



# An axiomatic foundation of conditional logit

Yves Breitmoser<sup>1</sup>

Received: 8 February 2019 / Accepted: 5 May 2020 / Published online: 3 June 2020  
© The Author(s) 2020

## Abstract

This paper considers a decision maker choosing from a set of options when options have multiple real-valued attributes. Assuming DM chooses all options with positive probability, four invariance assumptions are necessary and sufficient for choice probabilities to take McFadden’s conditional logit form: independence of irrelevant alternatives, translation invariance, presentation independence and context independence. Variations on these assumptions yield generalized logit and contextual logit models. This shows that even specific logit models have behavioral foundations in simple invariance assumptions involving observables only, which therefore are directly testable.

**Keywords** Conditional logit · Axiomatic foundation

**JEL Classification** D01 · C13

## 1 Introduction

Economic analyses typically rest on preference assumptions, and the resulting necessity to understand preferences inspired a large body of work developing methods to infer preferences from choice. The main difficulty is that choice is inherently stochastic, implying that preferences are not directly revealed. Structural modeling attempts to control for noise in the choice process, using explicit models of stochastic choice, and has a wide range of applications in empirical and behavioral analyses (for a “primer,” see Wilcox 2008). In order to apply models of stochastic choice, however, researchers need to specify the formal link between choice propensities and observables, as demonstrated by Axiom 4 in McFadden’s (1974) seminal characterization of conditional logit. This suggests that models of stochastic choice cannot be applied without making functional-form assumptions that are not directly testable (see, e.g., Keane 2010a; Nevo and Whinston 2010).

---

✉ Yves Breitmoser  
yves.breitmoser@uni-bielefeld.de

<sup>1</sup> Bielefeld University, Universitätsstr. 25, 33615 Bielefeld, Germany

The present paper seeks to contribute to this discussion by demonstrating that specific models of stochastic choice, several widely used logit models with linear links between propensities and observables, have axiomatic foundations in invariance assumptions and therefore do not require functional-form assumptions. These invariance assumptions involve solely observable attributes of options and are directly testable. Extensions to nonlinear link functions that are additively separable in observables (such as CES utilities) are equally possible.

Formally, I consider a decision maker (DM) choosing from a finite set of options. Each option is characterized by an observable attribute vector (say, payoffs in different states of the world, or prices, quantities and qualities of products), and each option is chosen with positive probability. Within this framework, and given an essentialness condition requiring that DM is not indifferent with respect to any of the attributes, IIA and two simple invariance assumptions, clarifying when choice responds to changes in attributes and when not, uniquely characterize the conditional logit model of McFadden (1974). On the one hand, " $\pi$ -invariance" requires choice to be invariant to translation of the attribute vector, and on the other, " $\pi$ -relevance" requires that choice responds "uniformly" to attribute changes other than translation (comprising "presentation independence" and "context independence," in a sense to be made precise). Empirical evidence seems to suggest that observed choice exhibits also a form of scaling invariance, however. After strengthening  $\pi$ -invariance and weakening  $\pi$ -relevance correspondingly, I obtain a multi-attribute generalization of the contextual utility model of Wilcox (2011).

Four points appear to be worth noting about these results. (1) All invariance assumptions solely involve choice probabilities and option attributes, which both are considered observable in applied work (for a foundation, see Gul et al. 2014) and thus directly testable. Applied work may therefore first test, for a given dataset, if the choice postulates underlying logit are satisfied, or if the postulates are satisfied after a transformation of attributes suggested by McFadden (1974), and only then apply logit if adequate. (2) The widely used conditional logit model, which assumes a linear utility function linking attributes and log-propensities of choice, is provided with a foundation void of the functional-form assumptions it has been criticized for (for discussion, see Keane 2010b; Rust 2010; Blundell 2010; Heckman and Urzua 2010). (3) For any given choice profile satisfying the invariance assumptions, the utility function linking attributes and log-propensities of choice is shown to be unique up to linear transformation, which is a novel uniqueness result that may be helpful in applications. (4) Observed behavior tends to exhibit a form of invariance to rescaling attributes that conditional logit does not accommodate, suggesting that alternate models such as the generalized form of contextual utility may indeed be more adequate in applied work.

To provide some context for the results, the general family of logit models is known to be characterized by positivity and IIA in the sense that choice probabilities then take the logit form for an unknown, potentially nonlinear utility function linking attributes and log-propensities of choice. My results provide testable conditions for this utility function to be uniquely linear in attributes or in pre-specified transformations thereof, as postulated by McFadden in his definition of conditional logit. The main condition is translation invariance, and interestingly, linear logit models

are already known to predict choice probabilities that are invariant to translation of option attributes (or, utilities in some work). This paper's contribution is therefore the result that translation invariance is sufficient if complemented by a condition clarifying when attribute changes are relevant, and the sufficiency finally removes the necessity to make functional-form assumptions.

The need for the additional condition arises, because translation invariance yields a functional equation and its solution includes non-trivial integration constants. These integration constants resemble prior probabilities as observed in the generalized logit model of Matejka and McKay (2015) and may be interpreted to represent presentation effects in choice (i.e., effects due to ordering and positioning of options). In order to formally represent and capture such presentation effects, I extend the standard framework of stochastic choice by explicitly distinguishing "contexts" of decisions, i.e., different mappings between options and attributes. This generalized framework, in turn, is the main difference to existing analyses of stochastic choice when options have multiple attributes, such as Gul et al. (2014), and allows us to establish the necessity and sufficiency of simple invariance assumptions in characterizing logit models with specific (here, linear) utility functions.

## 2 Related literature

We may distinguish at least three classes of approaches in the characterization of logit models. The original approach due to Luce (1959) demonstrates that a general family of nonlinear logit models is characterized by positivity and independence of irrelevant alternatives (IIA). That is, if choice probabilities satisfy positivity and IIA, then there exists a utility function  $v$  such that choice probabilities take the multinomial logit form. This utility function is not known, however, and applied research needs to specify a utility function. In order to provide a foundation for such work, McFadden (1974) introduces the conditional logit model, explicitly linking option attributes and choice probabilities without resorting to an unknown utility function. McFadden achieves this by introducing two additional axioms that fix the exponential structure (Axiom 3) and the additive separability (Axiom 4), but both assumptions involve non-observable utilities rendering them untestable (for discussion, see Breitmoser 2018).

The third class of approaches seeks to model the choice process explicitly in order to establish conditions such that the resulting choice probabilities take the logit form. All of these approaches involve non-observable entities as well, rendering direct tests impossible. For example, logit can be formulated as random utility model (Thurstone 1927; Block and Marschak 1960), but this involves non-observable utilities, a specific functional form assuming additive separability of utilities and perturbations, and it requires unobservable utility perturbations to be identically and independently distributed as extreme-value type I. Logit can also be characterized as the outcome of rational inattention (Matejka and McKay 2015), but this relies on assumptions involving non-observable utilities and assumptions relating non-observable costs of information acquisition to Shannon's measure of entropy. Logit's foundation in an additive perturbed utility representation (e.g., Fudenberg

et al. 2015) requires DM to maximize the difference of expected utility and perturbation costs and requires the unobservable perturbation costs to be proportional to the Shannon entropy. Finally, in another recent paper, Woodford (2014) characterizes logit choice as the solution to a specific optimization problem if a certain unobservable parameter ( $\rho$ ) is equal to 1.

