



ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna  
DEPARTMENT OF ECONOMICS

## **Intertemporal discrete choice**

Daniele Pennesi

*Quaderni - Working Paper DSE N°1061*



# Intertemporal discrete choice\*

Daniele Pennesi<sup>†</sup>

February 2016

## Abstract

The discounted logit is widely used to estimate time preferences using data from field and laboratory experiments. Despite its popularity, it exhibits the "problem of the scale": choice probabilities depend on the scale of the value function. When applied to intertemporal choice, the problem the scale implies that logit probabilities are sensitive to the temporal distance between the choice and the outcomes. This is a failure of an intuitive requirement of stationarity although future values are discounted geometrically. As a consequence, patterns of choice following from the structure of the logit may be attributed to non-stationary discounting. We solve this problem introducing the discounted Luce rule. It retains the flexibility and simplicity of the logit while it satisfies stationarity. We characterize the model in two settings: dated outcomes and consumption streams. Relaxations of stationarity give observable restrictions characterizing hyperbolic and quasi-hyperbolic discounting. Lastly, we discuss an extension of the model to recursive stochastic choices with the present bias.

KEYWORDS: Discrete Choice, Intertemporal Choice, Quasi-hyperbolic Discounting

JEL CLASSIFICATION: D01

---

\*Some results contained in the current work appeared in a paper circulated under the title "The intertemporal Luce rule".

<sup>†</sup>University of Bologna, Department of Economics, Piazza Scaravilli 2, 40126, Bologna (Italy). Email: [daniele.pennesi@unibo.it](mailto:daniele.pennesi@unibo.it)

# 1 Introduction

The discounted multinomial logit (DML) is the most common model of probabilistic choice among those used in the estimation of time preference (e.g. [Chabris et al., 2008](#); [Louie and Glimcher, 2010](#); [Tanaka et al., 2010](#)). Despite its popularity, the DML may be inappropriate to study intertemporal choice due to the "problem of the scale": choice probabilities depend on the scale of the value function ([Fosgerau and Bierlaire, 2009](#)). Consider the probability of choosing now an outcome  $x$  at time  $t$  over an alternative  $y$  at time  $s$ .<sup>1</sup> Should this probability remain unchanged when both rewards are equally delayed? A positive answer is motivated by an intuitive form of stationarity borrowed from deterministic choice ([Fishburn and Rubinstein, 1982](#)): choice probabilities should be *independent* of the temporal distance between the choice and the dates of the rewards. The only quantity that matters is the relative time distance between the rewards, i.e.  $|t - s|$ . Therefore, a stationary stochastic choice rule should assign the same probability to the choice of  $(x, t)$  over  $(y, s)$  and the choice of  $(x, t + r)$  over  $(y, s + r)$ . Although the discounted multinomial logit discounts future values *geometrically*, it fails to satisfy an even weaker version of this intuitive form of stationarity<sup>2</sup> (see Section 2). Choice probabilities in the discounted multinomial logit do *depend* on the temporal distance between the choice and the delivery of the rewards. As a consequence, the estimation of time preferences may be distorted. For example, attributing to quasi-hyperbolic discounting patterns of choice probabilities that follows from the structure of the discounted multinomial logit, rather than violations of geometric discounting (see Section 2).

We propose a new model, the discounted Luce rule, that solves the stationarity problem while maintaining the flexibility and simplicity of the discounted logit. The model is completely characterized through testable restrictions on choice probabilities. Differently from the discounted logit, it allows a stark identification of the observational consequences of hyperbolic and quasi-hyperbolic discounting, since they correspond to certain violations of stationarity. We axiomatize the model in two settings: dated rewards and consumption streams. Discrete choice over dated rewards  $(x, t)$ , meaning

---

<sup>1</sup>According to the DML this probability is  $P_{Logit}((x, t) \geq (y, s)) = \frac{e^{\delta^t v(x)}}{e^{\delta^t v(x)} + e^{\delta^s v(y)}}$ .

<sup>2</sup>More precisely, in the discounted multinomial logit the ratio between the probability of choosing  $(x, t)$  over  $(y, s)$  is different from the ratio between the probability of choosing  $(x, t + r)$  over  $(y, s + r)$ .

a reward  $x$  delivered at time  $t$ , is interesting for two reasons: first, in this setting all the properties characterizing the model are intuitive and easy to test. Second, choice among dated rewards includes the Multiple Price List method (Coller and Williams, 1999), the workhorse of experimental and field studies devoted to the elicitation of time preferences.<sup>3</sup> In the dated reward setting, the probability of choosing  $(x, t) \in A$  according to the discounted Luce rule is given by the relative discounted value of  $(x, t)$  in  $A$ :

$$P((x, t), A) = \frac{\delta^t v(x)}{\sum_{(y, s) \in A} \delta^s v(y)}$$

with  $\delta \in (0, 1]$  representing the discount factor. The model is completely characterized by stochastic stationarity and stochastic impatience. The former implies that choice probabilities are not sensitive to the temporal distance between the choice and the timing of rewards. Generalizations of the previous rule accounting for quasi-hyperbolic and hyperbolic discounting are provided through a relaxation of the stationarity axiom. Quasi-hyperbolic discounting, for example, predicts that the probability of choosing an immediate reward over a delayed one decreases when both rewards are equally delayed (Section 3.1). The discounted logit may generate the same behavior while discounting future values geometrically.

The more general setting of consumption streams,  $\mathbf{x} = (x_0, x_1, \dots, x_T)$ , allows us to compare the properties characterizing the discounted Luce rule to axiomatizations in the deterministic setting of additively separable discounted utility with geometric (Koopmans, 1960) or quasi-hyperbolic discounting (Hayashi, 2003; Montiel Olea and Strzalecki, 2014). In this case, the probability of choosing a consumption stream  $\mathbf{x}$  from a set  $A$  of alternatives is given by the relative present value of  $\mathbf{x}$  in  $A$ :

$$P(\mathbf{x}, A) = \frac{\sum_{t=0}^T \delta^t v(x_t)}{\sum_{\mathbf{y} \in A} \sum_{t=0}^T \delta^t v(y_t)}$$

Beyond an intuitive requirement of stationarity adapted to this setting, the main novelty characterizing the previous rule is a property called *separability*: the relative probability of choosing a consumption stream  $\mathbf{x}$  over  $\mathbf{y}$  in a set  $A$  is equal to the sum

---

<sup>3</sup>Among the others: Harrison et al. (2002) Andersen et al. (2008), Tanaka et al. (2010), Halevy (2015).

of the relative probabilities of choosing its "components", properly defined (see Section 4). Separability is the stochastic counterpart of additive separability in deterministic choice. Once stationarity is relaxed, we give observable conditions to characterize quasi-hyperbolic discounting:

$$P(\mathbf{x}, A) = \frac{v(x_0) + \beta \sum_{t=1}^T \delta^t v(x_t)}{\sum_{\mathbf{y} \in A} \left[ v(y_0) + \beta \sum_{t=1}^T \delta^t v(y_t) \right]}$$

The properties are comparable to the ones used in deterministic choice (e.g. in [Montiel Olea and Strzalecki, 2014](#)). The restrictions characterizing stochastic choice with quasi-hyperbolic discounting can inform applied works estimating time preferences from field data (for example durable good adoptions, [Chevalier and Goolsbee, 2009](#); [Dubé et al., 2014](#)) and works estimating recursive models of stochastic choice with quasi-hyperbolic discounting (see Section 6). These are structural econometric models that retain the recursive dynamic structure of [Rust \(1987\)](#), but allow quasi-hyperbolic discounting of the continuation value. Their behavioral characterization has not been studied yet. Only recently, [Fudenberg and Strzalecki \(2015\)](#); [Matějka et al. \(2015\)](#) studied the particular case of geometric discounting and logit choice probabilities. Therefore, our results can be helpful in understanding the implicit restrictions entailed by the use of such models. In the recursive quasi-hyperbolic discounted model, the individual stochastically selects in each period, an immediate consumption  $x_t$  and a continuation plan  $A_{t+1}$ . The value of an action at time  $t$ ,  $a_t = (x_t, A_{t+1})$  is equal to:

$$U_t(x_t, A_{t+1}) = v(x_t) + \beta \delta \mathbb{E} \left[ \max_{a_{t+1} \in A_{t+1}} U_{t+1}(a_{t+1}) + \epsilon_{a_{t+1}} \right]$$

and the choice from the menu at time  $t$  is probabilistic:

$$P_t(a_t, A_t) = \text{Prob} \left( U_t(a_t) + \epsilon_{a_t} \geq \max_{b_t \in A_t} U_t(b_t) + \epsilon_{b_t} \right)$$

The individual correctly anticipates the shocks to her future utility and discounts the continuation value quasi-hyperbolically. Non-experimental data on job search ([Paserman, 2008](#)), insecticide treated nets' adoption ([Tarozzi and Mahajan, 2011](#)), drug

compliance (An et al., 2014), mammography decisions (Fang and Wang, 2015) and cellphone usage (Yao et al., 2012), have been used to separately identify time preferences  $\delta$  and the present bias factor  $\beta$ . The recursive Luce rule corresponds to the previous model when the error terms are i.i.d. and distributed according to a Fréchet (see Section 2.1). We show a simple application of the recursive Luce rule with quasi-hyperbolic discounting to the purchase of a durable good and we study the elasticity of demand with respect to a transitory or a permanent price shock. We also discuss a possible extension of the model that relaxes the IIA axiom and retains the stationarity properties. Allowing the dependence of the value function on a parameter, as in the mixed logit model, the discounted mixed Luce rule has the same stationary properties of the Luce rule and can accommodate realistic substitution patterns among elements that are excluded by the IIA. Lastly, we study the elasticity and cross-elasticity of choice probability when an element of a consumption stream varies. We find that choice probabilities of the discounted Luce rule contain relevant information concerning elasticities and cross-elasticities. For example, the probability of choosing a consumption stream in a set is inversely related to the sum of the elasticities of its components.

The paper is organized as follows: after a review of the relevant literature, Section 2 introduces the shortcoming of using the discounted multinomial logit. In Section 3 we provide an axiomatization of the discounted Luce rule when choices are over dated outcomes. We then relax stationarity to pin down the observable restrictions of hyperbolic and quasi-hyperbolic discounting. In Section 4 we extend the model to consumption streams. Elasticity and cross-elasticity of choice probabilities is studied in Section 5. Section 6 discusses an extension of the model to recursive stochastic choice. Section 7 proposes a version of the model that relaxes the independence of irrelevant alternatives assumption.

