Let μ be a random discounting representation which is uniform over $\{D^{\omega} \mid \omega \in [0, 1]\}$.

[18] Technically, Decreasing Impatience is the intertemporal analog of the extremeness axiom from Gul and Pesendorfer (2006). See Appendix B for a precise statement of this relationship.

[19] One may wonder what restrictions on discount functions Stochastic Stationarity by itself implies. For every discount function D and set of time periods $J \subset T$, let $(\tilde{D}_t)_{t \in J}$ denote the normalized discount rates at each $t \in J$. Then Stochastic Stationarity is equivalent to $(\tilde{D}_t)_{t \in J}$ having the same distributions as $(\tilde{D}_{t+s})_{t \in J}$ for all $s \in T$ and $J \subset T$.

[20] For an example of random discounting that satisfies Decreasing Impatience but is not exponential, suppose $D(t) = \delta^{t^2}$ for some δ. This shows that Stochastic Stationarity and Decreasing Impatience are independent axioms.

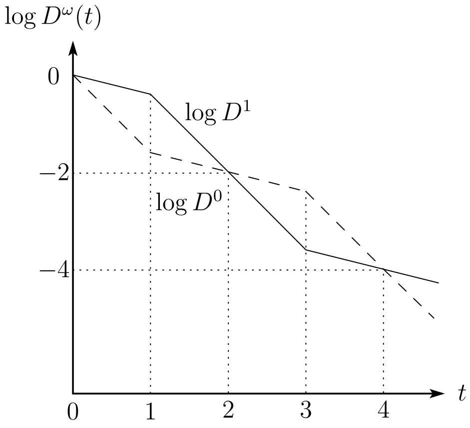

Fig. 4.1. Nonexponential μ which satisfies Stochastic Stationarity.

Proposition 2. *(i) μ is not exponential but satisfies Stochastic Stationarity; (ii) μ violates Decreasing Impatience, and its random choice cannot have a random exponential representation.*

Proof. See Appendix A.5. □

The formal proof is in the Appendix. Given Proposition 1, Fig. 4.1 clearly shows that μ violates Decreasing Impatience. We now provide a sketch of why μ satisfies Stochastic Stationarity. If t is even, then $(D^{\omega}(t), D^{\omega}(t+1), \dots) = e^{-t}(D^{\omega}(0), D^{\omega}(1), \dots)$. Then, for any $f, g \in H$, $D^{\omega} \cdot \left(u \circ f^t\right) \geq D^{\omega} \cdot \left(u \circ g^t\right) \Leftrightarrow D^{\omega} \cdot (u \circ f) \geq D^{\omega} \cdot (u \circ g)$. Note that this equivalence is captured in Fig. 4.1 by the fact that the slope of $\log D^{\omega}$ is the same between periods from 0 to 1 and periods from 3 to 4. Therefore, when t is even, each realization D^{ω} predicts no violation of the deterministic stationarity axiom.

If t is odd, then $(D^{\omega}(t), D^{\omega}(t+1), \dots) = e^{-t}(D^{1-\omega}(0), D^{1-\omega}(1), \dots)$. Then, for any $f, g \in H$, $D^{\omega} \cdot \left(u \circ f^t\right) \geq D^{\omega} \cdot \left(u \circ g^t\right) \Leftrightarrow D^{1-\omega} \cdot (u \circ f) \geq D^{1-\omega} \cdot (u \circ g)$. Note that this equivalence is captured in Fig. 4.1 by the fact that the slope of $\log D^1$ between periods from 0 to 1 is the same as the slope of $\log D^0$ between periods from 1 to 2. Therefore when t is odd, D^{ω} predicts a violation of the deterministic stationarity axiom if and only if $D^{1-\omega}$ predicts the opposite direction of the violation. The two reversals cancel each other out and "on average", deterministic stationarity is satisfied implying that Stochastic Stationarity is satisfied. Hence (i) holds. To see why (ii) holds suppose by way of contradiction that μ induces a random choice that has a random exponential representation. Since the discounting function at periods 2 and 4 are deterministic, choice must be deterministic at those periods and also at other periods such as 1 and 3, yielding a contradiction.

For the random quasi-hyperbolic discounting model, we need to weaken Stochastic Stationarity. In particular, suppose that Stochastic Stationarity holds only when menus are delayed by at least one period.

Axiom *(Weak Stochastic Stationarity). For any $F \in \mathcal{K}$ and $t \geq 1$,*

$$\rho_{F^1}\left(f^1\right) = \rho_{F^t}\left(f^t\right).$$

The deterministic version of this axiom has appeared in Hayashi (2003) and Olea and Strzalecki (2014). In our model, Weak Stochastic Stationarity along with Decreasing Impatience exactly characterize random quasi-hyperbolic discounting. As mentioned above, Decreasing Impatience is unnecessary if choices are deterministic.

Theorem 4. *Let ρ be represented by (μ, u). Then μ is quasi-hyperbolic if and only if ρ satisfies Weak Stochastic Stationarity and Decreasing Impatience.*

Proof. See Appendix A.4. □

5. Comparative statics

We now present some comparative statics for our random discounting model. First, for any RCR ρ with a Random Discounting representation, let b_ρ and w_ρ denote its best and worst consumptions respectively. For any $a \in [0, 1]$, we can define the lottery $p_\rho^a := ab_\rho + (1-a)\, w_\rho$. Note that p_ρ^a is a normalized utility that allows us to compare valuations across random choices with different tastes.

Consider two consumption streams f and g such that f provides a lower payoff than g in time period t_1 but a higher payoff than g in a later period $t_2 > t_1$. In any other time period, f and g are the same. Thus, f and g differ at two time periods and f is more back-loaded than to g. We say one RCR is *stochastically more patient* than another if the probability that the first chooses consumption stream f over g is always greater than the second.

Definition. ρ is *stochastically more patient* than τ if for any $f, g, f', g' \in H$, $a_1 < b_1$, $a_2 > b_2$ and $t_1 < t_2$ such that $f(t_i) = p_\rho^{a_i}$, $g(t_i) = p_\rho^{b_i}$, $f'(t_i) = p_\tau^{a_i}$, $g'(t_i) = p_\tau^{b_i}$ for $i \in \{1, 2\}$ and $f(s) = g(s)$, $f'(s) = g'(s)$ for all $s \notin \{t_1, t_2\}$, then

$$\rho(f, g) \geq \tau\left(f', g'\right).$$

Given two discount distributions μ and ν, let $\mu \gg \nu$ denote the fact that for all $t_1, t_2 \in T$ such that $t_1 < t_2$ the distribution of $D(t_2)/D(t_1)$ under μ first-order stochastically dominates (FOSD) its distribution under ν. This exactly captures the ordering of distributions of discount factors according to the level of patience. We now have the following result.

Proposition 3. *Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Then $\mu \gg \nu$ if and only if ρ is stochastically more patient than τ.*

Proof. Without loss of generality, we can normalize u and v such that $u\left(b_\rho\right) = v(b_\tau) = 1$ and $u\left(w_\rho\right) = v(w_\tau) = 0$. Define $f, g, f', g' \in H$, $a_1 < b_1$, $a_2 > b_2$ and $t_1 < t_2$ as in the definition of more stochastic patience. Thus,

$$\begin{aligned}
& \rho(f, g) \geq \tau\left(f', g'\right) \\
\Leftrightarrow \quad & \mu\{D \in \mathcal{D} \mid D(t_1)a_1 + D(t_2)a_2 \geq D(t_1)b_1 + D(t_2)b_2\} \\
\geq \; & \nu\{D \in \mathcal{D} \mid D(t_1)a_1 + D(t_2)a_2 \geq D(t_1)b_1 + D(t_2)b_2\} \\
\Leftrightarrow \quad & \mu\{D \in \mathcal{D} \mid D(t_1)(b_1 - a_1) \leq D(t_2)(a_2 - b_2)\} \\
\geq \; & \nu\{D \in \mathcal{D} \mid D(t_1)(b_1 - a_1) \leq D(t_2)(a_2 - b_2)\},
\end{aligned}$$

Since $a_1 < b_1$ and $a_2 > b_2$, the result follows. □

Note that this immediately implies the following result that allows us to perform FOSD comparisons of exponential discount distributions using random choice.

Corollary 1. *Let ρ and τ be represented by (μ, u) and (ν, v) respectively where both μ and ν are exponential. Then μ FOSD ν if and only if ρ is stochastically more patient than τ.*

Proof. Follows immediately from Proposition 3 above. □

One may wonder if it would be possible to generalize our definition of greater stochastic patience. Under deterministic choice, our notion of greater patience is equivalent to exhibiting a greater preference for f over g whenever f single-crosses g from below, that is, there exists some t^* such that f gives a lower (higher) payoff than g when $t \leq t^*$ $(t \geq t^*)$.[21] In stochastic choice however, this equivalence fails. In other words, it is no longer true that $\mu \gg \nu$ if and only if $\rho(f, g) \geq \tau(f, g)$ for consumption steams f and g where f single-crosses g from below. The following example illustrates.

Example 1. Let $D_1 = \left(1, \frac{1}{2}, \frac{1}{2}, 0, \dots\right)$, $D_2 = (1, 1, 0, 0, \dots)$, $D_1' = \left(1, \frac{1}{2}, 0, 0, \dots\right)$ and $D_2' = \left(1, 1, \frac{1}{2}, 0, \dots\right)$ Suppose $\mu = \frac{1}{2}\delta_{D_1} + \frac{1}{2}\delta_{D_2}$ and $\nu = \frac{1}{2}\delta_{D_1'} + \frac{1}{2}\delta_{D_2'}$. Let ρ and τ be represented by (μ, u) and (ν, v) respectively. It is easy to check that $\mu \gg \nu$. Consider $f, g \in H$ such that

$$u \circ f = (0, 1, 1, 0, \dots)$$
$$u \circ g = \left(\frac{5}{4}, 0, 0, 0, \dots\right)$$

and note that f single-crosses g from below. However, note that

$$\rho(f, g) = 0 < \frac{1}{2} = \tau(f, g).$$

6. Extension

6.1. Random vN-M utility

In this section, we consider a general model where there is randomness in both discounting and utilities. Idiosyncratic shocks to the economy may change the agent's perception of future consumption (i.e. discounting function) as well as his taste (i.e. vNM utility). One important question is to ask whether it is possible to distinguish the randomness of the discounting function from the randomness of the vNM utility. To address the question, we first provide an axiomatic characterization of the general model and then show that this distinction is possible.

We introduce new primitives as follows. Let U denote the set of vNM utilities on ΔX. Although utilities are random, we still need to assume that there exists some universally worst consumption. For instance, there may be shocks to risk aversion but the agent still prefers more money to less. Fix some outcome $w \in X$ and let $U^* \subset U$ denote the set of non-constant utilities such that $u(x) \geq u(w)$ for all $x \in X$. The existence of a worst consumption w also allows us to define delayed streams.

[21] See Benoit and Ok (2007).

Definition. ρ is said to have a *General Random Discounting Representation* if there exists a regular measure π on $\mathcal{D} \times U^*$ such that[22]

$$\rho_F(f) = \pi \left\{ (D, u) \in \mathcal{D} \times U^* \;\middle|\; \sum_t D(t)\left[u(f(t)) - u(g(t))\right] \geq 0 \text{ for all } g \in F \right\}.$$

In this case, we say ρ is represented by π. Since tastes are random in a General Random Discounting representation, Initial Determinism clearly cannot be satisfied. Nevertheless, the following condition ensures that w is a worst consumption.

Axiom *(Worst).* $\rho(f1w, w) = 1$ *for all* $f \in H$.

