DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

ON MULTIPLE DISCOUNT RATES

Christopher P. Chambers
UC San Diego

Federico Echenique
California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 1418

May 2016

On multiple discount rates

Christopher P. Chambers   Federico Echenique

Abstract

We propose a theory of intertemporal choice that is robust to specific assumptions on the discount rate. One class of models requires that one utility stream be chosen over another if and only if its discounted value is higher for all discount factors in a set. Another model focuses on an average discount factor. Yet another model is pessimistic, and evaluates a flow by the lowest available discounted value.

JEL classification numbers: D90,D70

Key words: Discounted utility; Stationarity; Decision Theory

On multiple discount rates

Christopher P. Chambers Federico Echenique

1 Introduction

Many important long-term decisions depend crucially on the rate used for discounting the future. Economists routinely discount the future stream of consequences of a plan or project, and use the discounted stream to make a decision. But it is very difficult to say precisely which discount rate should be used. The discount rate depends on ethical and empirical considerations on which economists and other experts disagree. Disagreements over the discount rate pose a severe problem because the conclusions of the analysis are extremely sensitive to the precise rate assumed. *To deal with disagreements over the discount rate, we develop a theory that is robust to specific assumptions on the discount rate.*

A case in point is the debate on climate change: the influential Stern review of climate change (Stern, 2007) makes a calculation of the effects of climate change over time, and recommends fairly drastic policy measures. Economists such as Partha Dasgupta, William Nordhaus, and Martin Weitzman take issue with Sterns' calculations. Specifically, they dispute his assumed discount rate. Hal Varian aptly summarizes the debate (Varian, 2006): "So, should the social discount rate be 0.1 percent, as Sir Nicholas Stern, who led the study, would have it, or 3 percent as Mr. Nordhaus prefers? There is no definitive answer to this question because it is inherently an ethical judgment that requires comparing the well-being of different people: those alive today and those alive in 50 or 100 years."

The problem is not only with Stern's review of climate change. Weitzman (2001) reports the results of a survey of over 2,000 economists, in which he asks them for the discount rate that they would use to evaluate long-term projects. The mean of the answers is 3.96 % with a standard deviation of 2.94 %. Weitzman runs the survey on a subsample of 50 very distinguished economists (including many who had won, or have since won, a Nobel prize). The mean of the answers is 4.09 % with a standard deviation of 3.07 %. So it is clear that there is a substantial disagreement among economists about the proper discount rate to use for discounting long-term streams. In fact, Weitzman concludes that: "The most critical single problem with discounting future benefits and costs is that no consensus now exists, or for that matter has ever existed, about what actual rate of interest to use."

In this paper, we explore the consequences of remaining agnostic about the specific discount rate that should be used. Our idea is to retain many of the advantages and conceptual foundations of discounting, while proposing a theory that is *robust* to specific assumptions about the discount rate.

We operationalize robustness in a couple of different ways. Think of choosing among sequences of real numbers: these could be consumption streams, utils, or monetary quantities computed from the costs and benefit of an economic project. In the sequel we often refer to utility streams for concreteness.

First, we describe a "dominance" criterion relating pairs of utility streams. The dominance ranking describes cases in which one stream can be said to be unambiguously better than another, independently of the discount rate. We say that a stream x "discounting dominates" a stream y if for every possible discount factor, the discounted sum along x is at least as great as the discounted sum along y. Concretely, we want to understand when it is the case that $\sum_{t=0}^{\infty} \delta^t x_t \geq \sum_{t=0}^{\infty} \delta^t y_t$ for all $\delta \in [0,1]$.[1]

The discounting dominance relation is useful for several reasons. A social planner evaluating multiple streams can use the ranking to filter out dominated streams. Everyone would agree that these streams should not be pursued. If we think of the different time periods as "generations," then discounting dominance refines the Pareto relation. The relation would then be used in guiding the choice of a social welfare function. Finally, discounting dominance distills the basic properties of discounting common to all discount factors, and broadens our theoretical understanding of the concept of discounting.

Our next result is motivated by the ideas of Weitzman (2001) and Jackson and Yariv (2015). Weitzman proposes that an average discount factor in every period best reflects economists' disagreements over which discount rate to use. Weitzman proposes an av-

[1]As we explain in Section 2.2, the results can be generalized to the case $\delta \in [a,b] \subseteq [0,1]$.

erage taken from a specific distribution, the Gamma distribution. Jackson and Yariv argue that by averaging individual exponential discount factors in each time period, one can explain behavior which exhibits present-bias. Indeed this is true, but their work leaves open the possibility that there are other behavioral predictions implied by the model.[2] We expand on Weitzman's idea by characterizing average discounting beyond the assumption of a specific distribution. We expand on Jackson-Yariv by characterizing all behavior consistent with their model, for an arbitrary set of individuals. Concretely, we characterize the weak orders $\succeq$ for which there exists a probability measure μ on $[0,1]$ such that $x \succeq y$ iff $\sum_{t=0}^{\infty}(\int_0^1 \delta^t d\mu(\delta))x_t \geq \sum_{t=0}^{\infty}(\int_0^1 \delta^t d\mu(\delta))y_t$.

Our third result mimics our first result on discounting dominance, but derives a set of discount factors endogenously. Observe that in the first result, we asked that one stream be better than another for all possible discount factors. Generally, however, some discount factors may appear unreasonable. Instead, one might ask that one stream be better than another for certain values of the discount factor. We characterize the model axiomatically (an exercise similar to Bewley (2002)), deriving the set of discount factors endogenously.

Concretely, we characterize the orderings $\succeq$ for which there exists $D \subseteq (0,1)$ such that $x \succeq y$ iff for all $\delta \in D$

$$\sum_{t=0}^{\infty} \delta^t x_t \geq \sum_{t=0}^{\infty} \delta^t y_t.$$

Such an ordering can be intepreted as the Pareto relation of a suitably well-behaved society of exponentially discounting agents.

Finally, our last result imagines that streams are evaluated according to a "worst-case" analysis. A set of discount factors is given, each normalized so that a constant stream is treated identically by each factor in the set. Any stream is judged by the *minimum* discounted payoff across all factors in the set. Concretely, we characterize the weak orders $\succeq$ with the property that $x \succeq y$ iff $U(x) \geq U(y)$, with

$$U(x) = \inf\{(1-\delta)\sum_{t=0}^{\infty} \delta^t x_t : \delta \in D\}.$$

The max-min representation is analogous to Huber (1981, Proposition 2.1 of Chapter 10.2) or Gilboa and Schmeidler (1989). In Section 3.1 we explain how our result differs

[2]They derive the notion of average discount factor endogenously, assuming a Pareto property and a utility which is additively separable across time.

from theirs.

The substantive properties, or axioms, behind the Pareto and max-min models relate to consumption smoothing. The first is a quasiconcavity property, stating roughly that smoother streams are better. As smoother streams reflect more "fair" streams in an intergenerational context; this property seems entirely natural. The second property is novel, and modifies the stationarity property of Koopmans (1960). Koopmans imagined that if two streams are ranked, that ranking would not change were a common utility appended to the initial period of each stream. The property is usually understood as a stationarity property: the agent anticipates his preference tomorrow to coincide with his preference today.

We argue that such a conclusion should not necessarily hold in a context in which there may be an innate preference for smoothing. For example, appending the common utility to each stream may reverse the already stated preference if one of the new streams becomes more smooth. We rectify this issue in the following way. We do not know what "more smooth" means, but we can at least say that a smooth utility stream (meaning a constant stream, or constant sequence) is smoother than anything else. Thus, if a stream x is at least as good as a smooth stream θ, this cannot be due to a preference for smoothing. In such a case, we would ask stationarity to hold; but we want to ensure that appending a new initial consumption cannot lead to new smoothing opportunities. We do so by requiring that the appended consumption is θ itself. Only in this case is the ranking preserved. We refer to this property as *stationarity.*

The two characterizations obviously differ in other aspects. For example, the Pareto representation requires an incomplete preference in general, and will give rise to status quo bias (as discussed, for example, by Bewley (2002)). It is also separable. The maxmin representation involves a complete preference, but may violate separability. These aspects are discussed in Section 3.2.

Related literature Our first two results use a version of Hausdorff's moment problem. We are not the first to note the relation between that problem and discounting in economics: Hara (2008) is an earlier application of (the continuous version of) the Hausdorff moment problem; *i.e.* Bernstein's Theorem. Bertsimas, Popescu, and Sethuraman (2000) use the Hausdorff moment problem in the context of pricing an asset whose moments are known.

An important motivation for our paper is the literature on multiple discount rates and the evaluation of long-term projects, see Weitzman (2001) and Jackson and Yariv (2015). In particular, the result on expected discount rates presented in Theorem 2 is motivated by two these papers. Jackson and Yariv consider utilitarian aggregation of discounted utilities, and Weitzman argues for the use of an expected discount rate (that he obtains through a survey of economists) like what we obtain in Theorem 2.

The Pareto and maxmin models are related to Bewley (2002), and Gilboa and Schmeidler (1989). We explain how we depart from these papers in Section 3.1. The paper by Wakai (2008) also considers a max-min representation over the discount factor, but in his model the discount factor may be different in each time period. In that sense, his model is closer to Gilboa-Schmeidler's multiple-prior version of max-min. Wakai's focus is on obtaining a dynamically consistent version of the model with multiple and time-varying discount rates.

The papers by Karni and Zilcha (2000), Higashi, Hyogo, and Takeoka (2009), Higashi, Hyogo, Tanaka, and Takeoka (2016), Pennesi (2015), and Lu and Saito (2016) all consider multiple but randomly chosen discount factors. This is of course quite different from our focus on robust conclusions with respect to a fixed set of discount factors.

2 Results

2.1 Definitions and notation

We use $\mathbf{N}$ to denote the set of non-negative integers, $\mathbf{N} = \{0, 1, \ldots\}$. The objects of choice are sequences $x = \{x_t\}_{t=0}^{\infty}$ of real numbers. For some results we assume that sequences lie in ℓ_1, the space of absolutely summable sequences. For the other results we assume sequences in ℓ_∞, the space of all bounded sequences. The space ℓ_1 is endowed with the norm $\|x\|_1 = \sum_{t\in\mathbf{N}} |x_t|$, while ℓ_∞ has norm $\|x\|_\infty = \sup\{|x_t| : t \in \mathbf{N}\}$. We denote by X the space of sequences under consideration, when we do not want to specify whether we are talking about ℓ_1 or ℓ_∞.