The present paper is most closely related to McFadden (1974), in its attempt to establish testable conditions for specific (linear) utility functions to form the link between observable attributes of options and choice propensities. This objective of McFadden has received renewed interest in recent work on the theoretical foundation of logit. Specifically, Ahn et al. (2018) characterize the linear logit model if only average choices are observable, and Allen and Rehbeck (2019) analyze stochastic choice more generally if attributes vary between observations rather than option sets, noting that the latter was Luce's original approach when defining independence of irrelevant alternatives. In this paper, we allow for both, variation of option sets and variation of attributes, showing that they jointly characterize linear logit models.<sup>1</sup>

### 3 Definitions

Decision maker DM chooses option  $x$  from menu  $B$ . Menu  $B$  is finite subsets of some set  $X$ , and the set of all finite subsets of  $X$  is denoted as  $P(X)$ . Each option  $x \in X$  is associated with an attribute vector via  $\pi : X \rightarrow \mathbb{R}^n$ , which may define payoffs in different states of the world (as in decision theory) or payoffs to different agents as a function of the option chosen (as in game theory), or product bundles and prices (as in consumer choice). I refer to the attribute mapping  $\pi$  as the context of DM's decision and to the pair  $(\pi, B)$  as a choice task. Given choice task  $(\pi, B)$ , the probability that DM chooses  $x$  is denoted as  $\Pr(x|\pi, B)$ .

The set of choice tasks  $(\pi, B)$  that can be constructed is  $\mathcal{D} = \Pi \times P(X)$ .  $\Pi$  denotes the set of attribute mappings  $\pi$  that may be constructed by changing attributes such as quantities or prices, or, for example, by permuting the attribute mapping (rearranging options in a shop or on a screen, or by relabeling states of the world or co-players in experiments). As indicated, a formal expression of variations of  $\pi$  will be necessary to address "presentation effects" that come with integration constants below. For the purpose of the present paper, I assume that attributes are exogenously given by an experimental design or the analyst, but note that this may in practice not always be trivial (Gul et al. 2014).

<sup>1</sup> To be clear, while these two studies appear to be the most closely related amongst the recent ones, the Luce model in particular and stochastic choice in general have been studied fairly comprehensively recently. To give just a few examples, Koida (2018) studies stochastic choice influenced by positioning of objects in menu, Ryan (2018) studies axiomatic characterizations of logit in choice under risk or uncertainty, and Echenique and Saito (2019) study generalized Luce models that relax positivity.

I assume that the set of choice tasks  $\mathcal{D}$  satisfies the following conditions. Throughout this paper,  $\mathbb{R}_+$  denotes the set of positive reals.<sup>2</sup>

**Assumption 1 (Framework)** The set of choice tasks  $\mathcal{D} = \Pi \times P(X)$  satisfies

- A1 *Transformability*  $a + b\pi \in \Pi$  for all  $\pi \in \Pi$  and all  $a \in \mathbb{R}^n, b \in \mathbb{R}_+$ ,
- A2 *Surjectivity* for all  $\pi \in \Pi$  and  $k \leq n$ , the image  $\pi_k[X] = \{\pi_k(x) | x \in X\}$  is a bounded, convex and non-singleton subset of  $\mathbb{R}$ ,
- A3 *Richness* for all  $k \leq n$ , there exist  $(\pi, B)$  and  $x, y \in B$  such that  $\pi_k(x) \neq \pi_k(y)$  and  $\pi_{-k}(x) = \pi_{-k}(y)$ .

Besides being standard assumptions in microeconomic theory and in recent analyses of stochastic choice (Gul et al. 2014; Fudenberg et al. 2015), these conditions help us ensure that the representation derived below is unique. Specifically, transformability ensures that we may discuss reactions to affine transformations of attributes, by ensuring that all affine transformations are well-defined objects. Surjectivity rules out scarce choice environments where the set of feasible attribute vectors is finite or even singleton; but it will be notationally convenient to know that  $\pi[X]$  is convex and bounded in all dimensions. These assumptions are satisfied in choice tasks typically of interest in behavioral work (such as in choice under risk or in games, as payoffs can be varied almost continuously up to exogenous bounds). Note that the attribute functions may still be fairly ill-behaved, violating smoothness, monotonicity and continuity for any number points. Richness finally ensures that we may discuss reactions to uni-dimensional variations of the attribute vector. This is straightforwardly satisfied in decision tasks typically of interest to analysts, for example by direct manipulations of prices or payoffs.

Within this framework, we assume DM’s choice profile  $\Pr$  adheres to the following postulates.

**Assumption 2 (Postulates on choice probabilities)** There exists  $\epsilon : \Pi \rightarrow \mathbb{R}_+$  such that for all  $(\pi, B)$ :

- P1 *Essentialness*  $\pi_k(x) \neq \pi_k(y)$  and  $\pi_{-k}(x) = \pi_{-k}(y)$  implies  $\Pr(x|\pi, B) \neq \Pr(y|\pi, B)$ ,
- P2 *Positivity*  $\Pr(x|\pi, \{x, y\}) \geq \epsilon_\pi$  for all  $x, y \in B$ ,
- P3 *IIA*  $\frac{\Pr(x|\pi, B)}{\Pr(y|\pi, B)} = \frac{\Pr(x|\pi, B')}{\Pr(y|\pi, B')}$  for all  $x, y \in B \cap B'$  and all  $\pi \in \Pi$ ,
- P4  *$\pi$ -Invariance*  $\Pr(x|\pi, B) = \Pr(x|\pi + r, B)$  for all  $r \in \mathbb{R}^n$ ,
- P5  *$\pi$ -Relevance* if  $\pi_x = \pi'_x$  and  $\pi_y = \pi'_y$ , then  $\Pr(x|\pi, \{x, y\}) = \Pr(x'|\pi', \{x', y'\})$ .

<sup>2</sup> With slight abuse of notation, I further identify all real numbers as constant functions such that addition and multiplication of a function with a real are well defined. Thus, for any  $\pi : X \rightarrow \mathbb{R}^n$  and any  $a \in \mathbb{R}^n, b \in \mathbb{R}$ ,  $\pi' = a + b\pi$  is equivalent to  $\pi'_k(x) = a_k + b_k \pi_k(x)$  for all  $k \leq n$  and  $x \in X$ . As usual, I use  $\pi_k(x)$  to denote the attribute  $k \leq n$  of option  $x$  and  $\pi_{-k}(x)$  to the list of all attributes but  $k$ .

