

# Discrete Choice under Risk with Limited Consideration

---

Levon Barseghyan  
Francesca Molinari  
Matthew Thirkettle

The Institute for Fiscal Studies  
Department of Economics, UCL

**cemmap** working paper CWP28/20

# Discrete Choice under Risk with Limited Consideration\*

Levon Barseghyan<sup>†</sup>

Francesca Molinari<sup>‡</sup>

Matthew Thirkettle<sup>§</sup>

June 5, 2020

## Abstract

This paper is concerned with learning decision makers' preferences using data on observed choices from a finite set of risky alternatives. We propose a discrete choice model with unobserved heterogeneity in consideration sets and in standard risk aversion. We obtain sufficient conditions for the model's semi-nonparametric point identification, including in cases where consideration depends on preferences and on some of the exogenous variables. Our method yields an estimator that is easy to compute and is applicable in markets with large choice sets. We illustrate its properties using a dataset on property insurance purchases.

**Keywords:** discrete choice, limited consideration, semi-nonparametric identification

---

\*We are grateful to Liran Einav, three anonymous referees, Abi Adams, Jose Apesteguia, Miguel Ballester, Arthur Lewbel, Chuck Manski, and Jack Porter for useful comments and constructive criticism. For comments and suggestions we thank the participants to the 2017 Barcelona GSE Summer Forum on Stochastic Choice, the 2018 Cornell Conference "Identification and Inference in Limited Attention Models", the 2018 Penn State-Cornell Conference on Econometrics and IO, the 2018 GNYMA Conference, the 2019 ASSA meetings, the IFS 2019 "Consumer Behaviour: New Models, New Methods" Conference, and to seminars at Stanford, Berkeley, UCL, Wisconsin, Bocconi, and Duke. Part of this research was carried out while Barseghyan and Molinari were on sabbatical leave at the Department of Economics at Duke University, whose hospitality is gratefully acknowledged. We gratefully acknowledge support from National Science Foundation grant SES-1824448 and from the Institute for Social Sciences at Cornell University.

<sup>†</sup>Department of Economics, Cornell University, lb247@cornell.edu.

<sup>‡</sup>Department of Economics, Cornell University, fm72@cornell.edu.

<sup>§</sup>Department of Economics, Cornell University, mkt68@cornell.edu.

# 1 Introduction

This paper is concerned with learning decision makers’ (DMs) preferences using data on observed choices from a finite set of risky alternatives with monetary outcomes. The prevailing empirical approach to study this problem merges expected utility theory (EUT) models with econometric methods for discrete choice analysis. Standard EUT assumes that the DM evaluates all available alternatives and chooses the one yielding the highest expected utility. The DM’s risk aversion is determined by the concavity of her Bernoulli utility function. The set of all alternatives – the *choice set* – is assumed to be observable by the researcher.

We depart from this standard approach by proposing a discrete choice model with unobserved heterogeneity in preferences and unobserved heterogeneity in consideration sets. Specifically, preferences satisfy the classic Single Crossing Property (SCP) of Mirrlees (1971) and Spence (1974), central to important studies of decision making under risk.<sup>1</sup> That is, the preference order of any two alternatives switches only at one value of the preference parameter.<sup>2</sup> Given her *unobserved* preference parameter, each DM evaluates only the alternatives in her *unobserved* consideration set, which is a subset of the choice set.

Our first contribution is to provide a general framework for point identification of these models. Our analysis relies on two types of observed data variation. In the first case, we assume that the data include a single (common) excluded regressor affecting the utility of each alternative. In the second case, we assume that each alternative has its own excluded regressor. In both cases, the excluded regressor(s) is independent of unobserved preference heterogeneity. When the excluded regressor(s) also has large support it becomes a “special regressor” (Lewbel, 2000, 2014). For reasons we explain, the case of the single common excluded regressor is the most demanding from an identification standpoint. Nonetheless, under classic conditions for semi-nonparametric identification of full-consideration discrete choice models (see, e.g., Lewbel, 2000; Matzkin, 2007), we obtain semi-nonparametric identification of the preference distribution given basically any consideration set formation process (henceforth, consideration process).<sup>3</sup> We prove identification of the consideration process for the widely used Alternative-specific Random Consideration (ARC) model of Manski (1977) and Manzini & Mariotti (2014). The identification argument is constructive and applicable

---

<sup>1</sup>E.g., Athey (2001); Apesteguia, Ballester, & Lu (2017); Chiappori, Salanié, Salanié, & Gandhi (2019)

<sup>2</sup>The EUT framework satisfies the SCP, which requires that if a DM with a certain degree of risk aversion prefers a safer lottery to a riskier one, then all DMs with higher risk aversion also prefer the safer lottery.

<sup>3</sup>The identification results are semi-nonparametric because we specify the utility function up to a DM-specific preference parameter. We establish nonparametric identification of the distribution of the latter.

beyond the ARC model. We establish identification results for preferences that do not require large support of the excluded regressor(s). We also show that identification of both preferences and the consideration process is attainable when consideration depends on preferences. In particular, we introduce (i) binary consideration types, and (ii) proportionally shifting consideration, which captures the notion that the DM’s attention probabilistically shifts from riskier to safer alternatives as her risk aversion increases.

We can significantly expand our results with alternative-specific excluded regressors. First, we can allow for essentially unrestricted dependence of consideration on preferences without assuming that the excluded regressors have large support. Second, we show that consideration can depend both on preferences and on some excluded regressors. We show this for two cases. In the first case, there is one alternative (the default) that is always considered. The probability of considering other alternatives can depend on the default-specific excluded regressor. This is a generalization of the models in Heiss, McFadden, Winter, Wuppermann, & Zhou (2016); Ho, Hogan, & Scott Morton (2017); Abaluck & Adams (2018), where the consideration process only allows for the possibility that either the default or the entire choice set is considered. We, however, allow for each subset of the choice set containing the default to have its own probability of being drawn and this probability can vary with the DM’s preferences. In the second case, we allow the consideration of each alternative to depend on its own excluded regressor, but not on the regressors of other alternatives (Goeree, 2008; Abaluck & Adams, 2018; Kawaguchi, Uetake, & Watanabe, 2020). We also allow for consideration to depend on preferences – a feature unique to our paper.

Our second contribution is to provide a simple method to compute our likelihood-based estimator. Its computational complexity grows polynomially in the number of parameters governing the consideration process (e.g., the choice set size in the ARC model).<sup>4</sup> Our method does not require enumerating all possible subsets of the choice set. If it did, the computational complexity would grow exponentially with the size of the choice set. Moreover, we compute the utility of each alternative only once for a given value of the preference parameter, gaining enormous computational advantage akin to that of importance-sampling methods.

Our third contribution is to elucidate the applicability and the advantages of our framework over the standard application of full consideration random utility models (RUMs) with additively separable unobserved heterogeneity (e.g., Mixed Logit). First, our model can generate

---

<sup>4</sup>The function evaluation time of the log-likelihood objective function grows linearly with the number of parameters. Provided that the objective function is locally concave, the local rate of convergence of the standard SQP program is quadratic. See, for example, Boggs & Tolle (1995).

zero shares for non-dominated alternatives. Second, the model has no difficulty explaining relatively large shares of dominated alternatives. Third, in markets with many choice domains, our model can match not only the marginal but also the joint distribution of choices across domains. Forth, our framework is immune to an important criticism by Apesteguia & Ballester (2018) against using standard RUMs to study decision making under risk. As these authors note, combining standard EUT with additive noise results in non-monotonicity of choice probabilities in the risk preferences, a clearly undesirable feature.

Our methodological innovation exploits restrictions implied by the SCP for the empirical choice frequencies and how these frequencies respond to changes in excluded regressor(s). Consider for example a market with three products,  $\{d_1, d_2, d_3\}$ , where alternative  $d_1$  is the riskiest and  $d_3$  is the safest, priced respectively at  $\{p_1, p_2, p_3\}$ . We seek to learn the distribution of the DM's risk aversion parameter denoted  $\nu$  and the parameters governing the consideration process. Our analysis rests on studying how an incremental change in the price of one alternative affects all choice frequencies. Assuming consideration is independent of preferences and prices, there are only two types of DMs who might respond to a change in  $p_1$ : DMs that are risk type  $\nu^{\{1,2\}}$  who are indifferent between  $d_1$  and  $d_2$ ; and DMs that are type  $\nu^{\{1,3\}}$  who are indifferent between  $d_1$  and  $d_3$ . If a DM of type  $\nu^{\{1,2\}}$  responds to an increase in  $p_1$ , she will abandon  $d_1$  in favor of  $d_2$ . Whether she responds to the change in price depends on her consideration set. For the response to occur, both  $d_1$  and  $d_2$  must be considered. Alternative  $d_3$  should not be considered if, at the current prices, DMs of type  $\nu^{\{1,2\}}$  prefer  $d_3$  to both  $d_1$  and  $d_2$ . In other words, we know not only the exact preferences of DMs that may respond to the price change, but also what their consideration sets must look like for this response to occur. Exploiting this insight yields our identification results.

Random preference models like the ones we consider are random utility models as envisioned by McFadden (1974) (for a textbook treatment see Manski, 2009). We show that our random preference models can be written as RUMs with unobserved heterogeneity in risk aversion and with an additive error that has a discrete distribution with support  $\{-\infty, 0\}$ . Then, it is natural to draw parallels with the Mixed (random coefficient) Logit model (e.g., McFadden & Train, 2000). In our setting, the Mixed Logit boils down to assuming that, given the DM's risk aversion, her evaluation of an alternative equals its expected utility summed with an unobserved heterogeneity term capturing the DM's idiosyncratic taste for unobserved characteristics of that alternative. However, in some markets it is hard to envision such characteristics: For example, many insurance contracts are identical in *all* aspects *except*

for the coverage level and price.<sup>5</sup> In other contexts, unobservable characteristics may affect choice mostly via consideration – as we model – rather than via “additive noise”.<sup>6</sup>

We show that the ARC model and the Mixed Logit generate several contrasting implications. First, the Mixed Logit generally implies that each alternative has a positive probability of being chosen, while the ARC model can generate zero shares by setting the consideration probability of a given alternative to zero. Second, the Mixed Logit satisfies a *Generalized Dominance property* that we derive: if for any degree of risk aversion alternative  $j$  has lower expected utility than either alternative  $k$  or  $l$ , then the probability of choosing  $j$  must be no larger than the probability of choosing  $k$  or  $l$ . The ARC model does not abide Generalized Dominance. Third, in the ARC model choice probabilities depend on the ordinal expected utility rankings of the alternatives, while in the Mixed Logit it depends on the cardinal ranking. As we show, this difference implies that choice probabilities are monotone in risk preferences in the ARC model, while in the Mixed Logit they are not (Apesteguia & Ballester, 2018).

Our empirical application is a study of households’ deductible choices across three lines of insurance: auto collision, auto comprehensive, and home (all perils). We aim to estimate the distribution of risk preferences and the consideration parameters; to assess the resulting fit of the models; and to evaluate the monetary cost of limited consideration. We find that the ARC model does a remarkable job at matching the distribution of observed choices, and because of its aforementioned properties, outperforms the Mixed Logit. Under the ARC model, we find that although households are on average strongly risk averse, they consider lower coverages more often than higher coverages. Finally, the average monetary loss per household resulting from limited consideration is \$49.

The rest of the paper is organized as follows. We describe the model of DMs’ preferences in Section 2, and study identification under a generic consideration process in Section 3. We present the ARC model and its identification in Section 4. In Section 5 we describe the computational advantages of our approach. Section 6 compares our model to the Mixed Logit. Section 7 presents our empirical application. Section 8 contextualizes our contribution relative to the extant literature and offers concluding remarks.

---

<sup>5</sup>E.g., employer provided health insurance, auto, or home insurance offered by a single company.

<sup>6</sup>E.g., a DM may only consider those supplemental prescription drug plans that cover specific medications.

## 2 Preferences

### 2.1 Decision Making under Risk in a Market Setting: An Example

Consider as an example the following insurance market, which mimics the setting of our empirical application. There is an underlying risk of a loss that occurs with probability  $\mu$  that may vary across DMs. A finite number of alternatives are available to insure against this loss. Conditional on risk type, i.e., given  $\mu$ , each alternative  $j \in \mathcal{D} \equiv \{1, \dots, D\}$  is fully characterized by the pair  $(d_j, p_j)$ . The first element is the insurance deductible, which is the DM's out of pocket expense in the case a loss occurs. Deductibles are decreasing with index  $j$ , and all deductibles are less than the lowest realization of the loss. The second element is the price (insurance premium), which also varies across DMs. For each DM there is a baseline price  $\bar{p}$  that determines prices for all alternatives faced by the DM according to the multiplication rule  $p_j = g_j \cdot \bar{p} + \delta$ . Lower deductibles provide more coverage and cost more, so that  $g_j$  is increasing with  $j$ . Both  $g_j$  and  $\delta$  are invariant across DMs. The lotteries that DMs face are  $L_j(x) \equiv (-p_j, 1 - \mu; -p_j - d_j, \mu)$ , where  $x \equiv \bar{p}$ . DMs are expected utility maximizers. Given initial wealth  $w$ , the expected utility of deductible lottery  $L_j(x)$  is

$$U_\nu(L_j(x)) = (1 - \mu) u_\nu(w - p_j) + \mu u_\nu(w - p_j - d_j),$$

where  $u_\nu(\cdot)$  is a Bernoulli utility function defined over final wealth states. We assume that  $u_\nu(\cdot)$  belongs to a family of utility functions that are fully characterized by a scalar  $\nu$  (e.g. Constant Absolute Risk Aversion (CARA), Constant Relative Risk Aversion (CRRA), or Negligible Third Derivative (NTD)), which varies across DMs.<sup>7</sup>

Given the risk type, the relationship between risk aversion and prices is standard. At sufficiently high  $\bar{p}$ , less coverage is always preferred to more coverage for all  $\nu$  on the support:  $U_\nu(L_1(x)) > U_\nu(L_2(x)) > \dots > U_\nu(L_D(x))$ . At sufficiently low  $\bar{p}$ , we have the opposite ordering for all  $\nu$  on the support:  $U_\nu(L_D(x)) > U_\nu(L_{D-1}(x)) > \dots > U_\nu(L_1(x))$ . At moderate prices, for each pair of deductible lotteries  $j < k$  there is a cutoff value  $c_{j,k}(x)$  in the interior of  $\nu$ 's support, found by solving  $U_\nu(L_j(x)) = U_\nu(L_k(x))$  for  $\nu$ . On the left of this cutoff the higher deductible is preferred and on the right the lower deductible is preferred. In other words,  $c_{j,k}(x)$  is the unique coefficient of risk aversion that makes the DM indifferent between  $L_j(x)$  and  $L_k(x)$ , known to the researcher at any given  $x$ . Those with lower  $\nu$

<sup>7</sup>Under CRRA, it is implied that DMs' initial wealth is known to the researcher. NTD utility is defined in Cohen & Einav (2007) and in Barseghyan, Molinari, O'Donoghue, & Teitelbaum (2013).

choose the riskier alternative  $L_j(x)$ , while those with higher  $\nu$  choose the safer alternative  $L_k(x)$ . Provided  $U_\nu(\cdot)$  is smooth in  $\nu$ ,  $c_{j,k}(x)$  is smooth in  $x$ . In fact, under CARA, CRRA, or NTD,  $c_{j,k}(x)$  is a continuously differentiable monotone function. The prices are such that, under CARA, CRRA, or NTD, whenever  $U_\nu(L_1(x)) > U_\nu(L_j(x))$  it is also the case that  $U_\nu(L_1(x)) > U_\nu(L_{j+1}(x))$ .<sup>8</sup> As we show below, this can be stated as  $c_{1,j}(x) < c_{1,j+1}(x)$ . That is, if the DM’s risk aversion is so low that she prefers the riskiest lottery to a safer one, then she also prefers it to an even safer one. Finally, there are no three-way ties. That is, for a given  $x$  there are no alternatives  $\{j, k, l\}$  such that  $U_\nu(L_j(x)) = U_\nu(L_k(x)) = U_\nu(L_l(x))$ .<sup>9</sup>

## 2.2 Preferences with Single Crossing Property

There is a continuum of DMs. Each of them faces a choice among a finite number of alternatives, i.e., a choice set, which is denoted  $\mathcal{D} = \{1, \dots, D\}$ . The number of alternatives is invariant across DMs. Alternatives vary by their utility-relevant characteristics and are distinguished by (at least) one characteristic,  $d_j \in \mathbb{R}$ ,  $j \in \mathcal{D}$ , which is DM invariant. This characteristic reflects the quality of alternative  $j$  (e.g., insurance deductible). When it is unambiguous, we may write  $d_j$  instead of “alternative  $j$ ”. Other characteristics may vary across DMs or across alternatives. Our analysis rests on the excluded regressor(s)  $x$ . To keep the notation as lean as possible, we state our assumptions and results implicitly conditioning on all remaining characteristics. Hence, alternative  $j$  is fully characterized by  $(d_j, x_j)$ . We consider two cases. In one case, all  $x_j$ ’s are perfectly correlated with a single (common) excluded regressor,  $x$ . In the other case, each  $x_j$  has its own variation, conditional on all other  $x_k$ ,  $k \neq j$ , and we let  $x = (x_1, x_2, \dots, x_D)$ .