Here, sequences of real numbers are interpreted as sequences of *utility values*. We assume that utility values are determined or known. Instead of endogenizing utility, our goal is to uncover methods of aggregating these utilities. Utility could correspond to an intergenerational (and interpersonally comparable) notion, or it could simply reflect period utility for a given agent.

When $x \in \ell_\infty$, and $m \in \ell_1$ is a positive sequence, then we use the notation

$$m \cdot x = \sum_{t \in \mathbf{N}} m_t x_t.$$

Countably additive probability measures on $\mathbf{N}$ are identified with sequences $m \in \ell_1$. Then $x \cdot m$ denotes the expectation of $x \in X$ with respect to the probability measure m.

The sequence $(1, 1, \ldots)$, which is identically 1, is denoted by $\mathbf{1}$. When $\theta \in \mathbf{R}$ is a scalar we often abuse notation and use θ to denote the constant sequence $\theta\mathbf{1}$. If x is a sequence, we denote by (θ, x) the concatenation of θ and x, the sequence takes the value θ for $t = 0$, and then x_{t-1} for each $t \geq 1$. Similarly, the sequence

$$(\underbrace{\theta, \ldots, \theta}_{T \text{ times}}, x)$$

takes the value θ for $t = 0, \ldots, T-1$ and x_{t-T} for $t \geq T$.

The notation for inequalities of sequences is: $x \geq y$ if $x_t \geq y_t$ for all $t \in \mathbf{N}$, $x > y$ if $x \geq y$ and $x \neq y$, and $x \gg y$ if $x_t > y_t$ for all $t \in \mathbf{N}$.

2.2 All discount factors

Here we address the following question. Given $x, y \in \ell_1$, when is it the case that for all $\delta \in [0, 1]$ we have $\sum_{t \in \mathbf{N}} \delta^t x_t \geq \sum_{t \in \mathbf{N}} \delta^t y_t$? In other words, when can we be certain that x will have a higher discounted value than y, independently of the discount rate? Observe that this relation is well-defined as $\sum_t \delta^t x_t \in \mathbf{R}$ for all $\delta \in [0, 1]$ and $x \in \ell_1$. Define $x \succeq^d y$ if for all $\delta \in [0, 1]$, $\sum_t \delta^t x_t \geq \sum_t \delta^t y_t$ (d for "discounting").

We can gain some insight from the seemingly trivial observation that $\delta^s(1-\delta)^t \geq 0$ when $\delta \in [0, 1]$ and $s, t \geq 0$. We shall see that nonnegativity of $\delta^s(1-\delta)^t$ for all $\delta \in [0, 1]$ and $s, t \geq 0$ *tells us (essentially) all the ways in which we may have* $x \succeq^d y$.

Let us focus on the case when $s = 0$. The Binomial Theorem implies that

$$0 \leq (1-\delta)^t = \sum_{i=0}^{t} (-1)^i \delta^t \binom{t}{i}.$$

There is an insight to be gleaned from this expression: First think of the case when $t = 1$. That $(1 - \delta)$ is nonnegative captures the idea of *impatience*: shifting one util from tomorrow to today results in a net gain of $1 - \delta$. In other words, the transformation $x + (1, -1, 0, \ldots)$ is always preferable to x, regardless of the value of δ.

Next, consider $t = 2$. Then

$$0 \leq (1 - \delta)^2 = 1 - 2\delta + \delta^2$$

reflects that "mean preserving spreads" are desirable. Shifting two utils from tomorrow, one of them added to today, and the other to the day after tomorrow, results in an increase of the discounted present value. It results in a net gain of $1 - 2\delta + \delta^2$. Again we can say that the transformation $x + (1, -2, , 1, 0, \ldots)$ is always preferable to x, regardless of the value of δ.

The case $t = 0$ also has an interpretation: that an increase in one util of today's consumption is desirable. In other words, the transformation $x + (1, 0, \ldots)$ is more valuable than x.

When we go beyond $t = 2$ there is no simple interpretation to the inequality $0 \leq (1 - \delta)^t$, but it is still true that each such inequality embodies a transformation of a sequence x into a sequence with a higher discounted value. For example, we know from $(1 - \delta)^3 \geq 0$ that $x + (1, -3, 3, -1, 0, \ldots)$ is more valuable than x. Similarly, from $(1 - \delta)^4 \geq 0$ we infer that $x + (1, -4, 6, -4, 1, 0, \ldots)$ is more valuable than x. Such conclusions follow from the discounting, even they have no simple economic explanation. The point is that such conclusions are independent of the specific value of the discount factor.

Now consider what happens when we take $s > 0$. This gives us related statements about modifying consumption further in the future. For example, consider $s = 1$, and the inequality

$$0 \leq \delta(1 - \delta) = \delta - \delta^2$$

. This reflects the statement that shifting a util from the day after tomorrow to tomorrow increases utility.

The main result of this section is that all transformations that give a higher discounted value, regardless of the discount factor, have this form. They are (limits of) linear combinations of the transformations defined by $0 \leq \delta(1 - \delta)^t$.

Thus motivated by the inequality $(1-\delta)^t = \sum_{i=0}^{t}(-1)^i \delta^t \binom{t}{i} \geq 0$, we introduce a class of transformations of a sequence x. The transformed x will have larger discounted present value, regardless of which value of delta one employs. The first transformation is denoted by

$$\eta(0,0) = (1,0,0,\ldots)$$

and involves adding one unit to x in period 0. It is clear that $x + \eta(0,0) \succeq^d x$. The second transformation is

$$\eta(0,1) = (1,-1,0,\ldots).$$

$x + \eta(0,1) \succeq x$ because regardless of the value of δ it is desirable to shift consumption to the present. The third transformation is

$$\eta(0,2) = (1,-2,1,0,\ldots).$$

As we have seen, $x + \eta(0,2) \succeq x$ because discounting favors the kind of mean preserving spreads captured by $\eta(0,2)$.

In general, we define a class of vectors, which we call *alternating binomial coefficients:* For $s,t \in \mathbf{N}$, let $\eta(s,t) \in l_\infty$ be defined as $\eta(s,t)_i = (-1)^{(i-s)}\binom{t}{i-s}$ for all $i \in \{s,\ldots,t\}$ and $\eta(s,t)_i = 0$ otherwise. We translate the transformation $\eta(0,t)$ by s to obtain $\eta(s,t)$: for example, $\eta(5,1)$ is a shift of consumption on date $t=6$ to $t=5$.

The relation between $\eta(s,t)$ and $\delta^s(1-\delta)^t$ should be clear: It follows from $0 \leq \delta^s(1-\delta)^t$ that $x + \lambda\eta(s,t) \succeq x$.

Define an *elementary transformation of order s* (for $s \in \{0,\ldots\}$) to be a vector of the form $\lambda\eta(s,t)$ for some t and $\lambda > 0$.

Theorem 1. *$y \succeq^d x$ if and only if for each $\epsilon > 0$, there is a finite collection of elementary transformations $\{\lambda_i \eta(s_i,t_i)\}$ for which*

$$\|(y-x) - \sum_i \eta(s_i,t_i)\|_1 \leq \epsilon.$$

An extension of Theorem 1 is possible.[3] Namely, suppose that it is agreed that the discount factor must lie in a compact interval $[a,b]$, rather than $[0,1]$. This would be the case, for example, if there were a lower bound on discounting future generations. Recall the discussion from the introduction, that $\delta^s(1-\delta)^t \geq 0$ for $s,t \geq 0$ generated essentially

[3]We thank Itai Sher for suggesting this question.

all statements that could be made about dominance. This is reflected in Theorem 1 by the fact that dominance arises from transformations of the form $\eta(s,t)$. Now, if we instead consider an arbitrary interval $[a,b]$ with $a<b$, we would instead investigate the expressions $(\delta-a)^s(b-\delta)^t$ for all possible s,t. For example, if $s=2, t=1$, we have the inequality

$$(\delta-a)^2(b-\delta)=a^2b-(a^2+2ab)\delta+(2a+b)\delta^2-\delta^3\geq 0,$$

implying an elementary transformation

$$\eta_a^b(s,t)=(a^2b,-(a^2+2ab),2a+b,-1,0,\ldots).$$

Observe that, again by the binomial theorem,

$$(\delta-a)^s(b-\delta)^t=\sum_{i=0}^{s}\sum_{j=0}^{t}\binom{s}{i}\binom{t}{j}\delta^{i+t-j}a^{s-i}b^j(-1)^{s+t-i-j}.$$

Hence, for all $k\leq s+t$, we have

$$\eta_a^b(s,t)_k=\sum_{\{(i,j):i+t=j+k\text{ and }(0,0)\leq(i,j)\leq(s,t)\}}\binom{s}{i}\binom{t}{j}a^{s-i}b^j(-1)^{s+t-i-j}.$$

If $k>s+t$, we have $\eta_a^b(s,t)_k=0$. Theorem 1 then holds as stated, replacing η with η_a^b (observe that $\eta_0^1=\eta$). Though we will not address it here, specific formulas can also be derived when it is known that the discount factor lie in a union of closed, compact intervals for example.

2.3 Axioms

We begin the remainder of our analysis by introducing a collection of axioms. We say that a binary relation is a *weak order* if it is complete and transitive, and an *ordering* if it is reflexive and transitive.

2.3.1 Standard axioms

We state some basic axioms that are either commonly used in the literature, or variations on commonly-used axioms.

The letters x, y and z refer to streams in X; θ is a constant stream. Unbound variables are universally quantified.