Essentialness requires that all dimensions of the attribute vector are relevant to DM. With respect to non-essential dimensions, the representation derived below would not be unique. Positivity allows that DM fails to maximize utility, however rarely, and captures the widely documented phenomenon that individual choice fluctuates and involves dominated options (Hey 2005). The above formulation of positivity requires that choice probabilities in all binary choice tasks are bounded below at a value strictly above zero, but this bound may be arbitrarily close to zero. Following McFadden (1974), positivity captures stochastic choice in a comparably mild manner, as an event occurring with zero probability is empirically indistinguishable from one occurring with positive but small probability.

Next,  $\pi$ -invariance requires the choice profile  $\text{Pr}$  to be invariant to translation of attribute mappings. While translation is directly testable, e.g., by varying show-up fees in experiments, I am not aware of studies directly testing it. A string of evidence suggesting that choice is translation invariant is provided by neuro-economic studies, which consistently find “adaptive coding” as I discuss in more detail when deliberating scaling invariance. Finally,  $\pi$ -relevance requires relative choice probabilities to be invariant across contexts if the option attributes are equivalent.

McFadden introduces the conditional logit model as a logit model where the log-propensities of choice are linear in option attributes (potentially after transforming attributes). This conditional logit model is a special case of the Luce model, which can be defined as follows:

**Definition 1** Choice profile  $\text{Pr}$  has a **Luce representation** if there exists  $V = \{V_\pi : X \rightarrow \mathbb{R}\}_{\pi \in \Pi}$  such that for all tasks  $(\pi, B) \in \mathcal{D}$  and options  $x \in B$ :

$$\text{Pr}(x|\pi, B) = V_\pi(x) / \sum_{y \in B} V_\pi(y). \quad (1)$$

Any such  $V$  is said to **admit a Luce representation**.

Based on this, we define conditional logit as follows (following McFadden 1974):

**Definition 2** The choice profile  $\text{Pr}$  has a **conditional logit representation** if there exists  $\lambda \in \mathbb{R}^n$  such that  $V$  with  $V_\pi(x) = \exp\{\lambda \cdot \pi_x\}$  for all  $x, \pi$  admits a Luce representation.

Note the abbreviated vector notation involving option attributes  $\pi_x \in \mathbb{R}^n$  for all  $x \in B$ .

Given Assumption 1, we will see that the above choice postulates are equivalent to the choice profile  $\text{Pr}$  taking the conditional logit form. That is, conditional logit is adequate if and only if the choice postulates are satisfied. Since the choice postulates solely involve observables, they are straightforwardly testable, which allows analysts to verify whether logit is an adequate model given their dataset. In practice, it may also be appropriate for analysts to determine whether the postulates are satisfied after invoking pre-specified transforms to the attributes and then run their analysis using these transforms, similar to power transforms such as Box–Cox transformations that

align data with normal distributions in other applied work. Such transforms allow the utility to simply be additively separable in attributes, which contains many well-known utility functions (most obviously CES utilities) as special cases.

### 4 Analysis

As shown by Luce, positivity and IIA imply that choice probabilities have a Luce representation, i.e., for each  $\pi$ , a propensity function  $V_\pi : X \rightarrow \mathbb{R}$  exists such that  $\Pr(x|\pi, B) = V_\pi(x) / \sum_{y \in B} V_\pi(y)$ . The following result also clarifies that the choice propensities  $V_\pi$  are context dependent, as IIA itself does not restrict choice across contexts  $\pi$ , and thus may involve arbitrary statistics of  $\pi$  (such as suprema or infima) as constants. Given context  $\pi$ , however,  $V_\pi$  can be expressed as a function of the option itself ( $x$ ) and its attributes  $\pi_x$ .

**Lemma 1** *Pr satisfies P2–P3  $\Rightarrow$  Pr has a Luce representation.*

The routine proof is relegated to “Appendix.” It should be clear that the functional forms of  $V_\pi$  are unrestricted and not directly observable by the analyst. There simply exists a family of unknown utility functions  $\{V_\pi\}$  admitting a Luce representation, and the question we ask is if there are testable conditions such that  $\{V_\pi\}$  assumes specific functional forms—focusing on the linear form assumed in McFadden’s definition of conditional logit.

Next assume that choice is also  $\pi$ -invariant, i.e., invariant to translations of contexts. On its own, this implies a comparably modest refinement of the set of propensity functions in relation to those compatible with just positivity and IIA. Here and in the following, writing “ $\text{cinf}\pi$ ” and later “ $\text{csup}\pi$ ”, I refer to  $\pi$ ’s **componentwise infimum and supremum**, respectively, over its domain ( $X$ ), i.e.,  $\text{cinf}\pi = (\inf_{x \in X} \pi_k(x))_{k \leq n}$  and  $\text{csup}\pi = (\sup_{x \in X} \pi_k(x))_{k \leq n}$ .

**Lemma 2** *Pr satisfies P2–P4  $\Rightarrow$  Pr has a Luce representation where for all  $r \in \mathbb{R}^n$ ,  $V_{\pi+r}$  is a linear transformation of  $V_\pi$ .*

The requirement of  $\pi$ -invariance has further implications once we take its counterpart  $\pi$ -relevance into account, but on its own, it poses no restriction on the functional form in our framework. This will be illustrated after the proof.

**Proof** By Lemma 1, there exists a collection of functions  $(V_\pi)_{\pi \in \Pi}$  such that  $\Pr(x|\pi, B) = V_\pi(x) / \sum_{y \in B} V_\pi(y)$  for all  $x, B, \pi$ . Now fix  $\pi \in \Pi$  and note that, given this representation of  $\Pr$ , by **P4** we obtain

$$\frac{V_\pi(x)}{\sum_{y \in B} V_\pi(y)} = \frac{V_{\pi+r}(x)}{\sum_{y \in B} V_{\pi+r}(y)} \quad \text{for all } r \in \mathbb{R}^n \text{ and } (\pi, B) \in \mathcal{D}. \tag{2}$$

By positivity (**P2**), the values of  $V_\pi$  and  $V_{\pi+r}$  are nonzero. Hence, there exists  $c \in \mathbb{R}$  such that  $V_{\pi+r} = c \cdot V_\pi$ . To see this, assume for contradiction that there is no such

constant. Then, there exist  $x, y$  such that  $V_{\pi+r}(x) = c_1 \cdot V_\pi(x)$  and  $V_{\pi+r}(y) = c_2 \cdot V_\pi(y)$  with  $c_1 \neq c_2$ . By **P4**,

$$\frac{V_\pi(x)}{V_\pi(y)} = \frac{V_{\pi+r}(x)}{V_{\pi+r}(y)} = \frac{c_1 \cdot V_\pi(x)}{c_2 \cdot V_\pi(y)}, \tag{3}$$

implying  $c_1 = c_2$ , the contradiction, i.e.,  $V_{\pi+r}$  is a linear transformation of  $V_\pi$ .  $\square$