## 1.1 Related literature

In the static setting, the multinomial logit and the Luce choice rule are equivalent. The latter has been introduced by Luce (1959). Recent works provided various foundations for the static multinomial logit based on bounded rationality (Mattsson and

Weibull, 2002), rational inattention (Matějka and McKay, 2015) and neuroscientific models of choice (Webb, 2015). Two recent models generalize the Luce rule to account for violations of the IIA: Gul et al. (2014) proposed a model similar to the nested logit, Echenique et al. (2014) included perception priorities. The axioms we introduce in the present work can be extended to their model, since stationarity and the IIA are not related. Concerning the dynamic setting, Fudenberg and Strzalecki (2015) axiomatized a general version of the recursive stochastic choice model in which larger menus may be disliked due to choice aversion. The individual stochastically chooses at each time an action and a continuation menu. The dynamic logit, widely used in applied works<sup>4</sup> (Rust, 1987; Hendel and Nevo, 2006; Gowrisankaran and Rysman, 2012; Chen et al., 2013), is a particular case of their model. The aim of their paper is different from ours since, we are interested in the effect of discount on choice probabilities and we focus on "static stochastic choice" over consumption streams. They are interested in the dynamics of stochastic choice. However, the notion of stationarity they use is comparable to ours. We show that it is weaker and it cannot distinguish the discounted logit from the discounted Luce rule (see Section 4). The axiomatic characterization of the quasi-hyperbolic Luce rule of Theorem 3 can be used to characterize (for example, adapting the axioms of Fudenberg and Strzalecki (2015)) a recursive model of stochastic choice, that allows for quasi-hyperbolic discounting. Such model is receiving an increasing attention in applied works (e.g. Paserman, 2008; Tarozzi and Mahajan, 2011; Yao et al., 2012; An et al., 2014; Fang and Wang, 2015). The interaction of discounting and stochastic choice has been understudied so far. Recently, Lu and Saito (2016) introduced a model where stochastic choice follows uncertainty about the discount function. They characterize geometric and quasi-hyperbolic discounting. Concerning critiques to the multinomial logit, Fosgerau and Bierlaire (2009) proposed a random utility model with multiplicative error that solves the scale problem. A more general critique of the use of stochastic choice models in the study of risk aversion and time preference comes from Apesteguía and Ballester (2015). They show that a large class of models including the logit is not monotone with respect to parameters measuring risk aversion or impatience. In other words, an increase in the risk aversion or impatience parameter is not necessarily followed by a larger probability of selecting a

---

<sup>4</sup>See (Aguirregabiria and Mira, 2010) for a review.

less risky or more impatient option. Our critique of the multinomial logit is different since it is not related to parametric restrictions. Moreover, the model we introduce in Section 7 belongs to the class of random parameter models proposed by [Apesteguiá and Ballester \(2015\)](#) as a solution to the monotonicity problem. In general, both papers highlight substantial flaws of the logit model that should be taken into account in applied works.

## 2 The "problem of the scale" and its consequences

Consider a multinomial logit model of choice from a general set  $A$  and two "value functions"  $u$  and  $u'$  with  $u' = ku$  for some  $k \geq 0$ , the "problem of the scale" ([Fosgerau and Bierlaire, 2009](#)) comes from the following inequality:

$$P_{Logit}^{u'}(a, A) = \frac{e^{ku(a)}}{\sum_{b \in A} e^{ku(b)}} \neq \frac{e^{u(a)}}{\sum_{b \in A} e^{u(b)}} = P_{Logit}^u(a, A)$$

The choice probabilities are sensitive to the scale of the value function. The problem is particularly relevant when we consider the discounted logit. Assume that the elements of  $A$  are dated outcomes and the individual has to choose today between a dated outcome  $(x, t)$ , meaning  $x$  delivered/payed at time  $t$ , and an alternative  $(y, s)$  and that  $u((x, t)) = \delta^t v(x)$ . For simplicity, if  $A = \{(x, t), (y, s)\}$ , we write  $P((x, t), A) = P((x, t) \geq (y, s))$ . The scale problem produces the following:

$$P_{Logit}((x, t) \geq (y, s)) = \frac{e^{\delta^t v(x)}}{e^{\delta^s v(y)} + e^{\delta^t v(x)}}$$

and when both payments are delayed by  $r > 0$  periods,

$$P_{Logit}((x, t+r) \geq (y, s+r)) = \frac{e^{\delta^{t+r} v(x)}}{e^{\delta^{s+r} v(y)} + e^{\delta^{t+r} v(x)}}$$

It is easy to see that

$$P_{Logit}((x, t) \geq (y, s)) \neq P_{Logit}((x, t+r) \geq (y, s+r))$$



the probability of choosing a dated outcome changes when *both* elements are *equally* postponed. Since delaying both outcomes by  $r > 0$  periods is equivalent to multiplying the value function by  $\delta^r$ , it is a "rescaling" of the value function  $v$ , and the scale problem implies that the probability of choosing  $(x, t)$  over  $(y, s)$  is different from the probability of choosing  $(x, t + r)$  over  $(y, s + r)$ . Therefore, choice probabilities are affected by the temporal distance between the decision time and the outcomes, even if it discounts future values geometrically.<sup>5</sup> This is a violation of "stationarity" in the sense of [Fishburn and Rubinstein \(1982\)](#) translated to the stochastic choice setting. To see the consequences, assume  $x, y \in \mathbb{R}$ ,  $x \leq y$  and  $v(x) > \delta^s v(y)$  for some  $s > 0$ , then

$$P_{Logit}((x, 0) \geq (y, s)) = \frac{e^{v(x)}}{e^{\delta^s v(y)} + e^{v(x)}} > \frac{e^{\delta^r v(x)}}{e^{\delta^{s+r} v(y)} + e^{\delta^r v(x)}} = P_{Logit}((x, r) \geq (y, s+r))$$

the probability of choosing an immediate/smaller payment over a later/larger payment decreases when both are equally postponed. This is the typical violation of stationarity attributed to the present bias ([Laibson, 1997](#)), however, it follows from the structure of the logit and not from the form of discounting, since future values are discounted geometrically. The same problem occurs if one considers discounting in continuous time. For example, [Tanaka et al. \(2010\)](#) estimated the parameters of the general discount using the following logistic function,

$$P_{Logit}((x, 0) \geq (y, t)) = \frac{1}{1 + \exp(x - y\beta(1 - (1 - \theta)kt)^{1/1-\theta})}$$

The discount function  $D(t) \triangleq \beta(1 - (1 - \theta)kt)^{1/1-\theta}$ , where  $k$  is the discount rate,  $\beta$  the present-bias coefficient and  $\theta$  the hyperbolic coefficient, encompasses geometric, hyperbolic and quasi-hyperbolic discounting.<sup>6</sup> Suppose  $\theta \rightarrow 1$  and  $\beta = 1$ , then (see [Tanaka et al., 2010](#)),  $D(t) = e^{-kt}$ , the standard geometric discount. It is immediate to note that, when both  $(x, 0)$  and  $(y, t)$  are postponed by  $r$  periods, the probability:

$$P_{Logit}((x, r) \geq (y, t+r)) = \frac{1}{1 + \exp(e^{-kr}(ye^{-kt} - x))} \neq \frac{1}{1 + \exp(ye^{-kt} - x)} = P_{Logit}((x, 0) \geq (y, t))$$

---

<sup>5</sup>The result is true for the discounted multinomial logit. Given a set  $A$ , assume all payments in  $A$  are postponed by  $r > 0$  periods and define  $A_r = \{(x, t+r) : (x, t) \in A\}$ , the probability of choosing  $(x, t+r)$  from  $A_r$  is  $P_{Logit}((x, t+r), A_r) = \frac{e^{\delta^{t+r} v(x)}}{\sum_{(y, s+r) \in A_r} e^{\delta^{s+r} v(y)}}$  and it is different from  $P_{Logit}((x, t), A) = \frac{e^{\delta^t v(x)}}{\sum_{(y, s) \in A} e^{\delta^s v(y)}}$ .

<sup>6</sup>For  $\theta \rightarrow 1$  and  $\beta = 1$ ,  $D(t) = e^{-kt}$ . For  $\theta = 2$  and  $\beta = 1$ ,  $D(t) = 1/(1 + kt)$ . For  $\theta \rightarrow 1$ ,  $D(t) = \beta e^{-kt}$ .

Again, choice probabilities are sensitive to the distance between the decision time and the payments.

The requirement of equality between the two probabilities  $P_{Logit}((x, t), A)$  and  $P_{Logit}((x, t + r), A_r)$ , where  $A_r = \{(x, t + r) : (x, t) \in A\}$ , may be excessively strong. The previous distortion, however, occurs even if we impose a much weaker condition. Consider the *relative* probability of choosing today  $(x, t)$  over  $(y, s)$ , again:

$$\frac{P_{Logit}((x, t), A)}{P_{Logit}((y, s), A)} = \frac{e^{\delta^t v(x)}}{e^{\delta^s v(y)}} \neq \frac{e^{\delta^{t+r} v(x)}}{e^{\delta^{s+r} v(y)}} = \frac{P_{Logit}((x, t + r), A_r)}{P_{Logit}((y, s + r), A_r)}$$

Though weaker, it is still violated by the discounted logit model. The *relative* probability of choosing  $(x, t)$  from  $A$  changes when *all* the rewards in  $A$  are *equally* delayed. Consider now a present biased individual and  $A = \{(x, 0), (y, t)\}$ . It is plausible to imagine that for a general discounted stochastic choice rule  $P$ :

$$\frac{P((x, 0), A)}{P((y, t), A)} \geq \frac{P((x, r), A_r)}{P((y, t + r), A_r)}$$

Confronted with a choice between a smaller/immediate reward and a bigger/later, the relative probability of choosing the smaller reward is greater when the payment is immediate, then the relative probability of choosing it when both payments are equally delayed. However, the relative probability according to the discounted Logit is:

$$\frac{P_{Logit}((x, 0), A)}{P_{Logit}((y, t), A)} = \frac{e^{v(x)}}{e^{\delta^t v(y)}}$$

When both payments are equally delayed, say by  $r$  periods, the ratio becomes:

$$\frac{P_{Logit}((x, r), A_r)}{P_{Logit}((y, t + r), A_r)} = \frac{e^{\delta^r v(x)}}{e^{\delta^{t+r} v(y)}}$$

but,  $\frac{e^{\delta^r v(x)}}{e^{\delta^{t+r} v(y)}} = \left(\frac{e^{v(x)}}{e^{\delta^t v(y)}}\right)^{\delta^r}$  and  $v(x) \geq \delta^t v(y)$  implies the same violation attributed to quasi-hyperbolic discounting, i.e.  $\frac{P_{Logit}((x, 0), A)}{P_{Logit}((y, t), A)} \geq \frac{P_{Logit}((x, r), A_r)}{P_{Logit}((y, t+r), A_r)}$ . Again, the inequality follows from the structure of the logit model rather than the form of discounting. We believe these distortions should be considered by any work using the discounted logit to estimate time preferences, especially when using the Multiple Price List method.

A different consequence of the lack of stochastic stationarity concerns the asymptotic behavior of the discounted multinomial logit. Consider the case in which all the payoffs in  $A$  are infinitely postponed. In that case:

$$\lim_{r \rightarrow \infty} P_{Logit}((x, t+r), A_r) = \lim_{r \rightarrow \infty} \frac{e^{\delta^{t+r}v(x)}}{\sum_{(y, s+r) \in A_r} e^{\delta^{s+r}v(y)}} = \frac{1}{|A|}$$

the discounted logit predicts that the asymptotic probability of choosing among infinitely delayed payoffs converges to the uniform distribution (see [Fudenberg and Strzalecki, 2015](#)). Asymptotic uniformity is difficult to reconcile with empirical evidence. For example, [Giglio et al. \(2015\)](#) found substantial long-run discount rates (more than 100 years) exploiting the features of the housing markets in the UK and Singapore, suggesting that individuals are sensitive to extremely delayed outcomes. Similarly, the choice of long-term environmental policies or investment in R&D, cannot be explained by the discounted logit. The stationarity property of the discounted Luce rule gives,

$$\lim_{r \rightarrow \infty} P((x, t+r), A_r) = \lim_{r \rightarrow \infty} \frac{\delta^{t+r}v(x)}{\sum_{(y, s+r) \in A_r} \delta^{s+r}v(y)} = P((x, t), A)$$

The probability of choosing an element is independent of the delay when all elements are delayed. A last consideration on the discounted logit concerns its asymptotic behavior when only one dated outcome is infinitely delayed. Suppose  $A = \{(x, t), (y, s)\}$  with  $s \neq t$  and let  $t \rightarrow \infty$ . Then,

$$\lim_{t \rightarrow \infty} P_{Logit}((x, t), A) = \lim_{t \rightarrow \infty} \frac{e^{\delta^t v(x)}}{e^{\delta^t v(x)} + e^{\delta^s v(y)}} = \frac{1}{1 + e^{\delta^s v(y)}} > 0$$

For the discounted logit, the probability of choosing a delayed outcome  $x$  is strictly positive, even when it is infinitely delayed. This is paradoxical if one consider the case of a very small  $\delta \approx 0$ . The probability of choosing an infinitely delayed  $x$ ,  $\lim_{t \rightarrow \infty} P_{Logit}((x, t), A) \approx \frac{1}{2}$ , although future has very low value. Differently, the discounted Luce rule for  $t \rightarrow \infty$ ,  $\lim_{t \rightarrow \infty} P((x, t), A) \rightarrow 0$ , as expected.