- *Monotonicity*: $x \geq y$ implies $x \succeq y$, and $x \gg y$ implies $x \succ y$.
- *Non-degeneracy*: There exist $x, y \in X$ for which $x \succ y$.
- *d-monotonicity*: $x \succeq^d y$ implies $x \succeq y$.
- *c-convexity*: For all $\lambda \in [0,1]$, if $x \succeq \theta$ and $y \succeq \theta$, then $\lambda x + (1-\lambda)y \succeq \theta$.
- *Translation invariance*: $x \succeq y$ implies $x + z \succeq y + z$.
- *c-Translation Invariance*: $x \succeq y$ implies $x + \theta \succeq y + \theta$.
- *Homotheticity*: For all $x, y \in X$ and all $\alpha \geq 0$, if $x \succeq y$, then $\alpha x \succeq \alpha y$.
- *Continuity*: $\{y \in X : y \succeq x\}$ and $\{y \in X : x \succeq y\}$ are closed.

Note that $x \succeq^d y$ presumes that $x, y \in \ell_1$. Inasmuch as the relation $\succeq^d$ can be interpreted as a basic "dominance" relation, d-monotonicity is as reasonable as monotonicity.

The c-convexity axiom imposes a preference for "smoothing" utility across time. In an intergenerational context, such a preference would naturally result from equity considerations. Note that, in the standard intertemporal choice model with discounted utility, smoothing is a consequence of the concavity of the utility function. There is no such concavity in our model. The streams under consideration are already measured in "utils" per period of time, and the standard intertemporal choice model is linear in utils. Our c-convexity axiom says that smoothing may be intrinsically desirable.

Translation invariance is usually understood as the requirement that the scale of utility across periods may matter, but there are no particular utility comparisons made across periods. c-Translation Invariance weakens translation invariance to allow for meaningful intertemporal comparison of utility. Note that Translation Invariance imposes separability across time (in the sense that if $x_t = y_t$ and $x'_t = y'_t$ for all $t \in E$, while $x_t = x'_t$ and $y_t = y'_t$ for all $t \in E^c = \mathbf{N} \setminus E$, then $x \succeq y$ implies $x' \succeq y'$). In contrast c-Translation Invariance does not impose such separability.

2.3.2 Novel axioms

Our first novel axioms are versions of the Koopmans (1960) stationarity property. Recall the notion of Koopmans (1960), which claims that stream x is at least as good as y iff it remains so preserved when an identical payoff is appended to the first period of each stream. Our axioms weaken Koopmans', in that they apply only when the second stream is smooth, and when the identical payoff appended coincides with this constant payoff.

Stationarity: *For all $t \in \mathbf{N}$ and all $\lambda \in [0,1]$,*

$$x \succeq \theta \text{ iff } \lambda x + (1-\lambda)(\underbrace{\theta, \ldots, \theta}_{t \text{ times}}, x) \succeq \theta.$$

Generally speaking, stationarity requires certain choices to be time-invariant. It requires that the comparison between two streams remains the same whether it is made today or in the future. We impose a form of stationarity that requires time-invariance of comparisons with constant, or smooth, streams. The reason is that postponing the decision has a natural interpretation in the case of smooth streams.

Suppose that a policy maker has to choose between two streams, x and a constant stream θ. Think of θ as a baseline, or status quo. Suppose also that she chooses x over θ, so she decides to switch and adopt x over the baseline θ. A postponed version of this decision problem would be to choose between (θ, x) and θ. Then we would argue that the policy maker should choose (θ, x) over θ because the decision problem of choosing between (θ, x) and θ is equivalent to postponing the decision of whether to switch to x for one period. The baseline θ is constant, and delivers θ in every period, so (θ, x) is the same as staying with the θ policy for one period and then switching to x.

A stronger version of stationarity (such as Koopmans') would demand that any decision is preserved if postponed. If our policy maker chooses x over y, then she would be required to choose (θ, x) over (θ, y) for any θ; that is, independently of history. But it is easy to imagine reasons for the decision to be reversed, and (θ, y) chosen over (θ, x).[4] Since (θ, y) is different from y we can imagine situations where θ in period 0 may "enhance" the value of y, for example if θ is a large positive value, and the stream y starts out poorly. The difference with our axiom, in which y is required to be the constant stream θ, is that (θ, y) is different from y. So in our case, we can justify the axiom by saying that if a policy maker is willing to switch from θ to x today, then she must be

[4] See also Hayashi (2016).

willing to switch tomorrow.

Finally, our stationarity axiom says more. Not only must the comparison of x and θ be the same as that between (θ, x) and θ, but this must also be true of the comparison of any lottery $\lambda x + (1-\lambda)(\theta, x)$ and θ. In particular, the only criteria for choosing between $\lambda x + (1-\lambda)(\theta, x)$ and θ must be the comparison of x with θ. The meaning is that there is no additional smoothing (or "hedging") motive in the comparisons of x with θ, now or in the future.

The following axiom, *compensation*, is a technical non-triviality axiom. Its purpose is to ensure that the future is never irrelevant. It is similar in spirit to Koopmans' sensitivity axiom (Postulate 2 of Koopmans (1960)).

Compensation: *For all t there are scalars $\bar{\theta}^t$, θ^t, and $\underline{\theta}^t$, with $\bar{\theta}^t > \theta^t > \underline{\theta}^t$, such that*

$$(\underbrace{\underline{\theta}^t, \ldots, \underline{\theta}^t}_{t \text{ times}}, \bar{\theta}^t, \ldots) \succeq \theta^t.$$

Compensation says that for any t there must exists three numbers: $\bar{\theta}^t > \theta^t > \underline{\theta}^t$, such that the worse outcome $\underline{\theta}^t$ for t periods is compensated by a better outcome $\bar{\theta}^t$ for all periods $t+1, \ldots$, relative to the smooth stream that gives the intermediate value θ^t in every period. The axiom ensures that no future period is irrelevant for the purpose of comparing utility streams.

Our last axiom is a formal, yet weak, expression of discounting. Roughly, it states that whenever a stream x is at least as good as a smooth stream θ, then the preference is always willing to wait "long enough" so that changes in x do not matter. Axioms along these lines were introduced by Villegas (1964) and Arrow (1974).

Continuity at infinity: *For all $x \in X$, all θ, if $\theta \succeq (x_0, \ldots, x_T, 0, \ldots)$ for all T, then $\theta \succeq x$.*

2.4 Average discount factor

Jackson and Yariv (2015) have shown that a society of individuals aggregating exponentially discounted preferences in a time-separable and Paretian fashion must socially discount according to a weighted sum of period individual discount factors. Here, we characterize the full implications of this model using Theorem 1. Clearly, such a society must have a discount factor which respects d-monotonicity and translation invariance. It turns out that these are essentially the *only* requirements imposed on the model.

Theorem 2. *A weak order $\succeq$ on ℓ_1 satisfies nondegeneracy, continuity, d-monotonicity, and translation invariance iff there is a Borel probability measure μ on $[0,1]$ such that $x \succeq y$ iff $\sum(\int_0^1 \delta^t d\mu(\delta))x_t \geq \sum(\int_0^1 \delta^t d\mu(\delta))y_t$.*

2.5 Pareto ordering

In this section, we address the question of an unobserved set of discount factors. When would it be the case that there is a collection of agents, each possessing her own discount factor, such that one stream dominates another whenever this is the case for every agent in the collection?

The question is different from the question addressed in Theorem 1. There, we were concerned with understanding the properties of a canonical dominance ordering, generated by all possible discount factors. Here, we wish to find the properties common to a family of dominance relations: those based on a (closed) set of nonzero discount factors.

Theorem 3. *An ordering $\succeq$ satisfies continuity, monotonicity, c-convexity, translation invariance, strong stationarity, compensation and monotone continuity iff there is a nonempty closed*[5] *set $D \subseteq (0,1)$ such that $x \succeq y$ iff for all $\delta \in D$*

$$(1-\delta)\sum_{t=0}^{\infty}\delta^t x_t \geq (1-\delta)\sum_{t=0}^{\infty}\delta^t y_t.$$

Furthermore, D is unique.

2.6 Max min

Theorem 4. *The preference relation $\succeq$ satisfies continuity, monotonicity, convexity, homotheticity, c-translation invariance indifference stationarity, compensation and monotone continuity iff there is a nonempty closed set $D \subseteq (0,1)$ such that $x \succeq y$ iff $U(x) \geq U(y)$, with*

$$U(x) = \min\{(1-\delta)\sum_{t=0}^{\infty}\delta^t x_t : \delta \in D\}.$$

[5]Closed means with respect to the standard Euclidean topology, and not with respect to the relative topology on $(0,1)$. So any closed set must exclude 0 and 1.

Furthermore, D is unique.

As noted earlier, indifference stationarity could be replaced by strong stationarity in Theorem 4 with no change to the results.

3 Discussion

3.1 On the proof of Theorems 3 and 4

Theorems 3 and 4 are obtained by first establishing a multiple prior representation, and then using the stationarity axiom to update some of the priors.

The proof of Theorem 3 relies on first obtaining a multiple prior representation as in Bewley (2002): there is a set of probability distributions M over $\mathbf{N}$, and $x \succeq y$ iff the expected value of x is larger than the expected value of y for all probability distributions in M. Similarly, the proof of Theorem 4 relies on a max-min multiple prior representation, as in Huber (1981) and Gilboa and Schmeidler (1989). We use the continuity at infinity axiom, and ideas from Villegas (1964), Arrow (1974), and Chateauneuf, Maccheroni, Marinacci, and Tallon (2005), to show that the measures in M are countably additive.

The main contribution in our paper is to use stationarity to show that M is the convex hull of *exponential* probability distributions. This is carried out in Lemma 8. The idea is to choose a subset of the extreme points of M (the exposed points of M; these are the extreme points that are the unique minimizers in M of some supporting linear functional), and show that when these priors are updated then they have the *memoryless* property that characterizes the exponential distribution.

Think of each $m \in M$ as representing the beliefs over when the world will end, and choose a particular extreme point m of M. We show that the stationarity axiom implies that for any time period $t \geq 0$, if m' is the belief $m \in M$ conditional (Bayesian updated) on the event $\{t, t+1, \ldots\}$ (that is, conditional on the event that the world does not end before time t), then $m' = m$. This means that m is the geometric distribution.

3.2 On Koopmans' axiomatization.

Koopmans (1960) is the first axiomatization of discounted utility. He relies on two crucial ideas: one is separability and the other is stationarity. Separability means two things.