Now, to illustrate, fix a one-dimensional attribute mapping  $\pi : X \rightarrow \mathbb{R}$  and assume  $\pi_x = 2$  and  $\pi_y = 0$ . Also consider  $\pi' = \pi + 8$ , which implies  $\pi'_x = 10$  and  $\pi'_y = 8$ . By  $\pi$ -invariance, the relative probability of choosing  $x$  over  $y$  is equal in both contexts  $\pi$  and  $\pi'$ . Two seemingly related invariances are not implied, and those prevent us from taking full advantage of  $\pi$ -invariance. On the one hand, assume there exist  $x', y' \in X$  with payoffs 10 and 8 in the original context  $\pi$ , i.e.,  $\pi_{x'} = 10$  and  $\pi_{y'} = 8$ . Translation invariance does not imply that the relative probability of choosing 10 ( $x'$ ) over 8 ( $y'$ ) in context  $\pi$  is equal to the one of choosing 2 ( $x$ ) over 0 ( $y$ ) in context  $\pi$ —although we know that choosing between 2 and 0 under  $\pi$  is equivalent to choosing between 10 and 8 in a different context  $\pi'$ . I refer to this phenomenon as “presentation effect”: The probability of choosing an option with a given outcome may depend on which option it is. Presentation effects may reflect labeling, ordering or positioning of options. They are compatible with  $V_\pi(x)$  as the option  $x$  itself is choice relevant, not just its attributes  $\pi_x$ . Presentation independence results if choice is invariant to permutation of options, and in order to express invariance to permutation, we need to distinguish contexts, since each permutation represents a different context  $\pi$ .

On the other hand, fix  $\pi''$  such that  $\pi''_x = 2$  and  $\pi''_y = 0$ , but  $\pi \neq \pi''$ . Hence,  $\pi''$  is not a translation of  $\pi$ , and choice propensities in contexts  $\pi$  and  $\pi''$  are unrelated by Lemma 2. Hence, the relative probabilities of choosing options  $x$  and  $y$  may well differ between these contexts, despite the equality of attributes and options assuming these attributes, which I will call “context dependence.” Strict context independence obtains if for all  $\pi, \pi' \in \Pi$  and all  $x, y \in X$ :

$$\pi_x = \pi'_x \quad \text{and} \quad \pi_y = \pi'_y \quad \Rightarrow \quad \Pr(x|\pi, \{x, y\}) = \Pr(x|\pi', \{x, y\}). \tag{4}$$

The postulate of  $\pi$ -relevance introduced above combines presentation and context independence, and with its help, we arrive at the conditional logit representation with a necessarily (log-)linear value function.

**Theorem 1** *Pr has a conditional logit representation with  $\lambda_k \neq 0$  for all  $k \leq n \Leftrightarrow$  Pr satisfies P1–P5. In addition, any  $V = \{V_\pi\}_{\pi \in \Pi}$  admitting a Luce representation is a collection of functions that are linear transformations of another.*

**Proof** First, we establish  $\Rightarrow$ . For any  $\pi$  and any  $k \leq n$ , consider any  $x, y$  such that  $\pi_k(x) \neq \pi_k(y)$  and  $\pi_{-k}(x) = \pi_{-k}(y)$  (A3). By assumption, Pr has a conditional logit representation with  $\lambda_k \neq 0$ , implying  $\Pr(x|\pi, B) \neq \Pr(y|\pi, B)$  for all  $B \supseteq \{x, y\}$ , and thus, essentialness is satisfied (**P1**). By surjectivity (A2), for any  $\pi$ , the image  $\pi[X]$  is

bounded, implying that  $Y_\pi = \{\lambda\pi_x \mid x \in X\}$  is bounded and that  $\Pr\{x|\pi, \{x, y\}\} \geq \epsilon_\pi$  for all  $x, y \in X$  with  $\epsilon_\pi := \inf Y_\pi / (\sup Y_\pi + \inf Y_\pi) > 0$  (yielding positivity, P2). The remaining properties (P3–P5) follow from the conditional logit representation.

The remainder establishes  $\Leftarrow$ .

*Step 1 (Representation independently of  $x$ ):*

Pick any  $\pi \in \Pi$  and  $x, y \in X$ . By **P5**,  $\pi_x = \pi_y$  implies  $\Pr(x|\pi, B) = \Pr(y|\pi, B)$  and thus  $V_\pi(x) = V_\pi(y)$ . Hence, choice propensities in any given context  $\pi \in \Pi$  solely depend on attributes, and we can define a function  $\tilde{V}_\pi : \mathbb{R}^n \rightarrow \mathbb{R}_+$  such that

$$\Pr(x|\pi, B) = \frac{\tilde{V}_\pi(\pi_x)}{\sum_{y \in B} \tilde{V}_\pi(\pi_y)} \quad \text{for all } x \in B, (\pi, B) \in \mathcal{D}. \tag{5}$$

Note that this does not rule out presentation effects entirely, but  $\pi_x$  contains the information required to implicitly represent presentation effects for any  $\pi$ .

*Step 2 (Narrow bracketing and presentation independence):*

Define  $x, y, x', y' \in X$  and  $\pi, \pi' \in \Pi$  such that (1)  $\pi' = \pi + r$ , (2)  $\pi'_y = \pi_y$ , and (3)  $\pi'_x = \pi_x$ , for some  $r \in \mathbb{R}^n$ , which is possible by surjectivity (A2). By **P5** (first equality) and **P4** (second equality),

$$\frac{\Pr(x'|\pi, \{x', y'\})}{\Pr(y'|\pi, \{x', y'\})} = \frac{\Pr(x|\pi', \{x, y\})}{\Pr(y|\pi', \{x, y\})} = \frac{\Pr(x|\pi, \{x, y\})}{\Pr(y|\pi, \{x, y\})}. \tag{6}$$

Using the representation from Eq. (5), for all  $r < (c\sup\pi - c\inf\pi)/2$  and all  $B \in P(X)$ ,

$$\frac{\tilde{V}_\pi(\pi_x)}{\sum_{y \in B} \tilde{V}_\pi(\pi_y)} = \frac{\tilde{V}_\pi(\pi_x + r)}{\sum_{y \in B} \tilde{V}_\pi(\pi_y + r)} \quad \text{for all } x \in B \text{ and } (\pi, B) \in \mathcal{D}. \tag{7}$$

Hence,

$$\frac{\tilde{V}_\pi(\pi_x)}{\tilde{V}_\pi(\pi_y)} = \frac{\tilde{V}_\pi(\pi_x + r)}{\tilde{V}_\pi(\pi_y + r)} \tag{8}$$

for all  $x, y$  and  $r$ , implying that there exists some function  $h_\pi : \mathbb{R}^n \rightarrow \mathbb{R}$  such that  $\tilde{V}_\pi(\pi_x + r) = \tilde{V}_\pi(\pi_x) \cdot h_\pi(r)$  for all  $x$  and  $r$ . Defining  $\hat{V}_\pi = \log \tilde{V}_\pi$  as well as  $\hat{h}_\pi = \log h_\pi$ , we obtain  $\tilde{V}_\pi(\pi_x + r) = \hat{V}_\pi(\pi_x) + \hat{h}_\pi(r)$ . By positivity,