The violation of stationarity of the discounted logit persists in the more general setting of consumption streams. If we consider the discounted logit, the probability of

choosing a consumption stream  $\mathbf{x} = (x_0, x_1, \dots, x_T)$  from a set  $A$  is

$$P_{Logit}(\mathbf{x}, A) = \frac{\exp\left(\sum_{t=0}^T \delta^t v(x_t)\right)}{\sum_{\mathbf{y} \in A} \exp\left(\sum_{s=0}^T \delta^s v(y_s)\right)}$$

An intuitive form of stationarity would require that a "shift forward" of consumption should not affect preferences. In other words, the probability of choosing  $\mathbf{x} = (x_0, x_1, \dots, x_{T-1}, z)$  over  $\mathbf{y} = (y_0, y_1, \dots, y_{T-1}, z)$  should not change if we "lag"  $\mathbf{x}$  and  $\mathbf{y}$  by one period (see [Fishburn, 1970](#), Def. 7.3). Namely, it should be equal to the probability of choosing  $\mathbf{x}_{+1} = (z, x_0, x_1, \dots, x_{T-1})$  over  $\mathbf{y}_{+1} = (z, y_0, y_1, \dots, y_{T-1})$ . However, in the discounted logit model the two probabilities differ. To see this, let  $A = \{\mathbf{x}, \mathbf{y}\}$  and  $A_{+1} = \{\mathbf{x}_{+1}, \mathbf{y}_{+1}\}$ , then:

$$P_{Logit}(\mathbf{x}, A) = \frac{\exp\left(\sum_{t=0}^{T-1} \delta^t v(x_t) + \delta^T v(z)\right)}{\exp\left(\sum_{t=0}^{T-1} \delta^t v(x_t) + \delta^T v(z)\right) + \exp\left(\sum_{t=0}^{T-1} \delta^t v(y_t) + \delta^T v(z)\right)}$$

it is equal to

$$P_{Logit}(\mathbf{x}, A) = \frac{1}{1 + \exp\left(\sum_{t=0}^{T-1} \delta^t (v(y_t) - v(x_t))\right)} \quad (1)$$

and  $P_{Logit}(\mathbf{x}_{+1}, A_{+1})$  is equal to

$$P_{Logit}(\mathbf{x}_{+1}, A_{+1}) = \frac{1}{1 + \exp\left(\sum_{t=1}^T \delta^s (v(y_{t-1}) - v(x_{t-1}))\right)} \quad (2)$$

It is easy to see that Eq. (1) and Eq. 2 are different. Lagging by one period both consumption streams  $\mathbf{x}$  and  $\mathbf{y}$  changes the likelihood of choosing them, contradicting an intuitive form of stationarity although discounting is geometric (see [Fishburn, 1970](#), Def. 7.3). Similarly to the dated outcome case, this may affect the elicitation of temporal preferences, for example confounding geometric and quasi-hyperbolic discounting. To the contrary, the discounted Luce rule satisfies stationarity (see Section 4).

## 2.1 Discounted random utility models

The discounted Luce rule is a particular case of the random utility model (Marschak, 1959) and there are two ways to see it. The first has been suggested by Fosgerau and Bierlaire (2009). It is well known that the multinomial logit can be derived from the additive random utility model (ARUM),  $P(a, A) = \text{Prob}(u(a) + \epsilon_a \geq \max_{b \in A} u(b) + \epsilon_b)$ , if the error terms are i.i.d. and distributed according to an extreme value distribution. To solve the scale problem, they proposed a *multiplicative* specification of the error terms, i.e.  $P(a, A) = \text{Prob}(u(a) \cdot \hat{\epsilon}_a \geq \max_{b \in A} u(b) \cdot \hat{\epsilon}_b)$ , with errors distributed according to a transformation of an extreme value distribution. In the case of the discounted Luce rule, this implies

$$P(\mathbf{x}, A) = \text{Prob} \left( \sum_{t=0}^T \delta^t v(x_t) \cdot \epsilon_{\mathbf{x}} \geq \max_{\mathbf{y} \in \mathbf{A}} \sum_{t=0}^T \delta^t v(y_t) \cdot \epsilon_{\mathbf{y}} \right)$$

Fosgerau and Bierlaire (2009) illustrated the practical implementation of the multiplicative error specification. An alternative way follows from Mattsson et al. (2014) and consists in changing the distribution of the error terms in the ARUM. The discounted Luce rule arises assuming error terms that are i.i.d. and distributed according to a Fréchet distribution<sup>7</sup> with  $\theta = 1$ :

$$P(\mathbf{x}, A) = \text{Prob} \left( \sum_{t=0}^T \delta^t v(x_t) + \tilde{\epsilon}_{\mathbf{x}} \geq \max_{\mathbf{y} \in \mathbf{A}} \sum_{t=0}^T \delta^t v(y_t) + \tilde{\epsilon}_{\mathbf{y}} \right)$$

Embedding the discounted Luce rule into the random utility framework greatly simplifies the empirical applicability of the model.

## 3 Discrete choice of dated outcomes

The interest in studying dated outcomes stems from the popularity of the Multiple Price List method (Coller and Williams, 1999; Andersen et al., 2008; Halevy, 2015) used to elicit time preferences in both laboratory and field experiments. The MPL is often coupled with logistic regressions to estimate discount rates (Chabris et al., 2008; Tanaka et al., 2010; Louie and Glimcher, 2010). However, Section 2 highlighted

---

<sup>7</sup>A random variable  $X$  is Fréchet distributed with parameters  $\theta, v > 0$ , written  $F(\theta, v)$ , if it has a distribution function  $F(x) = [-(v/x)^\theta]$  for all  $x > 0$ .

the possible drawbacks of using the discounted logit in this setting. We propose a model, the discounted Luce rule that addresses the stationarity problem and retains the flexibility of the logit. The properties that characterize the discounted Luce rule are intuitive and directly comparable with the axioms of deterministic intertemporal choice of dated outcomes (Fishburn and Rubinstein, 1982). Two axioms: Stochastic Impatience and Stationary Independence of Irrelevant Alternatives completely characterize the model.

Let  $Z$  be a finite set of alternatives and  $X \subseteq Z \times \{0, \dots, T\}$ , be a finite set of dated alternatives, i.e.  $(x, t) \in X$ , is a reward  $x$  at time  $t$ . A choice set is an element of  $\mathcal{A} = 2^X \setminus \{\emptyset\}$ . Choice probabilities are functions  $P : \mathcal{A} \times \mathcal{A} \rightarrow [0, 1]$ , for all  $A \in 2^X$ . Throughout the first part of the work, *we assume that all choice probabilities are strictly positive*,<sup>8</sup> i.e. for all  $x \in Z$ ,  $t \in \{0, \dots, T\}$  and all  $A \in \mathcal{A}$ ,  $P((x, t), A) > 0$ .

The standard Luce choice axiom or Independence of Irrelevant Alternatives (IIA) in this setting is the following:

(IIA). For all  $A, B \in \mathcal{A}$  and  $(x, t), (y, s) \in A \cap B$ ,

$$\frac{P((x, t), A)}{P((y, s), A)} = \frac{P((x, t), B)}{P((y, s), B)}$$

The relative probability of choosing a dated outcome  $(x, t)$  over  $(y, s)$  is constant across menus. The IIA axiom is necessary and sufficient for the existence of a positive random scale<sup>9</sup>  $u : X \rightarrow \mathbb{R}_{++}$ , such that

$$P((x, t), A) = \frac{u(x, t)}{\sum_{(y, s) \in A} u(y, s)}$$

For a given set  $A \in \mathcal{A}$  and  $r \geq 0$ , let define  $A_r \triangleq \{(x, t + r) : (x, t) \in A\}$ ,  $A_r$  contains all the rewards in  $A$  postponed by  $r$  periods. Following the discussion in Section 2, the Stochastic Stationarity Axiom imposes the following:

---

<sup>8</sup>As pointed out in Fudenberg and Strzalecki (2015), a small positive probability is empirically indistinguishable from a zero probability.

<sup>9</sup>A random scale is unique up to positive multiplication, i.e. if  $u$  and  $u'$  define the same probabilities, there exists a  $a > 0$  such that  $u = au'$ .

(SSA). For all  $A \in \mathcal{A}$  and  $t, s, r \geq 0$ ,

$$\frac{P((x, t), A)}{P((y, s), A)} = \frac{P((x, t+r), A_r)}{P((y, s+r), A_r)}$$

The relative probability of choosing  $x$  at time  $t$  over  $y$  at time  $s$ , does not vary when both payments are equally delayed. The condition holds for the sets  $A_r$  and says nothing concerning the interaction with dated outcomes that can be added or subtracted to the set. In other words, without assuming IIA, the relative probability of choosing  $(x, t)$  over  $(y, s)$  from  $A = \{(x, t), (y, s)\}$ , can be different from the probability of choosing  $(x, t+r)$  over  $(y, s+r)$  from  $B = \{(x, t+r), (y, s+r), (z, q+r)\}$ . So the SSA does not restrict possible interactions among outcomes. Consider now the following Stationary version of the IIA:

(SIIA). For all  $A, B \in \mathcal{A}$  and  $t, s, r \geq 0$ ,

$$\frac{P((x, t), A)}{P((y, s), A)} = \frac{P((x, t+r), B)}{P((y, s+r), B)}$$

The SIIA axiom implies the IIA axiom for  $r = 0$  and for  $B = A_r$  it implies the SSA.

Next lemma shows the opposite implications:

**Lemma 1.** *The SIIA holds, if and only if, the IIA and the Stochastic Stationarity Axiom hold.*

Assuming the SIIA is equivalent to assume both stochastic stationarity and the IIA axiom. In this work, we will retain the IIA and we will study the role of stationarity and the consequences of its weakening. Section 7 discusses a simple extension of the discounted Luce rule that relaxes the IIA. The next condition imposes a stochastic form of impatience: if two rewards have the same probability of being selected when one is payed later, the equality is broken in favour of the latter when both are delivered at the same date.

(Stochastic Impatience). For all  $x, y \in Z$  and  $t \geq 0$ , if  $P((x, t), A) = P((y, t+1), A)$

then  $P((x, t), B) \leq P((y, t), B)$ .

The next theorem characterizes the discounted Luce rule:

**Theorem 1.** *The SIIA axiom and Stochastic Impatience hold, if and only if, choice probabilities are represented by the discounted Luce rule:*

$$P((x, t), A) = \frac{\delta^t v(x)}{\sum_{(y, s) \in A} \delta^s v(y)}$$

for some random scale  $v : Z \rightarrow \mathbb{R}_{++}$  and  $\delta \in (0, 1]$ .

The IIA axiom, implied by the SIIA, gives choice probabilities the Luce's relative weight form. The Stochastic Stationarity part of the SIIA imposes separable and geometric discounting. Lastly, Stochastic Monotonicity allows us to interpret  $\delta$  as a discount factor. Due to the great amount of empirical and theoretical research in non-geometric discounting, the rest of the section focuses on violations of the SSA axiom, while maintaining IIA. Relaxing the latter and maintaining stationarity represents an interesting line for future research.

As a final note, we show how to elicit time preferences with the discounted Luce rule. Consider the following ratios  $\frac{P((x, t+1), A)}{P((y, 0), A)}$  and  $\frac{P((y, 0), B)}{P((x, t), B)}$  for some  $t \geq 0$  and  $x, y \in Z$ . It is immediate to see that  $\frac{P((x, t+1), A)}{P((y, 0), A)} \cdot \frac{P((y, 0), B)}{P((x, t), B)} = \delta$ . The discount factor  $\delta$  can be inferred directly from choice probabilities.