With this condition, we can define delayed streams as before and Impatience is well-defined. The next axiom ensures that utilities are time-invariant. It states that constant consumption streams are always chosen over streams with time-varying payoffs.[23]

Axiom *(Time Invariance). For* $t \in T$, *suppose that* $f(s) \in \{p, q\} \subset \Delta X$ *for all* $f \in F$ *and all* $s \leq t$. *If* $p, q \in F$,[24] *then*

$$\rho_{Ftw}(\{ptw, qtw\}) = 1$$

Finally, we assume a nondegeneracy condition for initial consumptions. Analogous to Nondegeneracy, this rules out the case where the agent is indifferent between all initial consumptions at time 0.

Axiom *(Initial Nondegeneracy).* $\rho_{F1w}(f1w) < 1$ *for some* F *and some* $f \in F$.

The following is the representation result for a General Random Discounting Representation.

Theorem 5. *ρ has a General Random Discounting Representation if and only if it satisfies Monotonicity, Linearity, Extremeness, Continuity, Initial Nondegeneracy, Worst, Time Invariance, and Impatience.*

Proof. See Appendix A.6. □

Finally, the following uniqueness result generalizes Theorem 1 to General Random Discounting. To see this, note that we can identify the utility shocks from streams that only have consumption at some fixed time period. We can then identify the discount shocks as any randomness above and beyond that generated by the utility shocks.

Theorem 6. *Let ρ and τ be represented by π and η respectively. Then $\pi = \eta$ if and only if $\rho(f, g) = \tau(f, g)$ for all $f, g \in H$.*

---

22 As before, regularity means that the random utilities of two consumption streams are either always or never equal.

23 Time Invariance is the random choice analog of the classic state-by-state independence condition in subjective expected utility.

24 Here, we use the convention where p and q refer to the constant consumption streams corresponding to their respective lotteries.

Proof. See Appendix A.7. □

This shows that the joint distribution of discounting and utilities can be recovered from random choice. Moreover, as before, binary choice data will suffice for this identification exercise. For instance, if we interpret the random choice as reflecting repeated choices of an individual, then we can detect when two agents exhibit the same randomness in discounting attitudes but one agent's utility is more random than that of the other. We can also be used to provide some measurement that captures the degree to which utilities vary across decision times.

6.2. Dynamic random choice

In the previous sections, we assumed that the agent's choices are static and made only at period 0. In this section, we study the agent's dynamic choice. We extend our primitive as follows. For each $t \in T$, let H^t denote the set of all consumption streams endowed with the product topology which yield the outcome w for each period $s \in T$ such that $s \leq t-1$. We denote by $\mathcal{K}^t$ the set of all menus of consumption streams endowed with the Hausdorff metric which yield the outcome w for each period $s \in T$ such that $s \leq t-1$.

Definition. For each $t \in T$, ρ^t is a function from $\mathcal{K}^t \to \Delta(H^t)$ such that $\rho^t_F(F) = 1$. We call ρ^t the *random choice rule (RCR) at period $t \in T$*.

The observable data set now consists of $\{\rho^t\}_{t\in T}$. The RCR ρ in the previous sections can be understood as ρ^0. As before, we interpret the random choice ρ^t as ex-ante choice observed at period t. Also as before, this can be interpreted as either an individual's random choice or aggregated random choice across population of agents. The latter interpretation applies in experimental settings where random choice corresponds to aggregated choices. For example, in Halevy (2015), subjects are asked to choose between a sooner but smaller consumption and a later but larger consumption at two different time periods (i.e., week 0, week 4). The aggregated choices are random because of the unobserved heterogeneity of subjects from the perspective of the outside observer. By aggregating choices across the subjects at the two time periods, we can elicit ρ^{week0} and ρ^{week4}. Note that in all these settings, we interpret the worst outcome as "no consumption".

For each $t \in T$, we can impose the same axioms on ρ^t as in the previous sections just by changing $\mathcal{K}$ to $\mathcal{K}^t$. One new axiom is a natural extension of the dynamic consistency axiom to the stochastic setting.

Axiom. *(Stochastic Dynamic Consistency) For any $t, s \in T$ such that $t < s$, for any $F \in \mathcal{K}^s$ and any $f \in F$,*

$$\rho^t_F(f) = \rho^s_F(f).$$

Since $F \in \mathcal{K}^s$, the agent's payoff is constant, namely zero, between period t and $s-1$. As a result, if the agent is dynamically consistent, then he should not change his choice at period s after making his choice at period t. Hence, we require that $\rho^t_F(f) = \rho^s_F(f)$.

Proposition 4. $\{\rho^t\}_{t\in T}$ *satisfies Stochastic Dynamic Consistency and for each $t \in T$, ρ^t satisfies the axioms in Theorems 2 and 3 (defined with $\mathcal{K}^t$ instead of $\mathcal{K}$) if and only if there exists a* *probability measure* μ *on* $[0,1]$ *and a vN-M function* u *on* ΔX *such that for all* $t \in T$, $F \in \mathcal{K}^t$, *and* $f \in F$

$$\rho^t_F(f) = \mu\left\{\delta \in [0,1] \;\middle|\; \sum_{s\in T: s\geq t} \delta^{s-t} u(f(s)) \geq \sum_{s\in T: s\geq t} \delta^{s-t} u(g(s)) \textit{ for all } g \in F\right\}. \tag{6.1}$$

Proof. See Appendix A.8. □

Here, we show that $\{\rho^t\}_{t\in T}$ satisfies Stochastic Dynamic Consistency. Fix any $t, s' \in T$ such that $t < s'$. Choose any $F \in \mathcal{K}^{s'}$ and any $f \in F$,

$$\begin{aligned}
\rho^t_F(f) &= \mu\left\{\delta \in [0,1] \;\middle|\; \textstyle\sum_{s\in T:s\geq t} \delta^{s-t} u(f(s)) \geq \sum_{s\in T:s\geq t} \delta^{s-t} u(g(s)) \text{ for all } g \in F\right\} \\
&= \mu\{\delta \in [0,1] \mid \textstyle\sum_{s\in T:s\geq s'} \delta^{s-t} u(f(s)) \geq \sum_{s\in T:s\geq s'} \delta^{s-t} u(g(s)) \text{ for all } g \in F\} \\
&\quad (\because u(h(s)) = 0 \text{ for all } h \in F \text{ and } s \in T \text{ such that } t \leq s \leq s'-1) \\
&= \mu\left\{\delta \in [0,1] \;\middle|\; \delta^{s'-t} \textstyle\sum_{s\in T:s\geq s'} \delta^{s-s'} u(f(s)) \geq \delta^{s'-t} \sum_{s\in T:s\geq s'} \delta^{s-s'} u(g(s)) \right. \\
&\qquad \left. \text{for all } g \in F\right\} \qquad (6.2) \\
&= \mu\left\{\delta \in [0,1] \;\middle|\; \textstyle\sum_{s\in T:s\geq s'} \delta^{s-s'} u(f(s)) \geq \sum_{s\in T:s\geq s'} \delta^{s-s'} u(g(s)) \right. \\
&\qquad \left. \text{for all } g \in F\right\} \\
&= \rho^{s'}_F(f).
\end{aligned}$$

To understand why Stochastic Dynamic Consistency is sufficient, note that by Theorem 1, it implies that the marginal distributions of the discount functions after a common time period are the same. Random exponential discounting then ensures that the distribution of δ must be the same.

Note that the same extension is impossible for the Random Discounting model defined in Section 2.2 where the utility is fixed and the discounting function is random and may not necessarily be exponential. This is because Stochastic Dynamic Consistency may be violated. To see this, note that the third equation of (6.2) may not hold with a general discounting function.

Appendix A. Proofs

Recall that $T = \{0, 1, \ldots, \infty\}$. For every $D \in [0,1]^T$, $f \in H$, and vN-M utility function u on ΔX, we use the condensed notation

$$D \cdot (u \circ f) := \sum_{t=0}^{\infty} D(t) u(f(t))$$

whenever the limit is well-defined, which may be infinite. Note that this converges for all $D \in \mathcal{D}$ since $\sum_{s>t} D(s) \to 0$ as $t \to \infty$ and u is bounded since X is finite. Given consumption streams $f, g \in H$ and $t \in T$, recall the spliced consumption stream

$$ftg(s) = \begin{cases} f(s) & \text{if } s < t, \\ g(s-t) & \text{if } s \geq t. \end{cases}$$

For any $F \in \mathcal{K}$, $Ftg = \{ftg \in H \mid f \in F\}$ denotes the spliced menu. Finally, recall that we use $\rho(f,g)$ to denote $\rho_{\{f,g\}}(f)$ for any $f, g \in H$.

A.1. Proof of Theorem 1

Let ρ and τ be represented by (μ, u) and (ν, v) respectively. Note that if part (3) is true, then $\rho_F(f) = \tau_F(f)$ for all $f \in H$ from the representation. Moreover, since $\rho(f,g) = \rho(g,f) = 1$ iff $\tau(f,g) = \tau(g,f) = 1$ iff f and g are tied, both RCRs have the same ties so $\rho = \tau$ and part (2) is true. Since part (2) implies part (1) trivially, we have that (3) implies (2) and (2) implies (1).

Hence, all that remains is to prove that part (1) implies part (3). Suppose (1) is true so $\rho(f,g) = \tau(f,g)$ for all $f, g \in H$. First, note that for any $p, q, r \in \Delta X$, $u(p) \geq u(q) \Leftrightarrow \mu\{D \in \mathcal{D} \mid u(p) \geq u(q)\} = 1 \Leftrightarrow \rho(p1r, q1r) = 1 \Leftrightarrow \tau(p1r, q1r) = 1 \Leftrightarrow \tau\{D \in \mathcal{D} \mid v(p) \geq v(q)\} = 1 \Leftrightarrow v(p) \geq v(q)$, so $u = \alpha v + \beta$ for some $\alpha > 0$ and a real number β. Without loss of generality, we can let $u = v$ and $w \in \Delta X$ be the worst stream for both ρ and τ. Fix some finite $J \subset T$ and let $f \in H$ be such that $f(t) = w$ for all $t \notin J$. Let $p \in \Delta X$ such that $u(p) = v(p) = a \in [0,1]$ and note that

$$\begin{aligned}
\mu\left\{D \in \mathcal{D} \;\middle|\; \sum_{t \in J} D(t)\, u(f(t)) \geq a\right\} &= \rho(f, p1w) \\
&= \tau(f, p1w) \\
&= \nu\left\{D \in \mathcal{D} \;\middle|\; \sum_{t \in J} D(t)\, v(f(t)) \geq a\right\}.
\end{aligned}$$

Since this is true for all $a \in [0,1]$ and such f, it must be that the distribution of $\sum_{t \in J} D(t)\, z(t)$ for all $z \in [0,1]^J$ must be the same under μ and ν. Note we can easily extend this for all $z \in \mathbb{R}^J_+$ by scaling so by the Cramer–Wold Theorem,[25] $(D(t))_{t \in J}$ has the same distribution under μ and ν.[26] Since this is true for all $J \subset T$, by Kolmogorov's Extension Theorem, $\mu = \nu$. This proves (3).

A.2. Proof of Theorem 2

A.2.1. Worst consumption stream is well-defined

We first prove that the worst consumption stream w is well-defined. First, we prove a technical lemma showing that under linearity, we can show the following.

Lemma 1. *If ρ satisfies Linearity, then $\rho(p1f, q1f) = \rho(p1g, q1g)$ for all $p, q \in \Delta X$ and $f, g \in H$.*

Proof. Let $r := \frac{1}{2}p + \frac{1}{2}q$ and note that

$$\frac{1}{2}(p1f) + \frac{1}{2}(q1g) = r1\left(\frac{1}{2}f + \frac{1}{2}g\right) = \frac{1}{2}(p1g) + \frac{1}{2}(q1f),$$

[25] See Billingsley (1986).