$$\frac{1 - \epsilon_\pi}{\epsilon_\pi} \geq \frac{\Pr(x|\pi, \{x, y\})}{\Pr(y|\pi, \{x, y\})} = \frac{\tilde{V}_\pi(x)}{\tilde{V}_\pi(y)} \geq \frac{\epsilon_\pi}{1 - \epsilon_\pi} \quad \text{for all } x, y \in X,$$

implying that  $\hat{V}_\pi = \log \tilde{V}_\pi$  is bounded above and below. By surjectivity (A2),  $\hat{V}_\pi$  is defined on sets of positive measure in  $\mathbb{R}$  for all dimensions  $k \leq n$  of the attribute vector. Thus, all solutions of this fundamental Pexider functional equation satisfy  $\hat{V}_\pi(\pi_x) = \lambda_\pi \cdot \pi_x + c_\pi$  for all  $x$ , with unique  $\lambda_\pi \in \mathbb{R}^n$  and some  $c_\pi \in \mathbb{R}$  (Aczél and Dhombres 1989, Corollary 10, p. 43, in conjunction with Theorem 8, p. 17). Using  $\tilde{V}_\pi = \exp \hat{V}_\pi$  and the relation of  $\tilde{V}_\pi$  to  $V_\pi$ , this yields  $V_\pi(x) = \exp\{\lambda_\pi \cdot \pi_x + c_\pi\}$  for all  $x \in X$ . Hence,

$$\Pr(x|\pi, B) = \frac{\exp\{\lambda_\pi \cdot \pi_x + c_\pi\}}{\sum_{y \in B} \exp\{\lambda_\pi \cdot \pi_y + c_\pi\}}. \tag{9}$$

Finally, again by P4 (translation invariance),

$$\Pr(x|\pi + r, B) = \frac{\exp\{\lambda_{\pi+r} \cdot (\pi_x + r) + c_{\pi+r}\}}{\sum_{y \in B} \exp\{\lambda_{\pi+r} \cdot (\pi_y + r) + c_{\pi+r}\}} = \frac{\exp\{\lambda_\pi \cdot \pi_x + c_\pi\}}{\sum_{y \in B} \exp\{\lambda_\pi \cdot \pi_y + c_\pi\}} = \Pr(x|\pi, B).$$

which implies  $\lambda_\pi = \lambda_{\pi+r}$  for all  $r \in \mathbb{R}^n$ .

*Step 3 (Context independence):*

Finally, we establish  $\lambda_\pi = \lambda_{\pi'}$  for all  $\pi, \pi'$ . For the purpose of contradiction, assume there exist  $\pi, \pi' \in \Pi$  such that  $\lambda_\pi \neq \lambda_{\pi'}$ . Since  $\pi_x = \pi'_x$  and  $\pi_y = \pi'_y$ , imply, by P5 and Eq. (9),

$$\frac{\Pr(x|\pi, B)}{\Pr(y|\pi, B)} = \frac{\Pr(x'|\pi', B)}{\Pr(y'|\pi', B)} \Rightarrow \frac{\exp\{\lambda_\pi \cdot \pi_x\}}{\exp\{\lambda_\pi \cdot \pi_y\}} = \frac{\exp\{\lambda_{\pi'} \cdot \pi_{x'}\}}{\exp\{\lambda_{\pi'} \cdot \pi_{y'}\}},$$

$\lambda_\pi \neq \lambda_{\pi'}$  can be satisfied only if there are no such  $x, x', y, y'$ . However, now consider the contexts  $\tilde{\pi} = \pi - \text{cinf}\pi$  and  $\tilde{\pi}' = \pi' - \text{cinf}\pi'$ . By A2 (surjectivity), these contexts overlap in the sense that there must exist  $x, y, x', y'$  such that  $\tilde{\pi}_x = \tilde{\pi}'_{x'}$  and  $\tilde{\pi}_y = \tilde{\pi}'_{y'}$ , implying

$$\frac{\Pr(x|\tilde{\pi}, B)}{\Pr(y|\tilde{\pi}, B)} = \frac{\Pr(x'|\tilde{\pi}', B)}{\Pr(y'|\tilde{\pi}', B)}$$

by P5 and  $\lambda_{\tilde{\pi}} = \lambda_{\tilde{\pi}'}$  by Eq. (9). As observed above, P4 implies  $\lambda_\pi = \lambda_{\tilde{\pi}}$  and  $\lambda_{\pi'} = \lambda_{\tilde{\pi}'}$ , the contradiction.

Hence, there exists  $\lambda \in \mathbb{R}$  such that  $\lambda_\pi = \lambda$  for all  $\pi$ , implying  $V_\pi(x) = \exp\{\lambda \cdot \pi_x + c_\pi\}$  for all  $x, \pi$  and that  $V_\pi$  has the conditional logit form defined up to linear transformation. □

### 5 Incorporating scaling invariance

A range of empirical and experimental studies suggests that  $\pi$ -relevance as introduced above may be too strict. Specifically, choice in many experiments appears to be invariant to scaling option attributes (i.e., to scaling payoffs of players), and this seems to contradict  $\pi$ -relevance. There is a caveat to this, however. The evidence suggesting that behavior is scale invariant stems from comparing observations between experiments (or, between experimental subjects). The most direct evidence that I am aware is provided by meta-analyses of experimental behavior, which consistently find that decisions are independent of the amounts of money at stake, for example in dictator games (Engel 2011), ultimatum games (Oosterbeek et al. 2004; Cooper and Dutcher 2011) and trust games (Johnson and Mislin 2011). An explanation for such scale invariance, and thus indirect evidence, is provided by the neuro-economic result called “adaptive coding” (see Tremblay and Schultz 1999, and the

recent survey of Camerer et al. 2017): The neuronal representation of option payoffs adapts to the range of feasible payoffs. The baseline activity of the cell encoding the value of a given object adapts to the minimum of the payoff range in a given context, and its peak activity adapts to the maximum of the payoff range. Thus, choice ends up being invariant to scaling, and to changes in background income as indicated above, which appears to falsify the strict form of  $\pi$ -relevance postulated before.<sup>3</sup>

However, there is little direct evidence for scale invariance within subjects (see, however, Wilcox 2011, 2015). That is, if subjects are presented pairs of decision problems that are equivalent up to scaling, in sufficiently quick succession such that the neuronal representation does not adapt, it is not clear that behavior would actually be scale invariant indeed. Intuitively, if a high-stake decision is immediately followed by a “seemingly trivial” low-stake decision, scale invariance might be violated. I am not aware of experimental evidence directly testing this intuition, but such tests are clearly conceivable.

In general, though, substantial rescaling of option attributes of decision problems presented in quick succession is rarely a concern in analyses. Instead, rescaling tends to be concern in analyses merging data obtained independently under various conditions, under varying treatments or in different experiments, and in such cases, where scale invariance seems confirmed by meta-analyses, one may wish to adapt  $\pi$ -relevance in order to acknowledge scale invariance. The following two choice postulates weaken  $\pi$ -relevance and strengthen  $\pi$ -invariance in order to additionally acknowledge such scale invariance.