First that $(\theta, x) \succeq (\theta', x)$ iff $(\theta, y) \succeq (\theta', y)$ for all y. Second, that $(\theta, x) \succeq (\theta, y)$ iff $(\theta', x) \succeq (\theta', y)$ for all θ'. It is easy to see that translation invariance implies separability, but c-translation invariance does not. So the Pareto model in Theorems 3 satisfies separability, but the following simple example illustrates that max-min model in Theorem 4 may violate separability: Let the preference relation $\succeq$ have a max-min representation with $D = \{1/5, 4/5\}$. Then $(0, 1, 0, \ldots) \succ (0, 0, 2, 0, \ldots)$ while $(5, 0, 2, \ldots) \succ (5, 1, 0, \ldots)$; a violation of separability. In light of some experimental evidence against separability (see Loewenstein (1987) and Wakai (2008)), it may be interesting to note that the max-min model does not impose it.

The second of Koopman's main axioms is stationarity. It says that $x \succeq y$ iff $(\theta, x) \succeq (\theta, y)$. It is probably obvious how his axiom differs from ours, but let us stress two aspects. In our stationarity axiom, stationarity is only imposed for comparisons with a smooth stream. As we explained in 2.3.2, our idea is that the smooth stream is a status quo, and that the comparison in the stationarity axiom can be phrased as postponing the decision to move away from the status quo.

The other way in which we depart from Koopmans is that our stationarity axiom requires that $\lambda x + (1 - \lambda)(\theta, x) \succeq \theta$ implies $x \succeq \theta$. The idea is again that the comparison between $\lambda x + (1 - \lambda)(\theta, x)$ and θ is based on the comparison between x and θ, but we should stress that this direction of the axiom is only really needed in Theorem 3. For the max-min model of Theorem 4 we can use the following version of stationarity instead:

Axiom 5 (Indifference stationarity)**.** *For all $t \in \mathbf{N}$ and all $\lambda \in [0, 1]$,*

$$x \sim \theta \Longrightarrow \lambda x + (1 - \lambda)(\underbrace{\theta, \ldots, \theta}_{t \text{ times}}, x) \sim \theta.$$

We cannot use indifference stationarity for the Pareto representation of Theorem 3 because $\succeq$ may be an incomplete ordering, which makes indifference unsuitable for our analysis. In the max-min model of Theorem 4, in contrast, we can base much of the analysis on the indifference relation $\sim$, and then Indifference stationarity can be used to replace stationarity in the theorem.

Finally, it is worth mentioning that all our models satisfy *impatience*, meaning that it is always desirable to obtain a positive outcome early; for example $(1, 0, \ldots) \succeq (0, 1, 0, \ldots)$. It is obvious that $(1, 0, \ldots) \succeq^d (0, 1, 0, \ldots)$, and therefore that the expected discounting model of Theorem 2 satisfies impatience. It is also true that the Pareto and max-min models satisfy impatience, but it is not the direct implication of any one of our axioms. Rather, impatience comes about because we obtain a multiple prior representation (see the discussion in 3.1), and stationarity and continuity at infinity imply discount factors that are in $(0, 1)$.

3.3 Non-convexity of D and Pareto optimality

The set of discount factors D in Theorems 3 and 4 does not need to be an interval in $(0, 1)$. The set of priors M is convex, but the set of discount factors does not need to be convex. Despite this lack of convexity, D has some of the same properties as the set of priors in models of multiple priors.

We will not spell out the details, but one can imagine an exchange economy in which ℓ_∞ is the commodity space, and with n agents, each of them with a set of discount factors D_i. Let M_i be the convex hull of the resulting exponential priors over $\mathbf{N}$. By results in Billot, Chateauneuf, Gilboa, and Tallon (2000) or Rigotti and Shannon (2005) the existence of smooth Pareto optimal outcomes relies on the existence of a point $m \in \bigcap_{i=1}^n M_i$.[6] It is then easy to show that there is $\delta \in \bigcap_{i=1}^n D_i$ such that m corresponds to the exponential distribution over $\mathbf{N}$ defined by δ.

4 Conclusion

This paper has suggested modeling robustness to discount factor by considering a set-valued concept. Our first result provides a language for discussing unambiguous ranking in the presence of discounting when the discount factor is either unknown, or there is disagreement about the discount factor. Our second result provides the implications for a discounter who places a probabilistic assessment on exponential discounting. Finally, our final two results describe an endogeneously derived set-valued concept.

5 Proof of Theorem 1

To establish the theorem, we need a preliminary definition.

[6] These results rely on a multi-set generalization of the separating hyperplane theorem, usually attributed to Dubovitskii-Milyutin. See *e.g.* Holmes (1975a), exercise 2.47.

Given $\gamma \in l_\infty$, define the *difference function* $\Delta_\gamma : \mathbf{N}^2 \to \mathbf{R}$ inductively as follows:

1. $\Delta_\gamma(0,t) = \gamma(t)$
2. $\Delta_\gamma(m,t) = (-1)^m[\Delta(m-1,t+1) - \Delta(m-1,t)]$.

Say that γ is *totally monotone* if for all $m,t \in \mathbf{N}$, $\Delta_\gamma(m,t) \geq 0$. Total monotonicity is basically the concept of infinite-order stochastic dominance, applied to a discrete environment. The class of totally monotone functions is a subset of l_∞ which we denote by $\mathcal{T}$.

Total monotonicity means for all t:

- $\gamma(t) \geq 0$
- $-\gamma(t+1) + \gamma(t) \geq 0$
- $\gamma(t+2) - 2\gamma(t+1) + \gamma(t) \geq 0$
- $-\gamma(t+3) + 3\gamma(t+2) - 3\gamma(t+1) + \gamma(t) \geq 0$
- $\gamma(t+4) - 4\gamma(t+3) + 6\gamma(t+2) - 4\gamma(t+1) - \gamma(t) \geq 0$

The inequalities are the same as $\eta(m,t) \cdot \gamma \geq 0$ for all $m,t \in \mathbf{N}$.

The following result is due to (Hausdorff, 1921), and is referred to as the *Hausdorff Moment Problem.*[7]

Proposition 6. *γ is totally monotone if and only if there is a Borel measure (*i.e.* non-negative measure on the Borel sets) μ on $[0,1]$ for which $\gamma(t) = \int_0^1 \delta^t \mu(\delta)$.*

Proof. (of Theorem 1) First, we establish that $x \succeq^d y$ if and only if for all $\gamma \in \mathcal{T}$, $\gamma \cdot x \geq \gamma \cdot y$.[8] For $\delta \in [0,1]$, $\gamma(t) = \delta^t$ is totally monotone by Proposition 6. So, if $\gamma \cdot x \geq \gamma \cdot y$ for all $\gamma \in \mathcal{T}$, then $x \succeq^d y$. Conversely, suppose that $x \succeq^d y$. Let $\gamma \in \mathcal{T}$. Then let μ be the Borel over $[0,1]$ associated with γ. Since $x \succeq^d y$, we know that $\sum_t \delta^t x_t \geq \sum_t \delta^t y_t$ for all $\delta \in [0,1]$; integrating with respect to μ obtains $\int_0^1 \sum_t \delta^t x_t d\mu(\delta) \geq \int_0^1 \sum_t \delta^t y_t d\mu(\delta)$. Now, $|\delta^t x_t| \leq |x_t|$ for all t, so $\int_0^1 \sum_t |x_t| d\mu(t) \leq \mu([0,1]) \sum_t |x_t|$. So by Fubini's Theorem (see

[7]Observe that this result is closely related to the characterization of belief functions as those capacities which are totally monotone, *e.g.* Shafer (1976).

[8]We use the notation $\gamma \cdot x = \sum_t \gamma(t) x_t$.

Theorem 11.26 of Aliprantis and Border (1999), $\int_0^1 \sum_t \delta^t x_t d\mu(t) = \sum_t \int_0^1 \delta^t x_t d\mu(\delta) = \gamma \cdot x$. Similarly, $\int_0^1 \sum_t \delta^t y_t d\mu(\delta) = \gamma \cdot y$, so that $\gamma \cdot x \geq \gamma \cdot y$.

Therefore, if $x \succeq^d y$ is false, there is a totally monotone γ for which $\gamma \cdot (x - y) < 0$. By renormalizing, we can choose γ so that $\gamma \cdot (y - x) \geq 1$. Now, it is simple to verify that γ is totally monotone if and only if $\gamma \cdot \eta(m,t) \geq 0$ for all $m, t \in \mathbf{N}$.[9] So $x \succeq^d y$ being false is equivalent to the consistency of the set of linear inequalities:

- $\gamma \cdot (y - x) \geq 1$
- $\gamma \cdot \eta(m,t) \geq 0$ for all $m, t \in \mathbf{N}$.

for some $\gamma \in l_\infty$.

Consider the set of vectors $(y - x, 1) \in \ell_1 \times \mathbf{R}$ and $(\eta(m,t), 0) \in \ell_1 \times \mathbf{R}$ for all (m,t); we can call this set $\mathcal{V}$. By the Corollary of p. 97 on Holmes (1975b), we may conclude that our inequality system is inconsistent if and only if $(0,1)$ is in the closed convex cone spanned by $\mathcal{V}$.

Therefore, we can conclude that for any $\epsilon > 0$, there is $(z,a) \in \ell_1 \times \mathbf{R}$, where (z,a) is in the convex cone spanned by $\mathcal{V}$ and for which $\|z\|_1 + |1 - a| < \epsilon$; which implies that each of $\|z\|_1 < \epsilon$ and $|1 - a| < \epsilon$. In particular, by taking a sufficiently close to 1, we can also guarantee that $\|\frac{1}{a} z\|_1 < \epsilon$.[10] The vector $\left(\frac{1}{a} z, 1\right)$ is in the convex cone spanned by $\mathcal{V}$.