**Assumption T0.** *The random variable (or vector)  $x$  has a strictly positive density on a set  $\mathcal{S} \subset \mathbb{R}$  ( $\mathcal{S} \subset \mathbb{R}^D$ ,  $\dim \mathcal{S} = D$ ).*

Each DM’s valuation of the alternatives is defined by a utility function  $U_\nu(d_j, x)$ , which depends on a DM-specific index  $\nu$  distributed according to  $F(\cdot)$  over a bounded support.<sup>10</sup>

**Assumption T1.** *The density of  $F(\cdot)$ , denoted  $f(\cdot)$ , is continuous and strictly positive on  $[0, \bar{\nu}]$  and zero everywhere else.*

---

<sup>8</sup>We formally verify this claim for our application in Appendix B.

<sup>9</sup>The identification argument of Barseghyan, Molinari, O’Donoghue, & Teitelbaum (2018) implies that there may exist only one pair of values  $(\mu, \nu)$  such that this three way tie attains.

<sup>10</sup>We assume that while  $\nu$  has bounded support, the utility function is well defined for any real valued  $\nu$ .



The DMs' draws of  $\nu$  are not observed by the researcher. We require that DMs' preferences are strictly monotone and satisfy the Single Crossing Property (SCP).

**Assumption T2** (Single Crossing Property). *For any two alternatives,  $d_j$  and  $d_k$ , there exists a continuously differentiable function  $c_{L,R} : \mathcal{S} \rightarrow \mathbb{R}_{[-\infty, \infty]}$  such that*

$$\begin{aligned} U_\nu(d_L, x) &> U_\nu(d_R, x) & \forall \nu \in (-\infty, c_{L,R}(x)) \\ U_\nu(d_L, x) &= U_\nu(d_R, x) & \nu = c_{L,R}(x) \\ U_\nu(d_L, x) &< U_\nu(d_R, x) & \forall \nu \in (c_{L,R}(x), \infty). \end{aligned}$$

where  $(L, R) = (j, k)$  or  $(L, R) = (k, j)$ . We refer to  $c_{L,R}(\cdot)$  as the cutoff between  $d_L$  and  $d_R$ .

The SCP implies that the DM's ranking of alternatives is monotone in  $\nu$ . In the context of risk preferences, if a DM with a certain level of risk aversion prefers a safer asset to a riskier one, then all DMs with higher risk aversion also prefer the safer asset. Since the cutoffs may be infinite, the SCP does not exclude dominated alternatives.<sup>11</sup>

**Definition 1** (Dominated Alternatives). *Given  $x$ , alternative  $d_j$  is dominated if there exists an alternative  $d_k$  such that  $\forall \nu \in \mathbb{R}$ ,  $U_\nu(d_k, x) > U_\nu(d_j, x)$ .*

We now establish some useful facts that follow from Assumption T2. First, the index  $L$  in  $c_{L,R}(\cdot)$  indicates the alternative that is preferred on the left of the cutoff. It is without loss of generality to assume  $L = \min(j, k)$  and  $R = \max(j, k)$  because of the following fact:

**Fact 1** (Natural Ordering of Alternatives). *Suppose Assumption T2 holds. Then alternatives can be enumerated such that as  $\nu \rightarrow -\infty$ ,  $U_\nu(d_1, x) > U_\nu(d_2, x) > \dots > U_\nu(d_D, x)$  for all  $x$  at which no alternative is dominated.*

We assume that alternatives are enumerated according to the Natural Ordering of Alternatives.<sup>12</sup> As the next fact shows, for high values of  $\nu$  the preference over the Natural Ordering of Alternatives is reversed.

**Fact 2** (Rank Switch). *Suppose Assumption T2 holds. Consider any  $x$  such that no alternative is dominated. As  $\nu \rightarrow \infty$ ,  $U_\nu(d_1, x) < U_\nu(d_2, x) < \dots < U_\nu(d_D, x)$ .*

<sup>11</sup>For choice under risk, this definition of dominance is equivalent to first order stochastic dominance.

<sup>12</sup>Under this enumeration,  $d_j$  will be ordered in either ascending or descending order. In our example from the previous section, since  $d_j$  refers to the deductible and  $\nu$  is the risk aversion coefficient, the natural ordering implies  $d_1 > d_2 > \dots > d_D$ .

The SCP also has implications for the relative position of the cutoffs. For readability, we state them for alternatives  $\{d_1, d_2, d_3\}$ , but they hold for any  $\{d_j, d_k, d_l\}$ ,  $j < k < l$ .

**Fact 3** (Simple Relative Order of Cutoffs). *Suppose Assumption T2 holds. Given  $x$ , if  $c_{1,2}(x) < c_{1,3}(x)$ , then  $c_{1,3}(x) < c_{2,3}(x)$  or both  $d_1$  and  $d_2$  dominate  $d_3$  ( $c_{1,3}(x) = c_{2,3}(x) = \infty$ ).*

The next fact concerns the relative order of cutoffs for non-dominated alternatives. Before stating it, it is convenient to define Never-the-First-Best Alternatives.

**Definition 2** (Never-the-First-Best). *Given  $x$ , alternative  $d_j$  is Never-the-First-Best in  $\mathcal{D}$  if for every  $\nu$  there exists another alternative  $d_k(\nu)$  in  $\mathcal{D}$  such that  $U_\nu(d_k(\nu), x) > U_\nu(d_j, x)$ .*

**Fact 4** (Cutoff Relative Order). *Suppose that Assumption T2 holds. If, given  $x$ , alternatives  $d_1$ ,  $d_2$ , and  $d_3$  are not dominated, then one and only one of the following cases holds:*

- (i)  $c_{1,2}(x) < c_{1,3}(x) < c_{2,3}(x)$  and  $d_2$  is the first best in  $\{d_1, d_2, d_3\}$ ,  $\forall \nu \in (c_{1,2}(x), c_{1,3}(x))$ ;
- (ii)  $c_{1,2}(x) > c_{1,3}(x) > c_{2,3}(x)$  and  $d_2$  is Never-the-First-Best in  $\{d_1, d_2, d_3\}$ ;
- (iii)  $c_{1,2}(x) = c_{1,3}(x) = c_{2,3}(x)$  and  $d_2$  is strictly worse than either  $d_1$  or  $d_3$  for all  $\nu$  except for  $\nu = c_{1,2}(x)$  where there is a three-way tie among these alternatives.

In Case (ii),  $U_\nu(d_1, x) > U_\nu(d_2, x)$  implies  $U_\nu(d_1, x) > U_\nu(d_3, x)$ , so  $d_2$  and  $d_3$  switch rank at a point where they are preferred to  $d_1$ . In Case (iii),  $U_\nu(d_1, x) > U_\nu(d_3, x)$  implies  $U_\nu(d_1, x) > U_\nu(d_2, x)$ , so  $d_2$  and  $d_3$  switch rank at a point where  $d_1$  is preferred to both of them.

### 3 Identification

The classic identification argument for discrete choice under full consideration rests on the following four pillars.

**Assumption I0.** *The random variable (or vector)  $x$  is **independent** of preferences.*

**Assumption I1.**  $\exists \mathcal{X} \subset \mathcal{S}$  s.t.  $c_{1,2}(x)$  **covers the support** of  $\nu$ :  $[0, \bar{\nu}] \subset \{c_{1,2}(x), x \in \mathcal{X}\}$ .

**Assumption I2.** *Consideration is **independent** of preferences.*

**Assumption I3.** *Consideration is **independent** of  $x$ .*

The last two conditions are vacuous in the standard full consideration model, while the first two are typically stated as data requirements. In this section we discuss how identification

works in the limited consideration case and the role of Assumptions I0-I3. We state our formal results for a generic consideration process.

### 3.1 Identification With Two Alternatives

Let the choice set be binary and suppose that the DM considers both alternatives. In addition, let  $x$  be a scalar so that there is a single excluded regressor. Under Assumptions I0-I3, any realization of  $x$  is associated with a single conditional moment in the data:

$$\Pr(d = d_1|x) = \int_0^{c_{1,2}(x)} dF = F(c_{1,2}(x)),$$

because the DM chooses  $d_1$  if and only if her preference parameter is less than  $c_{1,2}(x)$ . The distribution  $F(\cdot)$  is non-parametrically identified, since for any  $\nu$  on the support there is an  $x$  such that  $\nu = c_{1,2}(x)$ .

We emphasize two points. First, given a family of utility functions, for any  $x$  the value of the cutoff can be solved for. Hence, the function  $c_{1,2}(x)$  (and its derivatives) can be treated as data. Second, Assumption I1 requires that the cutoff reaches both ends of the support: there exist  $x^0$  and  $x^1$  such that  $F(c_{1,2}(x^0)) = 0$  and  $F(c_{1,2}(x^1)) = 1$ .

Turning to limited consideration, suppose that  $d_2$  is always considered, while  $d_1$  is considered with a constant probability  $\varphi_1 \leq 1$ . Then,  $d_1$  is chosen when it is considered and it is preferred to  $d_2$ , yielding:

$$\Pr(d = d_1|x) = \varphi_1 F(c_{1,2}(x)) \quad \text{and} \quad \frac{d\Pr(d = d_1|x)}{dx} = \varphi_1 f(c_{1,2}(x)) \frac{dc_{1,2}(x)}{dx}. \quad (1)$$

At first glance, it appears that the distribution of preferences is identified up to a constant. Yet, at the boundary of the support  $\Pr(d = d_1|x^1) = \varphi_1 F(c_{1,2}(x^1)) = \varphi_1$ , so that  $\varphi_1$  is identified.<sup>13</sup> Once  $\varphi_1$  is known, the distribution  $F(\cdot)$  is identified by varying  $c_{1,2}(x)$  over the support of  $\nu$ , just as it is in the full consideration case. We now explore what happens to identification if Assumptions I0–I3 are not satisfied.

**Assumption I0 fails:** the variation in  $x$  is not independent of preferences. Then  $F(\cdot)$  is not non-parametrically identified under either full or limited consideration.

---

<sup>13</sup>The same logic applies to the case with infinite support, but one needs to rely on limit arguments.

**Assumption I1 fails:** the variation in  $x$  is such that  $c_{1,2}(x)$  only covers an interval  $[\nu^l, \nu^u] \subset [0, \bar{\nu}]$ . Then the data provide no information about preferences outside of the interval  $[\nu^l, \nu^u]$ . Inside the interval, the conditional distribution  $F(\nu|\nu \in [\nu^l, \nu^u]) = \frac{F(\nu) - F(\nu^l)}{F(\nu^u) - F(\nu^l)}$  is identified under both limited and full consideration. The consideration probability (and hence the scale of  $F(\cdot)$ ) is partially identified and satisfies the bounds  $\Pr(d = d_1|x^u) \leq \varphi_1 \leq 1$ , where  $x^u$  is such that  $c_{1,2}(x^u) = \nu^u$ . Point identification can be attained if an additional assumption is maintained to pin down the scale of  $F(\cdot)$ . For example, one can simply assume full consideration and set  $\varphi_1 = 1$ .<sup>14</sup>

**Assumption I2 fails:** when  $\varphi_1$  depends on preferences and this dependence is arbitrary, then identification breaks down completely as there is one data moment to identify two unknown objects. However, since we assume – as it is common in the econometrics literature – that the density function of  $\nu$  is continuous and strictly positive, identification is possible for some types of dependence between consideration and preferences. Suppose  $\varphi_1$  is a piece-wise constant function of  $\nu$ , i.e., there are two consideration types:

$$\varphi_1(\nu) = \begin{cases} \underline{\varphi}_1, & \forall \nu \in [0, \nu^*) \\ \bar{\varphi}_1, & \forall \nu \in [\nu^*, \bar{\nu}] \end{cases},$$

where  $\nu^*$  is an unobserved breakpoint. We show that  $\underline{\varphi}_1$ ,  $\bar{\varphi}_1$ , and  $\nu^*$  are identified. First, the product  $\varphi_1(\nu)f(\nu)$  is identified under Assumptions I0, I1, and I3, since

$$\frac{d \Pr(d = d_1|x)}{dx} = \frac{d}{dx} \left( \int_0^{c_{1,2}(x)} \varphi_1(\nu) dF \right) = \varphi_1(\nu) f(\nu) \frac{dc_{1,2}(x)}{dx} \quad (2)$$

at  $\nu = c_{1,2}(x)$ . The product  $\varphi_1(\nu)f(\nu)$  is discontinuous only at the point  $\nu^*$ . Thus, the breakpoint is identified by continuously varying  $c_{1,2}(x)$  across  $[0, \bar{\nu}]$ . Next, the ratio  $\frac{\underline{\varphi}_1}{\bar{\varphi}_1}$  is identified by the ratio of the right and left derivatives of  $\Pr(d = d_1|x)$  at the breakpoint  $x^*$  ( $\nu^* = c_{1,2}(x^*)$ ).<sup>15</sup> The quantity  $F(\nu^*)$  is identified by the ratio:

$$\frac{\Pr(d = d_1|x^*)}{\Pr(d = d_1|x^1) - \Pr(d = d_1|x^*)} = \frac{\underline{\varphi}_1}{\bar{\varphi}_1} \cdot \frac{F(\nu^*)}{1 - F(\nu^*)}.$$

---

<sup>14</sup>Alternatively, one can make an equal tail assumption such as  $F(\nu^l) = 1 - F(\nu^u)$ . This equal tail assumption implies the two expressions for  $\Pr(d = d_1|x^l)$  and  $\Pr(d = d_1|x^u)$  define two equations in the two unknowns  $\varphi_1$  and  $F(\nu^u)$ . It is easy to show that these equations have a unique solution.

<sup>15</sup>Indeed, by Assumption I0, we have that  $\frac{\lim_{x \nearrow x^*} \frac{d \Pr(d=d_1|x)}{dx}}{\lim_{x \searrow x^*} \frac{d \Pr(d=d_1|x)}{dx}} = \frac{\underline{\varphi}_1 f(\nu^*) \frac{dc_{1,2}(x^*)}{dx}}{\bar{\varphi}_1 f(\nu^*) \frac{dc_{1,2}(x^*)}{dx}} = \frac{\underline{\varphi}_1}{\bar{\varphi}_1}$ .

Hence,  $\underline{\varphi}_1$  and  $\bar{\varphi}_1$  are identified. Identification of  $F(\cdot)$  on the entire support follows from Assumption I1. The same argument above applies if the probability of considering an alternative discretely jumps in  $x$  (i.e., **Assumption I3 fails**). Concretely, suppose there is a breakpoint in  $\varphi_1(x)$  at  $x^*$  and let  $\nu^* = c_{1,2}(x^*)$ . The breakpoint  $x^*$  is identified by the point of discontinuity in Equation (2), and the rest follows.

To summarize the case of the binary choice set, the only seemingly real difference in identification is that without large support the scale of the preference distribution  $F(\cdot)$  is partially identified under limited consideration, while it is assumed to be known under full consideration. Additional data moments (i.e., a larger choice set) are needed to progress further.

### 3.2 Single Common Excluded Regressor

This section offers a formal analysis for the case of a single common excluded regressor. We provide a test of limited consideration that does not require large support and identification results for the preference distribution, both with and without Assumption I2.