**Assumption 3** (Alternative postulates on choice probabilities)

P6 Strong  $\pi$ -invariance:  $\Pr(\cdot|\pi, B) = \Pr(\cdot|a + b \pi, B)$  for all  $a \in \mathbb{R}^n$  and  $b \in \mathbb{R}_+^n$

P7 Weak  $\pi$ -relevance:  $\pi$ -relevance if  $c\text{sup}\pi - \text{cinf}\pi = c\text{sup}\pi' - \text{cinf}\pi'$

If we adopt these postulates instead of **P4** and **P5**, then it follows immediately that the scale of the attribute range must factor out. That is,  $\Pr$  satisfies P1–P3 and P6 if and only if  $\Pr$  has Luce representation with  $V_\pi$  being linear transformations of some function ( $\odot$  is used to denote componentwise division of vectors)

$$\tilde{V}_\pi(\pi_x \odot (c\text{sup}\pi - \text{cinf}\pi)),$$

where  $\tilde{V}_\pi = \tilde{V}_{a+b\pi}$  for all  $a \in \mathbb{R}^n, b \in \mathbb{R}_+^n$ . Additionally invoking P7 yields a multi-attribute variation of the contextual logit model proposed by Wilcox (2011).

<sup>3</sup> In an interesting recent paper, Stevenson et al. (2019) demonstrate how a weak form of scale invariance called “divisive normalization” (also motivated by neuro-economic evidence) partially characterizes a model of stochastic choice that violates IIA but is otherwise comparable to the contextual logit model characterized below. Two key differences are that unavailable options (called “context” below) do not matter in the divisive normalization model and that the characterization is subject to an unknown utility function  $v$  that we seek to characterize simultaneously.

**Definition 3** The choice profile  $\Pr$  has a **contextual logit representation** if there exists  $\lambda \in \mathbb{R}^n$  such that  $V$  with  $V(x) = \exp\{\lambda \cdot [\pi_x \circ (\text{csup}\pi - \text{cinf}\pi)]\}$  for all  $x, \pi$  admits a Luce representation.

**Theorem 2**  $\Pr$  has a contextual logit representation with  $\lambda_k \neq 0$  for all  $k \leq n \Leftrightarrow \Pr$  satisfies P1–P3 and P6–P7.

The formal proof is relegated to “Appendix,” as it largely resembles that for conditional logit. In conclusion, let me briefly discuss a few points related to Theorem 2. To begin with, let me clarify the relation to contextual logit as defined by Wilcox (2011). Wilcox considers binary choice from lotteries, where lotteries are distributions over three outcomes  $z_1 < z_2 < z_3$ . These three outcomes are called the context of the decision; the specification of the lotteries may change while the context  $(z_1, z_2, z_3)$  is held constant. Lotteries over these outcomes are denoted  $S$  and  $T$ , and utilities from lotteries are denoted  $V(S)$  and  $V(T)$ . Given this, Wilcox defines the contextual logit representation as

$$\Pr(S) = H\left(\lambda \cdot \frac{V(S) - V(T)}{V(z_3) - V(z_1)}\right) \quad \text{with} \quad H(x) = 1/(1 + \exp\{-x\})$$

$V(z_3)$  and  $V(z_1)$  denote the utilities from the degenerate lotteries yielding the maximal outcome and minimal outcome (respectively) in context  $(z_1, z_2, z_3)$  with probability 1. Given the logit specification of  $H$ , this is equivalent to

$$\Pr(S) = \frac{\exp\{\lambda f(S)\}}{\exp\{\lambda f(S)\} + \exp\{\lambda f(T)\}} \quad \text{with} \quad f(y) = \frac{V(y)}{V_{\max} - V_{\min}},$$

where  $V_{\max} := V(z_3)$  and  $V_{\min} := V(z_1)$ . The above definition of contextual logit extends Wilcox’ definition, who considered a decision maker caring about one option attribute (expected utility), straightforwardly to multiple attributes.

One concern about applying the contextual logit model may be that the attributes of unavailable options matter. In Wilcox’ formulation, the “context” is defined by the degenerate lotteries that yield either outcome with certainty, which may be unavailable to DM yet tend to be observable by the analyst, but in general, the context of a decision may be subjective and therefore unobservable by the analyst (see, for example, Panizza et al. 2019). In order to capture choice in line with the contextual logit model defined above, it suffices to find some measure for the scale of the choice task and rescale attributes correspondingly, but to the extent the scale is subjective, rescaling may not be trivial in all applications.

Finally, invariance with respect to heterogeneous scaling across dimensions, i.e., with respect to scaling the attribute vector by a vector  $b \in \mathbb{R}^n$ , can be shown (analogously to the proof of Theorem 1) to yield a so-called strict utility model:  $V(\pi(x)) = \prod_k \pi_k(x)^{\lambda_k}$ . To see this, note that scaling invariance is equivalent to

translation invariance of logarithmized attributes.<sup>4</sup> Along these lines, many more models of stochastic choice may be found to have behavioral foundations in invariance assumptions.

**Acknowledgements** Open Access funding provided by Projekt DEAL. I thank the editor, an associate editor, a very constructive and patient reviewer, Niels Boissonnet, Friedel Bolle, Herbert Dawid, Nick Netzer, Martin Pollrich, Frank Riedel, Sebastian Schweighofer-Kodritsch, Felix Weinhardt, Georg Weizsäcker and audiences in Bielefeld, Heidelberg, at the BERA workshop in Berlin, and at THEEM 2016 in Kreuzlingen for many helpful comments. Financial support of the DFG (Project BR 4648/1 and CRC TRR 190) is greatly appreciated.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix

### Appendix 1: Proof of Lemma 1

To prove that IIA implies Luce, note first that  $\Pr(x|\pi, \{x, y\})$  is in general a function of  $x, y, \pi_x, \pi_y$ . By positivity, it is possible to define  $V(x, y, \pi_x, \pi_y) := \Pr(x|\pi, \{x, y\}) / \Pr(y|\pi, \{x, y\})$ , and thus by IIA (see McFadden 1974, p. 109, for details),

$$\Pr(x|\pi, B) = \frac{V(x, y, \pi_x, \pi_y)}{\sum_{x' \in B} V(x', y, \pi_{x'}, \pi_y)} \quad \text{for all } x, y \in B \text{ and all } B \in P(X). \quad (10)$$

Since this holds true for all  $x, y \in B$  and all  $B \in P(X)$ , it does so in particular for all possible benchmarks  $y \in X$ . Hence, the odds of choosing  $x$  over  $x'$  are constant for any pair of benchmark options  $y, y' \in X$ :

$$\frac{\Pr(x|\pi, B)}{\Pr(x'|\pi, B)} = \frac{V(x, y, \pi_x, \pi_y)}{V(x', y, \pi_{x'}, \pi_y)} = \frac{V(x, y', \pi_x, \pi_{y'})}{V(x', y', \pi_{x'}, \pi_{y'})} \quad \text{for all } x, x', y, y' \in B \text{ and all } B \in P(X).$$

