Decision Conflict, Power Logit, and the Deferral Outside Option

Georgios Gerasimou*†
University of Glasgow

March 16, 2026

Abstract

Decision makers often opt for the deferral outside option when they find it difficult to make an active choice. Contrary to existing logit models with an outside option where the latter is assigned a fixed value exogenously, this paper introduces and studies a class of logit models where that option's value is menu-dependent, may be determined endogenously, and could be interpreted as proxying the varying degree of decision difficulty at different menus. We focus on the *power logit* special class of these models. We show that these explain some observed choice-deferral effects that are caused by hard decisions, including non-monotonic "roller-coaster" choice-overload phenomena that are regulated by the presence or absence of a clearly dominant feasible alternative. We illustrate the usability, novel insights and explanatory gains of the proposed framework for duopolistic modelling and empirical discrete choice analysis.

Keywords:
Decision difficulty; outside option; choice deferral; quality/price competition; discrete choice; estimation.

*Email address: Georgios.Gerasimou@glasgow.ac.uk

†This paper originates in—and supersedes—work originally circulated as *"The decision-conflict and multicriteria logit"* that was originally posted online in May 2020. The author has benefited from discussing this work with many scholars, and is particularly thankful to Miguel Ballester, Jörg Stoye, Dimitris Christelis, Irina Merkurieva and Levent Ülkü for their comments; to Sudeep Bhatia and Tim Mullett for sharing the data that are analysed in Section 5; the Economics Department at Oxford University for their hospitality in October 2019; and audiences at Seminars in Economic Theory (2021), Heriott-Watt (2021), SAET 2021, BRIC 2022 (Prague), Glasgow, Durham Economic Theory Conference 2023, EEA-ESEM 2023 (Barcelona) and Oligo 2024 (Crete).

1 Introduction

It is a well-established fact that people often "choose not to choose" when they find it hard to compare the active-choice alternatives available to them, even when all these alternatives are individually considered "good enough" and are paid attention to. Real-world examples of such behavior include: (i) employees who operated within an "active decision" pension-savings environment and did not sign up for one of the plans that were available to them within, say, a day, week or month of first notice, possibly even opting for indefinite non-enrolment;[1] (ii) patients who, instead of choosing *"immediately"* one of the active treatments that were recommended to them against a medical condition, delayed making such a choice—often at a health cost—due to *"facing a treatment dilemma"*;[2] (iii) doctors who were willing to prescribe the single available drug to treat a medical condition but were not prepared to prescribe anything when they had to decide from the expanded set that contained one more drug, because *"the difficulty in deciding between the two medications led some physicians to recommend not starting either"* (Redelmeier and Shafir, 1995).[3] Even though in many such cases people do end up making an active choice *eventually*, the following question arises for the decision analyst: *Should the observed individual's initial avoidance/delay to make such a choice be ignored as non-relevant?* Put more concretely, should the decision of a medical doctor who eventually prescribes one of the two possible medications after several reminders be thought of as similar to their decision to prescribe that medication without delay in a different situation? In this paper we regard the decision to avoid/delay making an active choice that is caused by such context-specific "decision conflict" as providing meaningful information about both the individual and the alternatives in question, regardless of which active choice—if any—was eventually made at that menu at a later time.

In their influential monograph, Janis and Mann (1977, p. 46) defined *"decision conflicts"* as the *"simultaneous opposing tendencies to accept and reject a given course of action" and identified "hesitation, vacillation,* [and] *feelings of uncertainty"* to be among their most prominent symptoms *"whenever the decision comes within the focus of attention"*.[4] Motivated by the relevance of hesitation-driven opt-out decisions for understanding preferences and explaining behavior, our specific goal in this paper is to model choice in the presence of a choice-avoidance/deferral outside option within a stochastic choice framework in ways that deviate as little as possible from existing well-understood

[1]Such behavior is documented in Carroll, Choi, Laibson, Madrian, and Metrick (2009), for example.

[2]See Knops, Ubbink, Legemate, Stalpers, and Bossuyt (2013, p. 78). See also O'Connor (1995) and Garvelink et al. (2019) for additional references and overview of the use of a *"decisional conflict scale"* in medical decision making that was developed *"to measure a person's perceptions of their uncertainty in making a choice about health care options, the modifiable factors contributing to uncertainty, and the quality of the decision made"*.

[3]Other works that find evidence associating decision difficulty with such choice paralysis in different environments include Tversky and Shafir (1992), Dhar (1997), Dhar and Simonson (2003), Danan and Ziegelmeyer (2006), Bhatia and Mullett (2016), and Costa-Gomes, Cueva, Gerasimou, and Tejiščák (2022).

[4]Pochon, Riis, Sanfey, Nystrom, and Cohen (2008) is a targeted study in the neuroscience literature on the brain regions that are activated when subjects face decision conflict.

modelling practices and, at the same time, make predictions that are in line with some findings from the empirical/experimental literature and evade existing models. We pursue this by extending in disciplined ways the foundational Luce (1959)/logit model and its econometric specification pioneered by McFadden (1973). Specifically, we propose the class of *decision-conflict logit* models which, in their most general form, are a straightforward but so far unexplored extension of the logit with an outside option that assign a menu-dependent value to that option while retaining the menu-invariance assumption on all active-choice alternatives. The relative value of the outside option at a menu in turn determines the probability of avoiding/deferring choice and can be interpreted as proxying decision difficulty.

We show that, despite its simplicity, this general model can be the starting point for richly structured special cases. In particular, we introduce and focus on the broad class of *power logit* models that are examples of of such cases where decision difficulty depends in intuitive ways on the logit values of all active-choice alternatives. In these models, decision difficulty could be thought of as driven by the agent's noisy resampling of the menu's elements. In the *quadratic logit* special case of this class of models, for example, such resampling takes the form of the choice probability of a market alternative emerging as the product of *two* logit probabilities according to a *single* value function/criterion. Intuitively, the agent is more likely to choose an active-choice alternative if and only if its value realizations according to this criterion are much larger than those of everything else feasible across *both* rounds of sampling. Conversely, the agent is more likely to avoid/defer choice when no alternative achieves such unanimous clear dominance. This model could therefore be thought of as capturing a hesitant decision maker who behaves as if they used an objective criterion to compare alternatives (e.g. sum or multiply each option's values across all relevant attributes) but is aware that their subjective evaluation according to this objective criterion may be imperfect, possibly due to cognitive limitations, thereby leading them to performing this task twice. This model and its power-logit generalization appear to be the first to provide a theory where the no-choice outside option is feasible and has an endogenously determined, menu-dependent value.

We further show that these structured models explain the following empirical phenomena that various studies in cognitive and consumer psychology have documented about decisions that allow agents to avoid/delay making an active choice:

(i) As alternatives become more similar in their overall appeal, decision difficulty and the probability of choice delay are increased. We will refer to this as the *"similarity-driven deferral effect"*.

(ii) The dominance-driven non-monotonic relation between menu expansion and the tendency to opt out, which we refer to as the *"roller-coaster" choice-overload effect*. This has implications for the interpretation and policy responses to so-called "too-much-choice" phenomena.

(iii) *"Relative-desirability"* effects, whereby holding constant the total value in a menu while increasing the value differences between the menu's alternatives increases the

probability of an active choice.

We illustrate the applicability of our analysis both in theoretical and empirical settings. In our theoretical application we analyse a duopolistic market where firms compete simultaneously in price and quality under the common-knowledge assumption that consumer demand is determined by the power-logit model where a product's value is defined, intuitively, by its quality/price ratio. We derive simple and economically interpretable intuitive closed-form solutions for all equilibrium variables in the model: price, quality, profits and a notion of consumer welfare that appears suitable in environments where consumers opt out due to indecisiveness or overload. A key feature of the (symmetric) equilibrium is that, as the power parameter capturing consumers' decision difficulty increases, firms increase their products' quality/price ratio and see their profits decreased, both because of the reduced profit margins and because of the lower share of consumers who buy any product. Intuitively, this is driven by each firm increasing its quality/price ratio in an effort to reduce the consumer's decision difficulty and mitigate the risk of losing them to the rival firm *or* of driving them out of the market altogether.

In our empirical application we first show how the classic assumptions and argument that underpin the discrete-choice formulation of the logit *without* an outside option (McFadden, 1973) must be modified and extended in order for both the quadratic logit and the more general power logit models to admit a similar discrete-choice formulation and be taken to the data for maximum-likelihood estimation of their respective parameters. We then show the potential fruitfulness of such analyses by estimating both the quadratic and power logit models on the deferral-permitting discrete-choice data with film decisions from the survey experiment of Bhatia and Mullett (2016), using the participants' subjective ratings of the different films as the explanatory variable. To assess the added value of the hereby proposed models on these data, we use standard criteria to evaluate their goodness of fit and compare them to those of baseline logit models with a fixed and inferior or a random outside option. Our analysis suggests that both the power and quadratic logit often perform better compared to either version of the baseline logit under these performance criteria, particularly in those situations where theory suggests they would do so. Hence, the proposed power-logit framework could be considered in the analysis of similar datasets whenever the researcher suspects that the observed opting-out/deferring behavior might be due to decision difficulty rather than the relative unattractiveness of the available active-choice alternatives.

2 Theory

2.1 Logit with A Menu-Dependent Outside Option

Let X be the grand choice set of finitely many *active-choice* alternatives with generic elements $a, b \in X$. Let $\mathcal{M} := \{A : \emptyset \neq A \subseteq X\}$ be the collection of all *menus* of such alternatives, and $\mathcal{B}$ its sub-collection that comprises all binary menus. The *outside option*

is denoted by $o \notin X$. We clarify that this is not a *status quo* option (e.g. a tenant's current rental agreement) which can, in principle, be compared to the other feasible alternatives (e.g. other housing options) on the same or a similar set of relevant attributes. Instead, this option is devoid of attributes and its value to the decision maker is unobservable to the analyst.[5] A *random free-choice model* on X is a function $\rho : X \times \mathcal{M} \to \mathbb{R}_+$ such that $\rho(a, A) \in [0, 1]$ for all $A \in \mathcal{M}$ and all $a \in A$; $\rho(a, A) = 0$ for all $A \in \mathcal{M}$ and all $a \notin A$; and $\sum_{a \in A} \rho(a, A) \leq 1$, where $\rho(o, A) := 1 - \sum_{a \in A} \rho(a, A) \leq 1$ is the probability of choosing the—always feasible—outside option at menu A. To simplify notation, for $A, B \in \mathcal{M}$ with $B \subseteq A$ we write $\rho(B, A) := \sum_{b \in B} \rho(b, A)$.

We start by introducing the *logit with a general outside option* as the model that comprises value functions $u : X \to \mathbb{R}_{++}$ and $D : \mathcal{M} \to \mathbb{R}_+$ such that, for every menu $A \in \mathcal{M}$ and active-choice alternative $a \in A$,

$$\rho(a, A) \;=\; \frac{u(a)}{\sum_{b \in A} u(b) + D(A)}, \tag{1}$$

where the pair (u, D) is unique up to a common positive linear transformation. In this model, u captures the menu-independent values of active-choice alternatives and $D(A)$ the menu-dependent value of the outside option. Like the baseline Luce model [see (6) below], all active-choice alternatives in (1) are assigned menu-independent values that determine their relative likelihood of being chosen. Unlike the baseline model, where this property also extends to the outside option, here the probability of making an active choice in the first place (equivalently, of avoiding/deferring this decision) is determined by the menu-dependent value of D.

Two properties characterize the class of models that can be represented in this way:

A1 (Positivity).
For all $A \in \mathcal{M}$ and all $a \in A$: $\rho(a, A) > 0$.

A2 (The Active-Choice Luce Axiom).

[5] Hensher, Rose, and Greene (2015, p.53) , for example, describe this distinction as follows: *"At this point, it is worthwhile considering choice situations in which there exists the possibility to 'choose not to choose', or to remain with some status quo alternative. Many choice situations present decision makers with examples of both types of alternatives. For example, a person can elect to stay at home and not see a movie if three potential movie alternatives showing at a local cinema at some preferred time do not appeal to them. Likewise, a decision maker facing the expiration of their rental agreement may elect to simply renew their current rental contract or move apartments, hence signing a new lease. In the case of a* ***no choice*** *alternative, the alternative labelled* 'none' *will be devoid of any attribute levels (e.g., there is no movie ticket price, no time spent at the cinema, etc., associated with going to the movies). The absence of attributes, however, does not mean that the decision maker is indifferent to that alternative. In the movie ex- ample, if the three movies on offer are romantic comedies, then staying at home and not attending any of them might be the most preferred option."* A formal distinction in the treatment of status-quo and choice-deferral outside options in a deterministic choice-theoretic framework is provided in Gerasimou (2016).

For all $A, B \in \mathcal{M}$ and all $a, b \in A \cap B$:

$$\frac{\rho(a,A)}{\rho(b,A)} = \frac{\rho(a,B)}{\rho(b,B)}$$

A1 is standard and allows for a crisp illustration of the main ideas that we put forward in this paper. A2 imposes the familiar kind of IIA-consistency only in the odds of pairs of active-choice alternatives, while allowing odds that involve such an alternative and the outside option to deviate from it. That is, $\frac{\rho(o,A)}{\rho(b,A)} \neq \frac{\rho(o,B)}{\rho(b,B)}$ is allowed by A2.

Proposition 1
ρ is a logit with a general outside option if and only if it satisfies A1-A2.

Indeed, adapting the arguments in Luce (1959) yields an equivalence between A1-A2 and the existence of a function $u : X \to \mathbb{R}_{++}$ such that, for every $A \in \mathcal{M}$ and $a \in A$,

$$\rho(a,A) = \big(1 - \rho(o,A)\big) \cdot \frac{u(a)}{\sum_{b \in A} u(b)}, \tag{2}$$

where

$$u(a) := \alpha \cdot \frac{\rho(a,X)}{\rho(z,X)} \tag{3}$$

for arbitrary and fixed $\alpha > 0$ and $z \in X$. It follows then that for every $A \in \mathcal{M}$ there is a unique $D(A) \geq 0$ that makes (1) true, with

$$D(A) \equiv \frac{\rho(o,A)}{1 - \rho(o,A)} \cdot \sum_{b \in A} u(b). \tag{4}$$

Finally, it is immediate that (u, D) and (u', D') represent the same ρ if and only if $u = \alpha u'$ and $D = \alpha D'$ for some $\alpha > 0$.

We now compare (1) to the baseline logit *with* an outside option (Anderson, Palma, and Thisse, 1992; Hensher, Rose, and Greene, 2015) and *without* such an option. We recall that a random choice model ρ on $X \cup \{o\}$ admits the former representation if there is a function $u : X \cup \{o\} \to \mathbb{R}_{++}$ such that, for all $A \in \mathcal{M}$ and $a \in A$,

$$\rho(a,A) = \frac{u(a)}{\sum_{b \in A} u(b) + u(o)}. \tag{5}$$

On the other hand, ρ admits a logit representation *without* an outside option if there exists some $u : X \to \mathbb{R}_{++}$ such that

$$\rho(a,A) = \frac{u(a)}{\sum_{b \in A} u(b)}, \tag{6}$$

The latter obviously implies $\sum_{a\in A}\rho(a,A)=1$ for all $A\in\mathcal{M}$, so that the opportunity to defer is either infeasible in this model or feasible but never acted upon. Thus, (1) includes (6) as a special case when $o\notin X$. In addition, (1) extends (5) but without nesting it unless A1-A2 and ρ operate on the enriched domain $X\cup\{o\}$.