### 3.1 Implications of non-geometric discounting

Geometric discounting of future rewards is normatively plausible, but it is often challenged by the experimental evidence of diminishing impatience (for example, [Thaler, 1981](#)). Well-known alternatives are the quasi-hyperbolic discounting of [Laibson \(1997\)](#) and the hyperbolic discounting of [Prelec \(2004\)](#). Deviations from geometric discounting necessarily induce violations of the Stochastic Stationarity Axiom, for example, let consider the quasi hyperbolic discounting model of [Laibson \(1997\)](#),  $1, \beta\delta, \beta\delta^2, \beta\delta^3 \dots$  for some  $\beta \in [0, 1)$  and  $\delta \in (0, 1]$ . With quasi-hyperbolic the trade-off between consumption in two consecutive periods is maximum at the present. It is plausible to imagine that, for a general stochastic choice rule, quasi-hyperbolic discounting implies the following violation of the SSA:

$$\frac{\text{Prob}((x, 0), A)}{\text{Prob}((y, 1), A)} > \frac{\text{Prob}((x, t), A_t)}{\text{Prob}((y, t+1), A_t)} \quad (3)$$



the relative probability of choosing  $x$  now over  $y$  tomorrow decreases when both outcomes are equally delayed. This is the stochastic counterpart of the present bias. The discounted Luce rule of Theorem 1 cannot accommodate the previous inequality. Therefore, we introduce the general discounted Luce rule. Let define a *discount function*  $D(t)$ , as a decreasing function  $D : \mathbb{N}_+ \rightarrow (0, 1]$ , with  $D(0) = 1$  and  $\lim_{t \rightarrow \infty} D(t) = 0$  then, the choice probabilities according to the generalized discounted Luce rule are:

$$P((x, t), A) = \frac{D(t)v(x)}{\sum_{(y,s) \in A} D(s)v(y)}$$

The following axiom is the required relaxation of the SSA that allows for general discounting of future consequences.

(*Weak SSA*). For all  $A \in \mathcal{A}$  and all  $t, r \geq 0$ ,

$$\frac{P((x, t), A)}{P((y, t), A)} = \frac{P((x, t+r), A_r)}{P((y, t+r), A_r)}$$

It imposes invariance of the relative probability of choosing between two payoffs, only when they are payed at *the same date*. Intuitively, this ratio is not influenced by intertemporal trade-offs, since both outcomes are delivered on the same date. Then we have the following result:

**Proposition 1.** *The IIA axiom, Weak SSA and Stochastic Impatience hold, if and only if, choice probabilities are represented by a generalized discounted Luce rule., i.e.*

$$P((x, t), A) = \frac{D(t)v(x)}{\sum_{(y,s) \in A} D(s)v(y)}$$

for some random scale  $v : Z \rightarrow \mathbb{R}_{++}$  and discount function  $D : \{0, \dots, T\} \rightarrow (0, 1]$ .

We imposed the static IIA axiom to give probabilities a simple structure and we relax stochastic stationarity to allow non-geometric discounting. The result is a flexible rule that accommodates common discount functions, such as the hyperbolic and the quasi-hyperbolic. Violations of the SSA can be related to the degree of impatience of  $D(t)$ , defined as

$$I(t) = \frac{D(t)}{D(t+1)}$$

we say that  $D(t)$  exhibits the *present bias* if  $I(0) > I(t)$  for all  $t > 0$ . We say that  $D(t)$  exhibits strict *diminishing impatience* if  $I(t) > I(t+1)$  for all  $t$ . Quasi-hyperbolic discounting,  $D(0) = 1, D(t) = \beta\delta^t$ , exhibits the present bias and does not exhibit strict diminishing impatience. Hyperbolic discounting,  $D(t) = \frac{1}{1+kt}$ , exhibits both. Then we have the following simple consequences:

- $\frac{P((x,0),A)}{P((y,1),A)} > \frac{P((x,t),A_t)}{P((y,t+1),A_t)}$ ,  $\forall t > 0$ , if and only if,  $D(t)$  exhibits the Present Bias.
- $\frac{P((x,t-1),A_{t-1})}{P((y,t),A_{t-1})} > \frac{P((x,t),A_t)}{P((y,t+1),A_t)}$ ,  $\forall t > 0$ , if and only if,  $D(t)$  exhibits strict Diminishing Impatience.

With DI, the relative probability of choosing an earlier over a later payoff is always greater than the same probability when both are delayed by an additional period.

One may be interested in distinguishing quasi-hyperbolic of [Laibson \(1997\)](#) from the general discount function  $D(t)$ . The next axiom contains the required restrictions:

(*Quasi-hyperbolic SSA*). For all  $A \in \mathcal{A}$ :

1. (Delayed SSA):  $\frac{P((x,t),A)}{P((y,s),A)} = \frac{P((x,t+r),A_r)}{P((y,s+r),A_r)}$ , for all  $t, s > 0, r \geq 0$ .
2. (Present Bias):  $\frac{P((x,0),A)}{P((y,t),A)} \geq \frac{P((x,r),A)}{P((y,t+r),A_r)}$ , for all  $t \geq 0, r > 0$ .
3. (Invariance): For all  $x, y \in Z$ ,  $\frac{P((x,0),A)}{P((y,0),A)} = \frac{P((x,t),A_t)}{P((y,t),A_t)}$ .

All the intuitive features of the quasi-hyperbolic discounting affect relative choice probabilities. For non-immediate outcomes, the relative choice probabilities are constant when outcomes are equally delayed, since the (delayed) SSA holds. However, the relative probability of choosing an immediate payment over a delayed one is strictly greater than the same proportion when both payments are equally delayed, this is the present bias. The last part, Present Weak SSA, imposes equality of relative probability only when the outcomes are payed at the same date (it is a weakening of the weak SSA). All the restrictions of the Quasi-hyperbolic SSA are easily observable in laboratory or fields experiments. Then we have:

**Proposition 2.** *The IIA axiom, Quasi-hyperbolic SSA and Stochastic Impatience hold, if and only if, there exists a positive ratio scale  $v : Z \rightarrow \mathbb{R}_{++}$  and a discount function  $D(t)$  such that:*

$$P((x, t), A) = \frac{D(t)v(x)}{\sum_{(y,s) \in A} D(s)v(y)}$$

and  $D(t) = \beta\delta^t$  if  $t > 0$  and  $D(0) = 1$ , for some  $\beta, \delta \in (0, 1]$ .

## 4 Consumption streams

The dated outcomes setting is quite restrictive and cannot be compared with axiomatizations of additively separable discounted utility with geometric discounting (Koopmans, 1960; Fishburn, 1970) or quasi-hyperbolic discounting (Hayashi, 2003; Montiel Olea and Strzalecki, 2014). In this section we fill the gap extending to finite consumption streams the intuitions of the previous section.

For  $T > 0$ , let  $Z^{T+1} = Z \times Z \times \dots \times Z$  be a  $T+1$  product of a finite set of alternatives. An element of  $Z^{T+1}$  represents a consumption stream  $\mathbf{x} = (x_0, x_1, \dots, x_T)$ . A choice set is an element of  $\mathcal{A} = 2^{Z^{T+1}} \setminus \{\emptyset\}$ . Given a set  $A$ , the discounted Luce rule probability of choosing  $\mathbf{x} \in A$  is given by

$$P(\mathbf{x}, A) = \frac{\sum_{t=0}^T \delta^t v(x_t)}{\sum_{\mathbf{y} \in A} \sum_{t=0}^T \delta^t v(y_t)}$$

for some  $\delta \in (0, 1]$ . The probability of selecting a given consumption stream is given by its relative weight in the choice set. The weight is a discounted sum of the values of its components.

For an arbitrary  $x \in Z$ , denote  $\mathbf{x}(t)$  the consumption stream  $\mathbf{x}(t) = (z, z, \dots, x, z, \dots, z)$ , i.e.  $x$  is paid at  $t$  and  $z$  otherwise. The next condition postulates the existence of special  $z \in Z$ :

(*Separability*). There exists  $z \in Z$  such that, for all  $\mathbf{x}, \mathbf{y} \in Z^{T+1}$ , with  $\mathbf{x}, \mathbf{y} \in A$ ,  $\mathbf{y}, \mathbf{x}_t(t) \in B$  for all  $t \geq 0$  and with  $P(\mathbf{y}, A) > 0$ ,  $P(\mathbf{y}, B) > 0$  implies,

$$\frac{P(\mathbf{x}, A)}{P(\mathbf{y}, A)} = \frac{\sum_{t=0}^T P(\mathbf{x}_t(t), B)}{P(\mathbf{y}, B)}$$

Separability implies that the relative probability of choosing  $\mathbf{x}$  from a menu  $A$  when  $\mathbf{y}$  is available, is equal to the sum of the probabilities of choosing its "components"  $\mathbf{x}_t(t) = (z, z, \dots, x_t, z, \dots, z)$ , relative to  $\mathbf{y}$ . To gain intuition, assume  $T = 1$ ,  $\mathbf{x} = (x_0, x_1)$ ,  $\mathbf{y} = (y_0, y_1)$ ,  $A = \{(x_0, x_1), (y_0, y_1)\}$  and  $B = \{(x_0, z), (z, x_1), (y_0, y_1)\}$ . Then

Separability implies

$$\frac{P((x_0, x_1), A)}{P((y_0, y_1), A)} = \frac{P((x_0, z), B) + P((z, x_1), B)}{P((y_0, y_1), B)}$$

Suggesting that the probability of selecting a consumption stream can be decomposed in the probability of selecting its components. Separability is the stochastic choice counterpart of additive separability in deterministic choice. Let  $\mathbf{z} = (z, z, \dots, z)$ , the first consequence of Separability is the following:

**Lemma 2.** *For all  $z \in Z$  satisfying Separability,  $P(\mathbf{z}, A) = 0$  for all  $A \in \mathcal{A}$  with  $\mathbf{x}, \mathbf{z} \in A$  and  $\mathbf{x} \neq \mathbf{z}$ .*

The elements  $z$  can be interpreted as being "nothing" and the probability of selecting them is zero whenever there is an alternative. Given the existence of special  $z \in Z$ , we can turn to the stationarity properties of the discounted Luce rule in the consumption stream setting. For each  $A \in \mathcal{A}$ , we define  $A_{+1} = \{(z, \mathbf{x}) : \mathbf{x} \in A\}$ , where the notation  $(z, \mathbf{x})$  indicates  $(z, \mathbf{x}) = (z, x_0, \dots, x_{T-1})$ . Differently, for  $\mathbf{x} \in Z^{T+1}$ ,  $(\mathbf{x}, z) = (x_0, x_1, \dots, x_{T-1}, z)$  and  $(z, \mathbf{x}, z) = (z, x_0, x_1, \dots, x_{T-2}, z)$ . As for the interpretation,  $(z, \mathbf{x})$  is a "shift forward" of  $(\mathbf{x}, z)$ . Following the intuitions of the dated outcome setting, we consider a choice rule to be stationary if (assume the probabilities at the denominator are strictly positive):

$$\frac{P((z, \mathbf{x}), A)}{P((z, \mathbf{x}'), A)} = \frac{P((\mathbf{x}, z), A_{+1})}{P((\mathbf{x}', z), A_{+1})} \quad (4)$$