[26] For each $z \in \mathbb{R}^J_+$ we can find $k \in \mathbb{Z}_{++}$ such that $z/k \in [0,1]^J$ because J is finite. Define $\mu(D \in D^J | D \cdot z \geq a) = \mu(D \in D^J | D \cdot (z/k) \geq a/k)$. Note that the definition does not depend on k.

$$\frac{1}{2}(q1f)+\frac{1}{2}(q1g)=q1\left(\frac{1}{2}f+\frac{1}{2}g\right)=\frac{1}{2}(q1g)+\frac{1}{2}(q1f).$$

By Linearity, this implies that

$$\begin{aligned}\rho(p1f,\ q1f)&=\rho\left(\frac{1}{2}(p1f)+\frac{1}{2}(q1g),\ \frac{1}{2}(q1f)+\frac{1}{2}(q1g)\right)\\&=\rho\left(\frac{1}{2}(p1g)+\frac{1}{2}(q1f),\ \frac{1}{2}(q1g)+\frac{1}{2}(q1f)\right)=\rho(p1g,\ q1g)\end{aligned}$$

as desired. □

We can now show that the worst consumption stream $w\in H$ is well-defined.

Lemma 2. *Suppose ρ satisfies Monotonicity, Linearity, Extremeness, Continuity and Initial Determinism. Then there exists a constant consumption stream $w\in H$ such that $\rho(f1g, w1g)=1$ for all $f,g\in H$.*

Proof. Fix some consumption lottery $r\in\Delta X$. Consider the random choice rule τ on ΔX such that for any finite set of lotteries $C\subset\Delta X$ and $p\in C$,

$$\tau_C(p)=\rho_{C1r}(p1r).$$

Note that by Initial Determinism, τ is deterministic. Hence, from Lu (2016), τ can be represented by a deterministic expected utility u on ΔX. Let $w\in\Delta X$ be some worst lottery according to u. Note that w exists as X is finite. Let $w\in H$ denote the constant consumption stream that yields w every period. From Lemma 1, this implies that for any $f,g\in H$, $\rho(f1g,w1g)=\rho(f1r,\ w1r)=\tau(f(0),w)=1$, as desired. □

A.2.2. *Sufficiency of Theorem 2*

In order to prove that a Random Discounting Representation exists, we first prove it exists for a subset of menus. For each finite $J\subset T$ such that $0\in J$, let H^J be the subset of streams such that $f(t)=w$ for all $t\notin J$, where the existence of w follows from Lemma 2. Let $\mathcal{K}^J\subset\mathcal{K}$ be the subset of menus that only contain streams in H^J. Hence, we can define a RCR ρ^J on $\mathcal{K}^J$ such that for all $F\in\mathcal{K}^J$ and $f\in F$,

$$\rho_F^J(f)=\rho_F(f).$$

By the same argument as in Lu (2016), for every finite J, we can find a measure ν^J on ΔJ and a vN-M utility u on ΔX such that for every $F\in\mathcal{K}^J$ and $f\in F$

$$\rho_F^J(f)=\nu^J\{p\in\Delta J\mid p\cdot(u\circ f-u\circ g)\geq 0\text{ for all }g\in F\}$$

Note that Initial Determinism and Time Monotonicity imply that this u is fixed and independent of J. We normalize $u:\Delta X\to[0,1]$ such that $u(w)=0$. Choose $f\in H$ such that $u(f(t))=1$ for some $t\in T$ and $f(s)=w$ for all $s\neq t$. Then by Impatience, for any J such that $\{t,t+1\}\subset J$, we have

$$1=\rho\left(f,f^1\right)=\rho^J\left(f,f^1\right)=\nu^J\{p\in\Delta J\mid p(t)\geq p(t+1)\}.$$

Hence, p is decreasing ν^J-a.s. for all finite J where $0\in J$. For any $J\subset T$ such that $0\in J$, let $\mathcal{D}^J\subset[0,1]^J$ be such that $D(0)=1$ for all $D\in\mathcal{D}^J$. We can define a measure μ^J on $\mathcal{D}^J$ such that for every $F\in\mathcal{K}^J$ and $f\in F$,

$$\rho_F^J(f) = \mu^J \left\{ D \in \mathcal{D}^J \;\middle|\; D \cdot (u \circ f - u \circ g) \geq 0 \text{ for all } g \in F \right\}.$$

We now extend this representation from any finite J to all of T by using Kolmogorov's Extension Theorem. Hence, we need to check for the following consistency condition. Let $0 \in S \subset J \subset T$. For any $F \in \mathcal{K}^S$ and $f \in F$, $\mu^S \left\{ D \in \mathcal{D}^S \mid D \cdot (u \circ f - u \circ g) \geq 0 \text{ for all } g \in F \right\} = \rho_F^S(f) = \rho_F(f) = \rho_F^J(f) = \mu^J \left\{ D \in \mathcal{D}^J \mid D \cdot (u \circ f - u \circ g) \geq 0 \text{ for all } g \in F \right\}$.

Let $f \in H^S$ and $p \in \Delta X$ such that $u(p) = a \in [0, 1]$. Since $p1w \in H^S$, we then have

$$\mu^S \left\{ D \in \mathcal{D}^S \;\middle|\; D \cdot (u \circ f) \geq a \right\} = \mu^J \left\{ D \in \mathcal{D}^J \;\middle|\; D \cdot (u \circ f) \geq a \right\}.$$

In other words, for all $z \in [0, 1]^S$, the distribution of $D \cdot z$ under μ^S is the same as that under μ^J. As in the proof of Theorem 1, we can easily extend this for all $z \in \mathbb{R}_+^S$ so by Cramer–Wold, it must be that μ^S is exactly the projection of μ^J on $\mathcal{D}^S$. Formally, if we let $\chi_{JS} : \mathcal{D}^J \to \mathcal{D}^S$ be the projection mapping from $\mathcal{D}^J$ to $\mathcal{D}^S$, then

$$\mu^S = \mu^J \circ \chi_{JS}^{-1}.$$

Hence, from Kolmogorov's Extension Theorem, we know there exists a measure μ on $\mathcal{D}^T$ such that for any finite $J \subset T$ and $F \in \mathcal{K}^J$,

$$\begin{aligned} \rho_F(f) &= \mu^J \left\{ D \in \mathcal{D}^J \;\middle|\; D \cdot (u \circ f - u \circ g) \geq 0 \text{ for all } g \in F \right\} \\ &= \mu \left\{ D \in \mathcal{D}^T \;\middle|\; D \cdot (u \circ f - u \circ g) \geq 0 \text{ for all } g \in F \right\}. \end{aligned}$$

Moreover, we can assume that μ is a measure on the Borel σ-algebra corresponding to pointwise convergence on the product topology (see exercise I.6.35 of Cinlar, 2011).

We now need to generalize the representation for all $F \in \mathcal{K}$. First, for every $f \in F \in \mathcal{K}$ and finite $t \in T$, define the following two sets of maximizing discount functions

$$\begin{aligned} \mathcal{N}(f, F) &:= \left\{ D \in \mathcal{D}^T \;\middle|\; D \cdot (u \circ f - u \circ g) \geq 0 \text{ for all } g \in F \right\}, \\ \mathcal{N}^t(f, F) &:= \left\{ D \in \mathcal{D}^T \;\middle|\; D \cdot (u \circ (ftw) - u \circ (gtw)) \geq 0 \text{ for all } g \in F \right\}. \end{aligned}$$

Note that $\mathcal{N}(f, F)$ is well-defined only if $D \cdot (u \circ f - u \circ g)$ is well-defined for all $f, g \in F$.

Lemma 3. *Suppose $D \cdot (u \circ f - u \circ g)$ is well-defined for all $f, g \in F$ and $D \in \mathcal{D}^T$. Then*

(1) $\rho_F(f) = \mu(\mathcal{N}(f, F))$ *for all* $f \in F$,
(2) $\mu \left\{ D \in \mathcal{D}^T \mid D \cdot (u \circ f - u \circ g) = 0 \right\} \in \{0, 1\}$ *for all* $f, g \in F$.

Proof. We first show that if the premise holds, then $\rho_F(f) \leq \mu(\mathcal{N}(f, F))$. In order to show this, we prove that $\limsup_t \mathbf{1}_{\mathcal{N}^t(f,F)}(D) \leq \mathbf{1}_{\mathcal{N}(f,F)}(D)$ for all $D \in \mathcal{D}^T$. Suppose $\limsup_t \mathbf{1}_{\mathcal{N}^t(f,F)}(D) = 1$ so for any $t \in T$, we can find some $t' > t$ where $D \in \mathcal{N}^{t'}(f, F)$ or

$$\sum_{s \leq t'} D(s) \cdot (u(f(s)) - u(g(s))) \geq 0$$

for all $g \in F$. Since $D \cdot (u \circ f - u \circ g)$ is well-defined for all $f, g \in F$ and $D \in \mathcal{D}^T$, this implies that

$$D \cdot (u \circ f - u \circ g) = \lim_t \sum_{s \leq t} D(s) \cdot (u(f(s)) - u(g(s))) \geq 0$$

for all $g \in F$ so $D \in \mathcal{N}(f, F)$. Hence, $\limsup_t \mathbf{1}_{\mathcal{N}^t(f,F)}(D) \leq \mathbf{1}_{\mathcal{N}(f,F)}(D)$. Recall that $Ftw = \{ftw | f \in F\}$. Now, by Fatou's Lemma,

$$\begin{aligned}\lim_t \rho_{Ftw}(ftw) &= \lim_t \mu\left(\mathcal{N}^t(f, F)\right) \\ &\leq \int_{\mathcal{D}^T} \limsup_t \mathbf{1}_{\mathcal{N}^t(f,F)}(D)\, \mu(dD) \\ &\leq \int_{\mathcal{D}^T} \mathbf{1}_{\mathcal{N}(f,F)}(D)\, \mu(dD) = \mu(\mathcal{N}(f, F))\end{aligned}$$

Since $Ftw \to F$, by Continuity, this implies that

$$\rho_F(f) = \lim_t \rho_{Ftw}(ftw) \leq \mu(\mathcal{N}(f, F)) \tag{A.1}$$

as desired.

Before completing the proof of part (1), we will now prove part (2). Fix $f, g \in F$ and note that if f and g are tied, then from equation (A.1), we have $1 = \rho(f, g) \leq \mu(\mathcal{N}(f, \{f, g\}))$ and $1 = \rho(g, f) \leq \mu(\mathcal{N}(g, \{f, g\}))$ so $\mu\left\{D \in \mathcal{D}^T \;\middle|\; D \cdot (u \circ f - u \circ g) = 0\right\} = 1$.

Now, suppose f and g are not tied. Let $r \in \Delta X$ be such that $u(r) = 1$. By linearity, we can assume without loss of generality that $\frac{1}{2}u(f(0)) + \frac{1}{2}u(g(0)) < u(r)$. For any $\varepsilon > 0$, let $p_\varepsilon \in \Delta X$ be such that $u(p_\varepsilon) = \frac{1}{2}u(f(0)) + \frac{1}{2}u(g(0)) + \varepsilon$ and define $h_\varepsilon \in H$ such that $h_\varepsilon(0) = p_\varepsilon$ and $h_\varepsilon(t) = \frac{1}{2}f(t) + \frac{1}{2}g(t)$ for all $t > 0$. Now, for all $D \in \mathcal{D}^T$,

$$\begin{aligned}D \cdot (u \circ f - u \circ h_\varepsilon) &= D \cdot \left(u \circ f - u \circ \left(\frac{1}{2}f + \frac{1}{2}g\right) - (\varepsilon, 0, 0, \ldots)\right) \\ &= \frac{1}{2}D \cdot (u \circ f - u \circ g) - \varepsilon,\end{aligned}$$

which is well-defined as $D \cdot (u \circ f - u \circ g)$ is well-defined.