To start, suppose each alternative in the choice set  $\{d_1, d_2, d_3\}$  is considered independently of each other, with constant probability  $\{\varphi_1, \varphi_2, \varphi_3\}$ , respectively. At least one alternative is always considered, resulting in a non-empty consideration set with probability equal to one. We assume that Assumptions I0-I3 hold. Suppose  $U_\nu(d_1, x) > U_\nu(d_2, x)$  implies  $U_\nu(d_1, x) > U_\nu(d_3, x)$  for all  $x$ . That is, by Fact 3,  $c_{1,2}(x) < c_{1,3}(x) < c_{2,3}(x)$ . Then

$$\begin{aligned} \Pr(d = d_1|x) &= \varphi_1 \left( \int_0^{c_{1,2}(x)} dF + \int_{c_{1,2}(x)}^{c_{1,3}(x)} (1 - \varphi_2)dF + \int_{c_{1,3}(x)}^{\bar{\nu}} (1 - \varphi_2)(1 - \varphi_3)dF \right) \quad (3) \\ &= \varphi_1 \left( \varphi_2 F(c_{1,2}(x)) + (1 - \varphi_2)\varphi_3 F(c_{1,3}(x)) + (1 - \varphi_2)(1 - \varphi_3) \right) \end{aligned}$$

and, similarly,

$$\Pr(d = d_3|x) = \varphi_3 \left( \varphi_1(1 - \varphi_2)(1 - F(c_{1,3}(x))) + \varphi_2(1 - F(c_{2,3}(x))) + (1 - \varphi_1)(1 - \varphi_2) \right).$$

The leading factor  $\varphi_1$  appears in Equation (3), since it is necessary for alternative  $d_1$  to be considered for it to be chosen. Given  $d_1$  is considered, it will be chosen either when: (i)  $d_1$  is the first best; (ii)  $d_1$  is the second best and  $d_2$  is not considered; or (iii)  $d_1$  is the least preferred alternative and both  $d_2$  and  $d_3$  are not considered.

Since there are two cutoffs,  $c_{1,2}(x)$  and  $c_{1,3}(x)$ , that enter the moment  $\Pr(d = d_1|x)$ , there is

not, in general, a one-to-one mapping between the moment and the preference distribution at one point on the support, as it was the case in Section 3.1.<sup>16</sup> That is, as  $x$  changes, the observed choice frequency of  $d_1$  may change because of two types of marginal DMs: those indifferent between  $d_1$  and  $d_2$ , and those indifferent between  $d_1$  and  $d_3$ . This is apparent in the following derivative:

$$\frac{d\Pr(d = d_1|x)}{dx} = \varphi_1 \left( \varphi_2 f(c_{1,2}(x)) \frac{dc_{1,2}(x)}{dx} + (1 - \varphi_2) \varphi_3 f(c_{1,3}(x)) \frac{dc_{1,3}(x)}{dx} \right). \quad (4)$$

### 3.2.1 Alternative $d_2$ is considered whenever $d_1$ is considered

To restore the one-to-one mapping we need restrictions on the consideration process. In our example, it is immediate to see that when  $\varphi_2 = 1$ , the second term on the RHS of Equation (4) disappears and we are back to Equation (1).<sup>17</sup> We obtain a similar result for a *generic* consideration process without relying on large support:

**Theorem 1.** *Suppose Assumptions I0, I2, I3, T0-T2 hold, and*

1. *The consideration process is such that  $d_1$  is considered with positive probability and whenever it is considered so is  $d_2$ ;*
2. *There exists  $\mathcal{X} \subset \mathcal{S}$  such that  $c_{1,2}(x)$ ,  $x \in \mathcal{X}$ , covers  $[\nu^l, \nu^u] \subset [0, \bar{\nu}]$  and  $\forall x \in \mathcal{X}$*

$$U_\nu(d_1, x) > U_\nu(d_2, x) \Rightarrow U_\nu(d_1, x) > U_\nu(d_j, x), \quad \forall j > 2.$$

*Then  $F(\nu | \nu \in [\nu^l, \nu^u])$  is identified.*

The theorem above uses the derivative of  $\Pr(d = d_1|x)$  to create the one-to-one mapping from data to the preference density function. In a more general case, the same can be achieved using the derivative of  $\Pr(d \in \{d_1, d_2, \dots, d_j\}|x)$ :

**Theorem 2.** *Suppose Assumptions I0, I2, I3, T0-T2 hold, and*

1. *The consideration process satisfies the following conditions:  $\exists j$  s.t. whenever alternatives  $d_k$  and  $d_l$ ,  $k \leq j < l$ , are both considered, so are  $d_j$  and  $d_{j+1}$ . In addition,  $d_j$  and*

---

<sup>16</sup>The corresponding equation for  $\Pr(d = d_3|x)$  does not help matters, as it brings about  $F(\cdot)$  evaluated at yet another cutoff,  $c_{2,3}(x)$ .

<sup>17</sup>It is also the case that if the utility of only the second alternative responded to  $x$ , then the second term in Equation (4) would disappear and we would be back to a situation similar to that with two alternatives discussed above.

$d_{j+1}$  have a positive probability of being considered together.

2. There exists  $\mathcal{X} \subset \mathcal{S}$  such that  $c_{j,j+1}(x)$ ,  $x \in \mathcal{X}$ , covers  $[\nu^l, \nu^u] \subset [0, \bar{\nu}]$  and  $\forall x \in \mathcal{X}$

$$\begin{aligned} U_\nu(d_j, x) > U_\nu(d_{j+1}, x) &\Rightarrow U_\nu(d_j, x) > U_\nu(d_k, x), & \forall k > j + 1, \\ U_\nu(d_{j+1}, x) > U_\nu(d_j, x) &\Rightarrow U_\nu(d_{j+1}, x) > U_\nu(d_k, x), & \forall k < j. \end{aligned}$$

Then  $F(\nu | \nu \in [\nu^l, \nu^u])$  is identified.

Under Condition 1 the choice set is split into “low quality” and “high quality” sets. Any subset of the low quality set can be considered. Any subset of the high quality set can also be considered. However, if a consideration set contains both high and low quality elements, then it must contain the “bridging” alternatives  $\{d_j, d_{j+1}\}$ .<sup>18</sup> Condition 2 requires that whenever a DM prefers  $d_j$  to  $d_{j+1}$ , she also prefers  $d_j$  to all high quality alternatives; and whenever a DM prefers  $d_{j+1}$  to  $d_j$ , she also prefers  $d_{j+1}$  to all low quality alternatives.

**Testing for limited consideration.** In our example with  $\{d_1, d_2, d_3\}$  and  $\varphi_2 = 1$ , we obtain two moments:

$$\Pr(d = d_1|x) = \varphi_1 F(c_{1,2}(x)) \quad \text{and} \quad \Pr(d = d_3|x) = \varphi_3 (1 - F(c_{2,3}(x))).$$

Both  $\Pr(d = d_1|x)$  and  $\Pr(d = d_3|x)$  identify the preference distribution up to scale. Suppose there exists a pair of realizations of the excluded regressor, denoted  $x$  and  $x'$ , such that  $c_{1,2}(x) = c_{2,3}(x') = \nu^*$  for some  $\nu^* \in (0, \bar{\nu})$ . Then,

$$\Pr(d = d_1|x) + \Pr(d = d_3|x') \leq \frac{\Pr(d = d_1|x)}{\varphi_1} + \frac{\Pr(d = d_3|x')}{\varphi_3} = F(\nu^*) + 1 - F(\nu^*) = 1,$$

with equality if both  $\varphi_1$  and  $\varphi_3$  equal one. Hence, there is limited consideration if  $\Pr(d = d_1|x) \neq 1 - \Pr(d = d_3|x')$ , i.e.,  $\Pr(d = d_1|x) \neq \Pr(d \in \{d_1, d_2\}|x')$ . The following theorem generalizes this result for a *generic* consideration process that can depend on preferences and the excluded regressor without relying on large support.

---

<sup>18</sup>In particular, a mixture of the following processes satisfies Condition 1: *I. Ascending* consideration:  $d_k$  is considered only if  $d_{k-1}$  is considered; *II. Descending* consideration:  $d_k$  is considered only if  $d_{k+1}$  is considered; *III. Pyramid* consideration:  $d_j$  is always considered,  $d_k$ ,  $k > j$ , is considered only if  $d_{k-1}$  is considered;  $d_k$ ,  $k < j$ , is considered only if  $d_{k+1}$  is considered; *IV.* All singletons. The only restriction is that the mixture cannot be entirely composed of *IV.* These will cover e.g. threshold models (Kimya, 2018), (partial) elimination-by-aspects (Tversky, 1972), extremeness aversion (Simonson & Tversky, 1992), edge aversion (Teigen, 1983; Christenfeld, 1995; Rubinstein, Tversky, & Heller, 1997; Attali & Bar-Hillel, 2003), and edge advantage (Nisbett & Wilson, 1977; Schelling, 1980; Dayan & Bar-Hillel, 2011)

**Theorem 3.** *Suppose Assumptions I0, T0-T2 hold. Suppose there exist  $x, x' \in \mathcal{S}$ , and sets  $\mathcal{L}, \mathcal{L}' \subset \mathcal{D}$  s.t. for some  $\nu^* \in [0, \bar{\nu}]$*

1.  *$\arg \max_{j \in \mathcal{D}} U_\nu(d_j, x) \in \mathcal{L}, \forall \nu \in [0, \nu^*)$ , and  $\arg \max_{j \in \mathcal{D}} U_\nu(d_j, x) \in \mathcal{D} \setminus \mathcal{L}, \forall \nu \in (\nu^*, \bar{\nu}]$*
2.  *$\arg \max_{j \in \mathcal{D}} U_\nu(d_j, x') \in \mathcal{L}', \forall \nu \in [0, \nu^*)$ , and  $\arg \max_{j \in \mathcal{D}} U_\nu(d_j, x') \in \mathcal{D} \setminus \mathcal{L}', \forall \nu \in (\nu^*, \bar{\nu}]$*

*If  $\Pr(d \in \mathcal{L}|x) \neq \Pr(d \in \mathcal{L}'|x')$ , then there is limited consideration.*

Condition 1 of the theorem requires that, given  $x$ , the first-best alternative belongs to  $\mathcal{L}$  for all DMs with  $\nu < \nu^*$  and to  $\mathcal{D} \setminus \mathcal{L}$  for all DMs with  $\nu > \nu^*$ . Condition 2 is the identical requirement, but given  $x'$  and stated for  $\mathcal{L}'$ . Under these conditions and full consideration, the probability of choosing an alternative in  $\mathcal{L}$  or, respectively,  $\mathcal{L}'$  should be  $F(\nu^*)$  in both cases. Thus, if  $\Pr(d \in \mathcal{L}|x) \neq \Pr(d \in \mathcal{L}'|x')$ , then there is a consideration process pushing DMs' choices away from  $\mathcal{L}$  and  $\mathcal{L}'$  at different rates.

Returning to our example, in addition to testing for limited consideration, the information provided by  $\Pr(d = d_3|x)$  allows us to identify the consideration parameters without requiring large support for the excluded regressor, i.e., we can relax Assumption I1. The quantity  $\frac{\varphi_1}{\varphi_3}$  is identified by the ratio of the derivatives at  $c_{1,2}(x) = c_{1,3}(x') = \nu^*$ :

$$\frac{d\Pr(d = d_1|x)}{dx} = \varphi_1 f(\nu^*) \frac{dc_{1,2}(x)}{dx} \quad \text{and} \quad \frac{d\Pr(d = d_3|x')}{dx} = -\varphi_3 f(\nu^*) \frac{dc_{2,3}(x')}{dx}.$$

Given that the ratio of consideration probabilities is identified, the ratio  $\frac{\Pr(d=d_1|x)}{\Pr(d=d_3|x')}$  identifies  $F(\nu^*)$  and consequently  $\varphi_1$  and  $\varphi_3$  are identified from these moments.

We can also leverage the additional moment  $\Pr(d = d_3|x)$  to identify some forms of dependence between consideration and preferences, i.e., to relax Assumption I2. When the consideration probabilities depend on preferences,

$$\Pr(d = d_1|x) = \int_0^{c_{1,2}(x)} \varphi_1(\nu) dF \quad \text{and} \quad \Pr(d = d_3|x) = \int_{c_{2,3}(x)}^{\bar{\nu}} \varphi_3(\nu) dF.$$

The ratio of the derivatives of these two moments yields  $\frac{\varphi_1(\nu)}{\varphi_3(\nu)}$ . More assumptions are required to obtain point identification of the  $\varphi_j(\nu)$ 's.<sup>19</sup> One possibility is the *proportionally shifting*

---

<sup>19</sup> For example, one can assume that the consideration functions are properly normalized mixtures of known base functions or that the consideration functions are piece-wise constants, as on page 10.



consideration process:

$$\begin{aligned}\varphi_1(\nu) &= \varphi_1(1 - \alpha(\nu)) \\ \varphi_3(\nu) &= \varphi_3(1 + \alpha(\nu)),\end{aligned}$$

where  $\varphi_1$ ,  $\varphi_3$ , and  $\alpha(\nu)$  are unknown, and  $\alpha(\nu)$  is normalized:  $\alpha(\bar{\nu}) = 0$  (or  $\alpha(0) = 0$ ). In the context of risk preferences, this amounts to assuming that the DM's consideration shifts away from riskier to safer alternatives as her level of risk aversion increases. Under large support, identification for this case is achieved as follows. First  $\frac{\varphi_1}{\varphi_3}$  is identified when  $x$  and  $x'$  are chosen such that  $c_{1,2}(x) = c_{2,3}(x') = 0$  (or  $= \bar{\nu}$ ). Once  $\frac{\varphi_1}{\varphi_3}$  is identified,  $\frac{1-\alpha(\nu)}{1+\alpha(\nu)}$  is known for all  $\nu$ ; hence,  $\alpha(\nu)$  can be solved for.

### 3.2.2 Alternative $d_2$ is not always considered whenever $d_1$ is considered

A more general case might arise when  $d_2$  is not always considered when  $d_1$  is. In this case, identification rests on large support and proceeds in four steps. Returning to our example, first we rewrite Equation (4) as

$$\frac{d\Pr(d = d_1|x)}{dx} = \hat{f}(c_{1,2}(x))\frac{dc_{1,2}(x)}{dx} + \phi\hat{f}(c_{1,3}(x))\frac{dc_{1,3}(x)}{dx}, \quad (5)$$

where  $\phi \equiv \frac{\varphi_1(1-\varphi_2)\varphi_3}{\varphi_1\varphi_2}$  and  $\hat{f}(\nu) \equiv \varphi_1\varphi_2f(\nu)$ , and a similar expression holds for  $\frac{d\Pr(d=d_3|x)}{dx}$ .

Second, under the large support assumption we can find  $x$  and  $x'$  such that  $c_{1,2}(x) < \bar{\nu} < c_{1,3}(x)$  and  $c_{1,2}(x) = c_{1,3}(x') < \bar{\nu} < c_{2,3}(x')$ . For any such pair,  $f(c_{1,3}(x)) = f(c_{2,3}(x')) = 0$ , and hence

$$\frac{d\Pr(d = d_1|x)}{dx} = \hat{f}(c_{1,2}(x))\frac{dc_{1,2}(x)}{dx} \quad \text{and} \quad \frac{d\Pr(d = d_3|x')}{dx} = -\phi\hat{f}(c_{1,3}(x'))\frac{dc_{1,3}(x')}{dx}.$$

The first equation identifies  $\hat{f}(\nu)$  for preference parameters near the far end of the support, while the ratio of the two equations identifies  $\phi$ .