2.2 Decision-Conflict Logit and Power Logit

We define the *power logit* model by the existence of a menu-independent *stimulus intensity* value function $\widehat{u}:X\to\mathbb{R}_{++}$ and a parameter $p\geq 1$ such that, for every menu A and alternative a in A,

$$\rho(a,A) \;=\; \left(\frac{\widehat{u}(a)}{\sum\limits_{b\in A}\widehat{u}(b)}\right)^{p} \tag{7}$$

Clearly, this model predicts $\rho(o,A)>0$ at every menu A if and only if $p>1$, and reduces to (6) at $p=1$.

The agent portrayed in (7) could be thought of as behaving according to the standard logit with a single valuation criterion but, possibly aware of their decision difficulty, also as if they sampled all alternatives *more than once* before making a decision. For example, in the *quadratic logit* case of special interest where $p=2$, the agent might be thought of as sampling the same menu *twice*. Because the resulting value realizations generally differ across these two rounds of sampling due to the postulated randomness, this individual would be more likely to choose an active-choice alternative if its perceived *signal/stimulus intensity* from *both* inspections, captured by the two value realizations of $\widehat{u}$, is relatively high, and as being more likely to avoid/defer choice when this is not true for any such alternative. When deciding which insurance plan to buy, for example, an agent whose behavior is approximated by the quadratic logit may review the top-rated plans from a service comparison website in the morning, receive some value stimuli/signals from each of them, and then go back and repeat this process in the evening. Assuming that the two sampling rounds are independent (admittedly, a demanding assumption), an insurance plan is more likely to be chosen at the end of this two-stage process if its relative stimulus/signal intensity is sufficiently high to make the product stand out despite the agent's hesitation.

The intuition in the more general case where $p\neq 2$ in (7) is analogous and admits a probabilistic explanation. Specifically, if the analyst a priori restricts p to lie between 1 and 2, then $p-1$ might be interpreted as the (exogenous) *probability* that the agent will engage in two rounds of sampling, equalling 1 in the limit where the quadratic logit decision process emerges with certainty. Similarly, if p is assumed to lie between 2 and 3, then $p-2$ could be thought of as the probability that the agent will perform three rounds of sampling, conditional on the analyst expecting them to do at least two. More generally, the power parameter p in this model could be viewed as reflecting the agent's propensity to engage in possibly multiple rounds of sampling.

That this model is a logit with a general outside option may not be obvious at first glance but quickly becomes so upon noticing that one can write

$$u(a) := \widehat{u}(a)^p \tag{8}$$

$$D(A) := \left(\sum_{b\in A}\widehat{u}(b)\right)^p - \sum_{b\in A}\widehat{u}(b)^p \tag{9}$$

At $p = 2$ these expressions admit the simpler and more easily interpretable form

$$u(a) := \widehat{u}(a)^2, \tag{10}$$

$$\begin{aligned} D(A) &:= \left(\sum_{b\in A}\widehat{u}(b)\right)^2 - \sum_{b\in A}\widehat{u}(b)^2 \\ &= 2\sum_{\substack{a,b\in A,\\ a\neq b}}\widehat{u}(a)\widehat{u}(b) \\ &= \sum_{a,b\in A} D(\{a,b\}), \end{aligned} \tag{11}$$

where the last step makes use of the notational convention

$$D(\{a,a\}) \equiv D(\{a\})$$

This clarifies that the quadratic logit $\rho \sim (\widehat{u})^2$ is an *additive* (u, D) model in the sense that the value of the outside option at every menu depends additively on the value of that option at each of its binary submenus. It also clarifies that the latter value takes a symmetric Cobb-Douglas form with respect to $\widehat{u}$. We will return to additivity later in this section but note here that the quadratic case where $p = 2$ is the only one where the (u, D) representation of (7) has this property.

We proceed by noting the following direct implication of the power-logit model:

A3 (Desirability & Complexity)
For all $A \in \mathcal{M}$*:* $\rho(A, A) = 1 \iff |A| = 1$.

To motivate the intuition behind (and the label assigned to) A3 we first recall that, as was clarified early on, our aim here is to model decision difficulty that is rooted in a fully attentive individual's potential inability to make some preference comparisons between otherwise desirable options. If a *single* such option was feasible to such an individual, therefore, one might expect that person to immediately choose that one option. If on the other hand there are at least two available options and the individual is not forced to make a choice immediately, then the experimental/empirical evidence suggests that there is at least *some* probability that this person's attempt to find a most preferred option and choose that option will not be fruitful reasonably quickly. To the extent

that this is so, a legitimate approach from the analyst's perspective would be to portray that decision maker as *deferring* choice with positive probability whenever at least one non-trivial comparison is required.

Imagine, for example, a patient like those reported on in Knops, Ubbink, Legemate, Stalpers, and Bossuyt (2013) who has been diagnosed with a life-threatening disease. Suppose that their doctor informs them that there is only one available treatment that can cure this disease, and asks whether they would like to sign up for this treatment. One would expect the patient to sign up immediately because there would be no benefit from delaying their only chance for a cure. Now suppose instead that the doctor tells the patient that there are two possible treatments: one with high efficacy but severe side effects, and another with milder side effects but lower cure rates. Even though either one of these treatments would have been chosen immediately if it was the only feasible one (see Tversky and Shafir, 1992; Redelmeier and Shafir, 1995; Dhar, 1997; Costa-Gomes, Cueva, Gerasimou, and Tejiščák, 2022, for example), here one might expect the patient to delay making such an active choice, perhaps until they think about the conflicting pros and cons and then ultimately determine which treatment would be best for them. Situations of this kind are compatible with and, in fact, motivate our modelling framework in this paper.

In the spirit of these examples, A3 postulates that an active choice is made with certainty *only at* singleton menus and, as such, formalises the behavioral mechanisms outlined above. Of course, one can easily think of situations where this axiom is descriptively invalid. Yet for analytical purposes it is a useful property because it allows for completely isolating the decision-difficulty channel to deferrals from other potential channels such as undesirability of the available alternatives or limited attention, which have quite distinct behavioral origins.

The following statement readily follows from the preceding analysis.

Observation 1
ρ satisfies A1–A3 if and only if it is a (u, D)-model with the property that

$$D(A) = 0 \iff |A| = 1 \tag{12}$$

We will refer to this special class of generalized logit models with a context-dependent outside option as the class of *decision-conflict logit* models, and to the menu function D that captures the varying appeal of opting out at different menus as the *decision cost* or *decision complexity* function. Justifying such a name for the function D given the requirement that it be zero-valued only at singletons may benefit from some additional explanation that supplements the preceding discussion. When the decision environment is such that avoidance/deferral is caused solely by decision difficulty instead of other factors (e.g. none of the active-choice alternatives is good enough, or none is considered due to limited-attention constraints), our decision maker is portrayed as not having any problem deciding between deferring or choosing the only available active-choice option: they do the latter. By contrast, the decision between deferring or choosing from two or

more such options is *at least* somewhat costly because of the effort that is necessary to make the relevant preference comparisons.

We will refer to both a decision-conflict logit $\rho = (u, D)$ and D as *monotonic* if

$$A \supset B \implies D(A) \geq D(B). \tag{13}$$

If $D(A) > D(B)$ is always true when $A \supset B$, then D and $\rho = (u, D)$ will be called *strictly* monotonic. In line with our intended interpretation of D as a complexity/cost function, the total number of pairs of distinct alternatives increases as a menu expands, hence so does the expected number of comparisons between alternatives that a fully-attentive individual needs to make. In expectation, therefore, decision difficulty also goes up in absolute terms when more alternatives are added to a menu. Importantly, however, this *does not* imply that deferring always becomes more likely once a menu is expanded when D is monotonic (we will return to this point soon). But Monotonicity does have a familiar general implication for *active*-choice alternatives, which in the standard random forced-choice environments was originally stated in Block and Marschak (1960):

Observation 2
If ρ is a monotonic decision-conflict logit, then $a \in B \subset A$ implies $\rho(a, B) > \rho(a, A)$.

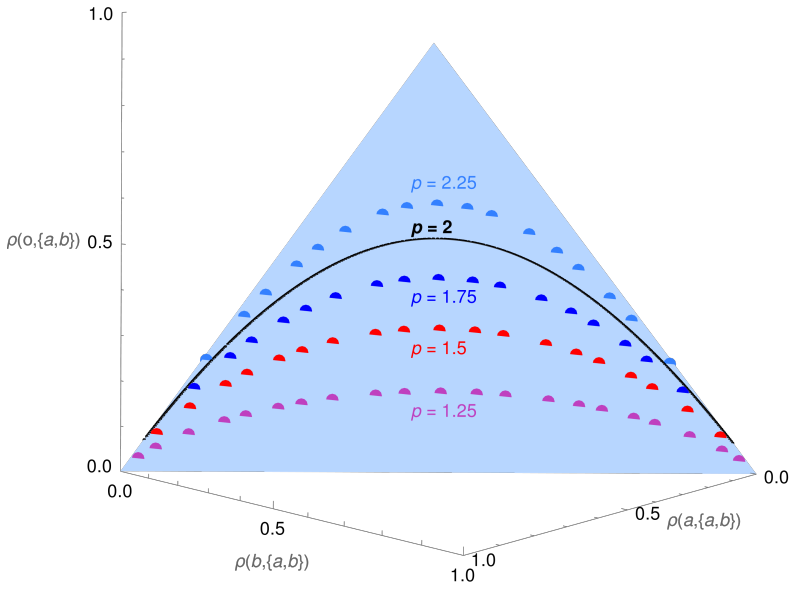
In particular, monotonic models satisfy what we will refer to as *active-choice regularity*, whereby the probability of such alternatives cannot increase when more options are added to a menu. Crucially, however, as we discuss and illustrate by example later, this property does not hold for the outside option.

When it comes to using a decision-conflict logit in suitable applications, the analyst must first decide whether to employ a special case where function D is set exogenously or one where it is determined endogenously. In the first case the choice might be dictated by the analyst's a priori assessment of the specific environment in question and could include, for example, defining D as the menu-cardinality function (Iyengar and Lepper, 2000; Iyengar, Huberman, and Jiang, 2004) or, if the alternatives have clearly identifiable attributes, some measure of similarity in attribute space (Spektor, Gluth, Fontanesi, and Rieskamp, 2019). The analyst's choice in the second case might instead be dictated by an agnosticism towards what is the most appropriate functional form for D, and by resorting instead to a general decision process where D is a function of the feasible options' u-values. The power-logit class of models is clearly of this kind.

2.3 Three Empirically Supported Behavioural Predictions

We proceed with an illustration of how a general (u, D) model, or even the more structured power logit, make predictions that help explain intuitively the three empirically documented choice-deferral phenomena mentioned in the Introduction. We start by noting that findings and arguments from the consumer-psychology literature reported in Dhar (1997), Sela, Berger, and Liu (2009), and Scheibehenne, Greifeneder, and Todd (2010),

Figure 1: The loci of binary choice probability distributions that are compatible with the power logit under different values of p.



among others, suggest that decision makers are sometimes more likely to avoid/delay choice when the feasible alternatives are perceived to be of similar value. The last authors noted, for example, that as the most attractive feasible options become more similar when new items are added to a menu, it can become more difficult for the decision maker to justify the choice of any particular option, which in turn would increase the likelihood of choice deferral. This is what we earlier referred to as *"similarity-driven deferral"*. In the same direction, but focusing on response times rather than deferral decisions, Bhatia and Mullett, 2018 recently reported evidence to suggest that choice between similarly attractive options is significantly correlated with longer response times.

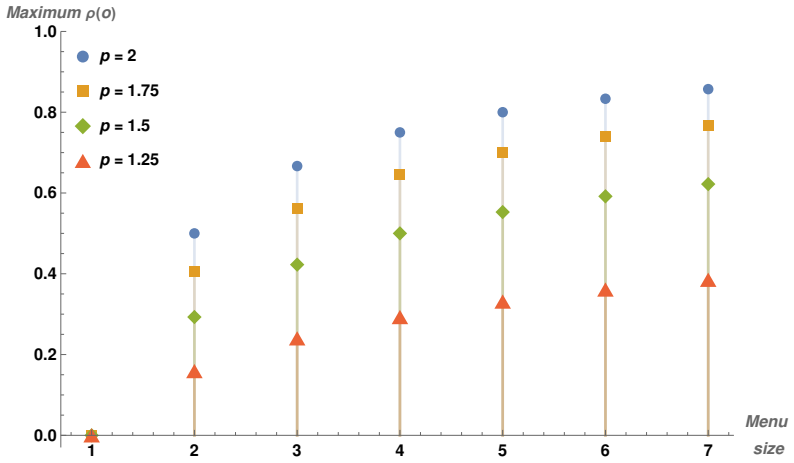
Our next result shows how the power logit predicts such an effect. More specifically, an interesting feature of this model is that its predicted probability of opting out at a menu as a function of the number of active-choice alternatives at that menu is bounded above in a simple way, and that upper bound is attained precisely when *all* feasible alternatives are of the same value.

Proposition 2
If ρ is a power logit $(\widehat{u}, p)$, then, for every menu A,

$$\begin{aligned} \rho(o, A) &\leq 1 - |A|^{1-p}, \\ \rho(o, A) &= 1 - |A|^{1-p} \iff \widehat{u}(a) = \widehat{u}(b) \quad \textit{for all } a, b \in A. \end{aligned}$$

In this model, therefore, an agent's decision difficulty at a menu, as revealed by the deferral probability at that menu, is maximized when all feasible active-choice alternatives are equally desirable, and this maximum difficulty is increasing in proportion to the total

Figure 2: Maximum probability of deferring as a function of menu size in the power logit model.