By symmetric argument, $D \cdot (u \circ g - u \circ h_\varepsilon) = \frac{1}{2}D \cdot (u \circ g - u \circ f) - \varepsilon$. For all positive number ε, define $F_\varepsilon = \{f, g, h_\varepsilon\}$. Then,

$$\begin{aligned}\mathcal{N}(f, F_\varepsilon) &= \left\{D \in \mathcal{D}^T \;\middle|\; D \cdot (u \circ f - u \circ g) \geq 2\varepsilon\right\}, \\ \mathcal{N}(g, F_\varepsilon) &= \left\{D \in \mathcal{D}^T \;\middle|\; D \cdot (u \circ g - u \circ f) \geq 2\varepsilon\right\}.\end{aligned}$$

Note that $\mathcal{N}(f, F_\varepsilon) \cap \mathcal{N}(g, F_\varepsilon) = \emptyset$ as $\varepsilon > 0$. Now, from equation (A.1) again, we have

$$\rho_{F_\varepsilon}(f) + \rho_{F_\varepsilon}(g) \leq \mu(\mathcal{N}(f, F_\varepsilon)) + \mu(\mathcal{N}(g, F_\varepsilon)) = \mu(\mathcal{N}(f, F_\varepsilon) \cup \mathcal{N}(g, F_\varepsilon)) \leq 1.$$

Consider a sequence of menus F_{ε_i} as $\varepsilon_i \to 0$. Suppose there are three menus F_{ε_i}, F_{ε_j}, and F_{ε_k} in this sequence that are not in $\mathcal{K}_0$. Since f and g are not tied, it must be that h_{ε_i}, h_{ε_j}, and h_{ε_k} are tied with f or g, respectively. Therefore, there exist $l, l' \in \{i, j, k\}$ such that h_{ε_l} and $h_{\varepsilon_{l'}}$ are tied with f or both of them are tied with g. Without loss of generality, we assume that h_{ε_j} and h_{ε_k} are both tied with f (the case for both tied with g is symmetric). Hence, h_{ε_j} and h_{ε_k} must be tied, so h_{ε_j} and $\frac{1}{2}f + \frac{1}{2}g$ must be tied. By Linearity, this implies that $r1w$ is tied with w, contradicting the representation from above. Hence, there cannot be more than two menus in this sequence that are not in $\mathcal{K}_0$. So we can always remove menus F_{ε_i} that are not in $\mathcal{K}_0$. Hence, we can assume that $F_{\varepsilon_i} \in \mathcal{K}_0$ for all i without loss of generality. By Continuity, we thus have that $1 = \rho(f, g) + \rho(g, f) = \lim_i \left(\rho_{F_{\varepsilon_i}}(f) + \rho_{F_{\varepsilon_i}}(g)\right) \leq \lim_i \mu\left(\mathcal{N}\left(f, F_{\varepsilon_i}\right) \cup \mathcal{N}\left(g, F_{\varepsilon_i}\right)\right)$. Hence,

$$\begin{aligned}
&\mu\left\{D \in \mathcal{D}^T \;\middle|\; D \cdot (u \circ f - u \circ g) = 0\right\} \\
&\quad = \lim_i \mu\left\{D \in \mathcal{D}^T \;\middle|\; -2\varepsilon_i < D \cdot (u \circ f - u \circ g) < 2\varepsilon_i\right\} \\
&\quad = 1 - \lim_i \mu\left(\mathcal{N}\left(f, F_{\varepsilon_i}\right) \cup \mathcal{N}\left(g, F_{\varepsilon_i}\right)\right) = 0.
\end{aligned}$$

This proves part (2) of the lemma.

We now return to the proof of part (1). Suppose that the inequality in equation (A.1) is strict for some $f \in F$. Let $F^* \subset F$ be the subset of streams in F that are not tied and $f \in F$. If we sum over all the non-tied streams F^*, then

$$1 = \sum_{g \in F^*} \rho_F(g) < \sum_{g \in F^*} \mu(\mathcal{N}(g, F)) \leq 1,$$

where the last inequality follows from part (2) as F^* contains no ties.

Since this cannot be true, it must be that $\rho_F(f) = \mu(\mathcal{N}(f, F))$ for all $f \in F$. This completes the proof for the lemma. □

We now complete the sufficiency proof. Let $r \in \Delta X$ such that $u(r) = 1$ and note that $wtr \to w$ as $t \to \infty$. Now, for every $D \in \mathcal{D}^T$,

$$S_t(D) := D \cdot (u \circ (wtr)) = \sum_{s \geq t} D(s)$$

is well-defined, which may be infinite. Hence, by part (1) of Lemma 3 and Continuity, we have $1 = \lim_t \rho(w, wtr) = \lim_t \mu\left\{D \in \mathcal{D}^T \mid S_t(D) = 0\right\}$, as $\{w, wtr\} \to \{w\}$. Since S is decreasing in t, $\lim_{t\to\infty} S_t$ is well-defined, although it could be infinite. Moreover, if $S_t(D) = 0$ for some $t \in T$, then $\lim_{t'\to\infty} S_{t'}(D) \leq S_t(D) = 0$. So for all $D \in \mathcal{D}^T$,

$$\limsup_t \mathbf{1}_{\{S_t(D)=0\}}(D) \leq \mathbf{1}_{\{\lim_{t\to\infty} S_t(D)=0\}}(D).$$

By Fatou's Lemma again,

$$\begin{aligned}
1 &= \lim_t \int_{\mathcal{D}^T} \mathbf{1}_{\{S_t(D)=0\}}(D)\, \mu(dD) \\
&\leq \int_{\mathcal{D}^T} \limsup_t \mathbf{1}_{\{S_t(D)=0\}}(D)\, \mu(dD) \\
&\leq \int_{\mathcal{D}^T} \mathbf{1}_{\{\lim_{t\to\infty} S_t(D)=0\}}(D)\, \mu(dD) = \mu\left\{D \in \mathcal{D}^T \;\middle|\; \lim_{t\to\infty} S_t(D) = 0\right\}.
\end{aligned}$$

Hence, $\lim_{t\to\infty} \sum_{s\geq t} D(s) = 0$ μ-a.s. Note this implies that $D \cdot (u \circ f - u \circ g)$ converges for all $f, g \in F$. Since D is decreasing μ-a.s. follows trivially from Impatience, $\mu(\mathcal{D}) = 1$. Hence, by part (1) of Lemma 3, we have for all $F \in \mathcal{K}$ and $f \in F$,

$$\rho_F(f) = \mu\{D \in \mathcal{D} \mid D \cdot (u \circ f) \geq D \cdot (u \circ g) \text{ for all } g \in F\}.$$

Moreover, the regularity of μ follows from part (2) of Lemma 3. We thus have a Random Discounting Representation as desired.

A.2.3. Necessity of Theorem 2

We now prove necessity of the axioms under a Random Discounting Representation. Note that Monotonicity, Linearity, Extremeness and Nondegeneracy follows by similar argument as in Lu (2016). To see Initial Determinism, note that if $f(t) = g(t)$ for all $t > 0$ and $f, g \in F$, then for any $f \in F$, $\rho_F(f) = \mu\{D \in \mathcal{D} \mid u(f(0)) \geq u(g(0))$ for all $g \in F\} \in \{0, 1\}$, as desired. To see Time Monotonicity, note that for $f \in F$, if $u(f(t)) \geq u(g(t))$ for all $g \in F$, $\rho_F(f) = \mu\{D \in \mathcal{D} \mid D \cdot (u \circ f - u \circ g) \geq 0$ for all $g \in F\} = 1$, as desired. To see Impatience, note that

$$\begin{aligned}\rho\left(f, f^t\right) &= \mu\left\{D \in \mathcal{D} \mid D \cdot \left(u \circ f - u \circ f^t\right) \geq 0\right\} \\ &= \mu\left\{D \in \mathcal{D} \mid \sum_{s \in T} (D(s) - D(s+t))\, u(f(s)) \geq 0\right\} = 1\end{aligned}$$

as D is decreasing μ-a.s.

Finally, we prove Continuity. Let $F_k \to F$ where $F_k, F \in \mathcal{K}_0$. Note that for any $f, g \in F_k$, f and g are not tied. Since μ is regular, this implies that $D \cdot (u \circ f - u \circ g) = 0$ with μ-measure zero. Now, define

$$\mathcal{I} := \bigcup_{f, g \in F_k \cup F} \{D \in \mathcal{D} \mid D \cdot (u \circ f) = D \cdot (u \circ g)\}$$

as the set of all discount functions that rank some $f, g \in F_k \cup F$ as the same. Note that $\mu(\mathcal{I}) = 0$ so if we let $\mathcal{D}^* := \mathcal{D} \backslash \mathcal{I}$, then $\mu(\mathcal{D}^*) = 1$. Let μ^* be the restriction of μ on $\mathcal{D}^*$. We will now define random variables $\xi_k : \mathcal{D}^* \to H$ and $\xi : \mathcal{D}^* \to H$ that have distributions ρ_{F_k} and ρ_F respectively. For each F_k, let $\xi_k : \mathcal{D}^* \to H$ be such that

$$\xi_k(D) := \arg\max_{f \in F_k} D \cdot (u \circ f)$$

and define ξ similarly for F. Note that these are well-defined because there exists a unique maximizer f for $D \in \mathcal{D}^*$. For any measurable set $E \subset H$,

$$\begin{aligned}\xi_k^{-1}(E) &= \left\{D \in \mathcal{D}^* \middle| \xi_k(D) \in E \cap F_k\right\} \\ &= \bigcup_{f \in E \cap F_k} \left\{D \in \mathcal{D}^* \middle| D \cdot (u \circ f) > D \cdot (u \circ g)\ \forall g \in F_k\right\}\end{aligned}$$

which is measurable. Hence, ξ_k and ξ are random variables. Note that

$$\begin{aligned}&\mu^* \circ \xi_k^{-1}(E) \\ &= \sum_{f \in E \cap F_k} \mu^* \left\{D \in \mathcal{D}^* \middle| D \cdot (u \circ f) > D \cdot (u \circ g)\ \forall g \in F_k\right\} \\ &= \sum_{f \in E \cap F_k} \mu\{D \in \mathcal{D} | D \cdot (u \circ f) \geq D \cdot (u \circ g)\ \forall g \in F_k\} \\ &= \rho_{F_k}(E \cap F_k) \\ &= \rho_{F_k}(E)\end{aligned}$$

so ρ_{F_k} and ρ_F are the distributions of ξ_k and ξ respectively. Note that for any $D \in \mathcal{D}^* \subset \mathcal{D}$, $D \cdot (u \circ f)$ is bounded and thus continuous in f. Hence, by the Maximum Theorem, $\xi_k(D) = \arg\max_{f \in F_k} D \cdot (u \circ f)$ is upper hemi-continuous in F_k. Since ξ_k is single-valued, ξ_k is continuous as a function of F_k. Since $F_k \to F$, $\xi_k \to \xi$ μ^*-a.s. Finally, since a.s. convergence implies convergence in distribution, $\rho_{F_k} \to \rho_F$ as desired.