Third, whenever  $\hat{f}(c_{1,3}(x))$  is known,  $\hat{f}(c_{1,2}(x))$  is uniquely pinned down by Equation (5). Because  $c_{1,2}(x) < c_{1,3}(x)$ ,  $\forall x$ , we can learn  $\hat{f}(\cdot)$  sequentially:

1. Take an  $x^1$  such that  $\hat{f}(c_{1,3}(x^1))$  is already known, learn  $\hat{f}(c_{1,2}(x^1))$ ;
2. Take  $x^2$  such that  $c_{1,3}(x^2) = c_{1,2}(x^1)$ , learn  $\hat{f}(c_{1,2}(x^2))$ ;

3. Let  $x^1 = x^2$ . Repeat Step 2 until the entire support has been covered, i.e.,  $c_{1,2}(x^2) \leq 0$ .

For this approach to work,  $c_{1,3}(x)$  cannot “catch up” to  $c_{1,2}(x)$  (i.e., as assumed,  $c_{1,2}(x) < c_{1,3}(x)$  whenever  $c_{1,2}(x)$  is on the support). This requires that DMs with preference coefficients on the support are not indifferent between more than two alternatives.<sup>20</sup> Last, integrating  $\hat{f}(\nu)$  over the entire support recovers the scale and the true density. Indeed,

$$\int_0^{\bar{\nu}} \hat{f}(\nu) d\nu = \varphi_1 \varphi_2 \int_0^{\bar{\nu}} f(\nu) d\nu = \varphi_1 \varphi_2$$

pins down  $\varphi_1 \varphi_2$ , and hence  $f(\cdot)$  is identified. We next generalize this strategy.

**Definition 3.** Let  $\mathcal{Q}(\mathcal{K})$  denote the probability that the consideration set takes realization  $\mathcal{K} \subset \mathcal{D}$ . Let  $\mathcal{O}(\mathcal{A}; \mathcal{B})$  be the probability that every alternative in set  $\mathcal{A}$  is in the consideration set, and every alternative in set  $\mathcal{B}$  is not:  $\mathcal{O}(\mathcal{A}; \mathcal{B}) \equiv \sum_{\mathcal{K}: \mathcal{A} \subset \mathcal{K}, \mathcal{B} \cap \mathcal{K} = \emptyset} \mathcal{Q}(\mathcal{K})$ .

**Theorem 4.** Suppose Assumptions I0, I2, I3, T0-T2 hold, and

1. The consideration process is such that there is positive probability  $d_1$  and  $d_2$  are considered together;
2. Assumption I1 holds for  $\mathcal{X} \subset \mathcal{S}$  s.t.  $\forall x \in \mathcal{X}$

$$U_\nu(d_1, x) > U_\nu(d_j, x) \Rightarrow U_\nu(d_1, x) > U_\nu(d_{j+1}, x), \quad \forall j > 1.$$

Then  $f(\cdot)$  is identified and so are  $\mathcal{O}(d_1; \emptyset)$  and  $\mathcal{O}(\{d_1, d_2\}; \emptyset)$ . For  $j > 2$ , if  $\Pr(d = d_j | x) > 0$  for some  $x$ , then  $\mathcal{O}(\{d_1, d_j\}; \{d_2, \dots, d_{j-1}\})$  is identified.

The first assumption of the theorem ensures that Equation (5) is informative. The second assumption implies that the cutoffs for alternative  $d_1$  are ordered:  $c_{1,j}(x) < c_{1,j+1}(x)$ . While Theorem 4 requires large support for the excluded regressor, it does not generally require it to exhibit variation that forces alternative  $d_1$  to go from being the first best to being the least preferred. Rather, the theorem requires that at one extreme of the support alternative  $d_1$  dominates all others. However, at the other extreme we only require that  $d_2$  is preferred to  $d_1$  for all DMs. On the one hand, Theorem 4 imposes the large support requirement on the excluded regressor, while Theorem 1 does not. On the other hand, Theorem 4 imposes less restrictions on the consideration process than does Theorem 1. In fact, identification is

---

<sup>20</sup>This situation arises because all cutoffs move together with the single common excluded regressor, e.g., price in our example from Section 2.1 and empirical application in Section 7.

attained for any consideration process that allows  $d_1$  and  $d_2$  to be considered together with positive probability. Moreover, if the probability of being considered together is zero for  $d_1$  and  $d_2$ , but positive for  $d_1$  and  $d_3$ , the theorem still holds as long as the assumptions hold for  $d_3$  instead of  $d_2$ . Theorem 4 identifies some features of the consideration process. Sometimes these features are sufficient for identifying the entire process. In particular, Theorem 4 yields identification of the ARC model, including the consideration process, as we formally show in Section 4.<sup>21</sup>

**Dependence between consideration and preferences.** We next generalize the example in Section 3.1 by allowing for high/low consideration types. We establish identification of the distribution of preferences for a generic consideration process.

**Definition 4.** Let  $\mathcal{Q}_\nu(\mathcal{K})$  denote the probability that a DM with preference parameter  $\nu$  draws consideration set  $\mathcal{K}$ , and  $\mathcal{O}_\nu(\mathcal{A}; \mathcal{B}) \equiv \sum_{\mathcal{K}: \mathcal{A} \subset \mathcal{K}, \mathcal{B} \cap \mathcal{K} = \emptyset} \mathcal{Q}_\nu(\mathcal{K})$ .

**Assumption I2.BCT** (Binary Consideration Types). For some unknown  $\nu^* \in (0, \bar{\nu})$ :

$$\mathcal{Q}_\nu(\mathcal{K}) = \begin{cases} \underline{\mathcal{Q}}(\mathcal{K}) & \text{if } \nu < \nu^* \\ \overline{\mathcal{Q}}(\mathcal{K}) & \text{if } \nu > \nu^* \end{cases}$$

where, for all  $\nu$ ,  $\sum_{\mathcal{K} \subset \mathcal{D}} \mathcal{Q}_\nu(\mathcal{K}) = 1$  and  $\mathcal{Q}_\nu(\mathcal{K}) \geq 0$ ,  $\forall \mathcal{K} \subset \mathcal{D}$ .

**Theorem 5.** Suppose Assumptions I0, I2.BCT, I3, T0-T2, and Condition 2 of Theorem 4 hold. Suppose Condition 1 of Theorem 4 holds for all  $\nu$ . Then  $f(\cdot)$  is identified and so is  $\mathcal{O}_\nu(\{d_1, d_2\}; \emptyset)$ . If  $\frac{d\Pr(d=d_1|x)}{dx}$  is discontinuous, then  $\nu^*$  is identified.

### 3.3 Alternative-specific Excluded Regressors

We now study the case with alternative-specific excluded regressors. We continue to assume that the choice set is  $\{d_1, d_2, d_3\}$ . However, now each alternative has its own regressor  $x_j$  that only affects the utility of alternative  $j$ :  $x = (x_1, x_2, x_3)$ . In addition, these regressors vary independently of one another. Suppose  $d_2$  is always considered. Let the consideration of  $d_j$  be a measurable function of  $x_j$  and preferences, continuous in its first argument:  $\varphi_j = \varphi_j(x_j, \nu)$ .

Identification is built on the following insight. Consider the change in the choice frequency of alternative  $d_1$  in response to an incremental change in  $x_2$  (e.g., a price increase for alternative

---

<sup>21</sup>See Barseghyan, Molinari, & Thirkettle (2019) for other examples of identified consideration processes.

$d_2$ ). The DMs who may switch to  $d_1$  as a result are those indifferent between  $d_1$  and  $d_2$  and consider them both. If these DMs prefer  $d_1$  and  $d_2$  to  $d_3$ , whether  $d_3$  is considered is irrelevant; otherwise, for the response to occur,  $d_3$  should not be considered. These two cases translate to the following statements: (i)  $c_{1,2}(x) < c_{1,3}(x) < c_{2,3}(x)$ ; and (ii)  $c_{2,3}(x) < c_{1,3}(x) < c_{1,2}(x)$  and  $d_3$  is not considered. No other ordering of cutoffs can occur by Fact 4. Then, the choice frequency relates to the model's primitives in these two cases as follows:

$$(i) \quad \Pr(d = d_1|x) = \int_0^{c_{1,2}(x)} \varphi_1(x_1, \nu) dF;$$

$$(ii) \quad \Pr(d = d_1|x) = \int_0^{c_{1,3}(x)} \varphi_1(x_1, \nu) dF + \int_{c_{1,3}(x)}^{c_{1,2}(x)} \varphi_1(x_1, \nu)(1 - \varphi_3(x_3, \nu)) dF.$$

The derivatives with respect to  $x_2$  of these expressions are

$$(i) \quad \frac{\partial \Pr(d = d_1|x)}{\partial x_2} = \varphi_1(x_1, \nu) f(c_{1,2}(x)) \frac{\partial c_{1,2}(x)}{\partial x_2}; \quad (6)$$

$$(ii) \quad \frac{\partial \Pr(d = d_1|x)}{\partial x_2} = (1 - \varphi_3(x_3, \nu)) \varphi_1(x_1, \nu) f(c_{1,2}(x)) \frac{\partial c_{1,2}(x)}{\partial x_2}.$$

Identification is attained by local variation in  $x$  around a point where DMs are indifferent between the three alternatives. For a given  $\nu$ , suppose there is an  $x = (x_1, x_2, x_3)$  such that  $c_{1,2}(x) = c_{1,3}(x) = c_{2,3}(x) = \nu$ , and suppose we observe an open neighborhood around this  $x$ . A small perturbation of  $x_3$  leaves  $U_\nu(d_1, x_1)$ ,  $U_\nu(d_2, x_2)$ , and hence  $c_{1,2}(x)$  unchanged. Taken in a direction that reduces  $U_\nu(d_3, x_3)$ , it generates the cutoff ordering of Case (i) above; and in the opposite direction Case (ii). The ratio of the expressions in Equation (6) identifies  $\varphi_3(x_3, \nu)$ . In a similar fashion, a small perturbation of  $x_1$  identifies  $\varphi_1(x_1, \nu)$ . Plugging these consideration probabilities into Equation (6) identifies  $f(\nu)$ . Alternative-specific variation yields identification without large support variation and without the independence Assumptions I2 and I3. Moreover, it is also possible to allow consideration of  $d_1$  (and  $d_3$ ) to depend on  $x_1$ ,  $\nu$ , and  $x_3$ . The key exclusion restriction in this case is that the consideration of  $d_2$  is independent of all components of  $x$ .

The intuition described above works in a variety of settings. In Section 4 we state a theorem that formalizes our identification results for the ARC model. Below we offer a theorem that applies to a basically unrestricted consideration process that allows for dependence on preferences but not on  $x$ . We then offer a corollary that along with dependence of preferences also allows for dependence on one alternative-specific excluded regressor.

**Definition 5** (Alternative-Specific Variation). *We say that there is alternative-specific variation if (i)  $x \in \mathbb{R}^D$ ; (ii)  $U_\nu(d_j, x)$  depends only on  $x_j$ :  $\frac{\partial U_\nu(d_j, x)}{\partial x_k} \neq 0 \Leftrightarrow k = j$ .*

**Theorem 6.** *Suppose Assumptions I0, I3, T0-T2 hold, there is alternative-specific variation, and the choice set contains at least three alternatives. Suppose*

1. *Each consideration set contains at least two alternatives and  $Q_\nu(\cdot)$  is measurable;*
2. *For a given value of  $\nu$ , there exists an  $x$  with an open neighborhood around it in  $\mathcal{S}$  s.t.*

$$U_\nu(d_1, x) = U_\nu(d_2, x) = \dots = U_\nu(d_D, x).$$

*Then  $f(\nu)$  is identified and so are  $Q_\nu(\mathcal{K})$ ,  $\forall \mathcal{K} \subset \mathcal{D}$ .*

The proof of the theorem works along the lines of the discussion above. Importantly, for any pair  $d_j$  and  $d_k$ , we only need the response of  $\Pr(d = d_j|x)$  to changes in  $x_k$  or the response of  $\Pr(d = d_k|x)$  to changes in  $x_j$ , but not both. Hence, identification relies on variation in  $D - 1$  excluded regressors.

The assumptions of the theorem above rule out singleton (and empty) consideration sets: identification is impossible with singleton consideration sets and arbitrary dependence on preferences, because any empirical choice frequency can be explained by such consideration sets. An alternative approach is to have one alternative – the “default” – to be always considered. Then, provided the fraction of DMs considering only this alternative is independent of  $\nu$ , consideration can depend on the excluded regressor of the default alternative:

**Definition 6.** *Let  $Q_\nu^{x_1}(\mathcal{K})$  denote the probability that, given  $x_1$ , the DM with preference parameter  $\nu$  draws consideration set  $\mathcal{K}$ .*

**Corollary 1.** *Suppose Assumptions I0, T0-T2 hold, there is alternative-specific variation, and the choice set contains at least three alternatives. Suppose*

- 1/ $x_1$ . *All consideration sets contain  $d_1$ , with  $Q_\nu^{x_1}(\{d_1\})$  independent of  $\nu$ :  $Q_\nu^{x_1}(\{d_1\}) = Q^{x_1}(\{d_1\}) < 1$ ,  $\forall \nu$ ; and  $Q_\nu^{x_1}(\cdot)$  are measurable functions, continuous in  $x_1$ ;*
- 2/ $x_1$ . *For a given value of  $x_1$  and each value of  $\nu \in [0, \bar{\nu}]$ , there exists an  $x_{-1} = (x_2, \dots, x_D)$  and an open neighborhood around  $x = (x_1, x_{-1})$  in  $\mathcal{S}$  s.t.*

$$U_\nu(d_1, x) = U_\nu(d_2, x) = \dots = U_\nu(d_D, x).$$

Then  $f(\nu)$  is identified and so are  $\mathcal{Q}_\nu^{x^1}(\mathcal{K})$ ,  $\forall \mathcal{K} \subset \mathcal{D}$ , for all  $\nu$  on the support.

## 4 The ARC Model

We now introduce a specific consideration process, while maintaining the preference structure, including the SCP, from Section 2.2. We refer to this model as the Alternative-specific Random Consideration (ARC) model (Manski, 1977; Manzini & Mariotti, 2014). Each alternative  $d_j$  appears in the consideration set with probability  $\varphi_j$  independently of other alternatives. For now, these probabilities are assumed to be the same across DMs, i.e., they do not depend on  $\nu$ .<sup>22</sup> Once the consideration set is drawn, the DM chooses the best alternative according to her preferences. To avoid empty consideration sets, following Manski (1977), we assume that at least one alternative whose identity is possibly unknown to the researcher is always considered.<sup>23</sup>

**Assumption ARC** (The Basic ARC Model). *The probability that the consideration set takes realization  $\mathcal{K}$  is*

$$\mathcal{Q}(\mathcal{K}) \equiv \prod_{k \in \mathcal{K}} \varphi_k \prod_{k \in \mathcal{D} \setminus \mathcal{K}} (1 - \varphi_k), \quad \forall \mathcal{K} \subset \mathcal{D}, \quad (7)$$

where  $\varphi_j > 0$ ,  $\forall j$ , and  $\exists d^*$  s.t.  $\varphi_{d^*} = 1$ .

By assuming  $\varphi_j > 0$ , we omit never-considered alternatives from the choice problem. Since a never-considered alternative is never compared to any other alternative, whether it is in the choice set or not does not affect the DMs problem at all. Hence, never-considered alternatives have no impact on what we can learn about preferences.<sup>24</sup>

We state identification results starting with the case of the single common excluded regressor. We first extend Theorem 4 to incorporate identification of the consideration parameters.

---

<sup>22</sup>As in Section 2.2, without loss of generality,  $\varphi_j$  can be interpreted as a function of exogenous characteristics such as advertisement. In such a case, all of the results below should be interpreted as conditional on given values of these characteristics.

<sup>23</sup>In the previous version of this paper (Barseghyan, Molinari, & Thirkettle, 2019) this completion rule is called Preferred Option(s). There we also provide identification results for other completion rules, including Coin Toss (if the empty consideration set is drawn, a non-empty consideration set is drawn uniformly at random), Default Option (there is a preset alternative that is chosen if the empty set is drawn), and Outside Option (the DM exits the market if the empty set is drawn).

<sup>24</sup>If an alternative is never chosen, it is w.l.o.g. to set its consideration parameter to zero, even if the said alternative is never chosen because it is dominated by always-considered alternatives (see Section 5).

In the next section we provide results for the case when consideration may depend on preferences.