The relative probability of choosing  $(z, \mathbf{x})$  over  $(z, \mathbf{y})$  remains unchanged after a shift forward of both. The property resembles the definition of stationarity of deterministic choice (see [Fishburn, 1970](#), Def. 7.3). Equation (4) holds for the discounted Luce rule if  $v(z) = 0$ , indeed:

$$\frac{P((z, \mathbf{x}), A)}{P((z, \mathbf{x}'), A)} = \frac{v(z) + \delta v(x_0) + \delta^2 v(x_1) + \sum_{t=3}^T \delta^t v(x_{t-1})}{v(z) + \delta v(x'_0) + \delta^2 v(x'_1) + \sum_{t=3}^T \delta^t v(x'_{t-1})} = \frac{P((\mathbf{x}, z), A_{+1})}{P((\mathbf{x}', z), A_{+1})}$$

but this follows from Lemma 2. Therefore, a discounted Luce rule satisfies the previous equality. Consider now the discounted multinomial logit. The probability of choosing

$\mathbf{x} \in A$  is:

$$P_{Logit}(\mathbf{x}, A) = \frac{\exp\left(\sum_{t=0}^T \delta^t v(x_t)\right)}{\sum_{\mathbf{y} \in A} \exp\left(\sum_{t=0}^T \delta^t v(y_t)\right)}$$

and the equality in Eq. (4) does not hold for the discounted logit:

$$\begin{aligned} \frac{P_{Logit}((z, \mathbf{x}), A)}{P_{Logit}((z, \mathbf{x}'), A)} &= \frac{\exp(v(z) + \sum_{t=1}^T \delta^t v(x_{t-1}))}{\exp(v(z) + \sum_{t=1}^T \delta^t v(x'_{t-1}))} \\ &\neq \frac{\exp(\sum_{t=0}^{T-1} \delta^t v(x_t) + \delta^T v(z))}{\exp(\sum_{t=0}^{T-1} \delta^t v(x'_t) + \delta^T v(z))} \\ &= \frac{P_{Logit}((\mathbf{x}, z), A_{+1})}{P_{Logit}((\mathbf{x}', z), A_{+1})} \end{aligned}$$

Therefore, according to the discounted logit, the relative probability of choosing a consumption stream  $\mathbf{x}$  over  $\mathbf{y}$  in a set  $A$ , changes when all the elements in  $A$  are "shifted" by one period. This form of stationarity is then able to tell apart the two models. We call it Stochastic Fishburn Stationarity:

(SFS). For all  $\mathbf{x}, \mathbf{x}' \in Z^{T+1}$ ,  $A \in \mathcal{A}$  with  $(\mathbf{x}, z), (\mathbf{x}', z) \in A$  and  $P((\mathbf{x}', z), A) > 0$ ,

$$P((z, \mathbf{x}'), A_{+1}) > 0:$$

$$\frac{P((\mathbf{x}, z), A)}{P((\mathbf{x}', z), A)} = \frac{P((z, \mathbf{x}), A_{+1})}{P((z, \mathbf{x}'), A_{+1})}$$

Fudenberg and Strzalecki (2015) proposed an alternative notion, called *stream stationarity*, that imposes the following:

$$P((\mathbf{x}, y), A) \geq P((\mathbf{x}', y), A) \iff P((y, \mathbf{x}), B) \geq P((y, \mathbf{x}'), B)$$

Stream stationarity is too weak to distinguish the discounted logit and the discounted Luce rule, since it is satisfied by both models.

Separability and the SFS axiom are the main innovations of the section, the following axioms are standard. Since we deal with zero probability events, we use a more general axiom than the IIA.<sup>10</sup>

---

<sup>10</sup>GIIA is equivalent to the original Luce choice axiom (see Luce, 1959), which is more general than the IIA.

(GIIA). For all  $A \in \mathcal{A}$  and  $\mathbf{x} \in A$ , there exists  $u : Z^{T+1} \rightarrow \mathbb{R}_+$ , such that:

$$P(\mathbf{x}, A) = \frac{u(\mathbf{x})}{\sum_{\mathbf{y} \in A} u(\mathbf{y})}$$

Lastly, we define a stochastic notion of impatience similar to the one in the delayed rewards setting:

(Stochastic Impatience). For all  $x, y \neq z \in Z$  and  $t \geq 0$ , if  $P(\mathbf{x}(t), A) = P(\mathbf{y}(t+1), A)$ , then  $P(\mathbf{x}(t), B) \leq P(\mathbf{y}(t), B)$ .

The next theorem characterizes the discounted Luce rule:

**Theorem 2.** *The GIIA axiom, SFS, Separability and Stochastic Impatience hold, if and only if, choice probabilities are represented by a discounted Luce rule, i.e.*

$$P(\mathbf{x}, A) = \frac{\sum_{t=0}^T \delta^t v(x_t)}{\sum_{\mathbf{y} \in A} \sum_{t=0}^T \delta^t v(y_t)}$$

for a ratio scale  $v : Z \rightarrow \mathbb{R}_+$  and  $\delta \in (0, 1]$ .

Also in this setting, we are interested in relaxing stationarity to determine the observable restrictions following quasi-hyperbolic discounting. Indeed, although  $v(z) = 0$ , quasi-hyperbolic discounting violates SFS. Let,

$$P(\mathbf{x}, A) = \frac{v(x_0) + \beta \sum_{t=1}^T \delta^t v(x_t)}{\sum_{\mathbf{y} \in A} \left[ v(y_0) + \beta \sum_{t=1}^T \delta^t v(y_t) \right]}$$

then SFS is violated:

$$\begin{aligned}
\frac{P((\mathbf{x}, z), A)}{P(\mathbf{x}', z), A)} &= \frac{v(x_0) + \beta \sum_{t=1}^{T-1} \delta^t v(x_t) + \beta \delta^T v(z)}{v(x'_0) + \beta \sum_{t=1}^{T-1} \delta^t v(x'_t) + \beta \delta^T v(z)} \\
&\neq \frac{v(z) + \beta \sum_{t=1}^T \delta^t v(x_{t-1})}{v(z) + \beta \sum_{t=1}^{T-1} \delta^t v(x'_{t-1})} \\
&= \frac{\sum_{t=1}^T \delta^t v(x_{t-1})}{\sum_{t=1}^{T-1} \delta^t v(x'_{t-1})} = \frac{P((z, \mathbf{x}), A_{+1})}{P((z, \mathbf{x}'), A_{+1})}
\end{aligned}$$

The following relaxation of SFS is parallel to that introduced in the delayed rewards setting and includes all the intuitive features of a present-biased Luce rule. They are comparable with the axioms characterizing quasi-hyperbolic discounting in deterministic choice (e.g. [Montiel Olea and Strzalecki, 2014](#)). It contains three properties: Delayed Stochastic Fishburn Stationarity, Present Bias and Invariance.

(*Quasi-hyperbolic Stationarity*).

1. (DSFS): For all  $A \in \mathcal{A}$ , with  $(z, \mathbf{x}), (z, \mathbf{x}') \in A$ , if  $P((z, \mathbf{x}', z), A) > 0$  and  $P((z, z, \mathbf{x}'), A_{+1}) > 0$ :

$$\frac{P((z, \mathbf{x}, z), A)}{P((z, \mathbf{x}', z), A)} = \frac{P((z, z, \mathbf{x}), A_{+1})}{P((z, z, \mathbf{x}'), A_{+1})}$$

2. (PB): For all  $A \in \mathcal{A}$ , with  $\mathbf{x}, (z, \mathbf{x}) \in A$ , if  $P((z, \mathbf{x}, z), A_{+1}) > 0$  and  $P((z, z, \mathbf{x}'), A_{+1}) > 0$ :

$$\frac{P((\mathbf{x}, z, z), A)}{P((z, \mathbf{x}, z), A_{+1})} \geq \frac{P((z, \mathbf{x}', z), A)}{P((z, z, \mathbf{x}'), A_{+1})}$$

3. (Invariance): For all  $x, y \in Z \setminus \{z\}$ ,  $t > 0$  and all  $A \in \mathcal{A}$ , if  $P(\mathbf{y}(0), A) > 0$  and  $P(\mathbf{y}(1), A_{+1}) > 0$ :

$$\frac{P(\mathbf{x}(0), A)}{P(\mathbf{y}(0), A)} = \frac{P(\mathbf{x}(1), A_{+1})}{P(\mathbf{y}(1), A_{+1})}$$

The DSFS imposes stationarity to non-immediate shifts. The present bias imposes a greater probability of choosing a stream of consumption when the immediate outcome has a positive value. Invariance excludes variations of the relative likelihood of choosing two outcomes when they are payed at the same date.

**Theorem 3.** *The GIIA axiom, Quasi-hyperbolic Stationarity, Separability and Stochastic Impatience hold, if and only if, choice probabilities are represented by a quasi-hyperbolic discounted Luce rule, i.e.*

$$P(\mathbf{x}, A) = \frac{v(x_0) + \beta \sum_{t=1}^T \delta^t v(x_t)}{\sum_{\mathbf{y} \in A} \left[ v(y_0) + \beta \sum_{t=1}^T \delta^t v(y_t) \right]}$$

The axioms characterizing the quasi-hyperbolic Luce rule can be used to extend the recursive axiomatization of [Fudenberg and Strzalecki \(2015\)](#) to a dynamic discrete choice model that accounts for quasi-hyperbolic discounting (see [Section 6](#) for a discussion). Such model has recently gained attention in applied works ([Paserman, 2008](#); [Tarozzi and Mahajan, 2011](#); [Yao et al., 2012](#); [An et al., 2014](#); [Fang and Wang, 2015](#)) but, its behavioral restrictions are not yet understood.

## 5 Elasticity and cross-elasticity

Elasticity measures how a variation in an observable factor, for example a component of the consumption stream  $x_t$ , affects choice probabilities ([Train, 2009](#)). The elasticities of the discounted Luce rule have similar properties to the elasticities of the logit, although they inherit the scale-free property of the model. [Fosgerau and Bierlaire \(2009\)](#) provides calculations that are valid for a Luce rule. Since we are studying the interaction of intertemporal preferences and stochastic choice, we will exploit the structure of the discounted Luce rule to study elasticity. We are interested in the elasticity of the probability of choosing  $\mathbf{x}$  when one element  $x_t$  of  $\mathbf{x}$  changes. Formally:

$$\mathcal{E}[P(\mathbf{x}, A); x_t] \triangleq \frac{\partial P(\mathbf{x}, A)}{\partial x_t} \frac{x_t}{P(\mathbf{x}, A)}$$



Given  $\mathbf{x} = (x_0, x_1, x_2, \dots)$ , first consider the expression  $\frac{\delta^t v(x_t)}{\sum_{s=0}^T \delta^s v(x_s)}$ , it can be interpreted as the "weight" of  $x_t$  if  $\mathbf{x} = (x_0, x_1, \dots, x_T)$ . Denote it as  $W(x_t, \mathbf{x}) = \frac{\delta^t v(x_t)}{\sum_{s=0}^T \delta^s v(x_s)}$ . With this notation, simple calculations give:

$$\mathcal{E}[P(\mathbf{x}, A); x_t] = \frac{x_t}{v(x_t)} \frac{\partial v(x_t)}{\partial x_t} W(x_t, \mathbf{x}) [1 - P(\mathbf{x}, A)]$$

Elasticity depends on the complementary probability of choosing  $\mathbf{x}$  from  $A$  and the relative weight of  $x_t$  in the consumption stream  $\mathbf{x}$ . Notice that elasticity does not depend on the scale of the value function. If we multiply  $v$  by a positive number  $k$ , the elasticity does not change. This is a stark difference with the multinomial logit. If the elements in the consumption stream are monetary payments and for a linear  $v(x_t) = kx_t$ ,  $\mathcal{E}[P(\mathbf{x}, A); x_t] = W(x_t, \mathbf{x}) [1 - P(\mathbf{x}, A)]$ . In the case of a linear  $v(x_t) = kx_t$ , a simple calculation shows that:

$$\sum_{x_t \in \mathbf{x}} \mathcal{E}[P(\mathbf{x}, A); x_t] = 1 - P(\mathbf{x}, A)$$

showing that the smaller is the elasticity of each component in  $\mathbf{x}$ , the higher is the probability of choosing  $\mathbf{x}$  from  $A$ . This follows from the "separate" contribution of each component to the value of a consumption stream. Discounting interacts with elasticity through  $W(x_t, \mathbf{x})$ , ceteris paribus, the further in time is  $x_t$ , the smaller is  $W(x_t, \mathbf{x})$ .