As a result, functions  $f(y, \pi_y)$  and  $V_1(x, \pi_x)$  exist such that  $V(x, y, \pi_x, \pi_y) = V_1(x, \pi_x) \cdot f(y, \pi_y)$  for all  $x, y \in X$ , and we can write, for all  $B \in P(X)$ ,  $x \in B$  and  $y \in X$ ,

<sup>4</sup> For a loosely related result involving scaling invariance, see Dagsvik (2018), who focuses on binary choice and continuity of choice probabilities in observables, however.

$$\Pr(x|\pi, B) = \frac{V(x, y, \pi_x, \pi_y)}{\sum_{x' \in B} V(x', y, \pi_{x'}, \pi_y)} = \frac{V_1(x, \pi_x)}{\sum_{x' \in B} V_1(x', \pi_{x'})}$$

Since this holds independently for each  $\pi$ , function  $V_1$  depends on  $\pi$ , and we can write  $V_\pi(x) = V_1(x, \pi_x)$  for all  $x, \pi$ . □

### Appendix 2: Proof of Theorem 2

$\Rightarrow$  is established as in the proof of Theorem 1. The remainder establishes  $\Leftarrow$ . We have to show that given P1–P3 and P6, P7 implies contextual logit. First, extend the domain of  $\pi$  to be a set function. Thus, P7 implies that for all  $\pi, \tilde{\pi} \in \Pi$ , all  $B, \tilde{B} \in P(X)$ , and all  $x \in B, y \in \tilde{B}$ ,

$$\pi(B) = \tilde{\pi}(\tilde{B}) \quad \text{and} \quad \pi_x = \tilde{\pi}_y \quad \Leftrightarrow \quad \Pr(x | \pi, B) = \Pr(y | \tilde{\pi}, \tilde{B}) \quad (11)$$

if  $\text{csup}\pi - \text{cinf}\pi = \text{csup}\pi' - \text{cinf}\pi'$ .

*Step 1 (Incorporating scaling invariance):*

Given the representation established in Lemma 2, we need to derive the additional implication of “scaling invariance” (i.e., the incremental effect of P6 over P4). As in Step 1 in the Proof of Theorem 1, we can demonstrate that there exist functions  $\{\hat{V}_\pi\}_\pi$  such that

$$\Pr(x|\pi, B) = \frac{\hat{V}_\pi(\pi_x)}{\sum_{y \in B} \hat{V}_\pi(\pi_y)}$$

for all  $x, \pi, B$ , and translating as well as rescaling these functions, it follows that there exist functions  $\{\tilde{V}_\pi\}_\pi$  such that

$$\Pr(x|\pi, B) = \frac{\tilde{V}_\pi((\pi_x - \text{cinf}\pi) \oslash (\text{csup}\pi - \text{cinf}\pi))}{\sum_{y \in B} \tilde{V}_\pi((\pi_y - \text{cinf}\pi) \oslash (\text{csup}\pi - \text{cinf}\pi))}$$

for all  $x, \pi, B$ . By P6,  $\Pr(x|\pi, B) = \Pr(x|\pi', B)$  if  $\pi' = a + b\pi$  for some  $a \in \mathbb{R}^n$  and  $b \in \mathbb{R}_+^n$ , and since  $(\pi_x - \text{cinf}\pi) \oslash (\text{csup}\pi - \text{cinf}\pi) = (\pi'_x - \text{cinf}\pi') \oslash (\text{csup}\pi' - \text{cinf}\pi')$  then, this yields

$$\frac{\tilde{V}_\pi((\pi'_x - \text{cinf}\pi') \oslash (\text{csup}\pi' - \text{cinf}\pi'))}{\sum_{y \in B} \tilde{V}_\pi((\pi'_y - \text{cinf}\pi') \oslash (\text{csup}\pi' - \text{cinf}\pi'))} = \frac{\tilde{V}_{\pi'}((\pi'_x - \text{cinf}\pi') \oslash (\text{csup}\pi' - \text{cinf}\pi'))}{\sum_{y \in B} \tilde{V}_{\pi'}((\pi'_y - \text{cinf}\pi') \oslash (\text{csup}\pi' - \text{cinf}\pi'))}$$

for all  $B \in P(X), x \in B$ . Hence,  $\tilde{V}_{\pi'}$  is a linear transformation of  $\tilde{V}_\pi$ .

*Step 2 (solving the functional equation):*

Now, fix  $\pi \in \Pi$  such that  $\text{csup}\pi - \text{cinf}\pi = \mathbf{1}$  and  $\text{cinf}\pi = \mathbf{0}$  (such  $\pi$  exists by transformability A1). Hence, using  $\tilde{V}$  as defined in Step 1,

$$\Pr(x|\pi, B) = \frac{\tilde{V}_\pi((\pi_x - \text{cinf}\pi) \oslash (\text{csup}\pi - \text{cinf}\pi))}{\sum_{y \in B} \tilde{V}_\pi((\pi_y - \text{cinf}\pi) \oslash (\text{csup}\pi - \text{cinf}\pi))} = \frac{\tilde{V}_\pi(\pi_x)}{\sum_{y \in B} \tilde{V}_\pi(\pi_y)},$$

i.e., conditional on the fixed context  $\pi$ , we may follow the arguments in the proof of Theorem 1 up to Eq. (9) and obtain (abusing the fraction sign to denote componentwise division  $\oslash$ )

$$\Pr(x|\pi, B) = \frac{\exp\{\lambda_\pi \cdot \pi_x + c_\pi\}}{\sum_{y \in B} \exp\{\lambda_\pi \cdot \pi_y + c_\pi\}} = \frac{\exp\left\{\frac{\lambda_\pi \cdot \pi_x - \text{cinf}\pi}{\text{csup}\pi - \text{cinf}\pi} + c_\pi\right\}}{\sum_{y \in B} \exp\left\{\frac{\lambda_\pi \cdot \pi_y - \text{cinf}\pi}{\text{csup}\pi - \text{cinf}\pi} + c_\pi\right\}},$$

with  $\lambda_\pi \in \mathbb{R}^n$  and  $c_\pi \in \mathbb{R}$ , where  $\lambda_{\pi+r} = \lambda_\pi$  for all  $r$ . As demonstrated in Step 1, this implies for any  $\pi'$  such that  $\pi' = a + b\pi$  for some  $a \in \mathbb{R}^n$  and  $b \in \mathbb{R}_+^n$ , still abusing the fraction sign to denote componentwise division  $\oslash$ ,

$$\begin{aligned} \Pr(x|\pi', B) &= \Pr(x|\pi, B) = \frac{\exp\{\lambda_\pi \cdot \pi_x + c_\pi\}}{\sum_{y \in B} \exp\{\lambda_\pi \cdot \pi_y + c_\pi\}} \\ &= \frac{\exp\left\{\frac{\lambda_\pi \cdot (\pi'_x - \text{cinf}\pi')}{\text{csup}\pi' - \text{cinf}\pi'} + c_{\pi'}\right\}}{\sum_{y \in B} \exp\left\{\frac{\lambda_\pi \cdot (\pi'_y - \text{cinf}\pi')}{\text{csup}\pi' - \text{cinf}\pi'} + c_{\pi'}\right\}} = \frac{\exp\left\{\frac{\lambda_\pi \cdot \pi'_x}{\text{csup}\pi' - \text{cinf}\pi'} + c_{\pi'}\right\}}{\sum_{y \in B} \exp\left\{\frac{\lambda_\pi \cdot \pi'_y}{\text{csup}\pi' - \text{cinf}\pi'} + c_{\pi'}\right\}} \end{aligned}$$

for all  $B \in P(X)$ ,  $x \in B$ . Hence, for any such  $\pi'$ ,

$$\Pr(x|\pi', B) = \frac{\exp\{\lambda_{\pi'} \cdot \pi'_x \oslash (\text{csup}\pi' - \text{cinf}\pi') + c_{\pi'}\}}{\sum_{y \in B} \exp\{\lambda_{\pi'} \cdot \pi'_y \oslash (\text{csup}\pi' - \text{cinf}\pi') + c_{\pi'}\}},$$

with  $\lambda_{\pi'} = \lambda_\pi$ .