**Theorem 7.** *Suppose Assumptions I0, I2-I3, T0-T2, ARC hold, and Assumption I1 holds for  $\mathcal{X}$  s.t.  $\forall x \in \mathcal{X}$*

$$U_\nu(d_1, x) > U_\nu(d_j, x) \Rightarrow U_\nu(d_1, x) > U_\nu(d_{j+1}, x), \quad \forall j > 1.$$

*Then  $f(\cdot)$  is identified and so are  $\varphi_1$  and  $\varphi_2$ . In addition, if  $\Pr(d = d_j|x) \neq 0$  for some  $x$ , then  $\varphi_j$  is identified.*

Identification in the ARC model with a single common excluded regressor but without large support is challenging. For reasons discussed in Section 3, there is no general way to pin down the distribution of the preference parameter, even up-to-scale. The collection of observed choice frequencies provides a (potentially infinite) number of equalities where the unknowns are the values of  $F(\cdot)$  and the consideration parameters. Combined with the monotonicity of  $F(\cdot)$ , this yields a system of moment inequalities. These could be fruitfully exploited as it is done in Barseghyan, Coughlin, Molinari, & Teitelbaum (2019).<sup>25</sup>

## 4.1 Preference-Dependent Consideration

We now relax the assumption that consideration is independent of preferences for the ARC model. We do so by considering two types of dependence. The first one adapts the binary consideration types to the ARC model.

**Assumption ARC.B** (ARC with Binary Consideration Types). *For some unknown  $\nu^* \in (0, \bar{\nu})$*

$$\varphi_j(\nu) = \begin{cases} \underline{\varphi}_j & \text{if } \nu < \nu^* \\ \bar{\varphi}_j & \text{if } \nu > \nu^* \end{cases},$$

*with  $\underline{\varphi}_{d^*} = \bar{\varphi}_{d^*} = 1$  for some  $d^*$ , and  $0 < \underline{\varphi}_j, \bar{\varphi}_j < 1$  for  $j \neq d^*$ .*

The second one adapts the proportionally shifting consideration:

**Assumption ARC.P** (ARC with Proportional Consideration). *The consideration process*

---

<sup>25</sup>Alternatively, point-identification can be achieved via an auxiliary assumption that one of the tails of  $F(\cdot)$  can be approximated arbitrarily closely by a finite collection of monotone smooth functions. Further details are available from the authors upon request.

follows the ARC model with  $\{D \geq 4 \ \& \ 1 \leq d^* \leq D\}$  or  $\{D = 3 \ \& \ d^* = 2\}$ , and

$$\varphi_j(\nu) = \begin{cases} \varphi_j(1 - \alpha(\nu)) & \text{if } j < d^* \\ 1 & \text{if } j = d^* \\ \varphi_j(1 + \alpha(\nu)) & \text{if } j > d^* \end{cases}$$

with  $\varphi_j$ 's and  $\alpha(\cdot)$  s.t.  $\alpha(\cdot)$  is continuous,  $\alpha(\bar{\nu}) = 0$ ,  $0 < \varphi_j(\nu) < 1$ ,  $\forall j \neq d^*$ ,  $\forall \nu \in [0, \bar{\nu}]$ .

With binary consideration, we show that identification of the distribution of preferences and all consideration probabilities attains with minimal strengthening of the assumptions in Theorem 5. With proportionally shifting consideration, we rely on additional data moments, i.e.,  $\Pr(d = d_j|x)$ ; and, to exploit those, we impose some testable regularity conditions on the order of the cutoffs of the relevant alternatives.

**Theorem 8.** *Suppose Assumptions I0, I3, T0-T2 hold, and Assumption I1 holds for  $\mathcal{X}$  s.t.  $\forall x \in \mathcal{X}$*

$$U_\nu(d_1, x) > U_\nu(d_j, x) \Rightarrow U_\nu(d_1, x) > U_\nu(d_{j+1}, x), \quad \forall j > 1,$$

and  $\exists x \in \mathcal{X}$  s.t.  $c_{j,k}(x) \leq 0$ ,  $\forall j, k$ ,  $j < k$ . Then  $f(\cdot)$  and  $\{\varphi_j(\cdot)\}_{j=1}^D$  are identified if

1. Assumption ARC.B holds and  $\frac{d\Pr(d=d_j|x)}{dx}$  is discontinuous for some  $j$ .
2. Assumption ARC.P holds and  $\forall x \in \mathcal{X}$

$$U_\nu(d_D, x) > U_\nu(d_j, x) \Rightarrow U_\nu(d_D, x) > U_\nu(d_{j-1}, x), \quad \forall j < D,$$

and  $\forall x \in \mathcal{X} \nexists \nu \in [0, \bar{\nu}]$  and  $\{j, k, l\}$  s.t.  $U(d_j, x) = U(d_k, x) = U(d_l, x)$ .

## 4.2 Identification with Alternative-specific Excluded Regressors

We now formally state the result developed in Section 3.3.

**Assumption ARC.AS.** *The consideration process follows the ARC model. The consideration probability of each alternative  $d_j$  is a measurable function of  $x_j$  and preferences:  $\varphi_j = \varphi_j(x_j, \nu)$ , continuous in the first argument. Default alternative  $d^*$  is s.t.  $\varphi_{d^*}(x_{d^*}, \nu) = 1$  for all  $x_{d^*} \in \mathcal{S}$  and for all  $\nu \in [0, \bar{\nu}]$ .*

**Theorem 9.** *Suppose Assumptions I0, T0-T2, ARC.AS hold. Suppose there is alternative-*



specific variation and the choice sets contain at least three alternatives. Suppose for a given value of  $\nu$  there exists an  $x = (x_1, x_2, \dots, x_D)$ , and an open neighborhood around it in  $\mathcal{S}$ , s.t.

$$U_\nu(d_1, x) = U_\nu(d_2, x) = \dots = U_\nu(d_D, x).$$

Then  $f(\nu)$  and  $\{\varphi_j(x_j, \nu)\}_{j=1}^D$  are identified.

## 5 Likelihood and Tractability

We now turn to the computational aspects of limited consideration models under the SCP and, in particular, of their likelihood function. Consider a generic consideration process. We denote the probability that alternative  $d_j$  is in the consideration set and every alternative in set  $\mathcal{B}$  is not by

$$\mathcal{O}_\nu^x(d_j; \mathcal{B}) \equiv \sum_{j \in \mathcal{K}; \mathcal{K} \cap \mathcal{B} = \emptyset} \mathcal{Q}_\nu^x(\mathcal{K}).$$

A computationally appealing way to write the likelihood function is to determine the probability that a DM with preference parameter  $\nu$  chooses alternative  $d_j$  conditional on  $x$ . Alternative  $d_j$  is chosen if and only if  $d_j$  is in the consideration set and every alternative that dominates  $d_j$  is not. Denote the set of alternatives that are preferred to  $d_j$  by

$$\mathcal{B}_\nu(d_j, x) \equiv \{k : U_\nu(d_k, x) > U_\nu(d_j, x)\}.$$

Then,

$$\Pr(d_j|x) = \int \Pr(d_j|x, \nu) dF = \int \mathcal{O}_\nu^x(d_j; \mathcal{B}_\nu(d_j, x)) dF. \quad (8)$$

The object on the RHS does not require evaluating the utility of each alternative within each possible consideration set. In fact,  $U_\nu(d_j, x)$  needs to be computed only once for each  $\nu$ ,  $d_j$ , and  $x$  to create  $\mathcal{B}_\nu(d_j, x)$ , which does not vary with the consideration set.<sup>26</sup> Hence the computational complexity lies in the mapping from the parameters governing the consideration process to  $\mathcal{O}_\nu^x(\cdot)$ 's. These, in turn, are summations that may not even require enumerating all possible consideration sets. To demonstrate this with a concrete example, we proceed

---

<sup>26</sup>The resulting computational gains are similar to those in importance sampling (e.g., Ackerberg, 2009).

with the basic ARC model. In this case, the RHS of Equation (8) is:

$$I(d_j|x) \equiv \varphi_j \int \prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) dF. \quad (9)$$

Given  $\{\varphi_j\}_{j=1}^D$ , the integrand  $\prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k)$  is piecewise constant in  $\nu$  with at most  $D-1$  breakpoints, corresponding to indifference points between alternatives  $j$  and  $k$ , i.e.,  $c_{j,k}(x)$ 's, that are computed only once for every  $x$ . There are at least two methods to compute this integral. First, for every  $d_j$  and  $x$ , we can directly compute the breakpoints and hence write  $I(d_j|x)$  as a weighted sum:

$$I(d_j|x) = \sum_{h=0}^{D-1} \varphi_j \left( (F(\nu_{h+1}) - F(\nu_h)) \prod_{k \in \mathcal{B}_{\nu_h}(d_j, x)} (1 - \varphi_k) \right),$$

where  $\nu_h$ 's are the sequentially ordered breakpoints augmented by the integration endpoints:  $\nu_0 = 0$  and  $\nu_D = \bar{\nu}$ . This expression is trivial to evaluate given  $F(\cdot)$  and breakpoints  $\{\nu_h\}_{h=0}^D$ . More importantly, since the breakpoints are invariant with respect to the consideration probabilities, they are computed only once for each  $x$ . This simplifies the likelihood maximization routine by orders of magnitude, as each evaluation of the objective function involves a summation over products with at most  $D$  terms. A second approach is to compute  $I(d_j|x)$  using Riemann approximation:

$$I(d_j|x) \approx \frac{\bar{\nu}}{M} \sum_{m=1}^M \varphi_j \left( f(\nu_m) \prod_{k \in \mathcal{B}_{\nu_m}(d_j, x)} (1 - \varphi_k) \right),$$

where  $M$  is the number of intervals in the approximating sum,  $\frac{\bar{\nu}}{M}$  is the intervals' length,  $\nu_m$ 's are the intervals' midpoints, and  $f(\cdot)$  is the density of  $F(\cdot)$ . Again, one does not need to evaluate the utility from different alternatives in the likelihood maximization. Instead, one *a priori* computes the utility rankings for each  $\nu_m$ ,  $m = 1, \dots, M$ . These rankings determine  $\mathcal{B}_{\nu_m}(d_j, x)$ . The likelihood maximization is now a standard search routine over  $\{\varphi_j\}_{j=1}^D$  and  $f(\cdot)$ . Our theory restricts  $f(\cdot)$  to the class of continuous and strictly positive functions. In practice, the search is over a class of non-parametric estimators for  $f(\cdot)$ .<sup>27</sup> If the density is parameterized, i.e.,  $f(\nu_m) \equiv f(\nu_m; \theta^f)$ , then the maximization is over  $\{\varphi_j\}_{j=1}^D$  and  $\theta^f$ . Finally, the midpoints are the same across all DMs, further reducing computational

---

<sup>27</sup>One could use a mixture of Beta distributions (Ghosal, 2001), as we do in Section 7.

burden.<sup>28</sup>

Allowing consideration to depend on preferences (or on  $x$ ) introduces only minimal adjustments to the likelihood function. Let, e.g., each consideration function be parameterized by  $\theta_j$ :  $\varphi_j(\nu) \equiv \varphi_j(\nu; \theta_j)$ . Then, at each  $\nu$ , we can simply substitute  $\varphi_j$  with the corresponding  $\varphi_j(\nu; \theta_j)$ , and the likelihood maximization is now over  $\{\theta_j\}_{j=1}^D$  and  $\theta^f$ . Given the desired level of parameterization – i.e., the dimensionality of the parameter vectors  $\theta_j$  and  $\theta^f$  – the computational complexity of the problem grows polynomially in  $D$ .

As a final remark, if alternative  $d_j$  is never chosen, then one can conduct estimation as if  $d_j$  were not in the choice set. Indeed, per Equation (9),  $\varphi_j$  contributes positively to the likelihood if and only if alternative  $d_j$  is chosen. When it is never chosen, it may only enter via the term  $(1 - \varphi_j)$ ; hence, the likelihood will be maximized by setting  $\varphi_j = 0$ . Therefore, setting  $\varphi_j = 0$  for all zero-share alternatives, regardless of why they were not chosen, has no impact on estimation. This too may speed up estimation.

## 6 Properties of Limited Consideration Models

### 6.1 The Standard RUM

We focus on a *standard* application of the RUM with full consideration in the context of our example in Section 2.1. The final evaluation of the utility that the DM derives from alternative  $j$  now includes a separately additive error term:

$$V_\nu(L_j(x)) = U_\nu(L_j(x)) + \varepsilon_j, \quad (10)$$

where, as before,  $\nu$  captures unobserved heterogeneity in preferences, and  $\varepsilon_j$  is assumed independent of the random coefficients (in this application,  $\nu$ ).

Typical implementations of this model further specify that  $\varepsilon_j$  is i.i.d. across alternatives (and DMs) with a Type 1 Extreme Value distribution, following the seminal work of McFadden (1974). This yields a Mixed Logit that differs from, for example, McFadden & Train (2000) because in the latter the random coefficient(s) enter the utility function linearly, while in

---

<sup>28</sup>Depending on the class of  $f(\cdot)$ , it may be more accurate to compute  $I(d_j|x)$  by substituting  $\frac{\bar{\nu}}{M}f(\nu_m)$  with  $F(\bar{\nu}_m) - F(\underline{\nu}_m)$ , where  $\bar{\nu}_m$  and  $\underline{\nu}_m$  are the endpoints of the corresponding interval.

the context of expected utility they enter nonlinearly. We now discuss two properties of the Mixed Logit that hinder its applicability in our context.

Coupling utility functions in the hyperbolic absolute risk aversion (HARA) family, for example CARA or CRRA, with a Type 1 Extreme Value distributed additive error, yields:

**Property 1** (Non-monotonicity in RUM). *In Model (10) with HARA preferences and  $\varepsilon_j$  i.i.d. Type 1 Extreme Value, as the DM's risk aversion increases, the probability that she chooses a riskier alternative declines at first, but eventually starts to increase (Apesteguia & Ballester, 2018).*<sup>29</sup>

To see why, consider two non-dominated alternatives  $d_j$  and  $d_k$  such that  $d_j$  is riskier than  $d_k$ . A risk neutral DM prefers  $d_j$  to  $d_k$ , and hence will choose the former with higher probability. As risk aversion increases, the DM eventually becomes indifferent between  $d_j$  and  $d_k$  and chooses either of these alternatives with equal probability. As risk aversion increases further, she prefers  $d_k$  to  $d_j$  and chooses the latter with lower probability. However, as risk aversion gets even larger, the expected utility of any lottery with finite stakes converges to zero. Consequently, the choice probabilities of all alternatives, regardless of their riskiness, converge to a common value.<sup>30</sup> Hence, at some point the probability of choosing  $d_j$  becomes increasing in risk aversion.

Next, we establish the relation between utility differences across two alternatives and their respective choice probabilities. Because our random expected utility model features unobserved preference heterogeneity, we work with an analog of the rank order property in Manski (1975) that is conditional on  $\nu$ :

**Definition 7.** (*Conditional Rank Order of Choice Probabilities*) *The model yields conditional rank order of the choice probabilities if for given  $\nu$  and alternatives  $j, k \in \mathcal{D}$ ,*

$$U_\nu(L_j(x)) > U_\nu(L_k(x)) \Rightarrow \Pr(d = d_j|x, \nu) > \Pr(d = d_k|x, \nu).$$

The standard Mixed Logit yields conditional rank ordering of the choice probabilities given  $\nu$ .<sup>31</sup> In turn, we show that the conditional rank order property implies the following upper

---

<sup>29</sup>See also Wilcox (2008).

<sup>30</sup>Recall that in the Mixed Logit the magnitude of the utility differences is tied to differences in (log) choice probabilities,  $U_\nu(L_k(x)) - U_\nu(L_j(x)) = \log(\Pr(d = d_k|x, \nu)) - \log(\Pr(d = d_j|x, \nu))$ , so that as  $\nu \rightarrow \infty$  the choice probabilities are predicted to be all equal.

<sup>31</sup>Manski (1975) establishes the rank order property for additive error random utility models (without

bound on the probability that suboptimal alternatives are chosen:

**Property 2.** (*Generalized Dominance*) Consider any  $x$ , alternative  $j$ , and set  $\mathcal{K} \subset \mathcal{D} \setminus \{j\}$  s.t. alternative  $j$  is never-the-first-best in  $\mathcal{K} \cup \{d_j\}$ . Then  $\Pr(d = d_j|x) < \sum_{k \in \mathcal{K}} \Pr(d = d_k|x)$ .

In the Mixed Logit, where the conditional rank order property holds, if alternative  $j$  is never-the-first best among  $\{d_j, d_k, d_l\}$ , then the probability of observing  $j$  is predicted to be less than the sum of the probabilities of observing  $k$  or  $l$ .