To simplify notation, write $w = \frac{1}{a} z$. Now, $(w, 1)$ is a finite combination of vectors of the form $(\lambda_i \eta(m_i, t_i), 0)$ and $(b(y - x), b)$. Clearly, it must be that $b = 1$, so we have $w = (y - x) + \sum_{i=1}^N \lambda_i \eta(m_i, t_i)$, which is what we wanted to show. $\square$

The extension mentioned after the statement of Theorem 1 follows from a generalization of Proposition 6. Specifically, it is known that for $\gamma : \mathbf{N} \to \mathbf{R}$, there is a Borel probability measure μ on $[a,b]$ for which $\gamma(t) = \int_0^1 \delta^t \mu(\delta)$ if and only if for every polynomial $P : \mathbf{R} \to \mathbf{R}$, given by $P(x) = \sum_{i=0}^n a_i x^i$ for which for all $x \in [a,b]$, we have $P(x) \geq 0$, it follows that $\sum_{i=0}^n a_i \gamma(i) \geq 0$ (see, *e.g.* Theorem 1.1 of Shohat and Tamarkin (1943)). Further, it is known that if P is a nonnegative polynomial on $[a,b]$, then it can be written as $P(x) = \sum_{(s,t) \in S} \lambda_{(s,t)} (x - a)^s (b - x)^t$ for some set of indices $S \subseteq \mathbf{N}^2$ and

[9] The proof uses Pascal's identity: $\binom{m-1}{i-(t+1)} + \binom{m-1}{i-t} = \binom{m}{i-t}$ to show (by induction on m) that $\gamma \cdot \eta(m,w) = \Delta_\gamma(m,t)$. See, *e.g.* Aigner (2007), p. 5.

[10] For example, let $\nu > 0$ so that $\nu^2 + \nu < \epsilon$, and take (z,a) so that $|\frac{1}{a}| < 1 + \nu$ and $\|z\|_1 < \nu$. Then $\|\frac{1}{a} z\|_1 \leq |\frac{1}{a}| \|z\|_1 < \nu^2 + \nu < \epsilon$.

$\lambda_{(s,t)} \geq 0$. A variant of this fact is due to Bernstein (1915), for the case $[-1, 1]$; see again Shohat and Tamarkin (1943), p. 8 who consider the case $[0, 1]$. The result then follows from renormalizing. Finally this leads to the result, as it implies that we only need to check nonnegativity of the polynomials $(x - a)^s(b - x)^t$ for each s, t.

6 Proof of Theorem 2

That the axioms are necessary is obvious. Conversely, suppose that the axioms are satisfied. Since ℓ_1 is separable (Theorem 15.21 of Aliprantis and Border (1999)), and since $\succeq$ is a continuous weak ordering, by Debreu's representation theorem (Debreu (1954)), there is a continuous utility function $U : \ell_1 \to \mathbf{R}$ which represents $\succeq$.

First note that $x \succeq y$ implies that $nx \succeq ny$ for any $n \geq 1$. The proof is by induction: suppose that $(n-1)x \succeq (n-1)y$, then using translation invariance twice we obtain that $nx = x + (n-1)x \succeq x + (n-1)y \succeq y + (n-1)y = ny$. In particular, this means that $nx \sim ny$ whenever $x \sim y$.

Second, we argue that there is a scalar θ such that $(\theta, 0, \ldots) \succ 0$. By nondegeneracy there is $x, y \in \ell_1$ with $x \succ y$. By translation invariance, we may without loss suppose that $y = 0$, so we obtain that $x \succ 0$. We may also suppose without loss that $x \geq 0$, since monotonicity (implied by d-monotonicity) implies $(x \vee 0) \succeq x$.[11] Now using d-monotonicity, note that $(\theta, 0, \ldots) \succeq x$ for any scalar $\theta \geq \|x\|_1$.

These two facts, that $nx \sim ny$ whenever $x \sim y$ and that $(\theta, 0, \ldots) \succ 0$, imply that if $\gamma > \gamma'$ then $(\gamma, 0, \ldots) \succ (\gamma', 0, \ldots)$. The reason is that $(\gamma, 0, \ldots) \succeq (\gamma', 0, \ldots)$ by monotonicity (again implied by d-monotonicity) and that $(\gamma, 0, \ldots) \sim (\gamma', 0, \ldots)$ would mean that $(n\gamma, 0, \ldots) \sim (n\gamma', 0, \ldots)$ for any $n \geq 1$. But if we choose n with $n(\gamma - \gamma') > \theta$ then $(n\gamma, 0, \ldots) \sim (n\gamma', 0, \ldots)$ would mean that $(\theta, 0, \ldots) \succ 0 \sim (n(\gamma - \gamma', \ldots)$ and contradict monotonicity.

Given that we have shown that $\gamma > \gamma'$ implies $U(\gamma, 0, \ldots) > U(\gamma', 0, \ldots)$ we may without loss of generality assume that $U(\gamma, 0, \ldots) = \gamma$ for $\gamma \in \mathbf{R}$.[12]

[11] The notation $(x \vee 0)$ refers to the element-by-element maximum.

[12] That this normalization is valid relies on the fact that $U(\{\theta, 0, 0, \ldots : \theta \in \mathbf{R}\}) = U(\ell_1)$. This latter property holds since, by continuity, for any $y \in \ell_1$ and any $\theta > 0$, there is $n > 0$ large so that $(\theta, 0, 0, \ldots) \succ \alpha n^{-1} y$. This can be seen to imply that $(n\theta, 0, 0, \ldots) \succ y$. Similarly, there is θ' for which $y \succ (\theta', 0, 0, \ldots)$. Hence by continuity there is θ^* for which $(\theta^*, 0, 0, \ldots) \sim y$, establishing the claim.

We claim that U is a linear functional. By definition of U, we know that for all $x, y \in \ell_1$, $U(U(x), 0, 0, \ldots) = U(x)$, so $x \sim (U(x), 0, 0, \ldots)$; and $y \sim (U(y), 0, 0, \ldots)$. Hence, we know that $x + y \sim (U(x), 0, 0, \ldots) + y \sim (U(x), 0, 0, \ldots) + (U(y), 0, 0, \ldots)$, by a double application of translation invariance. Conclude that $U(x+y) = U((U(x)+U(y), 0, 0, \ldots))$. Since $U((U(x) + U(y), 0, 0, \ldots)) = U(x) + U(y)$, we have that $U(x + y) = U(x) + U(y)$. That $U(\alpha x) = \alpha U(x)$ for any $\alpha \in \mathbf{R}$ follows from the preceding and the continuity of U. Hence, U is a continuous, monotone linear functional representing $\succeq$. Moreover, $U(1, 0, 0, \ldots) = 1$.

The dual space of ℓ_1 coincides with l_∞ (Theorem 12.28 of Aliprantis and Border (1999)), so that there is some bounded function $\gamma : \mathbf{N} \to \mathbf{R}$ for which for all $x \in \ell_1$, $U(x) = \sum_t \gamma(t) x_t$. Observe now that by d-monotonicity and Theorem 1, for each $s, t \in \mathbf{N}$, we have $U(\eta(s, t)) \geq 0$. In other words,

$$0 \leq U(\eta(s, t)) = \gamma \cdot \eta(s, t) = \Delta_\gamma(s, t)$$

for all $s, t \in \mathbf{N}$. Thus, γ is totally monotone. The result then follows from Proposition 6 and the fact that $U(1, 0, 0, \ldots) = 1$.

7 Proof of Theorems 4 and 3

The following lemma is useful.

Lemma 7. *The function $m : [0, 1) \to \ell_1$ given by $m(\delta) = (1 - \delta)(1, \delta, \delta^2, \ldots)$ is norm-continuous.*

Proof. First, we show that the map $d : [0, 1) \to \ell_1$ given by $d(\delta) = (1, \delta, \delta^2, \ldots)$ is continuous. The result will then follow as $m(\delta) = (1 - \delta) d(\delta)$.[13]

So, let $\delta_n \to \delta^*$. Then $\|d(\delta_n) - d(\delta^*)\|_1 = \sum_t |\delta_n^t - (\delta^*)^t|$. Observe that for each t, $|\delta_n^t - (\delta^*)^t| \to 0$. By letting $\hat{\delta} = \sup_n(\delta_n) < 1$, we have that for each t, $|\delta_n^t - (\delta^*)^t| \leq \max\{|(\delta^*)^t|, |\hat{\delta}^t - (\delta^*)^t|\}$, since the expression $|\delta^t - (\delta^*)^t|$ increases monotonically when δ moves away from δ^*. And observe that $\sum_t \max\{|(\delta^*)^t|, |\hat{\delta}^t - (\delta^*)^t|\} < +\infty$. Conclude by the Lebesgue Dominated Convergence Theorem (Theorem 11.20 of Aliprantis and Border (1999)) that $\|d(\delta_n) - d(\delta^*)\|_1 \to 0$. $\square$

[13] The latter is easily deemed continuous. By a simple application of the triangle inequality, if $\delta_n \to \delta^*$, we have $\|(1 - \delta_n) d(\delta_n) - (1 - \delta) d(\delta)\|_1 \leq |(\delta - \delta_n)| \|d(\delta_n)\|_1 + (1 - \delta) \|d(\delta_n) - d(\delta)\|_1$.

Lemma 8, following, characterizes cones in ℓ_∞ which are the set of streams which have nonnegative discounted payoff for every discount factor in some (endogenously determined) closed set of discount factors. The lemma is the main building block in both the maxmin representation, and the Bewley style representation. In each environment, the cone of vectors deemed at least as good as 0 must be a cone of this type. From there, it is a matter of translating the properties of the cone into the properties of the preference $\succeq$.

The lemma uses similar ideas to those of Villegas (1964), Arrow (1974), and Chateauneuf, Maccheroni, Marinacci, and Tallon (2005) to obtain countably additive measures. Villegas and Arrow show the existence of countably additive priors in Savage's subjective expected utility model. Chateauneuf et. al show that the set of priors in the α-maximin model is countably additive.

The main novelty in the lemma lies in using the boundary property 4 to show that the measures supporting the cone take the exponential form. This is achieved essentially by updating the supporting measures and by showing the "memoryless" property of the exponential distribution.