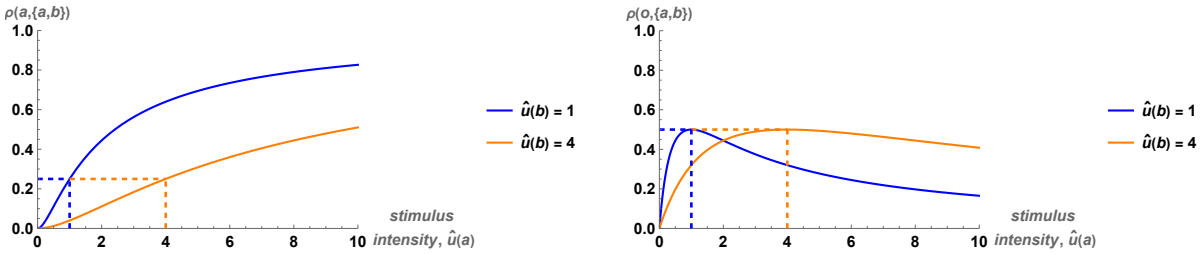
number of such alternatives at a decreasing rate (Figure 2).[6]

We now turn to the power-logit model's comparative statics in the important class of binary menus. Figure 3 illustrates, with a quadratic-logit example, the general pattern in the behavior of $\rho(a, \{a, b\})$ and $\rho(o, \{a, b\})$ as the stimulus intensity of a changes while that of b is held fixed. Interestingly, the monotonic increase of $\rho(a, \{a, b\})$ in $\widehat{u}(a)$ occurs at an increasing rate as this value approaches the $\frac{\widehat{u}(b)}{2}$ stimulus-intensity threshold from below than when $\widehat{u}(a)$ increases monotonically beyond $\frac{\widehat{u}(b)}{2}$. Intuitively, the inflection-point stimulus intensity value $\frac{\widehat{u}(b)}{2}$ that dissects $\rho(a, \{a, b\})$—viewed as a function of $\widehat{u}(a)$—into convex and concave regions suggests that marginal improvements in the appeal of a lead to more rapid market share increases when this alternative is still "catching up" with b than when it has become sufficiently close to (or surpassed) it in attractiveness. On the other hand, $\rho(o, \{a, b\})$ is a strictly concave function of $\widehat{u}(a)$ and, consistent with Proposition 2, attains its maximum value of $\frac{1}{2}$ when $\widehat{u}(a) = \widehat{u}(b)$. Thus, the model's novel prediction here, of relevance both from a consumer-welfare and a seller-profit perspective, is that minimal re-designing of a menu that consists of equally attractive alternatives is more likely to be effective at reducing opt-out behavior if one of the original alternatives *becomes less rather than more appealing*, other things equal.

Now, since, as was discussed previously, it is not generally true that avoiding/deferring becomes more likely as menus expand even for monotonic decision-conflict logit models,

[6] A clarifying remark may be due at this point. Equal values (specifically, utilities) between two or more alternatives are—in most of economic theory—associated with positive indifference, which in turn is interpreted as suggesting that the individual in question would be equally happy with any of these alternatives. By contrast, the influential *drift diffusion model* in neuroeconomics (Krajbich, Armel, and Rangel, 2010; Baldassi, Cerreia-Vioglio, Maccheroni, Marinacci, and Pirazzini, 2020; Fudenberg, Newey, Strack, and Strzalecki, 2020), which originates in the psychology literature (Ratcliff and McKoon, 2008), and related experimental evidence that have been of increasing visibility and interest in the economics literature lately make the opposite predictions/observations. The latter in turn are broadly in line with the general predictions of the power-logit model that we focus on in this paper. Considering the different motivations, methodological frameworks and intended interpretations in the two literatures, however, the seeming discrepancy is in our view more an issue of semantics than it is one of substance. In any case, our use of the term "value" rather than "utility" in reference to the terms appearing in logit formulae is partly motivated by this issue.

Figure 3: Comparative statics in the quadratic logit when one of the two alternatives becomes more attractive.



it is naturally of interest to understand when, exactly, such behavior is to be expected in this environment. The general idea in answering this question is that, even if decision difficulty increases in absolute terms when new alternatives are introduced, when these new alternatives are sufficiently better than the pre-existing ones their added value will offset the elevated decision cost and will ultimately result in a higher probability of making an active choice at the larger menu. To state this more formally we will abuse notation slightly by letting

$$u(S) \;:=\; \sum_{s\in S} u(s) \tag{14}$$

stand for the total Luce value at menu $S \in \mathcal{M}$.

Observation 3
If $\rho = (u, D)$ is a decision-conflict logit, then for any $A, B \in \mathcal{M}$ such that $A \supset B$:

$$\rho(o,A) \leq \rho(o,B) \iff \underbrace{\frac{D(A)-D(B)}{D(B)}}_{\substack{\textit{marginal cost}\\ \textit{from menu expansion}}} \leq \underbrace{\frac{u(A)-u(B)}{u(B)}}_{\substack{\textit{marginal benefit}\\ \textit{from menu expansion}}} . \tag{15}$$

This eloquent equivalence clarifies that the choice probability of opting out will decrease following menu expansion if and only if the marginal benefit of this expansion, as measured by the percentage increase in total value, exceeds its marginal cost, as measured by the percentage increase in decision complexity. This is a distinctive property of decision-conflict logit models. It clarifies that they do not belong to the random-utility class[7] with an outside option, and enables them to explain simply the non-monotonic and dominance-driven effect that menu expansion has been known to exert on the probability of deferring (Scheibehenne, Greifeneder, and Todd, 2010; Chernev, Böckenholt, and Goodman, 2015), which we earlier referred to as the *"roller-coaster choice overload"* effect.

Indeed, citing several studies in consumer psychology, the meta-analysis in Chernev, Böckenholt, and Goodman (2015) notes that *"it has been shown that consumers are more*

[7]See Apesteguia, Ballester, and Lu (2017), Stoye (2019), Strzalecki (2024) and references therein.

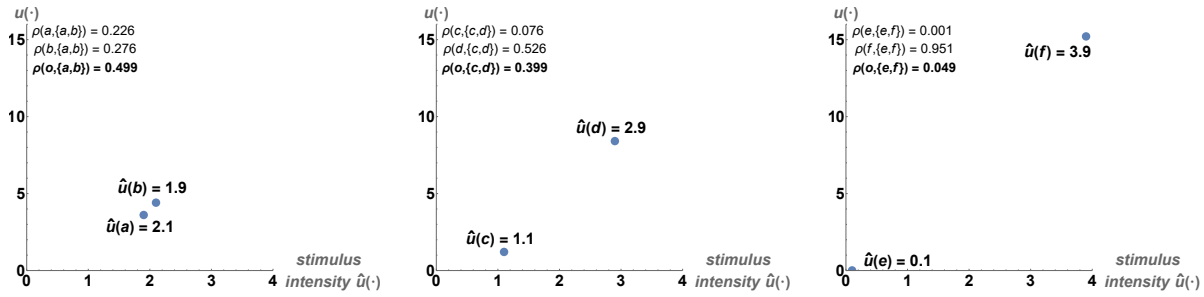
Table 1: Illustration of "roller-coaster" choice-overload predictions with the quadratic logit.

Option	$\widehat{u}$	$\rho(\cdot,\{a,b\})$	$\rho(\cdot,\{a,b,c\})$	$\rho(\cdot,\{a,b,c,d\})$
a	10	0.980	0.250	0.007
b	0.1	0.001	0.000	0.000
c	9.9	–	0.245	0.007
d	100	–	–	0.694
o	–	**0.019**	↗ **0.505** ↘	**0.292**

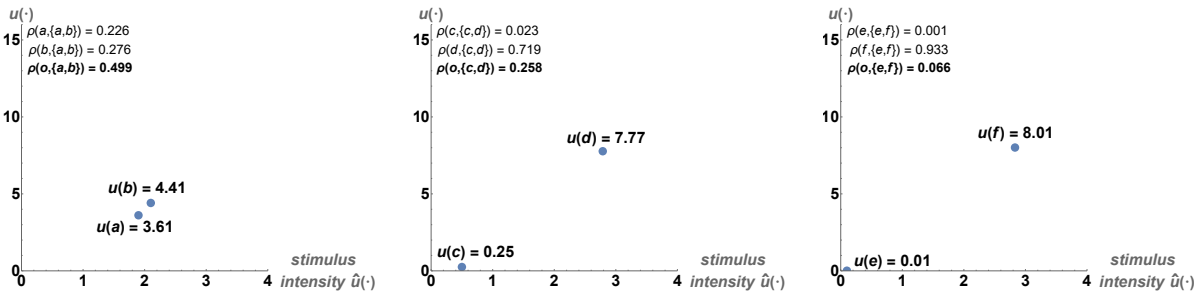
likely to make a purchase from an assortment when it contains a dominant option than when such an option is absent" (p. 338). This finding is important for the interpretation and policy responses to choice-overload phenomena of the kind that were first reported in Iyengar and Lepper (2000). To our knowledge, the decision-conflict logit is the first random-choice model that predicts this dominance-driven emergence and disappearance of choice-overload effects, and it does so without imposing any undesirability or inattention constraints. Table 1 illustrates an example such effect that is predicted by the quadratic logit model.

Figure 4: Illustration of relative-desirability effect predictions of the quadratic logit.

Case 1: the $\widehat{u}$-sum is constant in all menus



Case 2: the u-sum is constant in all menus



Finally, as the next result establishes, the power-logit also predicts another important choice-deferral phenomenon, known as the "relative-desirability" effect (Dhar, 1997; White, Hoffrage, and Reisen, 2015; Bhatia and Mullett, 2016). This refers to situations where, choosing the outside option becomes more likely in binary menus as the available options become more equally desirable, other things equal.

Proposition 3
If $\rho = (u, D) = (\widehat{u}, p)$ is a power logit, then for any $a, b, c, d \in X$ where $\widehat{u}(a) + \widehat{u}(b) = \widehat{u}(c) + \widehat{u}(d)$ or $u(a) + u(b) = u(c) + u(d)$ is true, the following is also true:

$$\rho(o, \{a, b\}) > \rho(o, \{c, d\}) \iff |\widehat{u}(a) - \widehat{u}(b)| < |\widehat{u}(c) - \widehat{u}(d)| \tag{16}$$

$$\iff |u(a) - u(b)| < |u(c) - u(d)| \tag{17}$$

This result, illustrated in Figure 4, is distinct from the similarity-driven deferral effect that was discussed in relation to Proposition 2 because it compares the probabilities of opting out at two distinct binary menus as a function of the absolute value/stimulus-intensity differences between the two active-choice alternatives, rather than focusing on when this probability is maximized within the same menu of any size. It clarifies, indeed, that the model predicts relative-desirability effects irrespective of whether the stimulus-intensity absolute difference of the two alternatives or that between their power-logit values—which emerge from the stimulus-intensity values via the (convex) power transformation—is used to assess relative desirability.

3 Power-Logit Duopolistic Competition in Price and Quality

We proceed with an illustration of the potential usefulness of the power-logit functional form in the analysis of oligopolistic markets when consumers potentially face comparison difficulties and may avoid/delay making an active choice.[8] To this end, we consider a market where two profit-maximizing firms compete for a single consumer (equivalently, a unit mass of consumers) by offering a product that is differentiated in quality, q_i, and price, p_i. Producing a product of quality q_i costs q_i to firm $i = 1, 2$, while $0 \leq q_i \leq p_i \leq I$ and $I > 0$ denotes consumer income. Furthermore, a consumer's value from product (q_i, p_i) coincides with that product's quality-price ratio:

$$u(q_i, p_i) = \frac{q_i}{p_i}. \tag{18}$$

This assumption further implies

$$u(q_i, p_i) \in [0, 1] \tag{19}$$

for all (q_i, p_i). Such a "value-for-money" specification imposes intuitive positive and negative dependences of u on quality and price, respectively, with the former being linear and the latter strictly convex. Moreover, while identifying value with quality-price ratios as in (18) rather than with quality-price differences $q_i - p_i$ appears to be a novel modelling assumption, it is consistent with some central implications of the behavioral choice model by Bordallo, Gennaioli, and Shleifer (2013) concerning consumer preferences for high

[8] Piccione and Spiegler (2012), Spiegler (2015), Bachi and Spiegler (2018) and Gerasimou and Papi (2018) have recently suggested distinct approaches to study such markets.

quality-price ratio products, even though that model starts from very different primitives and features a quality-price difference value function instead.

The two firms choose their products' quality and price levels simultaneously and under complete information. The market share of product (q_i, p_i) at menu/strategy profile $\big((q_1,p_1),(q_2,p_2)\big)$ is determined by the power logit model

$$\rho\big((q_i,p_i),\{(q_j,p_j)\}_{j=1}^2\big) \;=\; \left(\frac{\frac{q_i}{p_i}}{\frac{q_i}{p_i}+\frac{q_j}{p_j}}\right)^s,$$

where $s \geq 1$ and $s = 1$ in the baseline special case where there is no decision difficulty. Under the above assumptions, each firm $i = 1, 2$ solves

$$\max_{0\leq q_i\leq p_i\leq I} \pi_i(q_i,p_i) \;:=\; (p_i-q_i)\cdot\rho\Big((q_i,p_i),\{(q_j,p_j)\}_{j=1}^2\Big) \tag{20}$$

The strategic trade-off in this model, which applies both when $s = 1$ and $s > 1$, is that each firm wishes to increase its quality/price ratio in order to expand its market share, while at the same time also wishing to decrease it in order to enlarge its profit markup.

Turning to consumer welfare, taking into account that decision conflict can potentially drive the consumer out of the market altogether, and that -by A3- this would be undesirable, we consider a utilitarian-like welfare measure that weighs the possible value levels at a given strategy profile by the probabilities that these values will actually be realized at that profile. We formalize this with the *consumer welfare* function $W : \mathbb{R}^4_{++} \to [0,1]$ defined by

$$W\big((q_i,p_i),(q_j,p_j)\big) \;:=\; \rho((q_i,p_i),\{(q_j,p_j)\}_{j=1}^2)\cdot u(q_i,p_i)+\rho((q_j,p_j),\{(q_j,p_j)\}_{j=1}^2)\cdot u(q_j,p_j).$$

This welfare indicator may be particularly relevant in cases where consumer surplus is equilibrium-invariant, as will turn out to be the case in the present environment.[9]

Perhaps surprisingly, this duopolistic model leads to the following simple and intuitive equilibrium predictions:

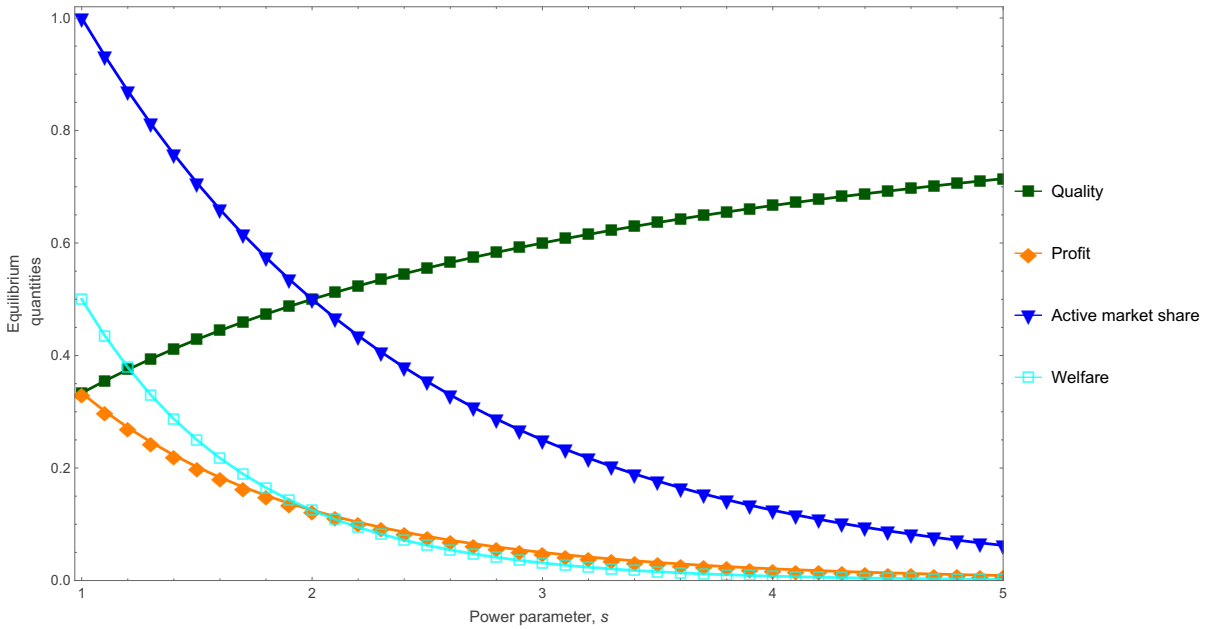
Proposition 4
The power-logit equilibrium is $(q_1^*,p_1^*) = (q_2^*,p_2^*) = \left(\frac{sI}{2+s}, I\right)$ *and is associated with equilibrium expected profits* $\pi_1^* = \pi_2^* = \frac{2^{1-s}}{2+s}I$ *and welfare* $W^* = 2^{1-2s}$.