## 6.2 Monotonicity in Limited Consideration Models

A model with consideration process independent of preferences yields predicted choice probabilities that are monotone in the preference parameter. We define monotonicity as follows:

**Property 3** (*Generalized Preference Monotonicity*). A model satisfies generalized preference monotonicity if for any  $\nu_1 < \nu_2$  and  $J \in \{1, 2, \dots, D\}$ :

$$\Pr\left(\bigcup_{j=1}^J d_j \mid x, \nu_1\right) \geq \Pr\left(\bigcup_{j=1}^J d_j \mid x, \nu_2\right).$$

In the context of risk preferences, Property 3 states that the probability of choosing one of the  $J$  riskiest alternatives declines as  $\nu$  increases. Since Property 3 is satisfied for any choice set under the SCP and full consideration, it is also satisfied under limited consideration:

**Proposition 10.** A model that satisfies the SCP (i.e., Assumption T2) and Assumption I2 satisfies Generalized Preference Monotonicity.

## 6.3 Ordinal Properties of Limited Consideration Models

In the Mixed Logit, the cardinality of the differences in the (random) expected utility of alternatives plays a crucial role in the determination of choice probabilities, as it interacts with the realization of the additive error. In contrast, in models that satisfy the SCP, the DMs' choices are determined by the ordinal expected utility ranking of the alternatives.

---

random coefficients) for a broader class of models that only require very weak restrictions on  $\varepsilon_j$ . Conditional on  $\nu$ , his results extend immediately to yield the conditional rank order property.

**Proposition 11.** *A model that satisfies the SCP (i.e., Assumption T2) exhibits the following scale invariance: multiplication of  $U_\nu(\cdot)$  by any positive function of  $\nu$  leaves the model's predictions unchanged.*

Limited consideration models satisfying the SCP can be recast as Ordinal Random Utility models (ORUM), with the key departure from standard RUMs being the nature of the additive error term. We illustrate this point with the following proposition:

**Proposition 12.** *(The ARC Model as ORUM) The basic ARC model is equivalent to an additive error random utility model with unobserved preference heterogeneity where all alternatives are considered, the DM's utility of each alternative  $j \in \{1, \dots, D\}$  is given by*

$$V_\nu(d_j, x) = U_\nu(d_j, x) + \varepsilon_j,$$

and  $\varepsilon_j$  is a random variable, independent of  $x$ ,  $\nu$ , and across alternatives, s.t.

$$\varepsilon_j = \begin{cases} 0 & \text{with probability } \varphi_j \\ -\infty & \text{with probability } (1 - \varphi_j). \end{cases}$$

Given Proposition 12 it is straightforward to establish:

**Proposition 13.** *The basic ARC model may violate the Conditional Rank Order Property and, hence, the Generalized Dominance.*

We conclude this section with Table 1, listing the differences across the Mixed Logit and various versions of the ARC model. The first two columns summarize the differences between the basic ARC model and the Mixed Logit. In the ARC model, the error terms are independent across alternatives, but in other limited consideration models this is not the case. See, e.g., the RCL model in Barseghyan, Molinari, & Thirkettle (2019). The next column reminds the reader that in our models the error terms can be correlated with preferences, and still be identified with a single common excluded regressor. Finally, the last column highlights the fact that with alternative-specific variation we may also have dependence of the error term on the excluded regressor(s). We intentionally do not fill in the properties of the last two models in the table, as whether these properties hold depends on the exact assumptions the researcher imposes on the consideration process.

**Table 1** Model Comparisons

	Mixed Logit	ARC		
		Basic	Binary Types Prop. Shifting	Unrestricted
<b>Error Distribution</b>				
Support	$\mathbb{R}$	$[-\infty, 0]$	$[-\infty, 0]$	$[-\infty, 0]$
Independent of $x$	Yes	Yes	Yes	No
Independent of $\nu$	Yes	Yes	No	No
Independent across alternatives	Yes	Yes	Yes	Yes
Identical across alternatives	Yes	No	No	No
<b>Properties</b>				
Monotonicity	No	Yes		
Conditional Rank Order Property	Yes	No		
Generalized Dominance	Yes	No		

## 7 Application

### 7.1 Data

We study households' deductible choices across three lines of property insurance: auto collision, auto comprehensive, and home all perils. The data come from a U.S. insurance company. Our analysis uses a sample of 7,736 households who purchased their auto and home policies for the first time between 2003 and 2007 and within six months of each other.<sup>32</sup> Table D.1 provides descriptive statistics for households' observable characteristics, which we use later to estimate households' preferences.<sup>33</sup> We observe the exact menu of alternatives available at the time of the purchase for each household and each line of coverage. The deductible alternatives vary across lines of coverage but not across households. Table D.2 presents the frequency of chosen deductibles in our data.

Premiums are set coverage-by-coverage as in the example from Section 2.1. Table D.5 reports the average premium by context and deductible, and Table 2 summarizes the premium distributions for the \$500 deductible. Premiums vary dramatically. The 99th percentile of

<sup>32</sup>The dataset is an updated version of the one used in Barseghyan et al. (2013). It contains information for an additional year of data and puts stricter restrictions on the timing of purchases across different lines. These restrictions are meant to minimize potential biases stemming from non-active choices, such as policy renewals, and temporal changes in socioeconomic conditions.

<sup>33</sup>These are the same variables that are used in Barseghyan et al. (2013) to control for households' characteristics. See discussion there for additional details.

**Table 2** Premium Quantiles for the \$500 Deductible

Quantiles	0.01	0.05	0.25	0.50	0.75	0.95	0.99
Collision	53	74	117	162	227	383	565
Comprehensive	29	41	69	99	141	242	427
Home	211	305	420	540	743	1,449	2,524

**Table 3** Claim Probabilities Across Contexts

Quantiles	0.01	0.05	0.25	0.50	0.75	0.95	0.99
Collision	0.036	0.045	0.062	0.077	0.096	0.128	0.156
Comprehensive	0.005	0.008	0.014	0.021	0.030	0.045	0.062
Home	0.024	0.032	0.048	0.064	0.084	0.130	0.183

the \$500 deductible is more than ten times the corresponding 1st percentile in each line of coverage.

Claim probabilities were derived using coverage-by-coverage Poisson-Gamma Bayesian credibility models applied to a large auxiliary panel (Barseghyan, Teitelbaum, & Xu, 2018). Predicted claim probabilities (summarized in Table 3) exhibit extreme variation: The 99th percentile claim probability in collision (comprehensive and home) is 4.3 (12 and 7.6) times higher than the corresponding 1st percentile. Finally, the correlation between claim probabilities and premiums for the \$500 deductible is 0.38 for collision, 0.15 for comprehensive, and 0.11 for home all perils. Hence, there is independent variation in both.<sup>34</sup>

## 7.2 Estimation Results

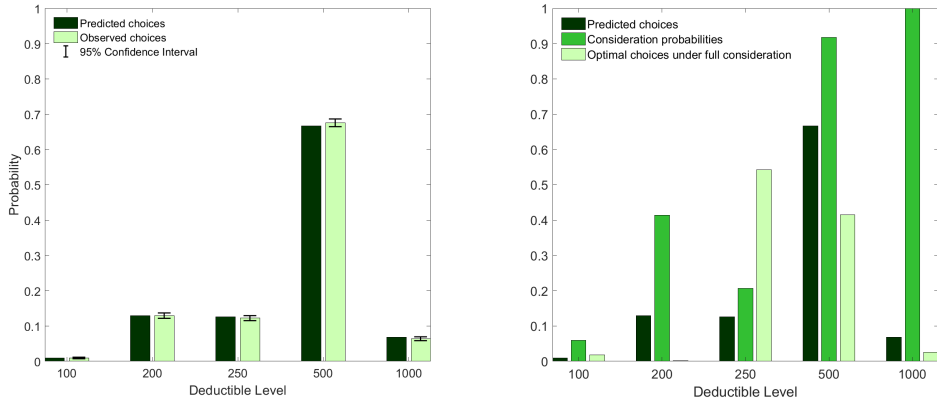
### 7.2.1 The basic ARC Model: Collision

We start by presenting estimation results in a simple setting where the only choice is the collision deductible and observable demographics do not affect preferences. To execute our estimation procedure we set  $\bar{\nu} = 0.02$ , which is conservative (see Barseghyan, Molinari, & Teitelbaum, 2016). We *ex post* verify that this does not affect our estimation by checking that the density of the estimated distribution is close to zero at the upper bound. We approximate  $F(\cdot)$  non-parametrically through a mixture of Beta distributions. In practice, however, both

<sup>34</sup>See Barseghyan et al. (2013) (see Cohen & Einav, 2007, in the context of Israeli auto insurance) for a detailed discussion of where such independent variation comes from.



**Figure 1: The ARC Model**



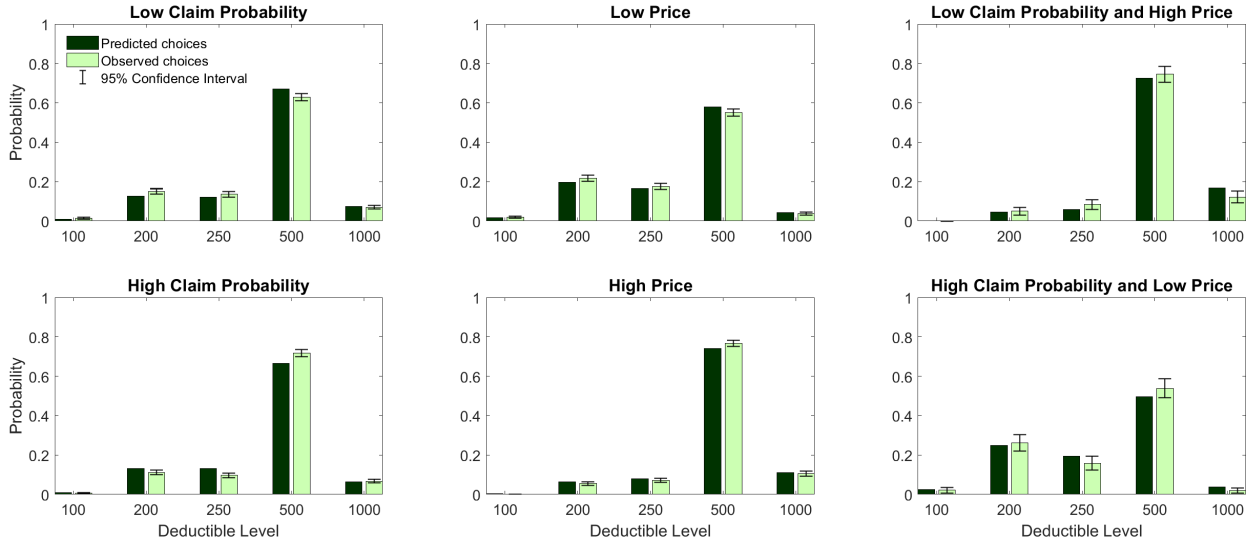
The first panel reports the distribution of predicted and observed choices. The second panel displays consideration probabilities and the distribution of optimal choices under full consideration.

AIC/BIC criteria indicate that a single component is sufficient for our analysis, resulting in total of seven parameters to be estimated. We let the data speak to the identity of the always-considered alternative.<sup>35</sup>

The estimated distribution and consideration parameters are reported in Table E.1. As the first panel in Figure 1 shows, the model closely matches the aggregate moments observed in the data. The second panel in Figure 1 illustrates side-by-side the frequency of predicted choices, consideration probabilities, and the distribution of households' first-best alternatives (i.e., the distribution of optimal choices under full consideration). Predicted choices are determined jointly by the preference induced ranking of deductibles and by the consideration probabilities: Limited consideration forces households' decision towards less desirable outcomes by stochastically eliminating better alternatives. The two highest deductibles (\$1000 and \$500) are considered at much higher frequency (1.00 and 0.92, respectively) than the other alternatives, suggesting that households have a tendency to regularly pay attention to the cheaper items in the choice set. Yet, the most frequent model-implied optimal choice under full consideration is the \$250 deductible, which is considered with low probability. In this application, assuming full consideration leads to a significant downward bias in the estimation of the underlying risk preferences. To see why, consider increasing the consideration probabilities for the lower deductibles to the same levels as the \$500 deductible. Holding risk preferences fixed, the likelihood that the lower deductibles are chosen increases and therefore the higher deductibles are chosen with lower probability. Average risk aversion must decline

<sup>35</sup>In fact, the estimation is run under the Coin Toss completion rule that nests the possibility that any alternative can be always considered. The data chooses  $\varphi_{1000} = 1$ .

**Figure 2:** The ARC Model: Conditional Distributions



to compensate for this shift and to push the likelihood function up. This is exactly the pattern we find when we estimate a near-full consideration model. In particular, we find that average risk aversion decreases by about 34% from 0.0036 to 0.0024 when all consideration parameters equal 0.9999.<sup>36</sup> To put these numbers into context, a DM with risk aversion equal to 0.0037 is willing to pay \$424 to avoid a \$1000 loss with probability 0.1, while a DM with risk aversion equal to 0.0027 is only willing to pay \$287 to avoid the loss.

The basic ARC model’s ability to match the data extends also to conditional moments. The first two panels of Figure 2 show observed and predicted choices for the fraction of households facing low and high premiums, respectively, and the next two panels are for households facing low and high claim probabilities.<sup>37</sup> Finally, the last two panels display households who face both low claim probabilities and high prices and *vice versa*. It is transparent from Figure 2 that the model matches closely the observed frequency of choices across different subgroups of households facing a variety of prices and claim probabilities, even though some of these frequencies are quite different from the aggregate ones.

The ARC model’s ability to violate Generalized Dominance is key in matching the data. In our dataset, because of the pricing schedule in collision, the \$200 is never-the-first best among  $\{\$100, \$200, \$250\}$ . It costs the same to get an additional \$50 of coverage by lowering

<sup>36</sup>We cannot assume that all consideration probabilities are equal to one, since the \$200 deductible is never the first best under full consideration and is chosen with positive probability.

<sup>37</sup>Low and high groups here are defined as households whose claim rate (or baseline price) are in the first and third terciles, respectively.

the deductible from \$250 to \$200 as it does to get an additional \$100 of coverage by lowering the deductible from \$200 to \$100. If a household’s risk aversion is sufficiently small, then it prefers the \$250 deductible to the \$200 deductible. If, on the other hand, the household’s level of risk aversion is such that it would prefer the \$200 deductible to the \$250 deductible, then it would also prefer getting twice the coverage for the same increase in the premium. That is, for any level of risk aversion, the \$200 deductible is dominated either by the \$100 deductible or by the \$250 deductible.<sup>38</sup> Yet, overall the \$200 deductible is chosen roughly as often as the \$100 and \$250 deductibles combined. More so, for certain sub-groups the \$200 deductible is chosen *much more often* than the \$100 and \$250 deductible combined. It follows that a model satisfying Generalized Dominance cannot rationalize these choices.

Next we relax the assumption that demographic variables do not influence risk preferences. The details of this step and the results are reported in Appendix E. Both consideration and preference estimates remain close to those reported above. We also estimate the model with binary types and the model with proportionally shifting consideration. While the results are intuitive, i.e., the relation between risk aversion and consideration of high coverages is positive, neither of these two models offer a significant improvement in fit and, in fact, we fail to reject the basic ARC model in favor of these using the likelihood ratio test.

### 7.2.2 The Mixed Logit Random Utility Model

As in the case of the ARC model, we assume that  $\nu$  is Beta distributed on  $[0, \bar{\nu}]$ , where  $\bar{\nu} = 0.02$ . The Mixed Logit satisfies the Conditional Rank Order Property and smoothly spreads households’ choices around their respective first bests. Consequently, it cannot match the observed distribution and, in particular, is unable to explain the relatively high observed share of the \$200 deductible. Table E.2 reports the estimation results and Figure E.3 compares the observed distribution of choices to the predicted choices. The predicted distribution is a much poorer fit relative to the ARC model. In fact, the Vuong (1989) test soundly rejects (at 1% level) the Mixed Logit in favor of the ARC model.