Lemma 8. *Let $P \subseteq \ell_\infty$. Suppose P satisfies the following properties.*

1. *P is a ℓ_∞-closed, convex cone.*
2. *There is $p \notin P$.*
3. *$\ell_\infty^+ \subseteq P$.*
4. *$p \in bd(P)$ implies $(0, 0, \ldots, 0, p) \in P$ and $p + (0, 0, \ldots, 0, p) \in bd(P)$.*
5. *For all $\theta \in [0, 1)$, there is T so that*
$$(\underbrace{1 - \theta, \ldots, 1 - \theta}_{T \text{ times}}, -\theta, -\theta, \ldots) \in P.$$
6. *For all T, $(\underbrace{0, \ldots, 0}_{T \text{ times}}, 1, \ldots) \in int(P)$.*

Then there is a nonempty closed $D \subseteq (0, 1)$ so that $P = \bigcap_{\delta \in D}\{x : \sum_t (1 - \delta)\delta^t x_t \geq 0\}$. Conversely, if there is such a set D, all of the properties are satisfied.

Proof. Establishing that if there is such a D, then the properties are satisfied is mostly simple: Let $M = \{m(\delta) : \delta \in D\}$, so that $P = \bigcap_{\delta \in D}\{x : m(\delta) \cdot x \geq 0\}$. Each set $\{x : m(\delta) \cdot x \geq 0\}$ is closed, and contains ℓ_∞^+, so (1) and (3) are satisfied. Property (2) is immediate as P contains no negative sequences.

For the other properties, note that Lemma 7 and the compactness of D imply that M is norm-compact. Observe that $x \in P$ iff $\inf_{\delta \in D}(1-\delta)\sum_t \delta^t x_t \geq 0$, and that moreover this infimum is achieved (by norm-compactness of M). Then, to see that (4) is satisfied, observe that if $x \in \mathrm{bd}(P)$, then there is $\delta \in D$ for which $m(\delta) \cdot x = 0$, and in particular then, $m(\delta) \cdot (\underbrace{0, \ldots, 0}_{T \text{ times}}, x) = 0$, and hence $m(\delta) \cdot (x + (\underbrace{0, \ldots, 0}_{T \text{ times}}, x)) = 0$. This means that $x + (0, \ldots, 0, x) \in \mathrm{bd}(P)$.

Properties (5) and (6) obtain as $0 < \inf D \leq \sup D < 1$. First, $m(\delta) \cdot (1-\theta, \ldots, 1-\theta, -\theta, \ldots) = (1-\delta^T) - \theta$. So $\theta < 1$ means that there is T such that $(1-\delta^T) - \theta \geq 0$ for all $\delta \in D$. Second, for any T, let $\varepsilon > 0$ be such that $\inf\{\delta^T : \delta \in D\} > \varepsilon$. Then if $m(\delta) \cdot (-\varepsilon, \ldots, -\varepsilon, 1-\varepsilon, \ldots) = \delta^T - \varepsilon \geq 0$ for all $\delta \in D$. This means that if $\|x - (0, \ldots, 0, 1, \ldots)\| < \varepsilon$ then $x \in P$.

We now turn to proving that properties (1)-(6) imply the existence of D as in the statement of the lemma.

Step 1: Constructing a set M of finitely additive probabilities on N as the polar cone of P.

Let $\mathrm{ba}(\mathbf{N})$ denote the bounded, additive set functions on $\mathbf{N}$, and observe that $(\ell_\infty, (ba)(\mathbf{N}))$ is a dual pair. Consider the cone $M^* \subseteq \mathrm{ba}(\mathbf{N})$ given by $M^* = \bigcap_{p \in P}\{x : x \cdot p \geq 0\}$. By Aliprantis and Border (1999) Theorems 5.86 and 5.91, $P = \bigcap_{x \in M^*}\{p : x \cdot p \geq 0\}$.[14] Since $\ell_\infty^+ \subseteq P$ (property (3)), we can conclude that $M^* \subseteq \mathrm{ba}(\mathbf{N})^+$. Moreover, there is nonzero $m \in M^*$ (by the existence of $p \notin P$, property 2.) For any such nonzero m, observe that since $m \geq 0$, it follows that $m(\mathbf{1}) > 0$.[15] Let $M = \{m \in M^* : m(\mathbf{1}) = 1\}$ and conclude that $P = \bigcap_{m \in M}\{p : x \cdot p \geq 0\}$.

Step 2: Verifying that all elements of M are countably additive, and that $m(\{T, \ldots\}) > 0$ for all $m \in M$.

[14]One needs to verify that P is weakly closed with respect to the pairing $(\ell_\infty, \mathrm{ba}(\mathbf{N}))$, but it is by Theorem 5.86 since $(ba)(\mathbf{N})$ are the ℓ_∞ continuous linear functionals by Aliprantis and Border (1999), Theorem 12.28.

[15]Otherwise, we would have $m(x) = 0$ for all $x \in [0, \mathbf{1}]$, which would imply $m = 0$.

We show now that each $m \in M$ is countably additive. Since for all $\theta \in [0,1)$, there is T so that $(\underbrace{1-\theta,\ldots,1-\theta}_{T \text{ times}}, -\theta, -\theta, \ldots) \in P$ (property (5)), it follows that for all $m \in M$, $m(\{0,\ldots,T-1\}) \geq \theta$. Conclude that $\lim_{t\to\infty} m(\{0,\ldots,t\}) = m(\mathbf{N})$, so that countable additivity is satisfied.[16] So we write $m(z) = m \cdot z$.

Since $(\underbrace{0,\ldots,0}_{T \text{ times}}, 1, \ldots) \in \text{int}(P)$ (property (6)), we can conclude that $m(\{T,\ldots\}) > 0$ for all $m \in M$.

Step 3: Establishing that M is weakly compact Countably additive and nonnegative set functions can be identified with elements of ℓ_1, so we can view M as a subset of ℓ_1. We show that M is weakly compact, under the pairing (ℓ_1, ℓ_∞).

We first show that M is *tight* as a collection of measures: for all $\varepsilon > 0$ there is a compact (finite) set $E \subseteq \mathbf{N}$ such that $m(E) > 1 - \varepsilon$ for all $m \in M$. So let $\varepsilon > 0$ and $\theta' \in (1-\varepsilon, 1)$. Then we know that there is T such that

$$(\underbrace{1-\theta',\ldots,1-\theta'}_{T \text{ times}}, -\theta', \ldots) \in P.$$

The set $E = \{0,\ldots,T-1\}$ works in the definition of tightness because for every $m \in M$, we have $m(\{0,\ldots,T-1\}) \geq \theta' > 1 - \varepsilon$.

The weak compactness of M then follows from a few simple identifications. Denote the set of countably additive probability measures on $\mathbf{N}$ by $\mathcal{P}(\mathbf{N})$, and the set of nonnegative summable sequences which sum to 1 by $\mathbf{1}(\mathbf{N})$. Observe that the weak* topology on $P(\mathbf{N})$ induced by the pairing $(\ell_\infty, P(\mathbf{N}))$ coincides with the weak topology on $\mathbf{1}(\mathbf{N})$ induced by the pairing $(\mathbf{1}(\mathbf{N}), \ell_\infty)$, when in the second instance we identify each $m \in P(\mathbf{N})$ with an element of $\mathbf{1}(\mathbf{N})$. By Lemma 14.21 of Aliprantis and Border (1999), since M is tight, its closure is compact in the first topology (and hence the second). But M is already closed, as the intersection of a collection of closed sets.[17] Therefore, we know that every net in M has a subnet which converges in the weak topology on $\mathbf{1}(\mathbf{N})$. Viewing now M as a subset of ℓ_1, we know that every net in M has a convergent subnet in the weak topology induced by the pairing (ℓ_1, ℓ_∞), which is what we wanted to show.

[16] For example, see Aliprantis and Border (1999), Lemma 9.9. Suppose $E_k \subset \mathbf{N}$ is a sequence of sets for which $\bigcap_k E_k = \varnothing$ and $E_{k+1} \subseteq E_k$. Then for each k, there is $t(k) \in \mathbf{N}$ such that $E_k \subseteq \{t(k), t(k)+1, \ldots\}$ and for which $t(k) \to \infty$. Without loss, take t to be nondecreasing. The result then follows as $m(E^k) \leq m(\{t(k), t(k)+1, \ldots, \}) \to 0$.

[17] Namely, the sets $\{m : p \cdot m \geq 0\}$ for $p \in P$ and $\{m : \mathbf{1} \cdot p = 1\}$.

Step 4: Characterizing exposed points of M. A point of M is *exposed* if there is a linear functional f with $f(m) < f(m')$ for all $m' \in M \setminus \{m\}$. We now show that any exposed point of M has the form $(1-\delta)(1, \delta, \delta^2, \ldots)$ for some $\delta \in [0,1]$. So, suppose that $m \in M$ is an exposed point. Then there exists $x \in \ell_\infty$ such that $x \cdot m < x \cdot m'$ for all $m' \in M \backslash \{m\}$. Clearly it is without loss to suppose that $x \cdot m = 0$.[18] Since $x \cdot m = 0$, it follows that x is on the boundary of P. Therefore, for any T, $x + (\underbrace{0, \ldots, 0}_{T \text{ times}}, x)$ is also on the boundary of P (property 4). Since $x + (\underbrace{0, \ldots, 0}_{T \text{ times}}, x)$ is on the boundary, it has a supporting hyperplane $m^x \in M^*$ passing through the origin, for which for all $y \in P$,

$$0 = m^x \cdot (x + (\underbrace{0, \ldots, 0}_{T \text{ times}}, x)) \leq m^x \cdot y.^{19}$$

We can choose m^x to be non-constant; so we can take $m^x \in M$. So there is $m^x \in M$ such that $0 = m^x \cdot ((\underbrace{0, \ldots, 0}_{T \text{ times}}, x) + x)$. But observe that, since $x \in P$ and $(\underbrace{0, \ldots, 0}_{T \text{ times}}, x) \in P$, $m^x \cdot x \geq 0$ and $m^x \cdot (0, \ldots, 0, x) \geq 0$. Then $0 = m^x \cdot (\underbrace{0, \ldots, 0}_{T \text{ times}}, x) + m^x \cdot x$ means that $m^x \cdot x = 0$ and $m^x \cdot (0, \ldots, 0, x) = 0$. But $m^x \cdot x = 0$ implies that $m^x = m$, as x was chosen to expose m. In turn, $m^x = m$ implies that $m \cdot (\underbrace{0, \ldots, 0}_{T \text{ times}}, x) = 0$ as well.