A.3. Proof of Proposition 1

A.3.1. Sufficiency Proposition 1

Let ρ be represented by (μ, u) and suppose satisfies Decreasing Impatience. Choose $r \in \Delta X$ such that $u(r) = 1$. Define $h \in H$ such that $h(2) = r$, and $h(s) = w$ for all $s \neq 2$. Also, for any $a \in (0, 1]$, define $f_a, g_a \in H$ such that $g_a = ah^{-1} + (1-a)w$ and $f_a = ag_a^{-1} + (1-a)w$. Hence, we can write down the utility streams for f_a, g_a, h as follows:

$$u \circ h = (0, 0, 1, 0, \dots), \; u \circ g_a = (0, a, 0, 0, \dots), \; u \circ f_a = \left(a^2, 0, 0, 0, \dots\right).$$

Moreover, for any $t \in T$, the utility streams for the t-delayed streams f_a^t, g_a^t, h^t are as follows:

$$u \circ h^t = (0, \dots 0, 0, 1, 0, \dots), \; u \circ g_a^t = (0, \dots 0, a, 0, 0, \dots),$$
$$u \circ f_a^t = \left(0, \dots a^2, 0, 0, 0, \dots\right),$$

where $h(2) = r$. Note that for $D \in \mathcal{D}$

$$D \cdot \left(u \circ h^t\right) = D(t+2), \; D \cdot \left(u \circ g_a^t\right) = D(t+1)\,a, \; D \cdot \left(u \circ f_a^t\right) = D(t)\,a^2.$$

Fix $t \in T$ and let $F_a^t := \left\{f_a^t, g_a^t, h^t\right\}$ for any $a > 0$.

Because of the regularity, note that either $D(t+1) = 0$ μ-a.s. or $D(t+1) > 0$ μ-a.s.[27] If $D(t+1) = 0$ μ-a.s., then Proposition 1 holds. So consider the latter case. Since discount functions are decreasing, we know that $D(t) \geq D(t+1) > 0$ μ-a.s. We will now show that μ-a.s.

$$\frac{D(t+1)}{D(t)} \leq \frac{D(t+2)}{D(t+1)}.$$

First, we show that there is at most one value $a \in (0, 1]$ such that g_a^t and h^t are tied. To see this suppose there exists $b \neq a$ such that g_b^t and h^t are tied. Then, since both g_a^t and g_b^t are tied with h^t, then $D(t+1)\,a = D(t+1)\,b$ which implies that $D(t+1) = 0$ μ-a.s. a contradiction. Next, we show that there is at most one value $a \in (0, 1]$ such that g_a^t and f_a^t are tied. To see this suppose there exists another $b \neq a$ such that g_b^t and f_b^t are tied. Then, we have $D(t+1) = D(t)\,a$ and $D(t+1) = D(t)\,b$ which implies that $D(t) = 0$ μ-a.s. again a contradiction.

Therefore, for almost all $a \in (0, 1]$, we have g_a^t is not tied with f_a^t nor h^t. Since F_a^t contains no ties, by Decreasing Impatience,

$$\begin{aligned} 0 &= \rho_{F_a^t}\left(g_a^t\right) \\ &= \mu\left\{D \in \mathcal{D} \;\middle|\; D(t+1)\,a \geq D(t+2) \text{ and } D(t+1)\,a \geq D(t)\,a^2\right\} \\ &= \mu\left\{D \in \mathcal{D} \;\middle|\; \frac{D(t+1)}{D(t)} \geq a \geq \frac{D(t+2)}{D(t+1)}\right\} \\ &= \mu\left\{D \in \mathcal{D} \mid X_t \geq a \geq Y_t\right\}, \end{aligned}$$

where we define $X_t := \frac{D(t+1)}{D(t)} \geq 0$ and $Y_t := \frac{D(t+2)}{D(t+1)} \geq 0$. Hence,

$$0 = \mu\left\{D \in \mathcal{D} \mid X_t \geq a \geq Y_t\right\}.$$

[27] Consider a consumption stream f such that $f(t+1) = r$ and $f(s) = w$ for all $s \neq t+1$. By Determinism, $\rho(f, w) \in \{0, 1\}$. If $\rho(w, f) = 0$, then $D(t+1) > 0$ μ-a.s. Otherwise, $D(t+1) = 0$ μ-a.s.

Since this is true for almost all $a > 0$, it must be that μ-a.s.

$$\frac{D(t+1)}{D(t)} = X_t \leq Y_t = \frac{D(t+2)}{D(t+1)}$$

as desired.

A.3.2. Necessity of Proposition 1

Let ρ be represented by (μ, u). Suppose $f = ag^{-1} + (1-a)w$ and $g = ah^{-1} + (1-a)w$. Note that if g is tied with either f or h, then $\rho_{\{f,g,h\}}(\{f,h\}) = 1$ trivially so suppose g is not tied with f nor h. We will show that $\rho_{\{f,g,h\}}(g) = 0$. Let T^+ be the set of $t \in T$ such that $D(t+1) > 0$ μ-a.s. Note that

$$\begin{aligned}
D \cdot (u \circ g - u \circ h) &= \sum_{t \in T} [D(t+1)a - D(t+2)] u(h(t+2)) \\
&\leq \sum_{t \in T^+} [D(t+1)a - D(t+2)] u(h(t+2)) \\
&= \sum_{t \in T^+} D(t) \frac{D(t+1)}{D(t)} \left(a - \frac{D(t+2)}{D(t+1)}\right) u(h(t+2)) \\
&\leq \sum_{t \in T^+} D(t) \frac{D(t+1)}{D(t)} \left(a - \frac{D(t+1)}{D(t)}\right) u(h(t+2))
\end{aligned}$$

since $\frac{D(t+2)}{D(t+1)} \geq \frac{D(t+1)}{D(t)}$ μ-a.s. for all $t \in T^+$. Note that

$$\frac{D(t+1)}{D(t)} \left(a - \frac{D(t+1)}{D(t)}\right) \leq a \left(a - \frac{D(t+1)}{D(t)}\right)$$

so

$$\begin{aligned}
D \cdot (u \circ g - u \circ h) &\leq \sum_{t \in T^+} D(t) a \left(a - \frac{D(t+1)}{D(t)}\right) u(h(t+2)) \\
&= \sum_{t \in T^+} \left(D(t) a^2 - D(t+1) a\right) u(h(t+2)) \\
&\leq \sum_{t \in T} \left(D(t) a^2 - D(t+1) a\right) u(h(t+2)) \\
&= D \cdot (u \circ f - u \circ g).
\end{aligned}$$

Thus, $D \cdot (u \circ f) \leq D \cdot (u \circ g)$ implies $D \cdot (u \circ g) \leq D \cdot (u \circ h)$ so $\rho_{\{f,g,h\}}(g) = 0$ as desired.

A.4. Proof of Theorems 3 and 4

We will now prove Theorems 3 and 4. We will prove them in reverse order as Theorem 3 follows easily from Theorem 4. Let ρ be represented by (μ, u)

A.4.1. Necessity of Theorem 4

Suppose μ is quasi-hyperbolic. We will show that ρ satisfies Decreasing Impatience and Weak Stochastic Stationary. Note that since μ is quasi-hyperbolic, for every $t > 0$

$$\frac{D(1)}{D(0)} = \beta\delta \leq \delta = \frac{D(t+1)}{D(t)}$$

Hence, ρ satisfies Decreasing Impatience by Proposition 1. We now prove Weak Stochastic Stationarity. Now, for any $t \geq 1$ and $f, g \in F$,

$$\begin{aligned} D \cdot \left(u \circ f^t - u \circ g^t\right) &= \sum_s \beta\delta^{s+t} \left[u\left(f\left(s\right)\right) - u\left(g\left(s\right)\right)\right] \\ &= \delta^{t-1} \sum_s \beta\delta^{s+1} \left[u\left(f\left(s\right)\right) - u\left(g\left(s\right)\right)\right] \\ &= \delta^{t-1} \left[D \cdot \left(u \circ f^1 - u \circ g^1\right)\right] \end{aligned}$$

Hence,

$$\begin{aligned} \rho_{F^t}\left(f^t\right) &= \mu\left\{D \in \mathcal{D} \mid D \cdot \left(u \circ f^t - u \circ g^t\right) \geq 0 \text{ for all } g^t \in F^t\right\} \\ &= \mu\left\{D \in \mathcal{D} \,\middle|\, D \cdot \left(u \circ f^1 - u \circ g^1\right) \geq 0 \text{ for all } g^1 \in F^1\right\} = \rho_{F^1}\left(f^1\right) \end{aligned}$$

so Weak Stochastic Stationarity is satisfied.

A.4.2. Necessity of Theorem 3

We now prove that if μ is exponential, then ρ must satisfy Decreasing Impatience and Stochastic Stationarity. Note that Decreasing Impatience follows immediately from the necessity proof of Theorem 4. To show Stochastic Stationarity, note that for any $t \in T$ and $f, g \in F$,

$$D \cdot \left(u \circ f^t - u \circ g^t\right) = \sum_{s \in T} \delta^{s+t} \left[u\left(f\left(s\right)\right) - u\left(g\left(s\right)\right)\right] = \delta^t \left[D \cdot \left(u \circ f - u \circ g\right)\right].$$

Hence,

$$\begin{aligned} \rho_{F^t}\left(f^t\right) &= \mu\left\{D \in \mathcal{D} \mid D \cdot \left(u \circ f^t - u \circ g^t\right) \geq 0 \text{ for all } g^t \in F^t\right\} \\ &= \mu\left\{D \in \mathcal{D} \mid D \cdot \left(u \circ f - u \circ g\right) \geq 0 \text{ for all } g \in F\right\} = \rho_F\left(f\right), \end{aligned}$$

so Stochastic Stationarity is satisfied.

A.4.3. Sufficiency of Theorem 4

We now prove the sufficiency of Theorem 4. First, we show the following lemma.

Lemma 4. *Let ρ be represented by (μ, u).*

(1) *If ρ satisfies Weak Stochastic Stationarity, then for all $t \geq 1$, $\mu\{D \in \mathcal{D} \mid D(1) = 0\} = \mu\{D \in \mathcal{D} \mid D(t) = 0\}$.*
(2) *If ρ satisfies Stochastic Stationarity, then for all $t \in T$, $\mu\{D \in \mathcal{D} \mid D(t) = 0\} = 0$.*

Proof. Suppose ρ is represented by (μ, u). Let $r \in \Delta X$ be such that $u(r) = 1$. We prove the two cases separately.

(1) First, suppose ρ satisfies Weak Stochastic Stationarity. Let $f \in H$ be such that $f(0) = r$ and $f(s) = w$ for all $s > 0$. Now, by Weak Stochastic Stationarity, for any $t \geq 1$, $\mu\{D \in \mathcal{D} \mid D(1) = 0\} = \rho\left(w^1, f^1\right) = \rho\left(w^t, f^t\right) = \rho\left(w, f^t\right) = \mu\{D \in \mathcal{D} \mid D(t) = 0\}$, as desired.

(2) Now, suppose ρ satisfies Stochastic Stationarity. If we let $h \in H$ be such that $h(0) = r$ and $h(s) = w$ for all $s > 0$, then by the same argument as above we have $\mu\{D \in \mathcal{D} \mid D(0) = 0\} = \rho(w, h) = \rho\left(w^t, h^t\right) = \mu\{D \in \mathcal{D} \mid D(t) = 0\}$. Since $D(0) = 1$, the result follows. □

Since ρ satisfies Weak Stochastic Stationarity, from Lemma 4 and the fact that μ is regular, we know that for all $t \geq 1$,

$$\mu\{D \in \mathcal{D} \mid D(t) = 0\} = \mu\{D \in \mathcal{D} \mid D(1) = 0\} \in \{0, 1\} \tag{A.2}$$

By this result, it suffices to consider the following two cases.

Case 1: $\mu\{D \in \mathcal{D} \mid D(1) = 0\} = 1$. Then by (A.2), $\mu\{D \in \mathcal{D} \mid D(t) = 0\} = 1$ for all $t \geq 1$. Then since T is countable, this implies that $D(t) = 0$ for all $t \geq 1$ μ-a.s. Hence, $\rho_F(f) = \mu\{D \in \mathcal{D} \mid u(f(0)) \geq u(g(0))$ for all $g \in F\}$, so μ is trivially quasi-hyperbolic with $\beta = 0$.