*Step 3 (Weak context and presentation independence):*

Let  $\Pi' \subset \Pi$  denote the set of contexts such that  $\pi \in \Pi'$  if and only if  $\text{csup}\pi - \text{cinf}\pi = \mathbf{1}$ . Given this, we can follow Step 3 in the proof of Theorem 1 to establish that there exists  $\lambda \in \mathbb{R}^n$  such that, for all  $\pi \in \Pi'$

$$\Pr(x|\pi, B) = \frac{\exp\{\lambda \cdot \pi_x + c_\pi\}}{\sum_{y \in B} \exp\{\lambda \cdot \pi_y + c_\pi\}}$$

for some  $c_\pi \in \mathbb{R}$  (which cancels out). The claimed extension to contexts  $\pi$  with  $\text{csup}\pi - \text{cinf}\pi \neq \mathbf{1}$  follows directly from P7. □

## References

- Aczél, J., Dhombres, J.G.: *Functional Equations in Several Variables*. Cambridge University Press, Cambridge (1989)
- Ahn, D.S., Echenique, F., Saito, K.: On path independent stochastic choice. *Theor. Econ.* **13**(1), 61–85 (2018)
- Allen, R., Rehbeck, J.: Revealed Stochastic Choice with Attributes. Available at SSRN 2818041 (2019)
- Block, H.D., Marschak, J.: Random orderings and stochastic theories of responses. *Contrib. Probab. Stat.* **2**, 97–132 (1960)
- Blundell, R.: Comments on: Michael P. Keane structural vs. atheoretic approaches to econometrics. *J. Econom.* **156**(1), 25–26 (2010)
- Breitmoser, Y.: The axiomatic foundation of logit. *Rationality and Competition Discussion Paper Series* 78 (2018)
- Camerer, C., Cohen, J., Fehr, E., Glimcher, P., Laibson, D.: *Neuroeconomics*, vol. 2, pp. 153–217. Princeton University Press, Princeton (2017)
- Cooper, D.J., Dutcher, E.G.: The dynamics of responder behavior in ultimatum games: a meta-study. *Exp. Econ.* **14**(4), 519–546 (2011)
- Dagsvik, J.K.: Invariance axioms and functional form restrictions in structural models. *Math. Soc. Sci.* **91**, 85–95 (2018)
- Echenique, F., Saito, K.: General luce model. *Econ. Theory* **68**(4), 811–826 (2019)
- Engel, C.: Dictator games: a meta study. *Exp. Econ.* **14**, 583–610 (2011)
- Fudenberg, D., Iijima, R., Strzalecki, T.: Stochastic choice and revealed perturbed utility. *Econometrica* **83**(6), 2371–2409 (2015)
- Gul, F., Natenzon, P., Pesendorfer, W.: Random choice as behavioral optimization. *Econometrica* **82**(5), 1873–1912 (2014)
- Heckman, J.J., Urzua, S.: Comparing iv with structural models: what simple iv can and cannot identify. *J. Econom.* **156**(1), 27–37 (2010)
- Hey, J.: Why we should not be silent about noise. *Exp. Econ.* **8**(4), 325–345 (2005)
- Johnson, N.D., Mislin, A.A.: Trust games: a meta-analysis. *J. Econ. Psychol.* **32**(5), 865–889 (2011)
- Keane, M.: A structural perspective on the experimentalist School. *J. Econ. Perspect.* **24**(2), 47–58 (2010a)
- Keane, M.P.: Structural vs. atheoretic approaches to econometrics. *J. Econom.* **156**(1), 3–20 (2010b)
- Koida, N.: Anticipated stochastic choice. *Econ. Theory* **65**(3), 545–574 (2018)
- Luce, R.: *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York (1959)
- Matejka, F., McKay, A.: Rational inattention to discrete choices: a new foundation for the multinomial logit model. *Am. Econ. Rev.* **105**(1), 272–98 (2015)
- McFadden, D.: Conditional logit analysis of qualitative choice models. In: Zarembka, P. (ed.) *Frontiers of Econometrics*, pp. 105–142. Academic Press, New York (1974)
- Nevo, A., Whinston, M.: Taking the dogma out of econometrics: structural modeling and credible inference. *J. Econ. Perspect.* **24**(2), 69–81 (2010)
- Oosterbeek, H., Sloof, R., Van De Kuilen, G.: Cultural differences in ultimatum game experiments: evidence from a meta-analysis. *Exp. Econ.* **7**(2), 171–188 (2004)
- Panizza, F., Vostroknutov, A., Coricelli, G.: Meta-context and choice-set effects in mini-dictator games. Working Paper (2019)
- Rust, J.: Comments on: structural vs. atheoretic approaches to econometrics by Michael Keane. *J. Econom.* **156**(1), 21–24 (2010)
- Ryan, M.: Uncertainty and binary stochastic choice. *Econ. Theory* **65**(3), 629–662 (2018)
- Steverson, K., Brandenburger, A., Glimcher, P.: Choice-theoretic foundations of the divisive normalization model. *J. Econ. Behav. Organ.* **164**, 148–165 (2019)
- Thurstone, L.: A law of comparative judgment. *Psychol. Rev.* **34**(4), 273–286 (1927)
- Tremblay, L., Schultz, W.: Relative reward preference in primate orbitofrontal cortex. *Nature* **398**(6729), 704–708 (1999)
- Wilcox, N.: Stochastic models for binary discrete choice under risk: a critical primer and econometric comparison. In: Cox, J.C., Harrison, G.W. (eds.) *Risk Aversion in Experiments, Research in Experimental Economics*, vol. 12, pp. 197–292. Emerald Group Publishing Limited, Bingley (2008)
- Wilcox, N.: Stochastically more risk averse: a contextual theory of stochastic discrete choice under risk. *J. Econom.* **162**(1), 89–104 (2011)

---

Wilcox, N.T.: Error and generalization in discrete choice under risk. Working Paper (2015)

Woodford, M.: Stochastic choice: an optimizing neuroeconomic model. *Am. Econ. Rev.* **104**(5), 495–500 (2014)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.