Thus, although the equilibrium pricing strategy features full surplus extraction irrespective of the value of the hesitation/resampling parameter s, the equilibrium quality level increases in s at the rate $\frac{s}{2+s}$. starting at the low of $\frac{I}{2}$ in the baseline case of logit market shares and no consumer hesitation ($s = 1$), and approaching I as s becomes large. An intuitive interpretation of this fact is that decision conflict inevitably introduces

[9] A related measure that identifies welfare with the proportion of consumers who make an active choice was studied in Spiegler (2015), while Gerasimou and Papi (2018) introduced an index that is similar to W but features instead the probability-weighted product variety that is associated with a strategy profile.

a third "competitor" into the market, the outside option, that becomes more "powerful" as s grows. The power logit predicts that the choice probability of the outside option goes down as the value of one of the two products is unilaterally increased, while the choice probability of the comparatively more appealing product simultaneously goes up during the process. This in turn creates incentives for each firm to unilaterally increase its quality level relative to the baseline logit case. But since increasing quality is costly, the above-mentioned strategic trade-off that is embedded in each firm's profit function eventually kicks in and halts this increase at the above symmetric-equilibrium level.

Figure 5: Power-logit equilibrium quantities in the duopolistic game as the power parameter varies (income, I, is normalized to 1).



Notably, while consumer surplus is zero in equilibrium because each firm's profits turn out to be strictly increasing in its product's price, consumer welfare changes in an interesting way as s varies. In particular, despite the increase in the attainable value level in equilibrium once firms best-respond to consumers' hesitation and resampling, welfare decreases in s. This decrease is caused by the fact that in the power logit with two equally attractive products the consumer is more/equally/less likely to defer than to make an active choice when $s > 2/s = 2/s < 2$ and, conditional on doing the latter, equally likely to choose either of the two available products (Proposition 2). The implication of this in the present environment is that the higher value that the consumer receives *in expectation* under the equilibrium with some decision conflict ($s > 1$) is not sufficiently high to offset the lower value that they receive *with certainty* under the equilibrium with no conflict ($s = 1$). The firms' profits, finally, also decrease when consumers are hesitant relative to the case where there they are not. This large decrease is intuitive and contributed by the reduced probability of the consumer choosing either product, as well as by the reduction in the firms' profit margins that is brought about by the improvement in quality. Figure 5 illustrates these facts graphically when I is normalized to 1.

4 Econometric Estimation

It is often the case in empirical applications that the choice frequencies available to the analyst are obtained from the choices made by a cross section of individuals who are presented with the same menu, rather than from a single decision maker's repeated choices at that menu. Random-utility based discrete choice estimation in those cases is often carried out under the assumption that the observable component of every individual's utility coincides, and that the error term in that model's formulation captures all individual heterogeneity that is unobserved to the analyst. Adopting and adapting this assumption to our non-random-utility environment, in this section (and in Appendix A) we show how the other assumptions and formal argument that underpin the discrete-choice formulation of the logit model without an outside option that was pioneered by McFadden (1973) can be modified to arrive at a similar discrete-choice version of the quadratic- and power-logit models. It is worth remarking that, as we show in Section 5, the use of otherwise standard discrete-choice datasets is sufficient towards estimating these models, as long as they are obtained from a "free choice" decision environment, i.e. one where individuals could choose the no-choice outside option, where the analyst observes both the active choices and those of the latter option.

4.1 Discrete Choice with the Quadratic Logit Model

We start by denoting the set of all quadratic-logit decision makers by $\{1,\ldots,n,\ldots,N\}$. Keeping the menu $A := \{a_1,\ldots,a_i,\ldots,a_k\} \subseteq X$ fixed throughout this and the next subsection, we proceed by recalling and breaking down the baseline assumptions of the discrete-choice formulation of the baseline logit in (6) as follows:

1. *Random utility* [structural assumption]: there is some function $u_n : X \to \mathbb{R}$ such that

$$u_n(a_i) \;=\; v_n(a_i) + \epsilon_{ni}, \tag{21}$$

where $v_n(a_i)$ and ϵ_{ni} are, respectively, the deterministic (observable to the analyst) and stochastic (unobservable) contributions to agent n's utility from good a_i. Denoting by x_{ni} and β, respectively, two m-vectors of observable product/consumer characteristics and estimable coefficients that capture their relative importance via the relationship specified by some function $g : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$, in applications it is assumed that

$$v_n(a_i) \;\equiv\; g(\beta; x_{ni}) \tag{22}$$

and, often, that the dependence of v_n on β and x_{ni} via g is linear-additive:

$$g(\beta; x_{ni}) \;=\; \beta \cdot x_{ni}, \tag{23}$$

where $\cdot$ denotes the inner product.
2. *Random utility maximization* [behavioral assumption]: for all $a_i \in A$,

$$\rho_n(a_i, A) \;=\; Pr\big(u_n(a_i) \geq u_n(a_j) \text{ for all } j \leq k\big). \tag{24}$$

3. *Gumbel noise* [distributional assumption]: the error term ϵ_{ni} is independently and identically distributed across i according to the standard Gumbel density

$$f(\epsilon_{ni}) \;=\; e^{-\epsilon_{ni}} e^{-e^{-\epsilon_{ni}}}. \tag{25}$$

As has been widely known since the seminal contribution of McFadden (1973),[10] these assumptions jointly imply the analytically convenient and famous form

$$\rho_n(a_i, A) \;=\; \frac{e^{v_n(a_i)}}{\sum\limits_{j=1}^{k} e^{v_n(a_j)}}. \tag{26}$$

We proceed by examining how the premises and conclusion of this classic discrete-choice logit model are affected and can be modified when we assume that decision maker n uses the single but noisy value criterion captured by u_n to sample the values of the alternatives in A twice, as per the quadratic special case of the power logit (focusing on the quadratic case here is done for simplicity of the exposition; we deal with the general case later). To this end, and recalling the interpretation that was put forward at the beginning of Section 2.3, we first note that maintaining the additivity and linearity assumption implies that at the end of the second round of sampling the individual has perceived two values for each alternative $a_i \in A$,

$$\begin{aligned} u_n^1(a_i) &= v_n(a_i) + \epsilon_{ni}^1, \\ u_n^2(a_i) &= v_n(a_i) + \epsilon_{ni}^2. \end{aligned}$$

These generally distinct values across the two rounds will vary according to the distribution of ϵ_{ni}. Such multiplicity of value realizations in turn implies that each alternative $a_i \in A$ is ultimately associated with a vector of values $\big(u_n^1(a_i), u_n^2(a_i)\big)$. With utility now being vector-valued, however, the utility-maximization behavioral assumption that underpins (26) is no longer applicable in an obvious way. To break this impasse we assume that the random utility maximization behavioral assumption is replaced by a *dominance* assumption whereby

$$\rho_n(a_i, A) \;=\; Pr\big(u_n^l(a_i) \geq u_n^l(a_j) \text{ for all } j \leq k \text{ and for } l = 1, 2\big). \tag{27}$$

Turning, finally, to the modification of the distributional assumption (25), to make it operational in the quadratic-logit framework we assume that the random errors ϵ_{ni}^1 and ϵ_{ni}^2 are independent across all alternatives $i \leq k$ *and* across the two sampling rounds

[10] Luce and Suppes (1965) and, indeed, McFadden (1973) also credit Eric W. Holman and Anthony A. J. Marley with this discovery.

$l \leq 2$. As was also anticipated in the discussion of Section 3.1, this is indeed a demanding simplifying assumption that we hope future studies will be able to relax.

With these assumptions in place we can now write

$$\begin{aligned}
\rho_n(a_i, A) &\equiv Pr(v_n(a_i) + \epsilon^1_{ni} \geq v_n(a_j) + \epsilon^1_{nj} \; \forall \; j \neq i) \times Pr(v_n(a_i) + \epsilon^2_{ni} \geq v_n(a_j) + \epsilon^2_{nj} \; \forall \; j \neq i) \\
&= Pr(\epsilon^1_{nj} \leq v_n(a_i) + \epsilon^1_{ni} - v_n(a_j) \; \forall \; j \neq i) \times Pr(\epsilon^2_{nj} \leq v_n(a_i) + \epsilon^2_{ni} - v_n(a_j) \; \forall \; j \neq i) \\
&\quad \int_{-\infty}^{\infty} \left(\prod_{j \neq i} e^{-e^{-(\epsilon^1_{ni} + v_n(a_i) - v_n(a_j))}} \right) e^{-\epsilon^1_{ni}} e^{-e^{-\epsilon^1_{ni}}} d\epsilon \\
&= \qquad\qquad \times \\
&\quad \int_{-\infty}^{\infty} \left(\prod_{j \neq i} e^{-e^{-(\epsilon^2_{ni} + v_n(a_i) - v_n(a_j))}} \right) e^{-\epsilon^2_{ni}} e^{-e^{-\epsilon^2_{ni}}} d\epsilon \\
&= \left(\frac{e^{v_n(a_i)}}{\sum_{j=1}^{k} e^{v_n(a_j)}} \right)^2, \qquad (28)
\end{aligned}$$

where each integral is k-dimensional, the first step makes use of the above behavioral, distributional and independence assumptions on ϵ^l_{ni}, while the last step follows from the derivation of the discrete-choice logit [see, for example, Train (2009, pp. 36-37 & 74-75)].

An important difference between the discrete-choice version of the logit *with* an outside option in (5) and its quadratic-logit counterpart is that in the former case the modeller specifies the value of that option *exogenously* (see Anderson, Palma, and Thisse, 1992; Hensher, Rose, and Greene, 2015), whereas in the latter case this value emerges endogenously as a function of the observable characteristics of all active-choice alternatives. Indeed, assuming now—and in the remainder of this section and the next—both (22) and (23), upon rewriting (28) as

$$\rho_n(a_i, A) = \frac{e^{2\beta \cdot x_{ni}}}{\sum_{j=1}^{k} e^{2\beta \cdot x_{nj}} + 2 \sum_{i \neq j} e^{\beta \cdot (x_{ni} + x_{nj})}} \qquad (29)$$

one observes that

$$u_n(a_i) \equiv e^{2\beta \cdot x_{ni}}, \qquad (30)$$

$$D_n(A) \equiv 2 \sum_{i \neq j} e^{\beta \cdot (x_{ni} + x_{nj})}. \qquad (31)$$

By contrast, in the baseline model we have

$$u_n(a_i) \equiv e^{\gamma \cdot x_{ni}}, \qquad (32)$$

$$u_n(o) \equiv e^{\gamma \cdot x_{no}}, \qquad (33)$$

where x_{no} is set by the analyst.

In Appendix A we consider the general case of the power logit model, formulate its likelihood function, and identify the first-order conditions on p and β for its maximization.

5 Proof-of-Concept Empirical Illustration

5.1 Free Choices from 100 Menus with Two Films

For our application we use the survey-experiment data with film choices that were collected by Bhatia and Mullett (2016). In that study, 58 subjects were initially asked to rate from 1 (least desirable) to 9 (most desirable)[11] the 100 most voted-on (hence most popular) films on the IMDB online platform (https://www.imdb.com) at the time. Following that, subjects were presented with 100 distinct binary menus with films that were drawn from that list, with the respective images presented side by side. In the free-choice treatment, subjects were asked to choose either the film positioned on the *left* or on the *right* of each menu, or to *defer* the decision (these choices were entered by clicking on the left, right and up keys, respectively).

Regarding the instructions that subjects received, the authors highlighted (p. 136) that these *"were created to avoid any suggestion of an explicit time limit (e.g. to suggest that participants should defer if they cannot decide quickly enough) or that deferral was a third comparable option (e.g. in the form of a status quo or default movie). More specifically, the instructions stated that if participants preferred the movie on the left/right then they should press the left/right arrow. If they could not make a decision about which of the two movies they preferred then they should press the up arrow instead."* In the forced-choice treatment, the same 100 menus were presented but deferral was not feasible.

The study featured a within-subject design and subjects were randomly assigned to start the experiment in either of the two treatments. There was no limit in the time subjects had available to make their 2×100 decisions.

5.2 Analysis

Although Bhatia and Mullett (2016) focused mainly on the relationship between choice deferral and response times, they also reported on the relationship between ratings and active-choice probabilities *conditional on an active choice being made*. Specifically, they found that the film with a higher rating, where relevant, is chosen 83% of the time (p. 137). Enabled in this way by the theoretical analysis of the previous sections, our focus here instead is on the *unconditional analysis* of the explanatory value of the subjects' own ratings on their subsequent active-choice *and* deferral decisions, and on comparing the results from this analysis when it builds either on the baseline logit with an outside

[11] A typo in Bhatia and Mullett, 2016 erroneously suggests that the highest rating was 7 instead.

option or on the hereby proposed power and quadratic logit.[12] In particular, on each of the 100 binary menus in this dataset (for which, we recall, 58 observations are available) we estimate and compare the goodness of fit of the models that we lay out below.