Case 2: $\mu\{D \in \mathcal{D} \mid D(1) = 0\} = 0$. Then by (A.2), $\mu\{D \in \mathcal{D} \mid D(t) = 0\} = 0$ for all $t \geq 1$. Since $D \geq 0$, $\mu\{D \in \mathcal{D} \mid D(t) > 0\} = 1$ for all $t \geq 1$. So $D(t) > 0$ μ a.s. for all $t \in T$. Hence, $D(t+1)/D(t)$ is well defined μ a.s. for all $t \in T$.

Choose $r \in \Delta X$ such that $u(r) = 1$. Define $h \in H$ such that $h(2) = r$, and $h(s) = w$ for all $s \neq 2$. Also, for any $a \in (0, 1]$, define $f_a, g_a \in H$ such that $g_a = ah^{-1} + (1-a)w$ and $f_a = ag_a^{-1} + (1-a)w$. Hence, we can write down the utility streams for f_a, g_a, h as follows:

$$u \circ h = (0, 0, 1, 0, \ldots),\ u \circ g_a = (0, a, 0, 0, \ldots),\ u \circ f_a = \left(a^2, 0, 0, 0, \ldots\right).$$

Moreover, for any $t \in T$, the utility streams for the t-delayed streams f_a^t, g_a^t, h^t are as follows:

$$u \circ h^t = (0, \ldots 0, 0, 1, 0, \ldots),\ u \circ g_a^t = (0, \ldots 0, a, 0, 0, \ldots),$$
$$u \circ f_a^t = \left(0, \ldots a^2, 0, 0, 0, \ldots\right),$$

where $h(2) = r$. Note that for $D \in \mathcal{D}$

$$D \cdot \left(u \circ h^t\right) = D(t+2),\ D \cdot \left(u \circ g_a^t\right) = D(t+1)\,a,\ D \cdot \left(u \circ f_a^t\right) = D(t)\,a^2.$$

Let $F_a^t := \left\{f_a^t, g_a^t, h^t\right\}$. We now consider two cases.

Subcase 2.1: Suppose there exists some $a > 0$ such that g_a^1 is tied with either f_a^1 or h^1. Consider the case in which g_a^1 is tied with h^1. Hence, $\rho\left(g_a^1, h^1\right) = 1 = \rho\left(h^1, g_a^1\right)$. By Weak Stochastic Stationarity, for all $t \in T$, $\rho\left(g_a^t, h^t\right) = \rho\left(g_a^1, h^1\right) = 1 = \rho\left(h^1, g_a^1\right) = \rho\left(h^t, g_a^t\right)$. Hence, for all $t \in T$, $1 = \mu\{D \in \mathcal{D} \mid D(t+1)\,a = D(t+2)\}$. Thus, if we let $\beta = \frac{D(1)}{a}$ and $\delta = a \leq 1$, then for all $t > 0$, we have μ-a.s.

$$D(t) = \frac{D(1)}{a} a^t = \beta\delta^t$$

The case for g_a^1 is tied with f_a^1 is symmetric. Finally, we show that $\beta \leq 1$ μ-a.s. By Proposition 1 again,

$$\beta\delta = D(1) = X_0 \leq Y_0 = \frac{D(2)}{D(1)} = \delta$$

so $\beta \leq 1$ μ-a.s. Hence μ is quasi-hyperbolic as desired.

Subcase 2.2: Now consider the second case where g_a^1 is not tied with f_a^1 nor h^1 for all $a > 0$. Note that by Weak Stochastic Stationarity, this implies that for all $t \geq 1$, g_a^t is not tied with f_a^t nor h^t.

Note that $\left\{f_a^{t+1}, g_a^{t+1}\right\} = a\left\{g_a^t, h^t\right\} + (1-a)\, w$. Hence, by Weak Stochastic Stationarity and Linearity $\rho\left(f_a^t, g_a^t\right) = \rho\left(f_a^{t+1}, g_a^{t+1}\right) = \rho\left(g_a^t, h^t\right)$ for any $t \geq 1$. This implies that for every $t \geq 1$,

$$\mu\left\{D \in \mathcal{D} \;\middle|\; D(t)\, a^2 \geq D(t+1)\, a\right\} = \mu\left\{D \in \mathcal{D} \mid D(t+1)\, a \geq D(t+2)\right\}$$

So $\mu\{D \in \mathcal{D} \mid X_t \leq a\} = \mu\{D \in \mathcal{D} \mid Y_t \leq a\}$, where $X_t := \frac{D(t+1)}{D(t)} \geq 0$ and $Y_t := \frac{D(t+2)}{D(t+1)} \geq 0$. By the inclusion-exclusion principle,[28] we have

$$\begin{aligned}
&\mu\{D \in \mathcal{D} \mid X_t \leq a \leq Y_t\} \\
&= \mu\{D \in \mathcal{D} \mid X_t \leq a\} + \mu\{D \in \mathcal{D} \mid a \leq Y_t\} - \mu\{D \in \mathcal{D} \mid X_t \leq a \text{ or } a \leq Y_t\} \\
&= \mu\{D \in \mathcal{D} \mid Y_t \leq a\} + \mu\{D \in \mathcal{D} \mid a \leq Y_t\} - \mu\{D \in \mathcal{D} \mid X_t \leq a \text{ or } a \leq Y_t\} \\
&= 1 - \mu\{D \in \mathcal{D} \mid X_t \leq a \text{ or } a \leq Y_t\} \\
&= \mu\{D \in \mathcal{D} \mid X_t \geq a \geq Y_t\}
\end{aligned}$$

where the third and fourth equalities hold because g_a^t is not tied with f_a^t nor h^t. Since ρ satisfies Decreasing Impatience, by Proposition 1, $X_t \leq Y_t$ μ-a.s. for all $t \in T$. Since this holds for any $a > 0$, it must be that for all $t \geq 1$,

$$\frac{D(t+1)}{D(t)} = X_t = Y_t = \frac{D(t+2)}{D(t+1)}$$

μ-a.s. If we let $\delta = \frac{D(2)}{D(1)} \leq 1$ and $\beta = \frac{D(1)^2}{D(2)}$, then for all $t > 0$,

$$D(t) = D(1)\left(\frac{D(2)}{D(1)}\right)^{t-1} = \beta\delta^t$$

μ-a.s. We can prove $\beta \leq 1$ μ-a.s. as in the previous case.

A.4.4. Sufficiency of Theorem 3

Now, suppose ρ satisfies Stochastic Stationary and Decreasing Impatience. From Lemma 4, we know that $D(t) > 0$ μ-a.s. for all $t \in T$. As in the sufficiency proof for Theorem 4, define the streams h, g_a, f_a and h^t, g_a^t, f_a^t such that for $D \in \mathcal{D}$,

$$D \cdot \left(u \circ h^t\right) = D(t+2),\;\; D \cdot \left(u \circ g_a^t\right) = D(t+1)\, a,\;\; D \cdot \left(u \circ f_a^t\right) = D(t)\, a^2.$$

Again we consider two cases.

Case 1: Suppose there exists some $a > 0$ such that g_a is tied with either f_a or h. Consider the case in which g_a is tied with h. Hence, $\rho(g_a, h) = 1 = \rho(h, g_a)$. By Stochastic Stationarity, for all $t \geq -1$, $\rho\left(g_a^t, h^t\right) = 1 = \rho\left(h^t, g_a^t\right)$. Hence, for all $t \in T$

$$1 = \mu\{D \in \mathcal{D} \mid D(t)\, a = D(t+1)\} = \mu\left\{D \in \mathcal{D} \;\middle|\; \frac{D(t+1)}{D(t)} = a\right\}.$$

[28] For any two events A and B, $\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cup B)$.

If we let $\delta = a$, then for all $t \in T$, we have μ-a.s.

$$D(t) = D(0)\, a^t = \delta^t$$

so μ is exponential as desired. As before, the case for g_a is tied with f_a is symmetric.

Case 2: Now consider the second case where g_a is not tied with f_a nor h for all $a > 0$. By Stochastic Stationarity, this implies that for all $t \in T$, g_a^t is not tied with f_a^t nor h^t. Let $X_t := \frac{D(t+1)}{D(t)}$ and $Y_t := \frac{D(t+2)}{D(t+1)}$ as before. Now, by the same argument as in the sufficiency proof for Theorem 4, Stochastic Stationarity and Linearity, imply that $\mu\{D \in \mathcal{D} \mid X_t \leq a \leq Y_t\} = \mu\{D \in \mathcal{D} \mid X_t \geq a \geq Y_t\}$. By the same argument as before, Proposition 1 implies that for all $t \in T$, $X_t \leq Y_t$ μ-a.s. Since this holds for any $a > 0$, it must be that for all $t \in T$,

$$\frac{D(t+1)}{D(t)} = X_t = Y_t = \frac{D(t+2)}{D(t+1)}$$

μ-a.s. If we let $\delta = D(1)$, then for all $t \in T$, $D(t) = D(1)^t = \delta^t$ μ-a.s. Since D is decreasing, $\delta \leq 1$ μ-a.s. Thus, μ is exponential as desired.

A.5. *Proof of Proposition 2*

Obviously, μ is not exponential. To show μ satisfies Stochastic Stationarity, choose any $d \in T$. Consider the case in which d is even and $d = 2n$. Then $(D^\omega(d), D^\omega(d+1), \ldots) = \exp(-2n)D^\omega$.[29] Therefore, for any $f, g \in H$,

$$(D^\omega(d), D^\omega(d+1), \ldots)\cdot u \circ f > (D^\omega(d), D^\omega(d+1), \ldots)\cdot u \circ g \Leftrightarrow D^\omega \cdot u \circ f > D^\omega \cdot u \circ g. \tag{A.3}$$

So for any $F \subset H$ and $f \in F$,

$$\begin{aligned} &\rho_{F^d}(f^d) \\ &= \mu\{D^\omega | (D^\omega(d), D^\omega(d+1), \ldots)\cdot u \circ f > (D^\omega(d), D^\omega(d+1), \ldots)\cdot u \circ g \\ &\qquad \text{for all } g \in F\} \\ &= \mu\{D^\omega | D^\omega \cdot u \circ f \geq D^\omega \cdot u \circ g \text{ for all } g \in F\} \qquad (\because \text{(A.3)}) \\ &= \rho_F(f). \end{aligned}$$

Consider the case in which d is odd and $d = 2n + 1$. Then $(D^\omega(d), D^\omega(d+1), \ldots) = \exp(-2n)D^{1-\omega}$.[30] Therefore, for any $f, g \in H$,

$$\begin{aligned} &(D^\omega(d), D^\omega(d+1), \ldots)\cdot u \circ f > (D^\omega(d), D^\omega(d+1), \ldots)\cdot u \circ g \\ &\Leftrightarrow D^{1-\omega} \cdot u \circ f > D^{1-\omega} \cdot u \circ g. \end{aligned} \tag{A.4}$$

[29] $(D^\omega(d), D^\omega(d+1), \ldots) = (\exp(-2n), \exp(-2n - \frac{1}{2} - \omega), \exp(-2(n+1)), \exp(-2(n+1) - \frac{1}{2} - \omega), \ldots) = \exp(-2n)(1, \exp(-\frac{1}{2} - \omega), \exp(-2), \exp(-\frac{5}{2} - \omega), \ldots) = \exp(-2n)D^\omega$.

[30] $(D^\omega(d), D^\omega(d+1), \ldots) = (\exp(-2n - \frac{1}{2} - \omega), \exp(-2(n+1)), \exp(-2(n+1) - \frac{1}{2} - \omega), \exp(-2(n+2)), \ldots) = \exp(-2n - \frac{1}{2} - \omega)(1, \exp(-\frac{1}{2} - (1-\omega)), \exp(-2), \exp(-\frac{5}{2} - (1-\omega)), \ldots) = \exp(-2n)D^{1-\omega}$.