### 7.2.3 The ARC Model: All Coverages

We now proceed with estimation of the full model. We consider two cases. In the first case, households’ risk preferences are invariant across lines of coverage, but consideration sets form

---

<sup>38</sup>This pattern is at odds not only with EUT but also many non-EU models (Barseghyan et al., 2016).

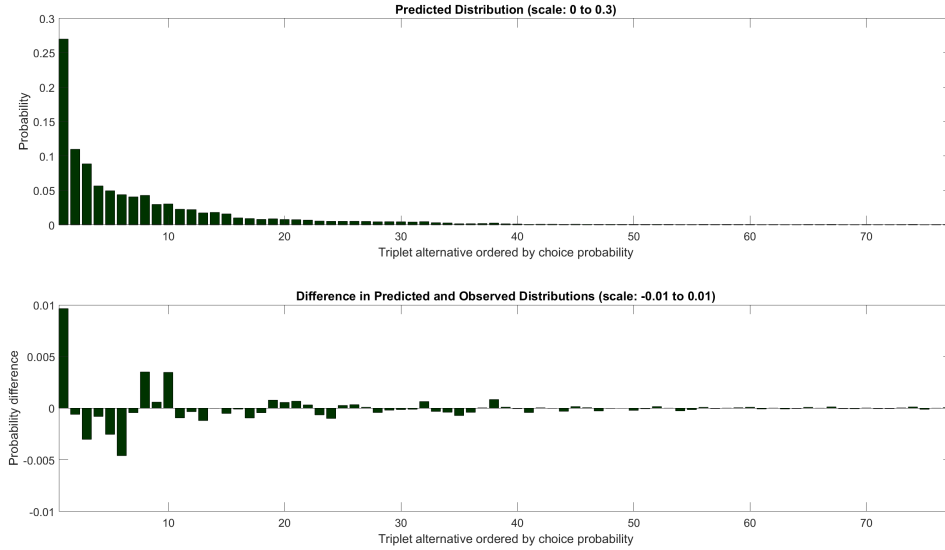
independently within each line of coverage. There are three sets of consideration parameters  $\{\varphi^{coll}, \varphi^{comp}, \varphi^{home}\}$  and the probability that alternative  $k$  is considered in one line of coverage (e.g. collision) is independent of the probability that alternative  $j$  is considered in another line of coverage (comprehensive or home). Hence, within each coverage, the households' problem is identical to that from the previous section. Just as in the case of collision coverage only, the model matches well the choice distributions within each line of coverage. However, the independence of consideration sets across lines of coverage implies that the model does not have the ability to match the joint distribution of choices. For example, the model predicts zero rank correlation across the lines of coverage and that 12% of households choose an alternative with a larger comprehensive deductible than collision deductible. The rank correlation ranges from 0.35 to 0.61 and only 0.2% of households choose a larger comprehensive deductible.

We next assume that households' consideration sets are formed over the entire deductible portfolio. There are 120 possible alternative triplets  $(d^{coll}, d^{comp}, d^{home})$ , each having its own probability of being considered. This model is flexible as it nests many rule of thumb assumptions such as only considering contracts with the same deductible level across the three contexts or only considering contracts with a larger collision deductible than comprehensive deductible. Figure 3 and Table E.4 present estimation results. The first panel of the figure shows the predicted distribution of choices across triplets, ranked in descending order by observed frequencies. The second panel plots the differences between predicted and observed choice distributions. Clearly, the predicted distribution is close to the observed distribution.

The largest difference between the predicted and observed shares equals 0.96 percentage points, which is for the (\$500, \$500, \$500) triplet that is chosen by 26% of the households. The integrated absolute error across all triplets is 4.61%. In our data, 43 out of 120 triplets are never chosen (these are omitted from Figure 3). As discussed in Section 5, the likelihood maximization implies that the consideration probabilities for these triplets must be zero, so that their predicted shares are zero. Hence, the likelihood maximization routine is faster and more reliable as we do not need to search for  $\varphi_j$  for these alternatives.

Another virtue of the ARC model is that it effortlessly reconciles two sides of the debate on stability of risk preferences (Barseghyan, Prince, & Teitelbaum, 2011; Einav, Finkelstein, Pascu, & Cullen, 2012; Barseghyan et al., 2016). On the one hand, households' risk aversion relative to their peers is correlated across lines of coverage, implying that households preferences have a stable component. On the other hand, analyses based on revealed preference

**Figure 3:** The ARC Model, Three Coverages



Triplets are sorted by observed frequency at which they are chosen. The first panel reports the predicted choice frequency and the second panel reports the difference in predicted and observed choice frequencies.

reject the standard models: under full consideration, for the vast majority of households one cannot find a level of (household-specific) risk aversion that justifies their choices simultaneously across all contexts. Limited consideration allows the model to match the observed joint distribution of choices, and hence their rank correlations.

The estimated risk preferences are similar to those estimated with collision only data, although the variance is slightly smaller. Turning to consideration, the triplet considered most frequently is the cheapest one: (\$1000, \$1000, \$1000). Its consideration probability is 0.81, while the next two most considered triplets are (\$500, \$500, \$1000) and (\$500, \$500, \$500). These are considered with probability 0.47 and 0.43, respectively. Overall, there is a strong positive correlation (0.54) between the consideration probability and the sum of the deductibles in a given alternative. We summarize once more the computational advantages of our procedure. First, estimation of our model remains feasible for a large choice set.<sup>39</sup> Second, the model’s parameters grow linearly with the size of the choice set – one parameter per an additional alternative. Third, enlarging the choice set does not call for new independent sources of data variation. For example, in our model whether there are five deductible

<sup>39</sup>In our setting, it is feasible to estimate an additive error RUM assuming the DMs consider each deductible triplet as a separate alternative (Figure E.5 and Table E.5). As the figure shows, the failure to match the data is evident. The Vuong test formally rejects it in favor of the ARC model.

alternatives or one hundred twenty does not make any difference either from an identification or an estimation stand point: with sufficient variation in  $\bar{p}$  and/or  $\mu$ , the model is identified and can be estimated. As a final remark, once the model is estimated, one can compute the average monetary cost of limited consideration. In our data it is \$49 (see Appendix C).

## 8 Discussion

The literature concerned with the formulation, identification, and estimation of discrete choice models with limited consideration is vast. However, to our knowledge, there is no previous work applying such models to the study of decision making under risk, except for the contemporaneous work of Barseghyan, Coughlin, et al. (2019). In particular, this paper is the first to exploit the SCP for identification purposes. As a result, several fundamental differences emerge between our work and existing papers. First, we achieve identification in the most challenging case where there is a single excluded regressor that affects the utility of all alternatives.<sup>40</sup> Second, we allow for consideration to depend on preferences. Third, with alternative-specific excluded regressors, this dependence can be essentially unrestricted and can be combined with dependence of consideration on (some of) the excluded regressors. Fourth, we scrutinize the large support assumption, show why it may be necessary, and when and how it is possible to make progress when it is not satisfied. Fifth, our approach comes with an easy to implement and computationally fast estimation strategy. Finally, we make a contribution specific to the study of decision making under risk by proposing a model that is immune from Apesteguia & Ballester (2018) criticism and features two sources of unobserved heterogeneity – risk aversion and limited consideration – whose distributions are identified. More generally, the paper establishes that, as long as the DMs’ preferences satisfy the SCP, allowing for limited consideration does not hinder the model’s identifiability or applicability. Hence, we view our framework as a stepping stone for studies of consumer behavior in markets where limited consideration may be present (one example is Coughlin, 2019, who builds on our framework to study consumer choice in Medicare Part D markets).

Papers that allow for limited consideration or more broadly for choice set heterogeneity can be classified in four groups. The first relies on auxiliary information about the composition or distribution of DMs’ choice sets, such as brand awareness (e.g., Draganska & Klapper,

---

<sup>40</sup>This setting is common in insurance markets, see, e.g., Cohen & Einav (2007); Einav et al. (2012); Sydnor (2010); Barseghyan et al. (2011, 2013); Handel (2013); Bhargava, Loewenstein, & Sydnor (2017).

2011; Honka & Chintagunta, 2017) or search activity (e.g., Honka & Chintagunta, 2017; De los Santos, Hortaçsu, & Wildenbeest, 2012; Kim, Albuquerque, & Bronnenberg, 2010; Honka, Hortaçsu, & Vitorino, 2017).<sup>41</sup> We do not require such information.

The second group attains identification via two-way exclusion restrictions, i.e., by assuming that some variables impact consideration but not utility and *vice versa*. A well-known example of this approach is Goeree (2008), who posits that advertising intensity affects the likelihood of considering a computer, but does not impact consumer preferences, while computer attributes such as CPU speed affect preferences but not consideration (see also van Nierop, Bronnenberg, Paap, Wedel, & Franses (2010) and Gaynor, Propper, & Seiler (2016)). Hortaçsu, Madanizadeh, & Puller (2017) create an exclusion restriction by exploiting the dynamic aspect of consumer choice.<sup>42</sup> The consumer’s decision to consider alternatives to her current service provider is a function of (her experiences with) the last period provider but not her next period provider (see also Heiss et al. (2016)). In contrast, we achieve identification with as little as one common excluded regressor and a single cross section.

The third group relies on restricting the consideration process to a specific class of models. Abaluck & Adams (2018) consider two such models (and their hybrid): a variant of the ARC and a “default specific” model (as in, e.g., Ho et al., 2017; Heiss et al., 2016) in which each DM’s consideration set comprises either a single default alternative or the entire feasible set. They assume that consideration and preferences are independent, and that each alternative has a characteristic with large support that is additively separable in utility and may only affect its own consideration but not the consideration of other alternatives.<sup>43</sup> They exploit violations of symmetry in the Slutsky matrix (i.e., in cross-alternative demand responses to prices) to detect limited consideration. Kawaguchi et al. (2020) study beverage purchases from vending machines, allowing advertisement to be a driver of consideration, but also to affect utility. Their approach is close to that of Goeree (2008), though they provide a formal argument for identification with large support and exclusion restrictions even when there is no choice set variation. A key assumption is that all beverages are considered with probability equal to one as the advertising intensity of each beverage becomes very large.

---

<sup>41</sup>For canonical cites see, e.g., Roberts & Lattin (1991) and Ben-Akiva & Boccara (1995).

<sup>42</sup>Time variation is used also in Crawford, Griffith, & Iaria (2020), who show that with panel data and preferences in the logit family, point identification of preferences is possible, without any exclusion restrictions, under the assumption that choice sets and preferences are independent conditional on observables and with restrictions on how choice sets evolve over time. These restrictions enable the construction of proper subsets of DMs’ true choice sets (‘sufficient sets’) that can be utilized to estimate the preference model.

<sup>43</sup>The exception is the “default” alternative, whose characteristic may trigger the consideration of the entire choice set.

The methods we propose relate to the papers in the third group in two aspects. First, we too *sometimes* require large support as a “fail safe” assumption, but only in the most challenging case of a single common excluded regressor. Second, we too rely on exclusion restrictions. The reliance on these assumptions is inescapable given the econometrics literature on point identification of discrete choice models. Our approach elucidates the identifying power of a single excluded regressor in models that satisfy the SCP and, in particular, the relative ranking of alternatives encapsulated in Facts 3 and 4 (see Lewbel & Yang, 2016, for related results for average treatment effects in ordered discrete choice models). We further exploit this structure to establish identification in models with substantially richer levels of unobserved heterogeneity, by allowing for dependence between consideration and preferences.

The fourth group of papers has a different goal than what we pursue here, as it provides partial rather than point identification results. Cattaneo, Ma, Masatlioglu, & Suleymanov (2019) propose a random attention model with homogeneous preferences, and they require that the probability of each consideration set is monotone in the number of alternatives in the choice problem. Their analysis yields testable implications and partial identification for preference orderings. Barseghyan, Coughlin, et al. (2019) study discrete choice models, where consideration may arbitrarily depend on preferences as well as on all observed characteristics. They show that such unrestricted forms of heterogeneity generally yield partial, but not point, identification of the preference distribution and obtain bounds on the distribution of consideration sets’ size. Finally, Dardanoni, Manzini, Mariotti, & Tyson (2020) consider a stochastic choice model with homogeneous preferences and heterogeneous cognitive types. They show how one can learn the moments of the distribution of cognitive types from a single cross section of aggregate choice shares.



## References

- Abaluck, J., & Adams, A. (2018). What do consumers consider before they choose? Identification from asymmetric demand responses. *NBER Working Paper, No. 23566*.
- Ackerberg, D. A. (2009). A new use of importance sampling to reduce computational burden in simulation estimation. *Quantitative Marketing and Economics, 7*(4), 343–376.
- Apestegua, J., & Ballester, M. A. (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy, 126*(1), 74–106.
- Apestegua, J., Ballester, M. A., & Lu, J. (2017). Single-crossing random utility models. *Econometrica, 85*(2), 661–674.
- Athey, S. (2001). Single crossing properties and the existence of pure strategy equilibria in games of incomplete information. *Econometrica, 69*(4), 861–889.
- Attali, Y., & Bar-Hillel, M. (2003). Guess where: The position of correct answers in multiple-choice test items as a psychometric variable. *Journal of Educational Measurement, 40*(2), 109–128.
- Barseghyan, L., Coughlin, M., Molinari, F., & Teitelbaum, J. C. (2019). Heterogeneous choice sets and preferences. *CEMMAP Working Paper, CWP37/19*.
- Barseghyan, L., Molinari, F., O’Donoghue, T., & Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. *American Economic Review, 103*(6), 2499–2529.
- Barseghyan, L., Molinari, F., O’Donoghue, T., & Teitelbaum, J. C. (2018). Estimating risk preferences in the field. *Journal of Economic Literature, 56*(2), 501–564.
- Barseghyan, L., Molinari, F., & Teitelbaum, J. C. (2016). Inference under stability of risk preferences. *Quantitative Economics, 7*(2), 367–409.
- Barseghyan, L., Molinari, F., & Thirkettle, M. (2019). Discrete choice under risk with limited consideration. *CEMMAP Working Paper, CWP08/19*.
- Barseghyan, L., Prince, J., & Teitelbaum, J. C. (2011). Are risk preferences stable across contexts? Evidence from insurance data. *American Economic Review, 101*(2), 591–631.
- Barseghyan, L., Teitelbaum, J. C., & Xu, L. (2018). Different contexts, different risk preferences? *SRRN Working Paper, No. 3200657*.
- Ben-Akiva, M., & Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Marketing Research, 12*(1), 9–24.
- Bhargava, S., Loewenstein, G., & Sydnor, J. (2017). Choose to lose: Health plan choices from a menu with dominated options. *Quarterly Journal of Economics, 132*(3), 1319–1372.
- Boggs, P. T., & Tolle, J. W. (1995). Sequential quadratic programming. *Acta numerica, 4*, 1–51.
- Cattaneo, M., Ma, X., Masatlioglu, Y., & Suleymanov, E. (2019). A random attention model. *Journal of Political Economy*, forthcoming.
- Chiappori, P.-A., Salanié, B., Salanié, F., & Gandhi, A. (2019). From aggregate betting data to individual risk preferences. *Econometrica, 87*(1), 1–36.
- Christenfeld, N. (1995). Choices from identical options. *Psychological Science, 6*(1), 50–55.