5.3 Multinomial Logit with a Fixed Outside Option

In line with existing practices (see, for example, pp. 411-414 in Hensher, Rose, and Greene, 2015), to estimate this model we treat the outside option as an explicit alternative with a fixed value that is common to all subjects.[13] Doing so leads to the following three-parameter multinomial logit specification:

$$P_n^{ML}(l,A) = \frac{e^{\beta_0^{l,A}+\beta_1^A rat.Left_n}}{1+e^{\beta_0^{l,A}+\beta_1^A rat.Left_n}+e^{\beta_0^{r,A}+\beta_1^A rat.Right_n}} \tag{34}$$

$$P_n^{ML}(r,A) = \frac{e^{\beta_0^{r,A}+\beta_1^A rat.Right_n}}{1+e^{\beta_0^{l,A}+\beta_1^A rat.Left_n}+e^{\beta_0^{r,A}+\beta_1^A rat.Right_n}} \tag{35}$$

$$P_n^{ML}(o,A) = \frac{1}{1+e^{\beta_0^{l,A}+\beta_1^A rat.Left_n}+e^{\beta_0^{r,A}+\beta_1^A rat.Right_n}} \tag{36}$$

The left-hand-side terms denote the estimated probabilities of subject n choosing "left", "right" or "defer" at binary menu A. On the right hand side, β_1^A and $\beta_0^{l,A}$, $\beta_0^{r,A}$ are, respectively, the estimated slope and intercept coefficients at menu A. The former captures the effect that a unitary increase in subject n's rating of the left (right) film–denoted here by $rat.Left$ ($rat.Right$)–has on the log-odds of choosing that film over deferring when the latter option's value is fixed. The option-specific intercepts $\beta_0^{l,A}$ and $\beta_0^{r,A}$ on the other hand capture the log-odds of choosing, respectively, the left and right film over deferring when the relevant film's rating is zero. Hence, including these terms in the estimation is essential for otherwise the prediction would be equal choice probabilities for "left", "right" and "defer" if both films had a zero rating. This, in turn, would go against the model's treatment of the outside option as any other alternative that is more likely to be chosen as the other feasible options become worse.

5.4 Multinomial Logit with a Randomly-Valued Outside Option

We also consider the variant of the preceding model where, instead of assuming a fixed common value ("rating") for the outside option, we allow it to vary across subjects and menus by randomizing uniformly over the permissible rating values.

[12] We recall that, as was clarified in Sections 2 and 3, the power logit and the baseline logit with an outside option are non-nested models.

[13] Under these two conditions the exact value of the outside option's "rating" is unimportant for this model's maximized log-likelihood and estimate of β_1^A, mattering only for the estimates of $\beta_0^{l,A}$ and $\beta_0^{r,A}$.

5.5 Quadratic Logit

As discussed in the previous subsection, estimating the quadratic logit amounts to estimating the parameter γ^A in

$$\begin{aligned} P_n^{QL}(l,A) &= \left(\frac{e^{\gamma^A \cdot rat.Left}}{e^{\gamma^A \cdot rat.Left} + e^{\gamma^A \cdot rat.Right}}\right)^2 \\ P_n^{QL}(r,A) &= \left(\frac{e^{\gamma^A \cdot rat.Right}}{e^{\gamma^A \cdot rat.Left} + e^{\gamma^A \cdot rat.Right}}\right)^2 \\ P_n^{QL}(o,A) &= 1 - P_n^{QL}(l,A) - P_n^{QL}(r,A) \end{aligned}$$

There are some important differences between this model and the multinomial logit with an outside option laid out above. First, unlike that model, the quadratic logit does not include any intercept terms. This is in line with the theoretical predictions of the general version of this model (Proposition 2), according to which all active-choice options are equally likely to be chosen when they have the same value. Including alternative-specific intercept terms here would go against this prediction as it would lead to generally distinct predicted probabilities for the left and right film when their ratings are identically equal to zero. Second, unlike β^A, the slope coefficient γ^A here captures the log-odds of *choosing one film over the other* (in particular, *not* of choosing one film over deferring) following a unitary change in the former film's rating. More specifically, given (2), (10), (11) and (30), a more appropriate interpretation of this coefficient is that it captures the relevant change in the log-odds of choosing one film over the other following a unitary increase in the former's rating *conditional on an active choice having been made*, while the unconditional change in these log-odds is obtained by multiplying them by $1-\rho(o,A)$. By contrast, (29) clarifies that the log-odds of choosing a film over deferring following a unitary increase in that option's rating is captured by $2\gamma^A$ instead.

5.6 Power Logit

Estimating this more general model now involves finding simultaneously optimal values for the slope coefficient θ^A *and* the power parameter p_A in

$$\begin{aligned} P_n^{PL}(l,A) &= \left(\frac{e^{\theta^A \cdot rat.Left}}{e^{\theta^A \cdot rat.Left} + e^{\theta^A \cdot rat.Right}}\right)^{p_A} \\ P_n^{PL}(r,A) &= \left(\frac{e^{\theta^A \cdot rat.Right}}{e^{\theta^A \cdot rat.Left} + e^{\theta^A \cdot rat.Right}}\right)^{p_A} \\ P_n^{PL}(o,A) &= 1 - P_n^{PL}(l,A) - P_n^{PL}(r,A), \end{aligned}$$

The parameter θ^A here admits an analogous interpretation to γ^A in the quadratic logit, while the term $p_A\theta^A$ is interpretable as the effect that a unitary change in a film's rating

has on the log-odds of choosing that film over deferring.

5.7 Model Estimation and Goodness-of-Fit Summary Comparisons

We perform a goodness-of-fit analysis and comparison of the four models that aim to assess their explanatory and predictive performance separately on each of the 100 menus.[14] To this end, we focus on the maximized log-likelihood value, the Akaike (AIC) and Bayesian (BIC) information criteria, and each model's proportion of correct predictions. In particular, denoting by $\widehat{L}_A$, k and N_A, respectively, a model's maximized log-likelihood value at menu A, the number of its parameters and its sample size, recall that $AIC = 2k - 2\log(\widehat{L}_A)$ and $BIC = k\log(N_A) - 2\log(\widehat{L}_A)$. The value of k is 3 for the two multinomial logit models with a fixed and random outside option, 2 for the power logit and 1 for the quadratic logit. The sample size is $N_A = 58$ in all four models and for each one of the 100 menus. In the prediction analysis we used the models' 100 menu-specific predicted choices per subject ($5800 = 100 \times 58$ in total) to subjects' actual choices at each menu. A model was taken to make a correct prediction for a given subject at a given menu if it predicted a weakly highest choice probability for the option that was actually chosen by that subject in that menu. The same principle was applied in a comprehensive cross-validation analysis where data from all but one of the 100 menus were pooled together and the resulting model estimates were used to make out-of-sample predictions at the excluded menu.

Table 2: Goodness-of-fit comparison of the four models' estimates at the 100 menus.

Model	Parameters	Log-likelihood	AIC	BIC	Correct predictions	
Logit with fixed and inferior outside option	3	85	75	60	2067	35.6%
Logit with random outside option	3	8	7	6	2112	36.4%
Power logit	2	7	16	22	2414	41.6%
Quadratic logit	1	0	2	12	1804	31.1%

Note: the random outside option in the second model was estimated on uniform-random integer values between 1 and 9.

Figure 7 plots the 100 pairs of power- and slope-parameter estimates that emerge from the power-logit model. The mean, median and standard deviation of the p estimates in those regressions are 1.51, 1.47 and 0.27, respectively. The slope-parameter estimates on the other hand have a mean, median and standard deviation of 0.43, 0.40 and 0.15, suggesting that the effect of a one-unit increase in a film's rating is an approximately

[14]The results presented in this subsection were obtained with code written in the R programming language (R Core Team, 2025, v4.5.2) with RStudio (RStudio Team, 2020), and utilising the "mlogit" (Croissant, 2020), "optimx" (Nash and Varadhan, 2011), "plyr" (Wickham, 2011) and "tidyverse" (Wickham et al., 2019) packages/libraries.

Table 3: Summary of cross-validation analysis.

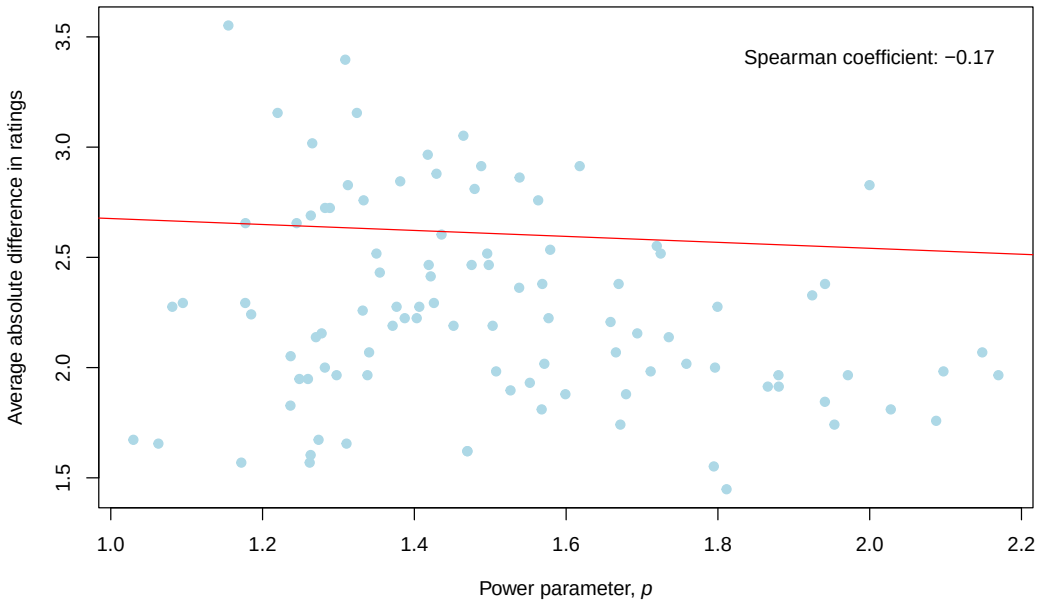
Model	Menus with strictly more correct predictions out-of-sample*	Average** parameter estimates
Logit with fixed and inferior outside option	68	$\overline{\beta}_0^L = -2.87$, $\overline{\beta}_0^R = -2.99$ $\overline{\beta}_1 = 0.59$
Power logit	17	$\overline{p} = 1.48$, $\overline{\theta} = 0.41$

*Both models predicted correctly at 15 menus. **Across all leave-one-menu-out estimations.

53% increase in the odds of choosing that film over the alternative. For comparison, the mean/median and standard deviation in the slope estimates corresponding to the baseline logit with a fixed outside option are 0.58 and 0.15, respectively, pointing to an approximately 78% increase in the above-mentioned odds.

Interestingly, there is a negative correlation (Spearman $\rho = -0.33$) between the $\widehat{p}$ and $\widehat{\theta}$ estimates in these data. The fact that $\widehat{p}$ tends to be lower at menus where $\widehat{\theta}$ is higher, however, can be interpreted intuitively through the lens of this model. Specifically, when $\widehat{p}$ is high, the deferral frequency also tends to be high. When deferrals are primarily caused by the relative undesirability of the two films, as per the logit with an outside option, a higher value of the slope parameter would be expected, in line with the $\overline{\beta}_1 > \overline{\theta}$ finding. This is so because, in this model, the marginal effect of a unitary change in a film's rating is more likely to be high when both films have a low rating. But when deferrals are not primarily due to undesirability but, instead, are mainly caused by decision difficulty, then relatively low values of $\widehat{\theta}$ could be observed not because of low but because of *similar* ratings, and by the harder comparison that such similarity entails.

Figure 6: Joint distribution of the power parameter in the 100 power-logit regressions and the average absolute differences in ratings at the respective menus.



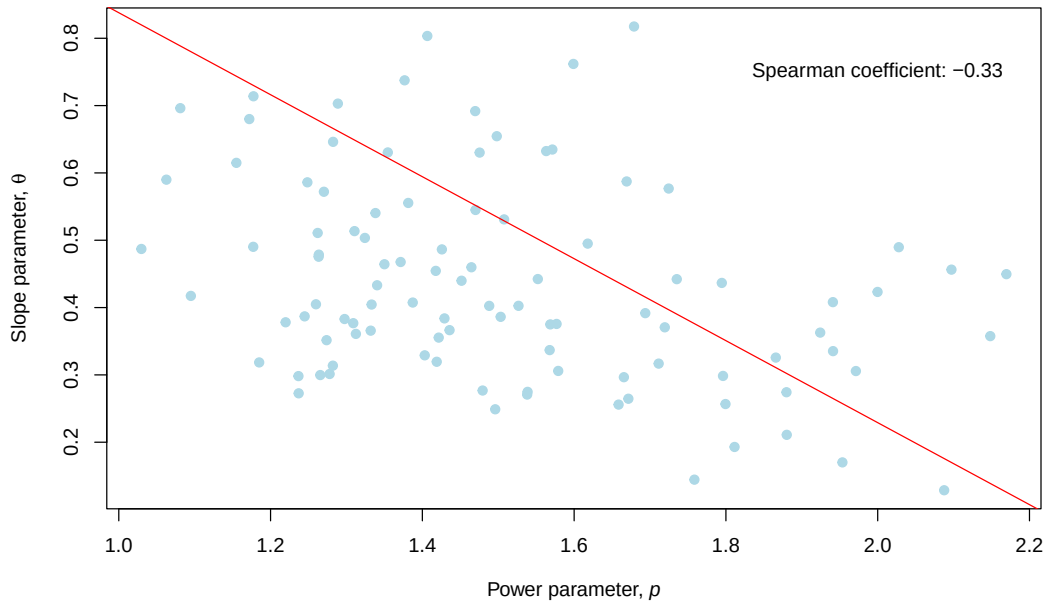
The potential presence of such a channel is further supported by the negative correlation (Spearman $\rho = -0.17$) between the $\widehat{p}$ estimates and average (across subjects) absolute differences in ratings at the respective menus. The mean, median and standard deviation of this variable at the 100 menus are 2.25, 2.21 and 0.45, respectively. The bottom-right quarter of the scatter plot in Figure 6 reveals the presence of 31 menus with an estimated p in excess of its median value of 1.47 and an average absolute difference in ratings between the two films at each of these menus below its median of 2.25. The mean and median estimates of the power-logit slope parameter $\widehat{\theta}$ at these 31 menus are 0.39, while the corresponding statistics in the remaining 69 menus are 0.46 and 0.42. The difference in the distribution of $\widehat{\theta}$ between these two groups is statistically significant ($p = 0.044$; two-sided Mann-Whitney test) and corroborates this intuition and theoretical prediction.

We now turn to the results of the goodness-of-fit comparison that is summarized in Table 2. The logit with an inferior outside option performs better than the other three models in most menus under each of the log-likelihood (85), AIC (75) and BIC (60) criteria. Together, the power and quadratic logit provide the best fits under AIC and BIC in 18 and 34 menus, respectively, followed by the logit with a random outside option (7 and 6 menus). Under the proportion-of-correct-predictions criterion, on the other hand, the power logit performs better (41.6%), followed by the baseline logit with a fixed or random outside option (35.6% and 36.4%, respectively) and by the quadratic logit (31.1%).[15] Our cross-validation analysis (Table 3) focuses on the two models that the preceding analysis suggests are the descriptively leading ones. It reveals that the baseline logit—but not the power logit—makes correct out-of-sample predictions regarding the option that is most likely to be chosen in nearly two thirds of all cases. Conversely, the power logit—but not the baseline logit—makes correct out-of-sample predictions in nearly one fifth of all cases. Finally, both models make correct out-of-sample predictions 15% of the time.