Let

$$m^T = \frac{(m(T-1), m(T), m(T+1), \ldots)}{m(\{T-1, \ldots\})} \in \ell_1.$$

(recall that we established that $m(\{T-1, \ldots\}) > 0$.) We shall first show that $m^T \in M$. Let $p \in P$. It is enough to show that $(\underbrace{0, \ldots, 0}_{T \text{ times}}, p) \in P$, as $m^T \cdot p = m \cdot (0, \ldots, 0, p) \geq 0$ and $p \in P$ is arbitrary. So let $0 \leq c = \inf\{p \cdot m' : m' \in M\}$, and note that $0 = \inf\{\cdot (p - c\mathbf{1}) : m' \in M\}$, the infimum being achieved at some $m' \in M$ by compactness of M. Then $p - c\mathbf{1} \in \mathrm{bd}(P)$. Property (4) implies that $(0, \ldots, 0, p - c\mathbf{1}) \in P$. Property (3) implies that $(0, \ldots, 0, p) \in P$.

Now, $m^T \cdot x = 0$ and x exposes m, so $m^T \in M$ implies that $m = m^T$. This equation ($m^T = m$ for all T) characterizes the geometric distribution: Let $h(s) = m(\{s, s+1, \ldots\})$.

[18] If $x \cdot m > 0$, observe that $x - (x \cdot m)\mathbf{1}$ satisfies $0 = (x - x \cdot m\mathbf{1}) \cdot m < (x - x \cdot m\mathbf{1}) \cdot m'$.

[19] That it has a supporting hyperplane follows from Aliprantis and Border (1999), Lemma 5.78. That the supporting hyperplane passes through zero follows as P is a cone. That m^x is in the polar cone to P follows by definition.

Then we have

$$\frac{h(s+t)}{h(t)} = \frac{m(\{t+s, t+s+1, \ldots\})}{m(\{t, t+1, \ldots\})}$$
$$= m(\{s, s+1, \ldots\}) = h(s).$$

Then we obtain $h(t) = h((t-1)+1) = h(t-1)h(1)$. Continuing by induction $h(t) = h(1)^t$. If we let $\delta = h(1) = m^*(\{1, 2, \ldots\})$, we have $m^*(\{t, \ldots\}) = \delta^t$ for all $t \geq 1$, and $m^*(\{0\}) = 1 - m^*(\{1, \ldots\}) = 1 - \delta$. Finally, observe $\delta > 0$ as $m(\{T, \ldots\}) > 0$ for all T.

So, conclude that each exposed point of M takes the form $(1-\delta)(1, \delta, \delta^2, \ldots)$ for some $\delta > 0$ (and clearly $\delta < 1$).

Step 5: Finalizing the characterization

Since we have established that M is weakly compact, a theorem of Lindenstrauss and Troyanski ensures that it is the weakly closed convex hull of its strongly exposed points (see Corollary 5.18 of Benyamini and Lindenstrauss (1998)); and, in particular then, of its exposed points. This then allows us to conclude that P has the desired form; let D denote the set of associated discount factors. By Lemma 7, we may take D to be closed. Moreover, $0 \notin \delta$, since for any $m \in M$ and any T, $m(\{T, \ldots,\}) > 0$. $\square$

7.1 Proof of Theorem 4

Let us denote $\{x : x \succeq 0\}$ by $\mathcal{U}(0)$. The theorem follows from an application of Lemma 8.

Lemma 9. *The set $\mathcal{U}(0)$ satisfies all of the properties listed in Lemma 8.*

Proof. Verification of most of these properties is simple. That $\mathcal{U}(0)$ is a closed convex cone follow from continuity, convexity, and homotheticity of $\succeq$. That $\ell_\infty^+ \subseteq \mathcal{U}(0)$ follows from monotonicity and continuity of $\succeq$. That there is $p \notin \mathcal{U}(0)$ follows from monotonicity, as $0 \succ -\mathbf{1}$.

Let us now show property 4 of Lemma 8, that $x \in \text{bd}(\mathcal{U}(0))$ implies $(0, \ldots, 0, x) \in \mathcal{U}(0)$ and $x + (0, \ldots, 0, x) \in \text{bd}(\mathcal{U}(0))$. Observe that, by continuity and monotonicity, $x \in \text{bd}(\mathcal{U}(0))$ if and only if $x \sim 0$: If $x \sim 0$, then for any ϵ, $x + \epsilon\mathbf{1} \succ x$ and $x \succ x - \epsilon\mathbf{1}$, so $x \in \text{bd}(\mathcal{U}(0))$. On the other hand, if $x \in \text{bd}(\mathcal{U}(0))$, then any open ball about x intersects both $\{y : y \succeq 0\}$ and $\{y : 0 \succ y\}$, so it follows by continuity that $x \sim 0$. So, to establish that property 4 holds, let $x \in \text{bd}(\mathcal{U}(0))$. Then $x \sim 0$, so weak stationarity implies that

$(0,0,\ldots,0,x) \sim 0$, and $(1/2)x + (1/2)(0,\ldots,0,x) \sim 0$. Then $(0,\ldots,0,x) \in \mathcal{U}(0)$, and, using homotheticity, $x + (0,\ldots,0,x) \sim 0$, so $x + (0,\ldots,0,x) \in \mathrm{bd}(\mathcal{U}(0))$.

Now turn to property 5. We show that for all $\theta \in [0,1)$, there is T so that

$$(\underbrace{1-\theta,\ldots,1-\theta}_{T \text{ times}}, -\theta, -\theta, \ldots) \in \mathcal{U}(0).$$

Suppose false, so that (by using c-additivity), there is some $\theta \in [0,1)$ such that for all T, $\theta \succ (\underbrace{1,\ldots,1}_{T \text{ times}}, 0, \ldots)$. Then monotone continuity implies $\theta \succeq \mathbf{1}$, contradicting monotonicity.

Finally, property 6 follows from compensation. For all T,

$$(\underbrace{\underline{\theta}^t - \theta^t, \ldots, \underline{\theta}^t - \theta^t}_{t \text{ times}}, \bar{\theta}^t - \theta^t, \ldots) \succeq 0$$

(using c-translation invariance). So monotonicity of $\succeq$ and $\underline{\theta}^t < \theta^t$ implies that $(0,\ldots,0,\bar{\theta}^t - \underline{\theta}^t, \ldots) \succ 0$. Homotheticity of $\succeq$ then implies that $(\underbrace{0,\ldots,0}_{T \text{ times}}, 1, \ldots) \succ 0$. Property 6 then follows from the continuity of $\succeq$. $\square$

We proceed to proving the theorem.

Let D be the set of discount factors provided by Lemma 8 for $\mathcal{U}(0)$. We claim that the function U defined by $U(x) = \min_{\delta \in D}(1-\delta)\sum_t \delta^t x_t$ represents $\succeq$. To this end, we first establish that $U(x) = 0$ if and only if $x \sim 0$. To see this, suppose $U(x) = 0$. Then, by definition of U, $x \in \mathcal{U}(0)$; thus $x \succeq 0$. To establish that $x \sim 0$ we rule out that $x \succ 0$. So suppose that $x \succ 0$. Then we would have by continuity of $\succeq$ that there is $\epsilon > 0$ small so that $x \succeq \epsilon\mathbf{1}$. But then $x - \epsilon\mathbf{1} \in \mathcal{U}(0)$ (by c-translation invariance), so that $U(x - \epsilon\mathbf{1}) \geq 0$, and then clearly $U(x) \geq \epsilon > 0$, a contradiction. So $U(x) = 0$ implies $x \sim 0$.

Conversely, suppose that $x \sim 0$. It follows that $x \in \mathcal{U}(0)$, from which we obtain $U(x) \geq 0$. If in fact $U(x) > 0$, then let $\varepsilon > 0$ be such that $U(x) \geq \epsilon$, and hence $U(x - \epsilon\mathbf{1}) \geq 0$. Thus $x - \epsilon\mathbf{1} \in \mathcal{U}(0)$, so that $x \succ x - \epsilon\mathbf{1} \succeq 0$, or $x \succ 0$, a contradiction.

So now let $x \in \ell_\infty$ be arbitrary. We claim that $x \sim U(x)\mathbf{1}$. But this follows directly from c-translation invariance, as $U(x - U(x)\mathbf{1}) = 0$ if and only if $x - U(x)\mathbf{1} \sim 0$ if and only if $x \sim U(x)\mathbf{1}$.

Finally the result follows a classical textbook argument; if $x \succeq y$, $x \sim U(x)\mathbf{1}$, and $y \sim U(y)\mathbf{1}$, it must be that $U(x) \geq U(y)$, otherwise we would have $U(y)\mathbf{1} \gg U(x)\mathbf{1}$, and hence $U(y)\mathbf{1} \succ U(x)\mathbf{1}$ by monotonicity. Conversely if $U(x) \geq U(y)$, we have by monotonicity that $U(x)\mathbf{1} \succeq U(y)\mathbf{1}$, so that $x \succeq y$.

The necessity of the axioms is straightforward and omitted. We only provide the calculations showing that the representation satisfies stationarity. Let $x \sim \theta$, so $\theta = U(x) = \min_{\delta \in D}(1-\delta)\sum_t \delta^t x_t$, where the minimum is achieved for some $\delta \in D$.

Let $z = \lambda x + (1-\lambda)(\theta, \ldots, \theta, x)$. Then for any δ

$$(1-\delta)\sum_t \delta^t z_t = \lambda(1-\delta^T)\theta + [\lambda + (1-\lambda)\delta^T](1-\delta)\sum_t \delta^t x_t.$$

But $\theta = \lambda(1-\delta^T)\theta + [\lambda + (1-\lambda))\delta^T]\theta$, so for $\delta \in D$, $(1-\delta)\sum_t \delta^t z_t \geq \theta$ if and only if $(1-\delta)\sum_t \delta^t x_t \geq \theta$. A similar statement holds for equalities. This implies that $U(z) = \theta$.