A different measure of sensitivity worth studying is the cross-elasticity: the variation of the probability of choosing  $\mathbf{x}$  from  $A$  when an element  $y_s$  of  $\mathbf{y} \in A$  changes. Formally,

$$\mathcal{E}[P(\mathbf{x}, A); y_s] \triangleq \frac{\partial P(\mathbf{x}, A)}{\partial y_s} \frac{y_s}{P(\mathbf{x}, A)}$$

Easy calculations show that:

$$\mathcal{E}[P(\mathbf{x}, A); y_s] = -\frac{\partial v(y_s)}{\partial y_s} \frac{y_s}{v(y_s)} \frac{\delta^s v(y_s)}{\sum_{\mathbf{y}' \in A} \sum_{r=0}^T \delta^r v(y'_r)}$$

Cross-elasticity depends negatively on the weight of  $y_s$  in the whole  $A$ . An increase in the probability of choosing  $\mathbf{y}$  negatively affects all the alternatives in  $A$ . In the case

of a linear  $v(x_t) = kx_t$ ,  $\mathcal{E}[P(\mathbf{x}, A); y_s] = -\frac{\delta^s v(y_s)}{\sum_{\mathbf{y}' \in A} \sum_{r=0}^T \delta^r v(y'_r)}$  and

$$\sum_{y_s \in \mathbf{y}} \mathcal{E}[P(\mathbf{x}, A); y_s] = -P(\mathbf{y}, A)$$

Choice probabilities provides information concerning elasticity and cross-elasticity also in the intertemporal setting.

## 6 Recursive Luce rule with quasi-hyperbolic discounting

We discuss in the final section an interesting application of our characterization of the discounted Luce rule with quasi-hyperbolic version. The properties characterizing the representations can be embedded into the axiomatization of [Fudenberg and Strzalecki \(2015\)](#) to model dynamic stochastic choice. Recursive structural models are widely used in econometrics ([Rust, 1987](#); [Aguirregabiria and Mira, 2010](#)) and their behavioral foundation has been studied only recently ([Fudenberg and Strzalecki, 2015](#); [Matějka et al., 2015](#)). In the recursive setting, in each period the individual chooses stochastically an immediate consumption and a continuation menu and she correctly anticipates future noise. To gain intuition, we briefly introduce the recursive setting of [Fudenberg and Strzalecki \(2015\)](#). There is a finite number of periods  $T > 0$ , in each period  $t$ , the choice is made from a menu  $A_t$ . We denote *actions*  $a_t, b_t$  the elements of the menu  $A_t$  and  $\mathcal{A}_t$  the space of all actions at time  $t$ . We define recursively the choice problem starting from the terminal node  $T$ , where the only actions are one-period outcomes, i.e.  $\mathcal{A}_T = Z$ . For a set  $S$ ,  $K(S)$  denotes the family of non-empty and finite subsets of  $S$ . We then define recursively  $\mathcal{M}_T = K(\mathcal{A}_T)$  and for all  $t < T$ ,  $\mathcal{A}_t = Z \times \mathcal{M}_{t+1}$  and  $\mathcal{M}_t = K(\mathcal{A}_t)$ . An action  $a_t$  in a menu  $A_t$  is a pair  $(x_t, A_{t+1})$  of immediate consumption  $x_t$  and a continuation menu  $A_{t+1}$ . According to recursive models of stochastic choice, the value of an action at time  $t$  is given by the following:

$$U_t(x_t, A_{t+1}) = v(x_t) + \delta \mathbb{E} \left[ \max_{a_{t+1} \in A_{t+1}} U_{t+1}(a_{t+1}) + \epsilon_{a_{t+1}} \right]$$

and the stochastic choice from the menu  $A_t$  is given by:

$$P_t(a_t, A_t) = \text{Prob} \left[ U_t(a_t) + \epsilon_{a_t} \geq \max_{b_t \in A_t} U_t(b_t) + \epsilon_{b_t} \right]$$

Assumptions on the distribution of the errors characterize different models of choice. A recent extension of the previous recursive model introduced quasi-hyperbolic discounting (Paserman, 2008; Tarozzi and Mahajan, 2011; Fang and Wang, 2015; An et al., 2014; Yao et al., 2012) and assumed logit choice probabilities from menus. In this case the value of an action  $(x_t, A_{t+1})$  is equal to:

$$U_t(x_t, A_{t+1}) = v(x_t) + \beta \delta \mathbb{E} \left[ \max_{a_{t+1} \in A_{t+1}} U_{t+1}(a_{t+1}) + \epsilon_{a_{t+1}} \right]$$

An axiomatic foundation of such model has not been proposed in the literature, however the characterization of the quasi-hyperbolic Luce rule in Theorem 3 can be used to axiomatize the previous model along the lines of Fudenberg and Strzalecki (2015). If we assume that errors  $\epsilon_{a_t}$  are i.i.d. for all  $a_t$  and  $t$  and distributed according to a Fréchet distribution with parameter  $\theta = 1$  (see Mattsson et al., 2014), time  $t$  choice probabilities are of the Luce form:  $P_t(a_t, A_t) = \frac{U_t(a_t)}{\sum_{b_t \in A_t} U_t(b_t)}$ . The discounted Luce rule is a particular case of the previous expression in which all the continuation menus are singletons. Indeed, Mattsson et al. (2014) again proved that with Fréchet errors, we can rewrite  $U_t(x_t, A_{t+1}) = v(x_t) + \beta \delta \left[ \sum_{a_{t+1} \in A_{t+1}} U_{t+1}(a_{t+1}) \right]$  and assuming that all continuation menus are singletons, we are back to the (quasi-hyperbolic) discounted Luce rule, i.e.

$$P_t(a_t, A_t) = \frac{v(x_t) + \beta \sum_{s=t+1}^T \delta^s v(x_s)}{\sum_{b_t \in A_t} \left[ v(y_t) + \beta \sum_{s=t+1}^T \delta^s v(y_s) \right]}$$

To gain intuition concerning the effect of quasi-hyperbolic discounting in the Luce model, we apply it to the adoption of a durable good, for example a smartphone (Gowrisankaran and Rysman, 2012). At time zero, the choice is between two smartphones,  $x, y$  with prices  $p_x, p_y$ . At time one, there is another smartphone  $z$  with price  $p_z$ . If  $x$  is bought at time zero, the time one choice is between keeping  $x$  or replacing it with  $z$  at price  $p_z$ . At time zero the choice is  $a_x = ((x, p_x), A_x)$  or

$a_y = (y, p_y), A_y)$  where  $A_x = ((x, 0), (z, p_z))$  and  $A_y = ((y, 0), (z, p_z))$ . The value  $V(A_x) = v(x) + v(z) - p_z$  and  $V(A_y) = v(y) + v(z) - p_z$ . So, at time zero, the probability of buying  $x$  is

$$P(a_x, A_0) = \frac{v(x) - p_x + \beta\delta V(A_x)}{v(x) - p_x + \beta\delta V(A_x) + v(y) - p_y + \beta\delta V(A_y)}$$

Suppose that at time one the individual may decide to switch, for example from  $x$  to  $y$  and vice-versa paying the relative price. Then, the problem becomes, at time zero the choice is  $a_x = ((x, p_x), A_x)$  or  $a_y = (y, p_y), A_y)$  where  $A_x = ((x, 0), (y, p_y))$  and  $A_y = ((y, 0), (x, p_x))$ . Now,  $V(A_x) = v(x) + v(y) - p_y$  and  $V(A_y) = v(y) + v(x) - p_x$ . The probability of buying  $x$  at time  $t$  is:

$$P(a_x, A_0) = \frac{v(x) - p_x + \beta\delta(v(x) + v(y) - p_y)}{v(x) - p_x + \beta\delta(v(x) + v(y) - p_y) + v(y) - p_y^0 + \beta\delta(v(y) + v(x) - p_x)}$$

We are interested to study the variation of choice probabilities with respect to a permanent or a temporary shock to the price of a smarhphone. A permanent shock is modeled assuming that the variation of the price at time zero implies a variation in price at time one as well. For a temporary shock, we treat the price at time one as a constant. Therefore, assume the price of  $x$  changes permanently, then

$$\mathcal{E}_{perm}[P(a_x, A_0), p_x] = -p_x(1 + \beta\delta)P(a_y, A_0)K(x)$$

where  $K(x) = \frac{1}{v(x) - p_x + \beta\delta(v(x) + v(y) - p_y)}$ . Differently, a temporary shock to the price of  $x$  at time zero does not affect its price at time one, therefore:

$$\mathcal{E}_{temp}[P(a_x, A_0), p_x] = -p_x P(a_y, A_0)K(x)$$

The permanent shock has a stronger effect than a temporary one, but it is mitigated by the present bias factor  $\beta$ . The probability of choosing  $x$  today decreases since the probability of selecting  $y$  increases in both periods. However, the magnitude of the variation in the second period is lowered by the present bias factor.

## 7 Relaxing the IIA: the mixed discounted Luce rule

Throughout the paper, we maintained the assumption that choice probabilities satisfy the IIA (or a weaker version). The IIA sometimes imposes unrealistic substitution patterns among the elements of a menu (see [Train, 2009](#)) and its violations are well-documented (for example, the similarity effect [Simonson \(1989\)](#) and the compromise effect [Simonson and Tversky \(1992\)](#)). We are not aware, however, of evidence supporting violations of the IIA in the intertemporal setting. Testing violations of the IIA in the dated rewards or consumption streams setting would be an interesting path for future research. We propose here a simple extension of the discounted Luce rule that allows for realistic substitution patterns relaxing the IIA axiom. It can be derived from the mixed logit (see [Train, 2009](#), Ch. 6). The mixed logit is a mixture of logit probabilities depending on a parameter  $\theta \in \Theta$ . Suppose that the value function of a logit model depends on a parameter  $\theta \in \Theta$  then, we can define the discounted mixed logit:

$$P_{MLogit}(\mathbf{x}, A) = \int_{\Theta} \frac{\sum_{t=0}^T e^{\delta^t v(x_t|\theta)}}{\sum_{y \in A} \sum_{t=0}^T e^{\delta^t v(y_t|\theta)}} f(\theta) d\theta$$

for some density  $f$ . It is easy to see that the failure of stationarity we discussed in [Section 2](#) occurs for the discounted mixed logit as well. Differently, stationarity holds for the discounted mixed Luce rule, defined as:

$$P_{MLuce}(\mathbf{x}, A) = \int_{\Theta} \frac{\sum_{t=0}^T \delta^t v(x_t|\theta)}{\sum_{y \in A} \sum_{t=0}^T \delta^t v(y_t|\theta)} f(\theta) d\theta$$

hence, the discounted mixed Luce rule is an extension of the discounted Luce rule that allows for substitution patterns and violation of the IIA, but retains the (stochastic) stationarity property of the model. We are not aware of any axiomatization of the mixed logit.

## 8 Conclusion

We study the interaction of discount and stochastic choice in two settings: dated outcomes and consumption streams. We propose a new model, the discounted Luce

rule, that solves the stationarity problem of the discounted logit. We characterize the observable restrictions on choice probabilities generated by geometric, as well as, hyperbolic and quasi-hyperbolic discounting. An extension to recursive choice problem is proposed with an application to the purchase of a durable good when discounting is quasi-hyperbolic.