Further light on the relevance of the behavioral channel that was discussed earlier can now be shed by comparing the models' fit in those menus where the average film ratings are high and low. This is relevant because the mechanism underpinning the logit with a fixed outside option suggests that choosing that option is more likely when the average rating is low. Intuitively, therefore, we would expect this model to provide a better fit in the latter group of menus than the power logit does. To this end, we compare the two models' AIC and BIC scores in the two groups of 50 menus with above- and below-median average total rating (the median value of this statistic is 11.44). In line with this intuition, the baseline logit performs better than the power logit in a higher proportion of the 50 menus with a low average rating than those with a high such rating, under both criteria (AIC: 92% vs 74%; BIC: 84% vs 60%), with the difference in proportions being

[15] The 5- and 4-percentage point differences in the rates of correct predictions between the power logit and the baseline logit with either an inferior or a random outside option are significant ($p < 0.001$ in both cases; p-values from 2-sided Fisher's exact tests).

Figure 7: Joint distribution of the power and slope parameters in the 100 power-logit regression estimates.



significant in both cases.[16]

These results suggest that the hereby proposed class of power-logit discrete-choice models with an endogenously determined menu-dependent value of the outside option can indeed provide meaningful explanatory gains relative to the baseline logit model with a fixed or random outside option. Moreover, these explanatory gains often occur in those decision environments where intuition and the theoretical analysis of previous sections would suggest that the proposed model should indeed perform better. We hope that this illustration will be helpful to the experimenter or empirical researcher who is interested in creating and analyzing similar free-choice datasets.

6 Related Literature

As was illustrated in the proof-of-concept empirical application of the previous section, and as was also explained in Section 2, standard discrete choice models with an outside option that are based on random-utility maximization treat this option just like any other alternative and predict that it is more likely to be chosen when its utility is higher than that of all feasible *active-choice* alternatives. Anderson, Palma, and Thisse (1992) and Hensher, Rose, and Greene (2015), for example, are textbook references that discuss this approach in detail. The models that we study in this paper differ radically from this (un-)desirability approach to modelling choice of the outside option, as they predict that every active-choice alternative is always chosen when it is the only feasible one (A3, Section 2). In addition, the more structured power-logit special case of this model predicts that the probability of opting out at larger menus increases as the feasible such alternatives become more equally appealing, in line with relevant empirical evidence.

[16]The respective p-values from two-sided Fisher's exact tests are $p = 0.03$ and $p = 0.01$.

Starting with Manzini and Mariotti (2014), several random choice models of limited attention—also logically distinct from the modelling framework that this paper proposes—have included an outside option as a model-closing assumption that requires this option to be chosen when no attention is paid to any of the feasible active-choice alternatives (Brady and Rehbeck, 2016; Aguiar, 2017; Aguiar, Boccardi, Kashaev, and Kim, 2023). Because of this assumption, deferring/opting out becomes less likely in these models as menus become bigger. Horan (2019) recently clarified how the deferral option can be removed from these models without affecting their general features and primary purpose, which is to explain active-choice decision making subject to cognitive/attention constraints.

We proceed with a more detailed comparison of the logit with a general outside option that is formulated in (1) and the intuitive generalization of the classic *nested logit* model (Ben-Akiva, 1973; McFadden, 1978) that was recently proposed and analysed in Kovach and Tserenjigmid (2022a). The latter assumes that the set of alternatives X can be partitioned into nests $X_1, \ldots, X_K$, and that there exist a non-negative function v^* on the collection $\bigcup_{i=1}^{K} 2^{X_i}$ and a strictly positive function u^* on X such that, for all $A \in \mathcal{M}$ and $a \in A \cap X_i$,

$$\rho(a, A) \;=\; \frac{v^*(A \cap X_i)}{\sum_{j \leq K} v^*(A \cap X_j)} \frac{u^*(a)}{\sum_{b \in A \cap X_i} u^*(b)} \tag{37}$$

That paper did not consider an outside option as it focused on explaining the kinds of canonical violations of A2 that motivated the original development of nested logit as a generalization of baseline logit. Yet the version of (37) that is closest to (1) emerges when: (i) X is expanded to $X \cup \{o\}$ and partitioned into the nests[17] $\{X_1 = \{o\}, X_2 = X\}$; (ii) the collection of menus is $\mathcal{M}^* := \{A \cup \{o\} : \emptyset \neq A \subseteq X\}$. In this case the choice probability of active-choice alternative a at decision problem $A^* \in \mathcal{M}^*$ reduces to

$$\rho(a, A^*) \;=\; \frac{v^*(A^* \setminus \{o\})}{v^*(\{o\}) + v^*(A^* \setminus \{o\})} \frac{u^*(a)}{\sum_{b \in A^* \setminus \{o\}} u^*(b) + u^*(o)} \tag{38}$$

Using the notational convention $A^* \equiv A \cup \{o\}$ for $A \in \mathcal{M}$, a little algebra shows that (1) and (38) become equivalent if and only if $u(a) \equiv u^*(a)$ for all $a \in X$ and $D(A) \equiv \dfrac{v^*(\{o\})u^*(o)}{v(A)}$ for all $A \in \mathcal{M}$. Thus, unless $v^*(\{o\}) = 0$ or $u^*(o) = 0$, equivalence between the logit with a general outside option and the generalized nested logit with a fixed outside option is possible *only if* $D(A) > 0$ for every $A \in \mathcal{M}$. The models that are representable as in (1) which do not impose this restriction at singleton menus. Moreover, (38) has two additional degrees of freedom compared to (1): one because u^* takes $|X \cup \{o\}| = |X| + 1$ values; and another because v^* takes $|\mathcal{M}^* \cup \{o\}| = |\mathcal{M}| + 1$ values. Therefore, even when $D(A) > 0$ for all $A \in \mathcal{M}$, (38) is not uniquely recoverable from (1). Thus, despite the

[17] See, for example, the top branch of the auto-mobile choice model in Figure 1 of Goldberg (1995).

structural similarity between (1) and (38)—which is perhaps best seen by contrasting the two multiplicative terms in generalized nested logit with the corresponding ones in (2)—the two models differ in some essential ways.

We note, finally, that also related to (1) but logically and interpretively distinct from it is the *focal* logit model of Kovach and Tserenjigmid (2022b) when a fixed outside option is introduced into the latter. That model's components comprise: (i) a menu-independent value function over alternatives u^{**} on $X \cup \{o\}$; (ii) a menu-dependent focus function F that assigns a consideration set $F(A^*)$ to every problem $A^* \in \mathcal{M}^*$; (iii) a menu-dependent focality bias function δ that gives a 'value boost' to alternatives in $F(A^*)$. Formally, the choice probability of active-choice alternative $a \in A^*$ in this model is given by

$$\rho(a, A^*) \;=\; \frac{u^{**}(a)\,(1 + \delta(A) \times \mathbb{1}\{a \in F(A^*)\})}{\sum_{b \in A^*} u^{**}(b)\,(1 + \delta(A) \times \mathbb{1}\{b \in F(A^*)\})}, \tag{39}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. Although (1) and (39) are distinct, they intersect in the special case where $u(a) \equiv u^{**}(a)$ for all $a \in X$; $u^{**}(o) \equiv 1$; $F(A^*) \equiv \{o\}$ for all $A^* \in \mathcal{M}^*$; and hence $\delta(A^*) \equiv D(A) - 1$ for all $A^* \in \mathcal{M}$ (recall that $A \equiv A^* \setminus \{o\}$).[18] The last restriction implies $D(\{a\}) > 0$ for all $a \in X$, hence $\rho(a, \{a\}) < 1$. The special cases of (1) that we studied here do not impose this restriction.

7 Concluding Remarks

Understanding the "easy" and "hard" parts of people's preference comparisons, as these are revealed by their active-choice or choice-avoidance/delay decisions, is important methodologically and also for practical applications such as effective choice architecture. The present paper contributes in this respect by introducing the tractable power logit model and its quadratic-logit special cases. These models—both of which belong to a general class of logit models with a context-dependent outside option—assume that people can avoid/delay making an active choice and are more likely to opt out in free-choice problems when it is harder for them to identify a best alternative from those available to them. This prediction is supported empirically and differs from the predictions of existing models where the outside option is chosen due to the undesirability of all feasible alternatives, limited attention, or other sources of bounded-rational behavior. In conjunction with the insights from the relevant decision-making literature, our analysis suggests that decision-conflict logit models can help theoretical and applied empirical economists think formally and perhaps more realistically about non-strategic as well as strategic situations where decision makers: (i) are presented sufficiently small menus, so that limited-attention considerations are not pertinent; (ii) consider all feasible active-choice alternatives to be desirable/good enough, so that any one of them would

[18] The author is grateful to Levent Ülkü for alerting him to this connection between (1) and (39).

be expected to be chosen if it were the only feasible item; (iii) find it difficult to compare these alternatives due to their complexity or due to potentially non-trivial trade-offs these generate; and (iv) are not forced to make an active choice.

Appendix: Proofs

Proof of Proposition 1. In the main text.

Proof of Proposition 2.

For the second claim, suppose $\widehat{u}(a) = \widehat{u}(b) := c$ for all $a, b \in A$. By (7), $\rho(o, A) = 1 - \sum_{a \in A} \left(\frac{\widehat{u}(a)}{\sum_{b \in A} \widehat{u}(b)} \right)^p = 1 - |A| \left(\frac{c}{|A|c} \right)^p = 1 - |A|^{1-p}$. Thus, $\rho(o, A)$ is independent of the specific $\widehat{u}$ values at A whenever these values coincide. This readily implies that, viewed as the function

$$\rho(o, A; \widehat{u}(a_1), \ldots, \widehat{u}(a_{|A|})) = \frac{\left(\sum_{i=1}^{|A|} \widehat{u}(a_i) \right)^p - \sum_{i=1}^{|A|} \widehat{u}(a_i)^p}{\left(\sum_{i=1}^{|A|} \widehat{u}(a_i) \right)^p}, \tag{40}$$

$\rho(o, A)$ has any $|A|$-vector of $\widehat{u}$ values $(c, \ldots, c)$ as a critical point that trivially satisfies both the first- and second-order conditions of local optimality. Yet, because the determinant of the Hessian matrix at any such point is zero, it is not immediately clear if this point is a local maximizer. To show that this is indeed so, by symmetry it suffices to consider marginal deviations in a single direction; say, an ϵ increase or decrease in $\widehat{u}(a_1)$. Since $\widehat{u}(a_i) = \widehat{u}(a_j) \equiv c > 0$, by assumption, this and (40) yield

$$\begin{aligned} \rho(o, A; c + \epsilon, c, \ldots, c) &= \frac{(|A|c + \epsilon)^p - (|A| - 1)c^p - (c + \epsilon)^p}{(|A|c + \epsilon)^p} \\ &= 1 - \frac{(|A| - 1)c^p + (c + \epsilon)^p}{(|A|c + \epsilon)^p}, \end{aligned}$$

Suppose to the contrary that this weakly exceeds $1 - |A|^{1-p}$. Without loss of generality, write $\epsilon := mc$ for some small $m > 0$ or $m < 0$. We have

$$\begin{aligned} |A|^{1-p} &\geq \frac{(|A| - 1)c^p + (c + \epsilon)^p}{(|A|c + \epsilon)^p} \\ &= \frac{(|A| - 1)c^p + (c + mc)^p}{(|A|c + mc)^p} \end{aligned}$$

To ease notation, write $n := |A|$. Rearranging, observe that the above is true if and only if $n^{1-p}(n + m)^p c^p \geq (n - 1)c^p + (1 + m)^p c^p$, which in turn is true if and only if $n^{1-p}(n+m)^p \geq n-1+(1+m)^p$. Rearranging further, we get $n \left(\frac{n + m}{n} \right)^p \geq n+1+(m+1)^p$

from which we finally obtain $\left(1+\frac{m}{n}\right)^p - \frac{n+1+(m+1)^p}{n} \geq 0$ Taking the limit as $m \to 0$ and rearranging leads to $n \geq n+2$, which is impossible. We have therefore established that the above critical point is indeed a local maximizer of $\rho(o, A)$.

We proceed toward showing that it is in fact a global maximizer, thereby concluding the proof. To this end, notice first that $\rho(o, A) < 1$, by strict positivity of $\widehat{u}$. Suppose to the contrary that there is a non-constant $|A|$-vector $(\widehat{u}(a_1), \ldots, \widehat{u}(a_{|A|})$ that satisfies the first-order conditions of optimality that are derived from (40). Differentiating and rearranging pins down these conditions to

$$\widehat{u}(a_i^*) \;=\; \left(\frac{\sum\limits_{j \neq i} \widehat{u}(a_j)}{\sum\limits_{j \neq i} \widehat{u}(a_j)^p} \right)^{\frac{1}{1-p}}, \qquad i = 1, \ldots, |A| \tag{41}$$

Solving this system leads to $\widehat{u}(a_1^*) = \widehat{u}(a_2^*) = \ldots = \widehat{u}(a_{|A|}^*)$, contradicting the supposed non-constancy of the postulated alternative local maximizer. It follows that $\rho(o, A)$ is maximized at any constant $|A|$-vector only. From this and the second claim that was established earlier it now follows that this maximum is indeed given by $1 - |A|^{1-p}$, as per the first claim. ■

Proof of Proposition 3.

To dispense with the absolute value sign, assume without loss of generality that $\widehat{u}(a) > \widehat{u}(b)$ and $\widehat{u}(c) > \widehat{u}(d)$. We will first show that (16) holds under either of the postulated conditions. Following that, we will show that (16) $\Leftrightarrow$ (17), also under either condition.