Let I be a uniform distribution over [0, 1]. Then for any $F \subset H$ and $f \in F$,

$$\begin{aligned}
&\rho_{F^d}(f^d) \\
&= \mu\{D^\omega | (D^\omega(d), D^\omega(d+1), \dots) \cdot u \circ f > (D^\omega(d), D^\omega(d+1), \dots) \cdot u \circ g \\
&\qquad \text{for all } g \in F\} \\
&= I(\{\omega | (D^\omega(d), D^\omega(d+1), \dots) \cdot u \circ f > (D^\omega(d), D^\omega(d+1), \dots) \cdot u \circ g \\
&\qquad \text{for all } g \in F\}) \\
&= I(\{\omega | D^{1-\omega} \cdot u \circ f > D^{1-\omega} \cdot u \circ g \text{ for all } g \in F\}) \qquad (\because \text{(A.4)}) \\
&= I(\{1-\omega | D^{1-\omega} \cdot u \circ f > D^{1-\omega} \cdot u \circ g \text{ for all } g \in F\}) \qquad (\because I \text{ is uniform}) \\
&= \mu\{D^\omega | D^\omega \cdot u \circ f > D^\omega \cdot u \circ g \text{ for all } g \in F\} \\
&= \rho_F(f).
\end{aligned}$$

Finally, we show that μ violates Decreasing Impatience. Fix $r \in \Delta X$ such that $u(r) = 1 > 0 = u(w)$. Define $h \in H$ such that $h(2) = r$, and $h(t) = w$ for all $t \in T$ such that $t \neq 2$. Fix $a \in (0, 1)$. Define $f, g \in H$ by $f = ag^{-1} + (1-a)w$ and $g = ah^{-1} + (1-a)w$.

Then for all $s \in T$

$$\begin{aligned}
&D^\omega \cdot u \circ g > D^\omega \cdot u \circ h \Leftrightarrow D_1^\omega a > D_2^\omega \Leftrightarrow a > \exp(-\tfrac{3}{2} + \omega), \\
&D^\omega \cdot u \circ g > D^\omega \cdot u \circ f \Leftrightarrow D_1^\omega a > a^2 \Leftrightarrow \exp(-\tfrac{1}{2} - \omega) > a.
\end{aligned}$$

Note that $\exp(-\frac{1}{2} - \omega) > \exp(-\frac{3}{2} + \omega)$ if and only if $\frac{1}{2} > \omega$. Let $a = \exp(-1)$. Then for all $\omega < \frac{1}{2}$, we have $\exp(-\frac{1}{2} - \omega) > \exp(-1) > \exp(-\frac{3}{2} + \omega)$. Hence, $\rho_{\{f,g,h\}}(g) = I\left(\omega < \frac{1}{2}\right) = \frac{1}{2}$. This contradicts Decreasing Impatience.

A.6. Proof of Theorem 5

First, we prove the following simple lemma.

Lemma 5. *Let $u, v \in U$ be non-constant and suppose $u(p) > u(q)$ implies $v(p) \geq v(q)$ for all $p, q \in \Delta X$. Then v is an affine transformation of u.*

Proof. Suppose there exist $p, q \in \Delta X$ such that $u(p) > u(q)$ and $v(p) = v(q)$. By the linearity of u and v, there exists $\varepsilon \in (0, 1)$ such that $u((1-\varepsilon)p + \varepsilon w) > u(q)$ and $v((1-\varepsilon)p + \varepsilon w) < v(q)$, which gives a contradiction. Therefore, we have $u(p) > u(q)$ implies $v(p) > v(q)$ for all $p, q \in \Delta X$. The converse can be proved in the same way, so we have $u(p) > u(q)$ if and only if $v(p) > v(q)$ for all $p, q \in \Delta X$. Therefore, v is an affine transformation of u. □

We now prove Theorem 5. For any finite $J \subset T$ such that $0 \in J$, define $H^J \subset H$ and $\mathcal{K}^J \subset \mathcal{K}$ as in the proof of Theorem 2. From Lu (2016), Monotonicity, Linearity, Extremeness and Continuity imply that we can find a measure ν^J on $\Delta J \times U^J$ such that for all $F \in \mathcal{K}^J$,

$$\rho_F^J(f) = \nu^J \left\{ (r, u) \in \Delta J \times U^J \;\middle|\; \sum_{t \in J} r(t) \left[u_t(f(t)) - u_t(g(t)) \right] \geq 0 \text{ for all } g \in F \right\}$$

Moreover, ν^J satisfies regularity.

For every $t \in T$ and $p \in \Delta X$, let $p_t \in H$ denote the stream such that $p_t(t) = p$ and $p_t(s) = w$ for all $s \neq t$. Let $J^* \subset J$ denote the set of time periods such that there exists $p, q \in \Delta X$ where p_t and q_t that are not tied. If $t \in J^*$, then $r(t)[u_t(p) - u_t(q)] \neq 0$ ν^J-a.s. which implies that $r(t) > 0$ ν^J-a.s. Moreover, this also implies that u_t is non-constant ν^J-a.s. Note that by Initial Nondegeneracy, $0 \in J^*$. By Worst, we also know that $u_0 \in U^*$.

Now, consider $J = \{0, \dots, n\}$ for some $n \in T$. Order $J^* = \{0, t_1, \dots, t_m\}$ and first consider t_1. We will show that for any $p, q \in \Delta X$, $u_0(p) > u_0(q)$ implies $u_{t_1}(p) \geq u_{t_1}(q)$ ν^J-a.s. Suppose otherwise, so

$$0 < \nu^J \left\{ (r,u) \in \Delta J \times U^J \,\middle|\, u_0(p) > u_0(q) \text{ and } u_{t_1}(p) < u_{t_1}(q) \right\}. \tag{A.5}$$

Consider streams $f = pt_1w$, $g = qt_1w$ and

$$h(t) = \begin{cases} p & \text{if } t < t_1 \\ q & \text{if } t = t_1 \\ w & \text{if } t > t_1 \end{cases}$$

Note that $f, g, h \in H^J$ and let $F = \{f, g, h\} \in \mathcal{K}^J$. Moreover, note that if h is tied with g, then $u_0(p) = u_0(q)$ ν^J-a.s. contradicting the strict inequality in (A.5). The case for if h is tied with f is symmetric so h is tied with neither f nor g. By Time Invariance,

$$0 = \rho_F(h) = \nu^J \left\{ (r,u) \in \Delta J \times U^J \,\middle|\, u_0(p) \geq u_0(q) \text{ and } u_{t_1}(p) \leq u_{t_1}(q) \right\}$$

contradicting inequality (A.5). This means that $u_0(p) > u_0(q)$ implies $u_{t_1}(p) \geq u_{t_1}(q)$ ν^J-a.s.

By the continuity of vNM utilities, we have $u_0(p) > u_0(q)$ implies $u_{t_1}(p) \geq u_{t_1}(q)$ for all $p, q \in \Delta X$, ν^J-a.s. Lemma 5 implies that u_{t_1} is an affine transformation of u_0 ν^J-a.s. We can repeat the above argument for all $t \in J^*$ to show that u_{t+1} is an affine transformation of u_t ν^J-a.s. Therefore, every u_t is an affine transformation of u_0 ν^J-a.s. for all $t \in J^*$.

Now, consider some $s \in J \backslash J^*$ and suppose there exists some $t \in J^*$ where $s < t$. Let $p \in \Delta X$ such that $u_0(p) > u_0(w)$ ν^J-a.s. Such p exists by Initial Nondegeneracy. By Impatience, we have

$$1 = \rho(p_s, p_t) = \nu^J \left\{ (r,u) \in \Delta J \times U^J \mid r(t)[u_t(w) - u_t(p)] \geq 0 \right\}.$$

Since $t \in J^*$, u_t is an affine transformation of u_0 so $u_t(p) > u_t(w)$ ν^J-a.s. This implies that $r(t) = 0$ ν^J-a.s. contradicting the fact that $t \in J^*$. Thus, if $s \in J \backslash J^*$, then for all $t \in J$ if $t > s$, then $t \in J \backslash J^*$.

As in the proof of Theorem 2, by Impatience, we can show that r is decreasing on J^*. Once $t \in J \setminus J^*$ appears, all $s > t$ belongs to $J \setminus J^*$. So we can thus set $r(s) = 0$ for all $s \in J \backslash J^*$. Moreover, r will still be decreasing over time.

To summarize, we can define a measure π^J on $\Delta J \times U^*$ such that for all $F \in \mathcal{K}^J$,

$$\rho_F^J(f) = \pi^J \left\{ (r,u) \in \Delta J \times U^* \,\middle|\, \sum_{t \in J} r(t)[u(f(t)) - u(g(t))] \geq 0 \text{ for all } g \in F \right\}$$

where r is decreasing in t π^J-a.s. The rest of the proof follows exactly as in the proof of Theorem 2 where we use Kolmogorov's Theorem to extend this to all menus in $\mathcal{K}$.

Finally, we prove the necessity of Time Invariance. Consider $f \in F$ such that $f(s) \in \{p, q\}$ for all $s \leq t$. Note that if $u(p) \geq u(q)$, then $\sum_{s \leq t} D(t)[u(p) - u(f(t))] \geq 0$. On the other hand, if $u(q) \geq u(p)$, then $\sum_{s \leq t} D(t)[u(q) - u(f(t))] \geq 0$. This implies that $\rho_{Ftw}(\{ptw, qtw\}) = 1$ as desired. The necessity of Worst and Initial Nondegeneracy are trivial.

A.7. Proof of Theorem 6

Note that if $\pi = \eta$, then $\rho(f, g) = \tau(f, g)$ for all $f, g \in H$ immediately from the representation. Thus, suppose ρ and τ agree on all binary choices. First, note that for any $p, q \in \Delta X$,

$$\begin{aligned} &\rho(p1g, q1g) = \pi\left\{(D, u) \in \mathcal{D} \times U^* \mid u(p) \geq u(q)\right\} \\ =\ &\tau(p1g, q1g) = \eta\left\{(D, u) \in \mathcal{D} \times U^* \mid u(p) \geq u(q)\right\} \end{aligned}$$

Thus, we can assume that utilities under both π and η have the same worst consumption w. Moreover, we can find some $b \in X$ such that $u(b) > u(w)$ both π and η-a.s. Without loss of generality, normalize the utilities so that $u(w) = 0$ and $u(b) = 1$.

Let J be a finite subset of T. Let $n = |X|$ be the number of prizes and let $H^n \subset H$ denote the set of streams f such that $f(t, x) \leq \frac{1}{n}$ for all $t \in T$ and $x \neq w$. Note that

$$\begin{aligned} \sum_t D(t)\, u(f(t)) &= \sum_t D(t) \sum_x [f(t)](x)\, u(x) = \sum_{t, x \neq w} D(t)\, u(x)\, f(t, x) \in [0, 1] \\ &= \sum_{t, x \neq w} w_{D,u}(t, x)\, f(t, x) \in [0, 1] \end{aligned}$$

where $w_{D,u}(t, x) := D(t)\, u(x)$. Then $w_{D,u} \in \left[0, \frac{1}{n}\right]^{J \times (n-1)}$ and we can think of f as corresponding to the vector $f \in \left[0, \frac{1}{n}\right]^{J \times (n-1)}$. We thus have for all $f \in H^n$,

$$\begin{aligned} &\rho(a(b1w) + (1-a)w,\ f) = \pi\left\{(D, u) \in \mathcal{D} \times U^* \;\middle|\; w_{D,u} \cdot f \leq a\right\} \\ =\ &\tau(a(b1w) + (1-a)w,\ f) = \eta\left\{(D, u) \in \mathcal{D} \times U^* \;\middle|\; w_{D,u} \cdot f \leq a\right\} \end{aligned}$$

Since this is true for all $f \in \left[0, \frac{1}{n}\right]^{J \times (n-1)}$, by using Cramer–Wold as in the proof for Theorem 1, we have the distribution of $w_{D,u}$ is the same under π as under η. Finally, note that $w_{D,u} = w_{D',u'}$ implies $(D, u) = (D', u')$ so $\pi = \eta$ as desired

A.8. Proof of Proposition 4

For each $t \in T$ and $F \in \mathcal{K}$, recall that we defined $F^t = \{f^t | f \in F\}$. For each $F \in \mathcal{K}^t$, we can define $F^{-t} = \{f^{-t} | f \in F\}$, where f^{-t} is an element of H such that $(f^{-t})^t = f$.