## Appendix: Proofs

*Proof.* Of Lemma 1. The SIIA clearly implies the other axioms. Now consider  $\frac{P((x,t),A)}{P((y,s),A)}$ , by the SSA  $\frac{P((x,t),A)}{P((y,s),A)} = \frac{P((x,t+r),A_r)}{P((y,s+r),A_r)}$  and by the IIA  $\frac{P((x,t+r),A_r)}{P((y,s+r),A_r)} = \frac{P((x,t+r),B)}{P((y,s+r),B)}$  for all  $B \in \mathcal{A}$  with  $(x, t+r), (y, s+r) \in B \cap A_r$ .  $\square$

*Proof.* Of Theorem 1. Necessity is straightforward. For the sufficiency, let  $r = 0$ , then the SIIA axiom implies the standard IIA. Hence, there exists  $u : X \rightarrow \mathbb{R}_{++}$  such that  $P((x, t), A) = \frac{u(x,t)}{\sum_{(y,s) \in A} u(y,x)}$ . Let assume,  $x = y, t = 1$  and  $s = 0$ , then

$$\frac{P((x, 1), A)}{P((x, 0), A)} = \frac{P((x, 1+r), B)}{P((x, r), B)}$$

or

$$\frac{u(x, 1)}{u(x, 0)} = \frac{u(x, r+1)}{u(x, r)}$$

The equality above implies  $u(x, t+1)u(x, 0) = u(x, 1)u(x, t)$ . Going back recursively gives  $u(x, t) = \left(\frac{u(x,1)}{u(x,0)}\right)^t u(x, 0)$ . Let define  $\delta_x = \frac{u(x,1)}{u(x,0)}$ , by Stochastic Impatience with  $t = 1, s = 0, \frac{u(x,0)}{u(y,0)} = 1$ , implies  $\frac{u(x,1)}{u(y,0)} \leq 1$ , hence  $\frac{u(x,1)}{u(y,0)} \leq \frac{u(x,0)}{u(y,0)} = 1$  and  $1 \geq \frac{u(x,1)}{u(x,0)} = \delta_x$ . Plugging  $u(x, t) = \delta_x^t u(x, 0)$  into the SIIA gives  $\frac{\delta_x^t}{\delta_y^s} = \frac{\delta_x^{t+r}}{\delta_y^{s+r}}$ , for  $s = 0$  and  $r = 1, \delta_x^t = \frac{\delta_x^{t+1}}{\delta_y}$  and it implies  $\delta_x = \delta_y$ . Hence, defining  $v(x) = u(x, 0)$  gives the result.  $\square$

*Proof.* Of Proposition 1. By the IIA axiom there exists  $v$  such that  $P((x, t), A) = \frac{v(x,t)}{\sum_{(y,s) \in A} v(y,x)}$ . Let consider  $t = 0$ , then

$$\frac{P((x, 0), A)}{P((y, 0), A)} = \frac{P((x, r), A_r)}{P((y, r), A_r)}$$

by IIA  $\frac{P((x,r),A_r)}{P((y,r),A_r)} = \frac{P((x,r),B)}{P((y,r),B)}$  for all  $B \in \mathcal{A}$  such that  $(x, r), (y, r) \in B$ . Then

$$\frac{u(x, 0)}{u(y, 0)} = \frac{u(x, r)}{u(y, r)}$$

or  $\frac{u(x,r)}{u(x,0)} = \frac{u(y,r)}{u(y,0)}$ . Since, it is true for all  $x, y, D(t)$  defined as  $D(t) \triangleq \frac{u(x,t)}{u(x,0)}$  is independent of  $x$ , by Stochastic Impatience  $0 \leq D(t) \leq 1$ , then, defining  $v(x) \triangleq u(x, 0)$  we have  $u(x, t) = D(t)v(x)$  and the result follows.  $\square$

*Proof.* Of Proposition 2. Necessity is straightforward. For sufficiency, by the IIA and

the Delayed SSA,

$$\frac{P((x, t), A)}{P((y, s), A)} = \frac{P((x, t+r), A_r)}{P((y, s+r), A_r)} = \frac{P((x, t+r), B)}{P((y, s+r), B)}$$

is equivalent to

$$\frac{u(x, t)}{u(y, s)} = \frac{u(x, t+r)}{u(y, s+r)}$$

for some ratio scale  $u : X \rightarrow \mathbb{R}_{++}$ . Let  $x = y$ ,  $t = 2$  and  $s = 1$ , then

$$\frac{u(x, 2)}{u(x, 1)} = \frac{u(x, r+2)}{u(x, r+1)}$$

hence  $u(x, r+2) = \frac{u(x,2)}{u(x,1)}u(x, r+1)$ , going back until  $r = 0$  implies  $u(x, r+2) = \frac{u(x,2)^{r+1}}{u(x,1)^{r+1}}u(x, 1)$ , now, let  $\delta_x = \frac{u(x,2)}{u(x,1)}$ , by S. Impatience  $\delta_x \in (0, 1]$ , then  $u(x, t) = \delta_x^{t-1}u(x, 1)$ . By Delayed SSA with  $t, s = 1$  and  $r = 1$ ,

$$\frac{P((x, 1), A)}{P((y, 1), A)} = \frac{P((x, 2), A_1)}{P((y, 2), A_1)} \implies \frac{u(y, 2)}{u(y, 1)} = \frac{u(x, 2)}{u(x, 1)}$$

for all  $x, y \in Z$ , hence  $\delta_x = \delta_y$  for all  $x, y \in Z$ . For  $t = 1, r = 1$ , Present Bias implies

$$\frac{u(x, 0)}{u(y, 1)} \geq \frac{u(x, 1)}{u(y, 2)}$$

or  $\frac{u(x,1)}{u(x,0)} \leq \frac{u(y,2)}{u(y,1)} = \delta$  hence  $u(x, 0)\delta \geq u(x, 1)$ , then for some  $\beta_x \leq 1$ ,  $\beta_x \delta u(x, 0) = u(x, 1)$ , then we have  $u(x, t) = \beta_x \delta^t u(x, 0)$  for  $t > 0$ . By Present Weak SSA and IIA,  $\frac{P((x,0),A)}{P((y,0),A)} = \frac{P((x,t),B)}{P((y,t),B)}$  implies  $\frac{u(x,0)}{u(y,0)} = \frac{\beta_x \delta^t u(x,0)}{\beta_y \delta^t u(y,0)}$  or  $\frac{\beta_x}{\beta_y} = 1$ , hence  $\beta_x = \beta_y$ , since it is true for arbitrary  $x, y$ , the result follows, defining  $v(x) = u(x, 0)$  and  $D(t) = \beta \delta^t$  for  $t > 0$  and  $D(0) = 1$ .  $\square$

*Proof.* Of Lemma 2. Take  $A = B$ , by Separability,  $P(\mathbf{z}, A) = \sum_{t=0}^T P(\mathbf{z}_t(t), A)$  for some  $A$  containing  $\mathbf{z}$ , but  $\mathbf{z}_t(t) = \mathbf{z}$  for all  $0 \leq t \leq T$ , hence,  $P(\mathbf{z}, A) = (T+1)P(\mathbf{z}, A)$  and this can be true only if  $P(\mathbf{z}, A) = 0$  since  $T > 0$ . In turn,  $P(\mathbf{z}, A) = 0$  implies  $u(z, z, \dots, z) = 0$  since  $z$  is unique.  $\square$

*Proof.* Of Theorem 2. By the GIIA, there exists a random scale  $u : Z^{T+1} \rightarrow \mathbb{R}_+$  such that  $P(\mathbf{x}, A) = \frac{u(x_0, x_1, \dots, x_T)}{\sum_{\mathbf{y} \in A} u(y_0, y_1, \dots, y_T)}$ . For an arbitrary  $x \in Z$ , let define  $v(x) = u(x, z, z, \dots) = u(\mathbf{x}(0))$  for some  $z$  satisfying Separability, moreover  $v(z) = 0$  for all  $z \in Z_0$ , where  $Z_0 = \{z \in Z : z \text{ satisfies Separability}\}$ . For an arbitrary  $x \in Z \setminus Z_0$ , let define  $\delta_x = \frac{u(\mathbf{x}(1))}{u(\mathbf{x}(0))}$ . By SFS and  $x, y \in Z$ ,

$$\frac{P(\mathbf{x}(1), A_{+1})}{P(\mathbf{y}(1), A_{+1})} = \frac{P(\mathbf{x}(0), A)}{P(\mathbf{y}(0), A)}$$

or equivalently,

$$\frac{P(\mathbf{x}(1), A_{+1})}{P(\mathbf{x}(0), A)} = \frac{P(\mathbf{y}(1), A_{+1})}{P(\mathbf{y}(0), A)} \implies \frac{u(\mathbf{x}(1))}{u(\mathbf{x}(0))} = \frac{u(\mathbf{y}(1))}{u(\mathbf{y}(0))}$$

and this implies  $\delta_x = \delta_y = \delta$  for all  $x, y \in Z$  and define  $\delta_z = \delta$  for all  $z \in Z_0$ . SFS implies

$$\frac{P(\mathbf{x}(1), A)}{P(\mathbf{x}(0), A)} = \frac{P(\mathbf{x}(2), A_{+1})}{P(\mathbf{x}(1), A_{+1})}$$

another application of SFS implies

$$\frac{P(\mathbf{x}(2), A_{+1})}{P(\mathbf{x}(1), A_{+1})} = \frac{P(\mathbf{x}(3), (A_{+1})_{+1})}{P(\mathbf{x}(2), (A_{+1})_{+1})}$$

equivalently,

$$\frac{u(\mathbf{x}(1))}{u(\mathbf{x}(0))} = \frac{u(\mathbf{x}(2))}{u(\mathbf{x}(1))} = \frac{u(\mathbf{x}(3))}{u(\mathbf{x}(2))}$$

By Impatience  $\delta \leq 1$ . The first equality implies  $u(\mathbf{x}(2)) = u(\mathbf{x}(1)) \cdot \frac{u(\mathbf{x}(1))}{u(\mathbf{x}(0))}$  and the second,  $u(\mathbf{x}(3)) = u(\mathbf{x}(2)) \cdot \frac{u(\mathbf{x}(1))}{u(\mathbf{x}(0))}$ , and together  $u(\mathbf{x}(3)) = u(\mathbf{x}(1)) \cdot \left(\frac{u(\mathbf{x}(1))}{u(\mathbf{x}(0))}\right)^2$ , repeating the same argument gives

$$u(\mathbf{x}(t)) = u(\mathbf{x}(1)) \cdot \left(\frac{u(\mathbf{x}(1))}{u(\mathbf{x}(0))}\right)^{t-1}$$

multiplying and dividing by  $u(\mathbf{x}(0))$ , gives

$$u(\mathbf{x}(t)) = u(\mathbf{x}(0)) \cdot \left(\frac{u(\mathbf{x}(1))}{u(\mathbf{x}(0))}\right)^t \quad (5)$$

in our notation this becomes  $u(\mathbf{x}(t)) = \delta^t v(x)$ . To conclude, we need to prove that  $u(x_0, x_1, x_2, \dots) = \sum_{t=0}^{\infty} \delta^t v(x_t)$ . To see this, consider  $P((x_0, x_1, \dots, x_T), A)$  for some  $A$ , Separability implies

$$\frac{u(x_0, x_1, \dots, x_T)}{u(y_0, y_1, \dots, y_T)} = \frac{\sum_{t=0}^T u(\mathbf{x}_t(t))}{u(y_0, y_1, \dots, y_T)}$$

then, by Eq. (5)

$$u(x_0, x_1, \dots, x_T) = \sum_{t=0}^T u(\mathbf{x}_t(t)) = \sum_{t=0}^T \delta^t v(x_t)$$