Starting with (16), consider first the case where $\widehat{u}(a) + \widehat{u}(b) = \widehat{u}(c) + \widehat{u}(d)$. Denote this common sum by s. We have $\rho(o, \{a, b\}) > \rho(o, \{c, d\}) \Leftrightarrow \frac{\widehat{u}(c)^p + \widehat{u}(d)^p}{s^p} > \frac{\widehat{u}(a)^p + \widehat{u}(b)^p}{s^p}$. This is equivalent to

$$\widehat{u}(c)^p + \widehat{u}(d)^p \;>\; \widehat{u}(a)^p + \widehat{u}(b)^p \tag{42}$$

Suppose to the contrary that

$$\widehat{u}(a) - \widehat{u}(b) \;\geq\; \widehat{u}(c) - \widehat{u}(d). \tag{43}$$

This and the postulated equality yield $\widehat{u}(a) \geq \widehat{u}(c)$. Furthermore, this and (42) jointly imply $\widehat{u}(a) > \widehat{u}(c)$ and $\widehat{u}(d) > \widehat{u}(b)$. Thus,

$$\widehat{u}(a) > \widehat{u}(c) > \widehat{u}(d) > \widehat{u}(b) \tag{44}$$

In view of (44), observe that the terms $\frac{\widehat{u}(a) - \widehat{u}(c)}{\widehat{u}(a) - \widehat{u}(b)}$ and $\frac{\widehat{u}(c) - \widehat{u}(b)}{\widehat{u}(a) - \widehat{u}(b)}$ are convex weights. Hence,

since $\widehat{u}(\cdot) \mapsto \widehat{u}(\cdot)^p$ is a strictly convex function, we have

$$\begin{aligned}
\left(\frac{\widehat{u}(a)-\widehat{u}(c)}{\widehat{u}(a)-\widehat{u}(b)}\right)\widehat{u}(b)^p + \left(\frac{\widehat{u}(c)-\widehat{u}(b)}{\widehat{u}(a)-\widehat{u}(b)}\right)\widehat{u}(a)^p &> \left[\left(\frac{\widehat{u}(a)-\widehat{u}(c)}{\widehat{u}(a)-\widehat{u}(b)}\right)\widehat{u}(b) + \left(\frac{\widehat{u}(c)-\widehat{u}(b)}{\widehat{u}(a)-\widehat{u}(b)}\right)\widehat{u}(a)\right]^p \\
&= \widehat{u}(c)^p, \qquad (45) \\
\left(\frac{\widehat{u}(a)-\widehat{u}(d)}{\widehat{u}(a)-\widehat{u}(b)}\right)\widehat{u}(b)^p + \left(\frac{\widehat{u}(d)-\widehat{u}(b)}{\widehat{u}(a)-\widehat{u}(b)}\right)\widehat{u}(a)^p &> \left[\left(\frac{\widehat{u}(a)-\widehat{u}(d)}{\widehat{u}(a)-\widehat{u}(b)}\right)\widehat{u}(b) + \left(\frac{\widehat{u}(d)-\widehat{u}(b)}{\widehat{u}(a)-\widehat{u}(b)}\right)\widehat{u}(a)\right]^p \\
&= \widehat{u}(d)^p \qquad (46)
\end{aligned}$$

Adding (45) to (46) and recalling that $\widehat{u}(a)+\widehat{u}(b) = \widehat{u}(c)+\widehat{u}(d) = s$ yields

$$\begin{aligned}
\widehat{u}(c)^p + \widehat{u}(d)^p &< \left(\frac{2\widehat{u}(a)-\widehat{u}(c)-\widehat{u}(d)}{\widehat{u}(a)-\widehat{u}(b)}\right)\widehat{u}(b)^p + \left(\frac{\widehat{u}(c)+\widehat{u}(d)-2\widehat{u}(b)}{\widehat{u}(a)-\widehat{u}(b)}\right)\widehat{u}(a)^p \\
&= \left(\frac{2\widehat{u}(a)-s}{\widehat{u}(a)-\widehat{u}(b)}\right)\widehat{u}(b)^p + \left(\frac{s-2\widehat{u}(b)}{\widehat{u}(a)-\widehat{u}(b)}\right)\widehat{u}(a)^p \\
&= \widehat{u}(b)^p + \widehat{u}(a)^p,
\end{aligned}$$

which contradicts (42). Thus,

$$\widehat{u}(a)-\widehat{u}(b) < \widehat{u}(c)-\widehat{u}(d) \qquad (47)$$

holds. Conversely, suppose (47) is true. This and the postulated equality together imply

$$\widehat{u}(c) > \widehat{u}(a) > \widehat{u}(b) > \widehat{u}(d) \qquad (48)$$

Applying the preceding convexity argument using (48) yields (42), thereby completing the proof that (16) holds under the first postulate.

We now show that (16) is true when $u(a)+u(b) = u(c)+u(d)$ or, equivalently,

$$\widehat{u}(a)^p + \widehat{u}(b)^p = \widehat{u}(c)^p + \widehat{u}(d)^p \qquad (49)$$

holds instead. Let t denote this common sum. We have $\rho(o,\{a,b\}) > \rho(o,\{c,d\}) \Leftrightarrow \frac{\widehat{u}(c)^p+\widehat{u}(d)^p}{\left(\widehat{u}(c)+\widehat{u}(d)\right)^p} > \frac{\widehat{u}(a)^p+\widehat{u}(b)^p}{\left(\widehat{u}(a)+\widehat{u}(b)\right)^p} \Leftrightarrow \frac{t}{\left(\widehat{u}(c)+\widehat{u}(d)\right)^p} > \frac{t}{\left(\widehat{u}(a)+\widehat{u}(b)\right)^p} \Leftrightarrow \left(\widehat{u}(a)+\widehat{u}(b)\right)^p > \left(\widehat{u}(c)+\widehat{u}(d)\right)^p$. This is true if and only if $\widehat{u}(a)+\widehat{u}(b) > \widehat{u}(c)+\widehat{u}(d)$, which is equivalent to

$$\widehat{u}(a)-\widehat{u}(c) > \widehat{u}(d)-\widehat{u}(b) \qquad (50)$$

Suppose to the contrary that

$$\widehat{u}(a)-\widehat{u}(b) \geq \widehat{u}(c)-\widehat{u}(d) \qquad (51)$$

From (50) and (51) we get $\widehat{u}(a) > \widehat{u}(c)$ and $\widehat{u}(b) < \widehat{u}(d)$. Thus,

$$\widehat{u}(a) > \widehat{u}(c) > \widehat{u}(d) > \widehat{u}(b) > 0 \qquad (52)$$

By (50), (51) and convexity of $\widehat{u}(\cdot) \mapsto \widehat{u}(\cdot)^p$ we have

$$\widehat{u}(a)^p - \widehat{u}(c)^p \;>\; \widehat{u}(d)^p - \widehat{u}(b)^p,$$

which contradicts (49). Hence, (47) holds. Conversely, suppose (47) is true and assume to the contrary that (50) is violated, i.e.

$$\widehat{u}(a) - \widehat{u}(c) \;\leq\; \widehat{u}(d) - \widehat{u}(b) \tag{53}$$

Rearranging (50),

$$\widehat{u}(a) - \widehat{u}(c) \;<\; \widehat{u}(b) - \widehat{u}(d) \tag{54}$$

By (53) + (54) we obtain $\widehat{u}(a) < \widehat{u}(c)$. This and (49) in turn imply $\widehat{u}(b) < \widehat{u}(d)$. Hence,

$$\widehat{u}(c) > \widehat{u}(a) > \widehat{u}(b) > \widehat{u}(d) > 0 \tag{55}$$

By (53) we have

$$\widehat{u}(c) - \widehat{u}(a) \;\geq\; \widehat{u}(b) - \widehat{u}(d) \tag{56}$$

Finally, (55), (56) and convexity of $\widehat{u}(\cdot) \mapsto \widehat{u}(\cdot)^p$ jointly lead to the same contradiction as above. This completes the proof that (16) holds under the second postulate as well.

We now show that (17) holds under either of the postulated conditions. That is, we verify that $\widehat{u}(a) - \widehat{u}(b) < \widehat{u}(c) - \widehat{u}(d) \Leftrightarrow \widehat{u}(a)^p - \widehat{u}(b)^p < \widehat{u}(c)^p - \widehat{u}(d)^p$. Suppose first that $\widehat{u}(a) + \widehat{u}(b) = \widehat{u}(c) + \widehat{u}(d)$. Let $\widehat{u}(a) - \widehat{u}(b) < \widehat{u}(c) - \widehat{u}(d)$ be true and assume to the contrary that

$$\widehat{u}(a)^p - \widehat{u}(b)^p \;\geq\; \widehat{u}(c)^p - \widehat{u}(d)^p \tag{57}$$

The former two assumptions imply $\widehat{u}(a) < \widehat{u}(c)$, $\widehat{u}(b) > \widehat{u}(d)$ and therefore

$$\widehat{u}(c) > \widehat{u}(a) > \widehat{u}(b) > \widehat{u}(d) \tag{58}$$

Using again the convexity argument that revolved around (45) and (46) we get

$$\widehat{u}(a)^p + \widehat{u}(b)^p > \widehat{u}(c)^p + \widehat{u}(d)^p \tag{59}$$

By (57) and (59) we now obtain $\widehat{u}(a) > \widehat{u}(c)$, which is a contradiction. Conversely, suppose $\widehat{u}(a)^p - \widehat{u}(b)^p < \widehat{u}(c)^p - \widehat{u}(d)^p$ and assume to the contrary that $\widehat{u}(a) - \widehat{u}(b) \geq \widehat{u}(c) - \widehat{u}(d)$. This and $\widehat{u}(a) + \widehat{u}(b) = \widehat{u}(c) + \widehat{u}(d)$ jointly imply $\widehat{u}(a) > \widehat{u}(c)$ and $\widehat{u}(b) < \widehat{u}(d)$. Thus, we have $\widehat{u}(a) > \widehat{u}(c) > \widehat{u}(d) > \widehat{u}(b)$. Using the above convexity argument once again we obtain $\widehat{u}(c)^p + \widehat{u}(d)^p < \widehat{u}(a)^p + \widehat{u}(b)^p$. Subtracting $\widehat{u}(a)^p - \widehat{u}(b)^p < \widehat{u}(c)^p - \widehat{u}(d)^p$ from this inequality yields $\widehat{u}(b) > \widehat{u}(d)$, a contradiction.

Finally, we establish (17) under the postulate

$$\widehat{u}(a)^p + \widehat{u}(b)^p \;=\; \widehat{u}(c)^p + \widehat{u}(d)^p \tag{60}$$

Let

$$\widehat{u}(a) - \widehat{u}(b) \;<\; \widehat{u}(c) - \widehat{u}(d), \tag{61}$$

and again assume to the contrary that (57) is true. By (57) + (54) we get $\widehat{u}(a) \geq \widehat{u}(c)$. This and (61) implies $\widehat{u}(b) > \widehat{u}(d)$. But $\widehat{u}(a) \geq \widehat{u}(c)$ and (60) also implies $\widehat{u}(b) \leq \widehat{u}(d)$. This is impossible. Conversely, suppose $\widehat{u}(a)^p - \widehat{u}(b)^p < \widehat{u}(c)^p - \widehat{u}(d)^p$. This and the postulated $\widehat{u}(a)^p + \widehat{u}(b)^p = \widehat{u}(c)^p + \widehat{u}(d)^p$ jointly imply $\widehat{u}(c) > \widehat{u}(a)$ and $\widehat{u}(b) < \widehat{u}(d)$. Together with the without-loss initial assumption whereby $\widehat{u}(a) > \widehat{u}(b)$ and $\widehat{u}(c) > \widehat{u}(d)$, this in turn implies $\widehat{u}(c) > \widehat{u}(a) > \widehat{u}(d) > \widehat{u}(b)$. Assume to the contrary that $\widehat{u}(a) - \widehat{u}(b) \geq \widehat{u}(c) - \widehat{u}(d)$. This is equivalent to $\widehat{u}(d) - \widehat{u}(b) \geq \widehat{u}(c) - \widehat{u}(a) > 0$. Rearranging (60), we also have $\widehat{u}(b)^p - \widehat{u}(d)^p = \widehat{u}(c)^p - \widehat{u}(a)^p$. Since $\widehat{u}(\cdot) \mapsto \widehat{u}(\cdot)^p$ is a strictly increasing function, it follows from the above that the left hand side of this equation is negative while the right hand positive. This is a contradiction. Thus, (17) holds in this case too. ■

Proof of Proposition 4.

Firm $i = 1, 2$ maximizes π_i with respect to q_i and p_i taking the choices of the other firm $j \neq i$ as given. Differentiating π_i with respect to p_i, q_i and simplifying we get

$$\begin{aligned}
\frac{\partial \pi_i}{\partial p_i} &= \frac{\left(\dfrac{p_j q_i}{(p_j q_i + p_i q_j}\right)^s (p_j q_i + q_j(p_i - p_i s + q_i s))}{p_j q_i + p_i q_j}, \\
\frac{\partial \pi_i}{\partial q_i} &= -\frac{\left(\dfrac{p_j q_i}{p_j q_i + p_i q_j}\right)^{(1+s)} (p_j q_i^2 + p_i q_j (q_i - p_i s + q_i s))}{p_j q_i^2}
\end{aligned}$$

Setting the two equations equal to zero yields the first-order conditions

$$p_i^* = \frac{q_i(p_j + q_j s)}{q_j(s-1)}, \tag{62}$$

$$q_i^* = \frac{p_i \sqrt{q_j}\sqrt{q_j + 4p_j s + 2q_j s + q_j s^2} - p_i q_j - p_i q_j s}{2p_j} \tag{63}$$

It can be checked upon rearranging these conditions in $\frac{q_i}{p_i}$ form (which, in particular, is a non-negative term) and simplifying that they cannot be satisfied simultaneously under the assumption that $p_i, q_i, s \geq 0$ and $I > 0$. This implies that there is no equilibrium where firms choose interior strategies. Since $q_i^* \leq p_i^*$ must hold, this fact and (62), (63) together imply either $p_i^* = 0$ or $p_i^* = I$. Because the latter (former) case is associated with a strictly positive (zero) profit, it follows that

$$p_i^* = I$$

for $i = 1, 2$. Since the problem is symmetric, by (63) and $p_i^* = I$ we get

$$q_i^* = \frac{1}{2}\left(\sqrt{q_j(4sI + q_j(1+s)^2)} - q_j(1+s)\right)$$

for $i = 1, 2$. Solving this system yields

$$q_i^* = \frac{sI}{2+s},$$

as claimed. The remaining assertions are verifiable by simple substitution. ■

Core References

Aguiar, Victor (2017). "Random Categorization and Bounded Rationality". In: *Economics Letters* 159, pp. 46–52.

Aguiar, Victor H., Maria Jose Boccardi, Nail Kashaev, and Jeongbin Kim (2023). "Random utility and limited consideration". In: *Quantitative Economics* 14, pp. 71–116.

Anderson, Simon P., Andrè de Palma, and Jacques-Francois Thisse (1992). *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.

Apesteguia, Jose, Miguel A. Ballester, and Jay Lu (2017). "Single-Crossing Random Utility Models". In: *Econometrica* 85, pp. 661–674.

Bachi, Benjamin and Ran Spiegler (2018). "Buridanic Competition". In: *Games and Economic Behavior* 107, pp. 298–315.

Baldassi, Carlo, Simone Cerreia-Vioglio, Fabio Maccheroni, Massimo Marinacci, and Marco Pirazzini (2020). "A Behavioral Characterization of the Drift Diffusion Model and its Multi-alternative Extension for Choice under Time Pressure". In: *Management Science* 66, pp. 5075–5093.

Ben-Akiva, Mosche E. (1973). "Structure of Passenger Travel Demand Models". PhD thesis. MIT.

Bhatia, Sudeep and Timothy L. Mullett (2016). "The Dynamics of Deferred Decision". In: *Cognitive Psychology* 86, pp. 112–151.

— (2018). "Similarity and Decision Time in Preferential Choice". In: *Quarterly Journal of Experimental Psychology* 71, pp. 1276–1280.

Block, H. D. and J. Marschak (1960). "Random Orderings and Stochastic Theories of Response". In: *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Ed. by I. Olkin. Stanford, CA: Stanford University Press, pp. 97–132.

Bordallo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2013). "Salience and Consumer Choice". In: *Journal of Political Economy* 121, pp. 803–843.

Brady, Richard L. and John Rehbeck (2016). "Menu-Dependent Stochastic Feasibility". In: *Econometrica* 84, pp. 1203–1223.

Carroll, Gabriel D., James J. Choi, David Laibson, Brigitte C. Madrian, and Andrew Metrick (2009). "Optimal Defaults and Active Decisions". In: *Quarterly Journal of Economics* 124, pp. 1639–1674.