Let $\rho(F) := \rho^t(F^t)$ for all $F \in \mathcal{K}$. Since ρ^t satisfies the axioms of Theorem 2, there exists a (μ^t, u^t) that represents it. Thus, for $s \geq t$,

$$\begin{aligned} \rho_F(f) &= \rho^t_{F^t}(f^t) = \rho^t_{F^s}(f^s) \\ &= \mu^t\left\{D \in \mathcal{D} \;\middle|\; \sum_{t'} D(t')\, u^t(f^s(t')) \geq \sum_{t'} D(t')\, u^t(g^s(t'))\ \forall g \in F\right\} \\ &= \mu^t_s\left\{D \in \mathcal{D} \;\middle|\; \sum_{t'} D(t')\, u^t(f(t')) \geq \sum_{t'} D(t')\, u^t(g(t'))\ \forall g \in F\right\} \end{aligned}$$

where μ_s^t is the marginal distribution of μ^t for $\left(D\left(t'\right)\right)_{t'\geq s}$. Note that the first and last equations follow from the definitions, the second from Stochastic Stationarity and the third from the representation. It then follows that ρ is represented by $\left(\mu_s^t, u^t\right)$.

By Stochastic Dynamic Consistency, we have $\rho^t(F) = \rho^s(F)$ for $t < s$ and $F \in \mathcal{K}^s$. Thus, $\rho_F(f) = \rho^s_{F^s}(f^s)$ so ρ is also represented by $\left(\mu_s^s, u^s\right)$. Theorem 1 then implies that $\mu_s^t = \mu_s^s$ and $u^t = \alpha u^s + \beta$ for $\alpha > 0$. Since ρ^t also satisfy the axioms of Theorem 3, μ^t is just a single-dimensional distribution. Thus, $\mu_s^t = \mu_s^s$ for all $s \geq t$ implies that the distribution of δ is the same for all ρ^t. Defining $\mu = \mu^0$ and $u = u^0$ yield the desired conclusion.

Next we assume the representation and show that $\{\rho^t\}_{t\in T}$ satisfies the axioms. In Section 6.2, we have shown that $\{\rho^t\}_{t\in T}$ satisfies Stochastic Dynamic Consistency. For each $t \in T$, ρ^t satisfies the axioms in Theorems 2 and 3 (defined with $\mathcal{K}^t$ instead of $\mathcal{K}$) as in the proof of Theorem 2 and 3.

Appendix B. Decreasing impatience and extremeness

In this section, we demonstrate a technical relationship between Decreasing Impatience and Extremeness. We show that Decreasing Impatience in the random exponential model plays an analogous role as Extremeness in the random expected utility model of Gul and Pesendorfer (2006).

Let X be a finite set and ΔX be the set of lotteries over X. Let C and C' be finite sets of lotteries. We say C is a *translate* of C' if and only if $C = C' + (p - q)$ for some $p \in C$ and $q \in C'$.[31] First, note that in the lottery setup, Stochastic Stationarity is equivalent to Linearity*, a weaker condition than Linearity.

Axiom *(Linearity*). $\rho_C(f) = \rho_{C'}\left(f'\right)$ if C and f are translates of C' and f' respectively.*

Clearly, Linearity implies Linearity*. There are random non-expected utility representations that yield random choice rules that satisfy Linearity* but not Extremeness. We now describe one such example. Let $X = \{x, y\}$ so we can associate each lottery with a point $p \in [0, 1]$. Let ω be uniformly distributed on $[0, 1]$ and let

$$u_\omega(p) := |p - \omega|,$$
$$v_\omega(p) := -|p - \omega|.$$

Consider a random utility that puts $\frac{1}{2}$ weight on u_ω and $\frac{1}{2}$ weight on v_ω. To show that this violates Extremeness, let $C = \left\{0, \frac{1}{2}, 1\right\}$. Since the mixed lottery $\frac{1}{2}$ is never chosen in C under u_ω, we have that

$$\rho_C\left(\frac{1}{2}\right) = \frac{1}{2}\cdot\mathbb{P}\left\{\omega \in [0,1] \;\middle|\; v_\omega\left(\frac{1}{2}\right) \geq \max\{v_\omega(0), v_\omega(1)\}\right\} = \frac{1}{4} > 0.$$

To show that this satisfies Linearity*, suppose $C = \{p_1, \dots, p_k\}$ with $p_1 < p_2 < \cdots < p_k$. Now, for each p_i such that $1 < i < k$, we have

$$\rho_C(p_i) = \frac{1}{2}\left(\frac{p_{i+1} - p_{i-1}}{2}\right)$$

[31] More explicitly, there exists $p \in C$ and $q \in C'$ such that $C = \left\{r + p - q \;\middle|\; r \in C'\right\}$.

which is unchanged if we translate C. For p_1, we have

$$\rho_C(p_1) = \frac{1}{2}\left(\frac{p_1+p_2}{2}\right) + \frac{1}{2}\left(1 - \frac{p_1+p_k}{2}\right) = \frac{1}{2}\left(1 - \frac{p_k-p_2}{2}\right)$$

which is again unchanged if we translate C. By symmetric argument, the same holds for p_k as well, so Linearity* is satisfied, but this is clearly not a random expected utility model.

By imposing Extremeness however, we are able to obtain a random expected utility representation. In other words, similar to Decreasing Impatience, Extremeness provides the additional restrictions to ensure the existence of a random utility representation with linear utilities.[32]

Proposition 5. *Suppose ρ has a random utility representation. Then ρ satisfies Linearity if and only if it satisfies Linearity* and Extremeness.*

Proof. Since Linearity implies Linearity*, all we need to show is that Linearity* and Extremeness imply Linearity. Let $C' = aC + (1-a)r$ for some $r \in \Delta X$ and $a \in (0,1)$. By Extremeness, we can only consider the extreme set of points of C without loss of generality. Suppose C has k extreme points. Hence, we can translate C' k-times such that each translated C'_i overlaps with an extreme point $p_i \in C$ and $\mathrm{conv}\left(C'_i\right) \subset \mathrm{conv}\,(C)$. Now, define

$$E := \bigcup_i C'_i$$

Note that $\mathrm{ext}\,(E) = C$. By Extremeness, we know that $\rho_E(C) = 1$. By Monotonicity, we have

$$\rho_E(p_i) \leq \rho_C(p_i)$$

for all $p_i \in C$. Since p_i are also the extreme points of E, by Extremeness, we also have that $\sum_i \rho_E(p_i) = 1$. Hence, it must be that $\rho_E(p_i) = \rho_C(p_i)$ for all i. Moreover, by Monotonicity again, we have that for each i,

$$\rho_{C'_i}(p_i) \geq \rho_E(p_i)$$

as $C'_i \subset E$ for all i. If we let $p'_i = ap_i + (1-a)r \in C'$, then by Linearity*, we have

$$\rho_{C'}\left(p'_i\right) = \rho_{C'_i}(p_i) \geq \rho_E(p_i) = \rho_C(p_i)$$

for all $p_i \in C$. Since this is true for all i, and by Extremeness again $\sum_i \rho_{C'}\left(p'_i\right) = 1$, it must be that $\rho_{C'}\left(p'_i\right) = \rho_C(p_i)$ for all i. Hence, Linearity is satisfied as desired. □

References

Agranov, M., Ortoleva, P., 2017. Stochastic choice and preferences for randomization. J. Polit. Econ. 125, 40–68.
Apesteguia, J., Ballester, M.A., 2018. Monotone stochastic choice models: the case of risk and time preferences. J. Polit. Econ. 126, 74–106.
Apesteguia, J., Ballester, M.A., Lu, J., 2017. Single-crossing random utility models. Econometrica 85, 661–674.
Ballinger, T.P., Wilcox, N.T., 1997. Decisions, error and heterogeneity. Econ. J. 107, 1090–1105.
Benoit, J.-P., Ok, E., 2007. Delay aversion. Theor. Econ. 2, 71–113.
Billingsley, P., 1986. Probability and Measure. John Wiley and Sons.

[32] While Proposition 5 shows how Extremeness ensures the existent of a random linear utility representation, it does not guarantee that the utilities in any random utility representation must be linear. In this sense, our Theorem 3 is stronger.

Camerer, C.F., 1989. Does the basketball market believe in the 'hot hand'? Am. Econ. Rev. 79, 1257–1261.
Chakraborty, A., Halevy, Y., 2017. Allais Meets Strotz: Remarks on the Relation Between Present Bias and the Certainty Effect. Tech. rep., Working paper.
Cinlar, E., 2011. Probability and Stochastics. Springer.
Frederick, S., Loewenstein, G., O'donoghue, T., 2002. Time discounting and time preference: a critical review. J. Econ. Lit. 40, 351–401.
Gul, F., Pesendorfer, W., 2006. Random expected utility. Econometrica 74, 121–146.
Halevy, Y., 2008. Strotz meets Allais: diminishing impatience and the certainty effect. Am. Econ. Rev. 98, 1145–1162.
Halevy, Y., 2015. Time consistency: stationarity and time invariance. Econometrica 83, 335–352.
Hayashi, T., 2003. Quasi-stationary cardinal utility and present bias. J. Econ. Theory 112, 343–352.
Higashi, Y., Hyogo, K., Takeoka, N., 2009. Subjective random discounting and intertemporal choice. J. Econ. Theory 144, 1015–1053.
Higashi, Y., Hyogo, K., Takeoka, N., Tanaka, H., 2016. Comparative impatience under random discounting. Econ. Theory 63, 1–31.
Koopmans, T.C., 1960. Stationary ordinal utility and impatience. Econometrica 28, 287–309.
Lu, J., 2016. Random choice and private information. Econometrica 84, 1983–2027.
McFadden, D., 2001. Economic choices. Am. Econ. Rev. 91, 351–378.
Meier, S., Sprenger, C.D., 2015. Temporal stability of time preferences. Rev. Econ. Stat. 97, 273–286.
Olea, J.L.M., Strzalecki, T., 2014. Axiomatization and measurement of quasi-hyperbolic discounting. Q. J. Econ. 129, 1449–1499.
Pennesi, D., 2015. The Intertemporal Luce Rule. Working paper, THEMA.
Regenwetter, M., Dana, J., Davis-Stober, C.P., 2011. Transitivity of preferences. Psychol. Rev. 118, 42.
Saito, K., 2011. Strotz meets Allais: diminishing impatience and the certainty effect: Comment. Am. Econ. Rev. 101, 2271–2275.
Saito, K., 2015. Preferences for flexibility and randomization under uncertainty. Am. Econ. Rev. 105, 1246–1271.
Saito, K., 2017. A Relationship Between Risk and Time Preferences. Tech. rep., Working paper.
Train, K., 2009. Discrete Choice Methods with Simulation. Cambridge University Press.
Tversky, A., 1969. Intransitivity of preferences. Psychol. Rev. 76, 31–48.