7.2 Proof of Theorem 3

We establish the sufficiency of the axioms first. Let $P = \{x \in \ell_\infty : x \succeq 0\}$. Translation invariance implies that $x \succeq y$ iff $x - y \succeq 0$. So $x \succeq y$ iff $x - y \in P$. If we can show that P satisfies the conditions of Lemma 8 then we are done, because if $D \subseteq (0,1)$ is as delivered by the lemma, then $x \succeq y$ iff $x - y \in P$ iff $\forall \delta \in D$ $\sum_{t=0}^{\infty}(1-\delta)\delta^t(x_t - y_t) \geq 0$.

Lemma 10. *The set P satisfies all of the properties listed in Lemma 8.*

Proof. We have $x \in P$ iff $\lambda x \succeq 0$ for any $\lambda > 0$ by homotheticity. So $x \in P$ implies that $\lambda x \in P$ for any $\lambda > 0$. Hence P is a cone. P is closed since $\succeq$ is continuous. That P is convex follows from the convexity of $\succeq$.

Monotonicity of $\succeq$ implies that the set of positive vectors is contained in P (property 3) and that $-\mathbf{1} \notin P$, so property 2 is satisfied.

Let $x \in \mathrm{bd}(P)$ and $T > 0$. Strong stationarity of $\succeq$ implies that $(\underbrace{0, \ldots, 0}_{T \text{ times}}, x) \in P$. So $x + (\underbrace{0, \ldots, 0}_{T \text{ times}}, x) \in P$ because P is a convex cone. To show that $x + (\underbrace{0, \ldots, 0}_{T \text{ times}}, x) \in \mathrm{bd}(P)$, let $\varepsilon > 0$ and x' be such that $\|x - x'\|_\infty < \varepsilon/2$ and $x' \notin P$. Note that

$$\|x + (0, \ldots, 0, x) - x' + (0, \ldots, 0, x')\|_\infty < \varepsilon.$$

We claim that $x' + (0, \ldots, 0, x') \notin P$. So suppose that $x' + (0, \ldots, 0, x') \in P$. Then $(1/2)x' + (1/2)(0, \ldots, 0, x') \in P$ as P is a cone. Thus $(1/2)x' + (1/2)(0, \ldots, 0, x') \succeq 0$, which by strong stationarity implies that $x' \succeq 0$, contradicting that $x' \notin P$.

The proof that P satisfies properties 5 and 6 are the same as the corrresponding proofs in Lemma 9 (in Lemma 9 we use c-translation invariance, which is weaker than what we assume for Theorem 3). $\square$

Now we turn to the necessity of the axioms. It is clear that continuity at infinity is necessary, as $\theta \geq (1-\delta)\sum_{t=0}^{T} x_t$ for all $\delta \in D$ implies that $\theta \geq (1-\delta)\sum_{t=0}^{\infty} x_t$ for all $\delta \in D$. Compensation is also a simple consequence of D being bounded away from 1.

Lemma 11. *Stationarity is necessary.*

Proof. Let $t > 0$ and $\lambda \in [0, 1]$. Let $z = \lambda x + (1-\lambda)(\underbrace{\theta, \ldots, \theta}_{t \text{ times}}, x) - \theta\mathbf{1}$. Then for any $\delta \in (0, 1)$

$$\begin{aligned}\sum_{\tau=0}^{\infty} \delta^\tau z_\tau &= \lambda \sum_{\tau=0}^{\infty} \delta^\tau (x_\tau - \theta) + (1-\lambda) \sum_{\tau=t}^{\infty} \delta^\tau (x_{\tau-t} - \theta) \\ &= [\lambda + (1-\lambda)\delta^t] \sum_{\tau=0}^{\infty} \delta^\tau (x_\tau - \theta)\end{aligned}$$

Note that $[\lambda + (1-\lambda)\delta^t] > 0$. So $(1-\delta)\sum_{\tau=0}^{\infty} \delta^\tau z_\tau$ for all $\delta \in D$ iff $(1-\delta)\sum_{\tau=0}^{\infty} \delta^\tau (x_\tau - \theta) \geq 0$ for all $\delta \in D$. $\square$

7.3 Uniqueness

The uniqueness argument is common to Theorem 3 and Theorem 4, so we put its proof here:

Proof. By Lemma 7, $m(D)$ and $m(D')$ are closed, as the continuous image of compact sets. Let M and M' be the closed convex hulls of $m(D)$ and $m(D')$, respectively. If

$\delta \in D' \setminus D$ then $m(\delta) \notin M$ (because no $m(\delta)$ can be written as a convex combination of some finite $m(\delta_1), \ldots, m(\delta_n)$).

Topologize $\Delta(\mathbf{N})$ with the weak*-topology on $\sigma(C_b(\mathbf{N}), \Delta(\mathbf{N}))$; that is, the weakest topology for which the map $\mu \mapsto x \cdot \mu$ is continuous for every $x \in C_b(\mathbf{N})$ (observe also that any such $x \in l_\infty$). By Lemma 14.21 of Aliprantis and Border (1999), each of M and M' is compact.

Since $m(\delta) \notin M$, there is a continuous linear functional x separating $m(\delta)$ from M (Theorem 5.58 of Aliprantis and Border (1999)). By Lemma 14.4 and Theorem 5.83 of Aliprantis and Border (1999), there is $x \in l_\infty$ for which $x \cdot m(\delta) < \inf_{m' \in M} x \cdot m'$. Let $y \in \mathbf{R}$ be given by $y = \frac{x \cdot m(\delta) + \inf_{m' \in M} x \cdot m'}{2}$ and observe that $(x - y) \cdot m(\delta) < 0 < \inf_{m' \in M} (x - y) \cdot m'$. Conclude that $0 \succ (x - y)$ and $(x - y) \succ' 0$. $\square$

References

Aigner, M. (2007): *A Course in Enumeration*. Springer.

Aliprantis, C. D., and K. C. Border (1999): *Infinite dimensional analysis: a hitchhiker's guide*. Springer, second edition.

Arrow, K. J. (1974): *Essays in the theory of risk-bearing*. North-Holland Amsterdam.

Benyamini, Y., and J. Lindenstrauss (1998): *Geometric Nonlinear Functional Analysis*, no. pt. 1 in American Mathematical Society colloquium publications. American Mathematical Soc.

Bernstein, S. N. (1915): "Sur la représentation des polynômes positifs," *Har'k. Mat. Obshch*, 14(5), 227–228.

Bertsimas, D., I. Popescu, and J. Sethuraman (2000): "Moment problems and semidefinite optimization," in *Handbook of semidefinite programming*, ed. by H. Wolkowicz, R. Saigal, and L. Vandenberghe, pp. 469–509. Springer.

Bewley, T. F. (2002): "Knightian decision theory. Part I," *Decisions in Economics and Finance*, 25(2), 79–110.

Billot, A., A. Chateauneuf, I. Gilboa, and J.-M. Tallon (2000): "Sharing Beliefs: Between Agreeing and Disagreeing," *Econometrica*, 68(3), 685–694.

Chateauneuf, A., F. Maccheroni, M. Marinacci, and J.-M. Tallon (2005): "Monotone continuous multiple priors," *Economic Theory*, 26(4), 973–982.

Debreu, G. (1954): "Representation of a preference ordering by a numerical function," in *Decision Processes*, ed. by R. Thrall, R. L. Davis, and C. H. Coombs, pp. 159–165. Wiley.

Gilboa, I., and D. Schmeidler (1989): "Maxmin expected utility with non-unique prior," *Journal of Mathematical Economics*, 18(2), 141–153.

Hara, C. (2008): "Complete monotonicity of the representative consumer's discount factor," *Journal of Mathematical Economics*, 44(12), 1321–1331.

Hausdorff, F. (1921): "Summationsmethoden und momentfolgen. I," *Mathematische Zeitschrift*, 9, 74–109.

Hayashi, T. (2016): "Consistent updating of social welfare functions," *Social Choice and Welfare*, 46, 1–40.

Higashi, Y., K. Hyogo, and N. Takeoka (2009): "Subjective random discounting and intertemporal choice," *Journal of Economic Theory*, 144(3), 1015 – 1053.

Higashi, Y., K. Hyogo, H. Tanaka, and N. Takeoka (2016): "Comparative impatience under random discounting," *Economic Theory*, p. forthcoming.

Holmes, R. B. (1975a): *Geometric functional analysis and its applications*. Springer.

——— (1975b): *Geometric functional analysis and its applications*. Springer-Verlag New York.

Huber, P. J. (1981): *Robust statistics*. Springer.

Jackson, M. O., and L. Yariv (2015): "Collective dynamic choice: the necessity of time inconsistency," *American Economic Journal: Microeconomics*, 7(4), 150–178.

Karni, E., and I. Zilcha (2000): "Saving behavior in stationary equilibrium with random discounting," *Economic Theory*, 15(3), 551–564.

Koopmans, T. C. (1960): "Stationary ordinal utility and impatience," *Econometrica*, 28(2), 287–309.

Loewenstein, G. (1987): "Anticipation and the valuation of delayed consumption," *The Economic Journal*, 97(387), 666–684.

Lu, J., and K. Saito (2016): "Random Intertemporal Choice," mimeo, Caltech.

Pennesi, D. (2015): "Uncertain discount and hyperbolic preferences," Universite de Cergy-Pontoise, THEMA Working Paper.

Rigotti, L., and C. Shannon (2005): "Uncertainty and Risk in Financial Markets," *Econometrica*, 73(1), 203–243.

Shafer, G. (1976): *A mathematical theory of evidence*, vol. 1. Princeton University Press, Princeton.

Shohat, J. A., and J. D. Tamarkin (1943): *The problem of moments*, no. 1. American Mathematical Soc.

Stern, N. (2007): *The Economics of Climate Change: the Stern Review*. Cambridge University Press.

Varian, H. R. (2006): "Recalculating the Costs of Climate Change," *New York Times*.

Villegas, C. (1964): "On Qualitative Probability σ-Algebras," *The Annals of Mathematical Statistics*, pp. 1787–1796.

Wakai, K. (2008): "A Model of Utility Smoothing," *Econometrica*, 76(1), 137–153.

Weitzman, M. L. (2001): "Gamma Discounting," *American Economic Review*, 91(1), 260–271.