Chernev, Alexander, Ulf Böckenholt, and Joseph Goodman (2015). "Choice Overload: A Conceptual Review and Meta-Analysis". In: *Journal of Consumer Psychology* 25, pp. 333–358.

Costa-Gomes, Miguel, Carlos Cueva, Georgios Gerasimou, and Matus Tejiščák (2022). "Choice, Deferral and Consistency". In: *Quantitative Economics* 13, pp. 1297–1318.

Danan, Eric and Anthony Ziegelmeyer (2006). "Are Preferences Complete? An Experimental Measurement of Indecisiveness Under Risk". In: *Working Paper*.

Dhar, Ravi (1997). "Consumer Preference for a No-Choice Option". In: *Journal of Consumer Research* 24, pp. 215–231.

Dhar, Ravi and Itamar Simonson (2003). "The Effect of Forced Choice on Choice". In: *Journal of Marketing Research* 40, pp. 146–160.

Fudenberg, Drew, Whitney K. Newey, Philipp Strack, and Tomasz Strzalecki (2020). "Testing the Drift-Diffusion Model". In: *Proceedings of the National Academy of Science* 117, pp. 33141–33148.

Garvelink, Mirjam M., Laura Boland, Krystal Klein, Don Vu Nguyen, Matthew Menear, Hilary L. Bekker, Karen B. Eden, Annie LeBlanc, Annette M. O'Connor, Dawn Stacey, and France Légaré (2019). "Decisional Conflict Scale Use over 20 Years: The Anniversary Review". In: *Medical Decision Making* 39, pp. 301–314.

Gerasimou, Georgios (2016). "Asymmetric Dominance, Deferral and Status Quo Bias in a Behavioral Model of Choice". In: *Theory and Decision* 80, pp. 295–312.

Gerasimou, Georgios and Mauro Papi (2018). "Duopolistic Competition with Choice-Overloaded Consumers". In: *European Economic Review* 101, pp. 330–353.

Goldberg, Pinelopi Koujianou (1995). "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry". In: *Econometrica* 63, pp. 891–951.

Greene, William H. and David A. Hensher (2010). *Modelling Ordered Choices*. New York: Cambridge University Press.

Hensher, David A., John M. Rose, and William H. Greene (2015). *Applied Choice Analysis*. New York: Cambridge University Press.

Horan, Sean (2019). "Random Consideration and Choice: A Case Study of Default Options". In: *Mathematical Social Sciences* 102, pp. 73–84.

Iyengar, Sheena S. and Mark R. Lepper (2000). "When Choice is Demotivating: Can One Desire Too Much of a Good Thing?" In: *Journal of Personality and Social Psychology* 79, pp. 995–1006.

Iyengar, Sheena Sethi, Gur Huberman, and Wei Jiang (2004). "How Much Choice is Too Much? Contributions to 401(k) Retirement Plans". In: *Pension Design and Structure: New Lessons from Behavioral Finance*. Ed. by Olivia S. Mitchell and Stephen P. Utkus. New York: Oxford University Press, pp. 83–96.

Janis, Irving L. and Leon Mann (1977). *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment*. New York: Free Press.

Knops, Anouk M., Dirk T. Ubbink, Dink A. Legemate, Lukas J. Stalpers, and Patrick M. Bossuyt (2013). "Interpreting Patient Decisional Conflict Scores: Behavior and Emotions in Decisions about Treatment". In: *Medical Decision Making* 33, pp. 78–84.

Kovach, Matthew and Gerelt Tserenjigmid (2022a). "Behavioral Foundations of Nested Stochastic Choice and Nested Logit". In: *Journal of Political Economy* 130, pp. 2411–2462.

— (2022b). "The Focal Luce Model". In: *American Economic Journal: Microeconomics* 14, pp. 378–413.

Krajbich, Ian, Carrie Armel, and Antonio Rangel (2010). "Visual Fixations and the Computation and Comparison of Value in Simple Choice". In: *Nature Neuroscience* 13, pp. 1292–1298.

Luce, R. Duncan (1959). *Individual Choice Behavior: A Theoretical Analysis*. New York, NY: Wiley.

Luce, R. Duncan and Patrick Suppes (1965). "Preference, Utility and Subjective Probability". In: *Handbook of Mathematical Psychology, Volume 3*. Ed. by R. D. Luce, R. R. Bush, and E. H. Galanter. New York: Wiley, pp. 249–410.

Manzini, Paola and Marco Mariotti (2014). "Stochastic Choice and Consideration Sets". In: *Econometrica* 83, pp. 1153–1176.

McFadden, Daniel (1973). "Conditional Logit Analysis of Qualitative Choice Behavior". In: *Frontiers in Econometrics*. Ed. by P. Zarembka. New York: Academic Press, pp. 105–142.

— (1978). "Modeling the Choice of Residential Location". In: *Spatial Interaction Theory and Planning Models*. Ed. by Anders Karlqvist, Jörgen W. Weibull, Lars Lundqvist, and Folke Snickars. Amsterdam: North-Holland, 75–96.

O'Connor, Annete M. (1995). "Validation of a Decisional Conflict Scale". In: *Medical Decision Making* 15, pp. 25–30.

Piccione, Michele and Ran Spiegler (2012). "Price Competition under Limited Comparability". In: *Quarterly Journal of Economics* 127, pp. 1–39.

Pochon, Jean-Baptiste, Jason Riis, Alan G. Sanfey, Leigh E. Nystrom, and Jonathan D. Cohen (2008). "Functional Imaging of Decision Conflict". In: *The Journal of Neuroscience* 28, pp. 3468–3473.

Ratcliff, Roger and Gail McKoon (2008). "The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks". In: *Neural Computation* 20, pp. 873–922.

Redelmeier, D. A. and E. Shafir (1995). "Medical Decision Making in Situations that Offer Multiple Alternatives". In: *Journal of the American Medical Association* 273, pp. 302–305.

Scheibehenne, Benjamin, Rainer Greifeneder, and Peter M. Todd (2010). "Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload". In: *Journal of Consumer Research* 37, pp. 409–425.

Sela, Aner, Jonah Berger, and Wendy Liu (2009). "Variety, Vice, and Virtue: How Assortment Size Influences Option Choice". In: *Journal of Consumer Research* 35, pp. 941–951.

Spektor, Mikhail S., Sebastian Gluth, Laura Fontanesi, and Jörg Rieskamp (2019). "How Similarity Between Choice Options Affects Decisions from Experience: The Accentuation-of-Differences Model". In: *Psychological Review* 126, pp. 52–88.

Spiegler, Ran (2015). "On the Equilibrium Effects of Nudging". In: *Journal of Legal Studies* 44, pp. 389–416.

Stoye, Jörg (2019). "Revealed Stochastic Preference: A One-Paragraph Proof and Generalization". In: *Economics Letters* 177, pp. 66–68.

Strzalecki, Tomasz (2024). *Stochastic Choice Theory*. Econometric Society Monographs. Cambridge: Cambridge University Press.

Train, Kenneth E. (2009). *Discrete Choice Methods with Simulation*. 2nd. Cambridge: Cambridge University Press.

Tversky, Amos and Eldar Shafir (1992). "Choice under Conflict: The Dynamics of Deferred Decision". In: *Psychological Science* 3, pp. 358–361.

White, Chris M., Ulrich Hoffrage, and Nils Reisen (2015). "Choice Deferral Can Arise from Absolute Evaluations or Relative Comparisons". In: *Journal of Experimental Psychology: Applied* 21, pp. 140–157.

Software References

Croissant, Yves (2020). "mlogit: Random Utility Models in R". In: *Journal of Statistical Software* 95.

Nash, John C. and Ravi Varadhan (2011). "Unifying Optimization Algorithms to Aid Software System Users: optimx for R". In: *Journal of Statistical Software* 43.9, pp. 1–14. DOI: 10.18637/jss.v043.i09.

R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC. Boston, MA. URL: http://www.rstudio.com/.

Wickham, Hadley (2011). "The Split-Apply-Combine Strategy for Data Analysis". In: *Journal of Statistical Software* 40.1, pp. 1–29. URL: https://www.jstatsoft.org/v40/i01/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Pedersen, Evan Miller, Stephan Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani (2019). "Welcome to the tidyverse". In: *Journal of Open Source Software* 4, p. 1686.

A Maximum-Likelihood Estimation in the Power Logit

We extend the quadratic-logit analysis of Section 4.1 and clarify the properties and estimation aspects pertaining to the discrete-choice version of the more general power-logit model,where

$$\rho_n(a_i, A) = \frac{e^{p\beta \cdot x_{ni}}}{\left(\sum_{j=1}^{k} e^{\beta \cdot x_{nj}}\right)^p} \tag{64}$$

$$\rho_n(o, A) = \frac{\left(\sum_{j=1}^{k} e^{\beta \cdot x_{nj}}\right)^p - \sum_{j=1}^{k} e^{p\beta \cdot x_{nj}}}{\left(\sum_{j=1}^{k} e^{\beta \cdot x_{nj}}\right)^p} \tag{65}$$

Following McFadden (1973) and the ensuing literature, we show how the vector β and scalar $p > 1$ in (64)-(65) can be estimated by minimizing the log-likelihood function that emerges from this model. To this end, let us write

$$\begin{aligned} Pr_{ni} &\equiv Pr_{ni}(\beta) \\ &:= \rho_n(a_i, A), \\ &\quad i \leq k, \end{aligned} \qquad \text{and} \qquad \begin{aligned} Pr_{no} &\equiv Pr_{no}(\beta) \\ &:= \rho_n(o, A) \\ &= 1 - \textstyle\sum_{i=1}^{k} Pr_{ni} \\ &> 0. \end{aligned}$$

Next, let us denote by y_n the n-th individual's observed decision at menu A. It is critical to distinguish between this decision being an active choice or choice of the outside option. To this end, we define the binary variables y_{ni}, $i = 1, \ldots, k$, and y_{no} by

$$y_{ni} := \begin{cases} 1, & \text{if } y_n = a_i \\ 0, & \text{otherwise} \end{cases} \qquad \text{and} \qquad y_{no} := \begin{cases} 1, & \text{if } y_n \neq a_i \text{ for all } a_i \in A \\ 0, & \text{otherwise} \end{cases}$$

to account for the former and latter cases, respectively. With these in place, the multinomial density for a given active-choice or opt-out decision made by agent n can now be written as

$$g_n(\beta, p) = Pr_{no}^{y_{no}} \prod_{i=1}^{k} Pr_{ni}^{y_{ni}}.$$

Assuming an exogenous sample and covariates x_{ni} for every agent $n \leq N$ and alternative $i \leq k$, the likelihood function that results from the N independent decisions is now given by

$$\begin{aligned} L(\beta, p) &= \prod_{n=1}^{N} g_n(\beta) \\ &= \prod_{n=1}^{N} \left(Pr_{no}^{y_{no}} \prod_{i=1}^{k} Pr_{ni}^{y_{ni}} \right) \end{aligned}$$

This leads to the log-likelihood function

$$\begin{aligned} LL(\beta, p) &= \sum_n \sum_i y_{ni} \ln Pr_{ni} + \sum_n y_{no} \ln Pr_{no} \\ &= \sum_n \sum_i y_{ni} \ln \left(\frac{e^{\beta \cdot x_{ni}}}{\sum_j e^{\beta \cdot x_{nj}}} \right)^p + \sum_n y_{no} \ln \left(\frac{\left(\sum_j e^{\beta \cdot x_{nj}} \right)^p - \sum_j e^{p\beta \cdot x_{nj}}}{\left(\sum_j e^{\beta \cdot x_{nj}} \right)^p} \right) \end{aligned}$$

Recalling that $\beta = (\beta^1, \ldots, \beta^m)$ and $x_{ni} = (x_{ni}^1, \ldots, x_{ni}^m)$, the first-order conditions of its maximization with respect to p and β are

$$
\begin{aligned}
\frac{\partial LL(\beta,p)}{\partial p} &= \sum_n \sum_i y_{ni}\left[\beta\cdot x_{ni} - \ln\left(\sum_j e^{\beta\cdot x_{nj}}\right)\right] \\
&- \sum_n y_{no}\left[\frac{\sum_j (\beta\cdot x_{nj}) e^{\widehat{p}\beta\cdot x_{nj}} - \sum_j e^{\widehat{p}\beta\cdot x_{nj}} \ln\left(\sum_j e^{\beta\cdot x_{nj}}\right)}{\left(\sum_j e^{\beta\cdot x_{nj}}\right)^{\widehat{p}} - \sum_j e^{\widehat{p}\beta\cdot x_{nj}}}\right] \\
&= 0, \\
\frac{\partial LL(\beta,p)}{\partial \beta^l} &= p\sum_n \sum_i y_{ni} x^l_{ni} - p\sum_n \sum_i y_{ni} x^l_{ni}\left(\frac{e^{\widehat{\beta}\cdot x_{ni}}}{\sum_j e^{\widehat{\beta}\cdot x_{nj}}}\right) \\
&- p\sum_n \sum_i y_{no} x^l_{ni}\left(\frac{e^{\widehat{\beta}\cdot x_{ni}}}{\sum_j e^{\widehat{\beta}\cdot x_{nj}}}\right) \\
&+ p\sum_n y_{no}\left[\frac{\left(\sum_j e^{\widehat{\beta}\cdot x_{nj}}\right)^{p-1}\left(\sum_j x^l_{nj} e^{\widehat{\beta}\cdot x_{nj}}\right) - \sum_j x^l_{nj} e^{p\widehat{\beta}\cdot x_{nj}}}{\left(\sum_j e^{\widehat{\beta}\cdot x_{nj}}\right)^{p} - \sum_j e^{p\widehat{\beta}\cdot x_{nj}}}\right] \\
&= 0, \qquad l = 1,\ldots,m
\end{aligned}
$$

Observing that $\sum_i y_{ni} + y_{no} = 1$ holds by construction and that $p > 0$ enters all their terms multiplicatively, the latter m first-order conditions simplify to

$$
\begin{aligned}
\sum_n \sum_i y_{ni} x^l_{ni} &= \sum_n \sum_i x^l_{ni}\left(\frac{e^{\widehat{\beta}\cdot x_{ni}}}{\sum_j e^{\widehat{\beta}\cdot x_{nj}}}\right) \\
&- \sum_n y_{no}\left[\frac{\left(\sum_j e^{\widehat{\beta}\cdot x_{nj}}\right)^{p-1}\left(\sum_j x^l_{nj} e^{\widehat{\beta}\cdot x_{nj}}\right) - \sum_j x^l_{nj} e^{p\widehat{\beta}\cdot x_{nj}}}{\left(\sum_j e^{\widehat{\beta}\cdot x_{nj}}\right)^{p} - \sum_j e^{p\widehat{\beta}\cdot x_{nj}}}\right]
\end{aligned}
$$

Thus, unlike the standard logit where the term appearing with a negative sign in the last equation is absent and where, by construction, the estimated $\widehat{\beta}$ ensures that empirical and average predicted frequencies of active-choice alternatives coincide (Train, 2009; Greene and Hensher, 2010), the presence of the said term here clarifies that this is no longer true in the power logit when deferral choices are present in the data.