□

*Proof.* Of Theorem 3. By the GIIA, there exists a random scale  $u : Z^{T+1} \rightarrow \mathbb{R}$  such that  $P(\mathbf{x}, A) = \frac{u(x_0, x_1, \dots, x_T)}{\sum_{\mathbf{y} \in A} u(y_0, y_1, \dots, y_T)}$ . For a given  $x \in Z$ , let define  $v(x) = u(x, z, z, \dots)$

for some  $z \in Z_0$ . For an arbitrary  $x \in Z \setminus Z_0$ , let define  $\delta_x = \frac{u(\mathbf{x}(2))}{u(\mathbf{x}(1))}$  By QSFS,

$$\frac{P(\mathbf{x}(2), A_{+1})}{P(\mathbf{y}(2), A_{+1})} = \frac{P(\mathbf{x}(1), A)}{P(\mathbf{y}(1), A)}$$

or equivalently,

$$\frac{P(\mathbf{x}(2), A_{+1})}{P(\mathbf{x}(1), A)} = \frac{P(\mathbf{y}(2), A_{+1})}{P(\mathbf{y}(1), A)} \implies \frac{u(\mathbf{x}(2))}{u(\mathbf{x}(1))} = \frac{u(\mathbf{y}(2))}{u(\mathbf{y}(1))}$$

and this implies  $\delta_x = \delta_y = \delta$  for all  $x, y \in Z \setminus Z_0$  and define  $\delta_z = \delta$  for all  $z \in Z_0$ . QSFS implies

$$\frac{P(\mathbf{x}(2), A)}{P(\mathbf{x}(1), A)} = \frac{P(\mathbf{x}(3), A_{+1})}{P(\mathbf{x}(2), A_{+1})}$$

another application of QSFS implies

$$\frac{P(\mathbf{x}(3), A_{+1})}{P(\mathbf{x}(2), A_{+1})} = \frac{P(\mathbf{x}(4), (A_{+1})_{+1})}{P(\mathbf{x}(3), (A_{+1})_{+1})}$$



equivalently,

$$\frac{u(\mathbf{x}(2))}{u(\mathbf{x}(1))} = \frac{u(\mathbf{x}(3))}{u(\mathbf{x}(2))} = \frac{u(\mathbf{x}(4))}{u(\mathbf{x}(3))}$$

The first equality implies  $u(\mathbf{x}(3)) = u(\mathbf{x}(2)) \cdot \frac{u(\mathbf{x}(2))}{u(\mathbf{x}(1))}$  and the second,  $u(\mathbf{x}(4)) = u(\mathbf{x}(3)) \cdot \frac{u(\mathbf{x}(2))}{u(\mathbf{x}(1))}$ , and together  $u(\mathbf{x}(4)) = u(\mathbf{x}(2)) \cdot \left(\frac{u(\mathbf{x}(2))}{u(\mathbf{x}(1))}\right)^2$ , repeating the same argument for an arbitrary  $t > 0$ , gives

$$u(\mathbf{x}(t)) = u(\mathbf{x}(2)) \cdot \left(\frac{u(\mathbf{x}(2))}{u(\mathbf{x}(1))}\right)^{t-2}$$

multiplying and dividing by  $u(\mathbf{x}(1))$ , gives

$$u(\mathbf{x}(t)) = u(\mathbf{x}(1)) \cdot \left(\frac{u(\mathbf{x}(2))}{u(\mathbf{x}(1))}\right)^{t-1} \quad (6)$$

By PB. with  $\mathbf{x}' = \mathbf{y}(0)$  and  $\mathbf{x} = \mathbf{x}(0)$ :

$$\frac{u(\mathbf{x}(0))}{u(\mathbf{x}(1))} \geq \frac{u(\mathbf{y}(1))}{u(\mathbf{y}(2))}$$

hence  $\frac{u(\mathbf{x}(1))}{u(\mathbf{x}(0))} \leq \frac{u(\mathbf{y}(2))}{u(\mathbf{y}(1))} = \delta$  then, there exists  $\beta_x \in [0, 1]$  such that  $u(\mathbf{x}(0)) = \beta_x \delta u(\mathbf{x}(1))$   
 Multiplying and dividing Eq. (6) by  $u(\mathbf{x}(0))$  and using the last fact gives

$$u(\mathbf{x}(t)) = u(\mathbf{x}(0)) \cdot \beta_x \left(\frac{u(\mathbf{x}(2))}{u(\mathbf{x}(1))}\right)^t$$

for all  $t > 0$ . In our notation  $u(\mathbf{x}(t)) = \beta_x \delta^t v(x)$  for  $t > 0$  and  $u(\mathbf{x}(0)) = v(x)$ . By Invariance,

$$\frac{P(\mathbf{x}(1), A_{+1})}{P(\mathbf{x}(0), A)} = \frac{P(\mathbf{y}(1), A_{+1})}{P(\mathbf{y}(0), A)}$$

and,

$$\frac{u(\mathbf{x}(1))}{u(\mathbf{x}(0))} = \frac{u(\mathbf{y}(1))}{u(\mathbf{y}(0))} \iff \frac{v(x)\beta_x\delta}{v(x)} = \frac{v(y)\beta_y\delta}{v(y)}$$

that implies  $\beta_x = \beta_y$  for all  $x, y \in Z \setminus Z_0$ . To conclude, we need to prove that  $u(x_0, x_1, x_2, \dots) = v(x_0) + \beta \sum_{t=1}^T \delta^t v(x_t)$ . To see this, consider  $P((x_0, x_1, \dots, x_T), A)$  for some  $A$  that satisfies the condition, then Separability implies

$$\frac{u(x_0, x_1, \dots, x_T)}{\sum_{\mathbf{y} \in A} u(y_0, y_1, \dots, y_T)} = \sum_{t=0}^T \frac{u(\mathbf{x}_t(t))}{\sum_{\mathbf{y} \in A} u(y_0, y_1, \dots)}$$

then,

$$u(x_0, x_1, \dots, x_T) = v(x_0) + \beta \sum_{t=1}^T \delta^t v(x_t)$$

□

## References

Aguirregabiria, V. and P. Mira (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics* 156(1), 38–67.

- An, Y., Y. Hu, and J. Ni (2014). Dynamic models with unobserved state variables and heterogeneity: time inconsistency in drug compliance. Technical report.
- Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2008). Eliciting risk and time preferences. *Econometrica* 76(3), 583–618.
- Apestequia, J. and M. A. Ballester (2015). Monotone stochastic choice models: The case of risk and time preferences. Economics Working Papers 1499, Department of Economics and Business, Universitat Pompeu Fabra.
- Chabris, C. F., D. Laibson, C. L. Morris, J. P. Schuldt, and D. Taubinsky (2008). Individual laboratory-measured discount rates predict field behavior. *Journal of Risk and Uncertainty* 37(2-3), 237–269.
- Chen, J., S. Esteban, and M. Shum (2013). When do secondary markets harm firms? *American Economic Review* 103(7), 2911–34.
- Chevalier, J. and A. Goolsbee (2009). Are durable goods consumers forward-looking? Evidence from college textbooks. *The Quarterly Journal of Economics* 124(4), 1853–1884.
- Coller, M. and M. B. Williams (1999). Eliciting individual discount rates. *Experimental Economics* 2(2), 107–127.
- Dubé, J., G. J. Hitsch, and P. Jindal (2014). The joint identification of utility and discount functions from stated choice data: An application to durable goods adoption. *Quantitative Marketing and Economics* 12(4), 331–377.
- Echenique, F., K. Saito, and G. Tserenjigmid (2014). The perception-adjusted Luce model. Technical report.
- Fang, H. and Y. Wang (2015). Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. *International Economic Review* 56(2), 565–596.
- Fishburn, P. C. (1970). *Utility theory for decision making*. New York: Wiley.
- Fishburn, P. C. and A. Rubinstein (1982). Time preference. *International Economic Review* 23(3), 677–694.
- Fosgerau, M. and M. Bierlaire (2009). Discrete choice models with multiplicative error terms. *Transportation Research Part B: Methodological* 43(5), 494–505.
- Fudenberg, D. and T. Strzalecki (2015). Dynamic logit with choice aversion. *Econometrica* 83, 651–691.
- Giglio, S., M. Maggiori, and J. Stroebel (2015). Very long-run discount rates. *The Quarterly Journal of Economics* 130(1), 1–53.
- Gowrisankaran, G. and M. Rysman (2012). Dynamics of consumer demand for new durable goods. *Journal of Political Economy* 120(6), 1173–1219.
- Gul, F., P. Natenzon, and W. Pesendorfer (2014). Random choice as behavioral optimization. *Econometrica* 82(5), 1873–1912.
- Halevy, Y. (2015). Time consistency: stationarity and time invariance. *Econometrica* 83(1), 335–352.

- Harrison, G. W., M. I. Lau, and M. B. Williams (2002). Estimating individual discount rates in Denmark: A field experiment. *American Economic Review* 92(5), 1606–1617.
- Hayashi, T. (2003). Quasi-stationary cardinal utility and present bias. *Journal of Economic Theory* 112(2), 343 – 352.
- Hendel, I. and A. Nevo (2006). Measuring the implications of sales and consumer inventory behavior. *Econometrica* 74(6), 1637–1673.
- Koopmans, T. C. (1960). Stationary ordinal utility and impatience. *Econometrica* 28, 287.
- Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics* 112(2), 443–77.
- Louie, K. and P. W. Glimcher (2010). Separating value from choice: delay discounting activity in the lateral intraparietal area. *The Journal of Neuroscience* 30(16), 5498–5507.
- Lu, J. and K. Saito (2016). Random intertemporal choice. Technical report.
- Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.
- Marschak, J. (1959). Binary choice constraints on random utility indicators. Cowles Foundation Discussion Papers 74, Cowles Foundation for Research in Economics, Yale University.
- Matějka, F. and A. McKay (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review* 105(1), 272–98.
- Matějka, F., J. Steiner, and C. Stewart (2015). Rational inattention dynamics: Inertia and delay in decision-making. CEPR Discussion Papers 10720.
- Mattsson, L. and J. W. Weibull (2002). Probabilistic choice and procedurally bounded rationality. *Games and Economic Behavior* 41(1), 61–78.
- Mattsson, L., J. W. Weibull, and P. O. Lindberg (2014). Extreme values, invariance and choice probabilities. *Transportation Research Part B: Methodological* 59, 81–95.
- Montiel Olea, J. L. and T. Strzalecki (2014). Axiomatization and measurement of quasi-hyperbolic discounting. *The Quarterly Journal of Economics* 129, 1449–1499.
- Paserman, M. D. (2008). Job search and hyperbolic discounting: Structural estimation and policy evaluation. *The Economic Journal* 118(531), 1418–1452.
- Prelec, D. (2004). Decreasing impatience: A criterion for non-stationary time preference and "hyperbolic" discounting. *Scandinavian Journal of Economics* 106(3), 511–532.
- Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica* 55(5), 999–1033.
- Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, 158–174.

- Simonson, I. and A. Tversky (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*.
- Tanaka, T., C. F. Camerer, and Q. Nguyen (2010). Risk and time preferences: linking experimental and household survey data from Vietnam. *American Economic Review* 100(1), 557–71.
- Tarozzi, A. and A. Mahajan (2011). Time inconsistency, expectations and technology adoption: The case of insecticide treated nets. *Economic Research Initiatives at Duke (ERID) Working Paper* (105).
- Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters* 8(3), 201–207.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.
- Webb, R. (2015). The dynamics of stochastic choice. *Available at SSRN 2226018*.
- Yao, S., C. F. Mela, J. Chiang, and Y. Chen (2012). Determining consumers' discount rates with field studies. *Journal of Marketing Research* 49(6), 822–841.



Alma Mater Studiorum - Università di Bologna  
DEPARTMENT OF ECONOMICS

Strada Maggiore 45  
40125 Bologna - Italy  
Tel. +39 051 2092604  
Fax +39 051 2092664  
<http://www.dse.unibo.it>