- Cohen, A., & Einav, L. (2007). Estimating risk preferences from deductible choice. *American Economic Review*, *97*(3), 745-788.
- Coughlin, M. (2019). Insurance choice with non-monetary plan attributes: Limited consideration in Medicare Part D. *Working paper, Cornell University*.
- Crawford, G. S., Griffith, R., & Iaria, A. (2020). A survey of preference estimation with unobserved choice set heterogeneity. *Journal of Econometrics*, forthcoming.
- Dardanoni, V., Manzini, P., Mariotti, M., & Tyson, C. J. (2020). Inferring cognitive heterogeneity from aggregate choices. *Econometrica*, *88*(3), 1269-1296.
- Dayan, E., & Bar-Hillel, M. (2011). Nudge to nobesity II: Menu positions influence food orders. *Judgment and Decision making*, *6*(4), 333-342.
- De los Santos, B., Hortaçsu, A., & Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, *102*(6), 2955-2980.
- Draganska, M., & Klapper, D. (2011). Choice set heterogeneity and the role of advertising: An analysis with micro and macro data. *Journal of Marketing Research*, *48*(4), 653-669.
- Einav, L., Finkelstein, A., Pascu, I., & Cullen, M. R. (2012). How general are risk preferences? Choice under uncertainty in different domains. *American Economic Review*, *102*(6), 2606-2638.
- Gaynor, M., Propper, C., & Seiler, S. (2016). Free to choose? Reform, choice, and consideration sets in the english national health service. *American Economic Review*, *106*(11), 3521-57.
- Ghosal, S. (2001). Convergence rates for density estimation with bernstein polynomials. *Annals of Statistics*, *29*(5), 1264-1280.
- Goeree, M. S. (2008). Limited information and advertising in the U.S. personal computer industry. *Econometrica*, *76*(5), 1017-1074.
- Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, *103*(7), 2643-2682.
- Heiss, F., McFadden, D., Winter, J., Wuppermann, A., & Zhou, B. (2016). Inattention and switching costs as sources of inertia in medicare part D. *NBER Working Paper, No. 22765*.
- Ho, K., Hogan, J., & Scott Morton, F. (2017). The impact of consumer inattention on insurer pricing in the medicare part d program. *The RAND Journal of Economics*, *48*(4), 877-905.
- Honka, E., & Chintagunta, P. (2017). Simultaneous or sequential? Search strategies in the U.S. auto insurance industry. *Marketing Science*, *36*(1), 21-42.
- Honka, E., Hortaçsu, A., & Vitorino, M. A. (2017). Advertising, consumer awareness, and choice: Evidence from the U.S. banking industry. *RAND Journal of Economics*, *48*(3), 611-646.
- Hortaçsu, A., Madanizadeh, S. A., & Puller, S. L. (2017). Power to choose? An analysis of consumer inertia in the residential electricity market. *American Economic Journal: Economic Policy*, *9*(4), 192-226.
- Kawaguchi, K., Uetake, K., & Watanabe, Y. (2020). Designing context-based marketing: Product recommendations under time pressure. *SSRN Working Paper, No. 2529294*.

- Kim, J. B., Albuquerque, P., & Bronnenberg, B. J. (2010). Online demand under limited consumer search. *Marketing Science*, 29(6), 1001-1023.
- Kimya, M. (2018). Choice, consideration sets, and attribute filters. *American Economic Journal: Microeconomics*, 10(4), 223-47.
- Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics*, 97(1), 145 – 177.
- Lewbel, A. (2014). An overview of the special regressor method. In J. Racine, L. Su, & A. Ullah (Eds.), *Oxford handbook of applied nonparametric and semiparametric econometrics and statistics* (p. 38 - 62). Oxford University Press.
- Lewbel, A., & Yang, T. T. (2016). Identifying the average treatment effect in ordered treatment models without unconfoundedness. *Journal of Econometrics*, 195(1), 1–22.
- Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics*, 3(3), 205 – 228.
- Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3), 229–254.
- Manski, C. F. (2009). Identification for prediction and decision. *Harvard University Press*.
- Manzini, P., & Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, 82(3), 1153-1176.
- Matzkin, R. L. (2007). Nonparametric identification. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of econometrics* (Vol. 6, pp. 5307 – 5368). Elsevier.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (p. 105-142). New York: Academic Press.
- McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447-470.
- Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2), 175–208.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological review*, 84(3), 231.
- Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, 32(1/2), 122–136.
- Roberts, J. H., & Lattin, J. M. (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research*, 28(4), 429–440.
- Rubinstein, A., Tversky, A., & Heller, D. (1997). Naive strategies in competitive games. In *Understanding strategic interaction* (pp. 394–402). Springer.
- Schelling, T. C. (1980). *The strategy of conflict*. Harvard university press.
- Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, XXIX, 281-95.
- Spence, A. M. (1974). Market signaling: Informational transfer in hiring and related screening processes. *Harvard University Press*, 143.
- Sydnor, J. (2010). (Over)Insuring modest risks. *American Economic Journal: Applied*

- Economics*, 2(4), 177-199.
- Teigen, K. H. H. (1983). Studies in subjective probability 1: Prediction of random events. *Scandinavian Journal of Psychology*, 24(1), 13–25.
- Tversky, A. (1972). Choice by elimination. *Journal of Mathematical Psychology*, 9(4), 341 - 367.
- van Nierop, E., Bronnenberg, B., Paap, R., Wedel, M., & Franses, P. H. (2010). Retrieving unobserved consideration sets from household panel data. *Journal of Marketing Research*, 47(1), 63-74.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57(2), 307–333.
- Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. *Risk aversion in experiments*, 12, 197–292.

# Appendices

## A Proofs

*Proof of Fact 4.* If  $c_{1,2}(x)$  is less than  $c_{2,3}(x)$ , then, for any DM with preference  $\nu$  s.t.  $c_{1,2}(x) < \nu < c_{2,3}(x)$ ,  $d_2$  is preferred to both  $d_1$  and  $d_3$ , i.e. we are in Case (i). If  $c_{1,2}(x) > c_{2,3}(x)$ , then  $d_2$  is either dominated by either  $d_1$  or  $d_3$ . The relative location of  $c_{1,3}(x)$  is established as follows. First, suppose  $c_{1,3}(x) < c_{1,2}(x) < c_{2,3}(x)$ . For any  $\nu \in (c_{1,3}(x), c_{1,2}(x))$  we have  $U_\nu(d_3, x) > U_\nu(d_1, x) > U_\nu(d_2, x) > U_\nu(d_3, x)$ , which is an obvious contradiction. Second, suppose  $c_{2,3}(x) < c_{1,2}(x) < c_{1,3}(x)$ . Then, for any  $\nu \in (c_{1,2}(x), c_{1,3}(x))$  we have  $U_\nu(d_1, x) > U_\nu(d_3, x) > U_\nu(d_2, x) > U_\nu(d_1, x)$ , which is an obvious contradiction. The remaining two possibilities are excluded following the same logic.  $\square$

We maintain that  $x$  has strictly positive density on  $\mathcal{S}$  (Assumption T0), its density is continuous (Assumption T1), and that preferences are continuous and strictly monotone in  $x$ . Therefore, whenever  $x$  is scalar and  $\mathcal{X}$  is such that  $c_{1,2}(x)$  covers  $[\nu^l, \nu^u]$ , it is equivalent to there existing an interval  $[x^l, x^u] \subset \mathcal{S}$  such that  $[\nu^l, \nu^u] = \{c_{1,2}(x) : x \in [x^l, x^u]\}$ .

*Proof of Theorem 1.* The second condition in the theorem implies that  $c_{1,2}(x) < c_{1,j}(x)$  for any  $t \in [0, 1]$  and any  $j \neq 1, 2$ . Suppose that  $c_{1,j}(x) < c_{1,j+1}(x)$  for all  $j$  (the same argument can be applied for any order of the cutoffs provided that  $c_{1,2}(x)$  leads). Then

$$\frac{d\Pr(d = d_1|x)}{dx} = \sum_{\mathcal{K} \subset \mathcal{D}: 1, 2 \in \mathcal{K}} Q(\mathcal{K}) f(c_{1,2}(x)) \frac{dc_{1,2}(x)}{dx} + \sum_{j=3}^D \sum_{\substack{\mathcal{K} \subset \mathcal{D}: \\ 1, j \in \mathcal{K}, \\ 2, \dots, j-1 \notin \mathcal{K}}} Q(\mathcal{K}) f(c_{1,j}(x)) \frac{dc_{1,j}(x)}{dx},$$

where  $Q(\mathcal{K})$  is the probability that the DM draws the consideration set  $\mathcal{K}$ . Note that  $Q(\mathcal{K})$  does not depend on  $x$  nor  $\nu$  by Assumptions I2-I3. The second term in the equation above is zero by the first condition in the theorem. Therefore

$$\begin{aligned} \frac{d\Pr(d = d_1|x)}{dx} &= \left( \sum_{\mathcal{K} \subset \mathcal{D}: 1, 2 \in \mathcal{K}} Q(\mathcal{K}) \right) f(c_{1,2}(x)) \frac{dc_{1,2}(x)}{dx} \\ &\equiv \alpha f(c_{1,2}(x)) \frac{dc_{1,2}(x)}{dx}. \end{aligned}$$

Consequently, the product  $\alpha f(c_{1,2}(x))$  can be written in terms of data:

$$\alpha f(c_{1,2}(x)) = \frac{\frac{d \Pr(d=d_1|x)}{dx}}{\frac{dc_{1,2}(x)}{dx}},$$

and hence variation in  $x$  guarantees that  $f(\nu)$  is identified up-to-scale on  $[c_{1,2}(x^l), c_{1,2}(x^u)] = [\nu^l, \nu^u]$ . It follows that  $F(\nu|\nu \in [\nu^l, \nu^u])$  is identified. When there is large support variation,  $\alpha f(\nu)$  is identified for all  $\nu \in [0, \bar{\nu}]$ ; hence  $\alpha = \int_0^{\bar{\nu}} \alpha f(\nu) d\nu$  and  $F(\nu)$  are identified.  $\square$

*Proof of Theorem 2.* Under Condition 1 of the Theorem there can only be three types of consideration sets. The first type are all possible subsets of  $\{d_1, \dots, d_j\}$ ; the second type are all possible subsets of  $\{d_{j+1}, \dots, d_D\}$ . The third type necessarily contains both  $d_j$  and  $d_{j+1}$ . The probability of choosing an alternative in  $\{d_1, \dots, d_j\}$  is one for the first type and zero for the second type. Hence,

$$\frac{d \Pr(d \in \{d_1, \dots, d_j\} | x)}{dx} = \sum_{\mathcal{K} \subset \mathcal{D}: j, j+1 \in \mathcal{K}} Q(\mathcal{K}) f(c_{j,j+1}(x)) \frac{dc_{j,j+1}(x)}{dx}.$$

The rest of the proof follows the same steps as in the proof of Theorem 1, except we now track  $c_{j,j+1}(x)$ .  $\square$

*Proof of Theorem 3.* For the purpose of obtaining a contradiction, suppose that there is full consideration. Then

$$\begin{aligned} \Pr(d \in \mathcal{L} | x) &= \Pr\left(\arg \max_{j \in \mathcal{D}} U_\nu(d_j, x) \in \mathcal{L} \mid x\right) \\ &= \Pr(\nu \in [0, \nu^*]) \\ &= F(\nu^*) \\ &= \Pr\left(\arg \max_{j \in \mathcal{D}} U_\nu(d_j, x') \in \mathcal{L}' \mid x'\right) \\ &= \Pr(d \in \mathcal{L}' | x'). \end{aligned}$$

This is a contradiction. Therefore there is limited consideration.  $\square$

**Lemma 1.** *Suppose Assumptions T0-T2 and I1 hold. Suppose  $c_{1,2}(x) < c_{1,j}(x)$ . Let  $\{x^t\}_{t=1}^\infty$  be s.t.  $c_{1,2}(x^t) = c_{1,j}(x^{t+1})$ . Then  $\exists T < \infty$  s.t.  $c_{1,2}(x^T) \leq 0$ .*

*Proof.* The cutoff  $c_{1,2}(x^t)$  is a strictly declining sequence. Suppose it is bounded away from zero. Then it converges to some  $\nu^\infty > 0$  such that  $\nu^\infty = c_{1,2}(x^\infty) = c_{1,j}(x^\infty)$  for some  $x^\infty \in \mathcal{X}$ , a contradiction.  $\square$

*Proof of Theorem 4.* The second condition in the theorem implies that the cutoffs are ordered:  $c_{1,j}(x) < c_{1,j+1}(x)$  for all  $x \in \mathcal{X}$ . Hence

$$\begin{aligned} \Pr(d = d_1|x) &= \sum_{j=2}^D \sum_{\substack{\mathcal{K} \subset \mathcal{D}: \\ 1, j \in \mathcal{K}, \\ 2, \dots, j-1 \notin \mathcal{K}}} Q(\mathcal{K})F(c_{1,j}(x)) + Q(\{d_1\}) \\ &= \sum_{j=2}^D \mathcal{O}(\{d_1, d_j\}; \{d_2, \dots, d_{j-1}\})F(c_{1,j}(x)) + \mathcal{O}(d_1; \emptyset) \\ &\equiv \sum_{j=2}^D \Lambda_j F(c_{1,j}(x)) + \Lambda_1, \end{aligned}$$

so that

$$\frac{d \Pr(d = d_1|x)}{dx} = \sum_{j=2}^D \Lambda_j f(c_{1,j}(x)) \frac{dc_{1,j}(x)}{dx}.$$

By Assumption I1, we can set  $x^u = x^{\bar{v}}$  s.t.  $c_{1,2}(x^{\bar{v}}) = \bar{v}$  and similarly  $x^l = x^0$  s.t.  $c_{1,2}(x^0) = 0$ . It may be the case that  $c_{1,\hat{D}}(x^0) < \bar{v}$  and  $c_{1,\hat{D}+1}(x^0) > \bar{v}$  for some  $\hat{D} \geq 2$ . Then,  $\forall j > \hat{D}$ ,  $\Lambda_j$  does not enter the expression for the derivative of  $\Pr(d = d_1|x)$ ,  $\forall x \in [x^0, x^{\bar{v}}]$ , because  $f(c_{1,j}(x)) = 0$ . Henceforth, we only consider the relevant alternatives for the derivative of  $\Pr(d = d_1|x)$ , namely  $j \leq \hat{D}$ .

Next, consider the derivative of  $\Pr(d = d_j|x)$ . By Fact 3, the term  $\Lambda_j$  is the leading coefficient on  $f(\cdot)$  for this derivative. There exists  $x^j \in \mathcal{X}$  such that  $c_{1,j}(x^j) = \bar{v}$ . Thus,

$$\lim_{x \nearrow x^j} \frac{d \Pr(d = d_j|x)}{dx} = -\Lambda_j f(\bar{v}) \frac{dc_{1,j}(x^j)}{dx}, \quad \forall j : 2 \leq j \leq \hat{D}.$$

Write the expression above for  $d_j$  and  $d_2$  and take the ratio. This identifies  $\Omega_j \equiv \frac{\Lambda_j}{\Lambda_2}$ , where  $\Lambda_2 \neq 0$  by the first assumption in the theorem. Rewrite the derivative of  $\Pr(d = d_1|x)$  as

follows:

$$\begin{aligned}
\frac{d\Pr(d = d_1|x)}{dx} &= \sum_{j=2}^{\hat{D}} \Lambda_j f(c_{1,j}(x)) \frac{dc_{1,j}(x)}{dx} \\
&= \sum_{j=2}^{\hat{D}} \frac{\Lambda_j}{\Lambda_2} \Lambda_2 f(c_{1,j}(x)) \frac{dc_{1,j}(x)}{dx} \\
&= \sum_{j=2}^{\hat{D}} \Omega_j [\Lambda_2 f(c_{1,j}(x))] \frac{dc_{1,j}(x)}{dx} \\
&= \sum_{j=2}^{\hat{D}} \Omega_j \hat{f}(c_{1,j}(x)) \frac{dc_{1,j}(x)}{dx},
\end{aligned}$$

where  $\hat{f}(\nu) \equiv \Lambda_2 f(c_{1,j}(x))$ . Equipped with  $\Omega_j$ , we can recover  $\hat{f}(\nu)$  sequentially. Note that  $\forall x$  s.t.  $c_{1,2}(x) \leq \bar{\nu}$  and  $c_{1,3}(x) > \bar{\nu}$ , the up-to-scale density  $\hat{f}(c_{1,2}(x))$  is identified. Indeed, it is the only unknown in the expression above. We proceed as follows.

First, let  $x^1$  be such that  $c_{1,3}(x^1) = \bar{\nu}$ . Then,  $\hat{f}(\cdot)$  is identified on  $[c_{1,2}(x^1), \bar{\nu}]$ .

Second, let  $x^2$  be such that  $c_{1,2}(x^2) = c_{1,3}(x^1)$ . Now  $\hat{f}(\nu)$  is identified on  $[c_{1,2}(x^2), \bar{\nu}]$  because in the expression for the derivative of  $\Pr(d = d_1|x)$  all cutoffs  $c_{1,j}(x)$ ,  $j > 2$ , lie on the part of the support where the up-to-scale density is known.

Repeating the step above,  $\hat{f}(\nu)$  is identified on  $[0, \bar{\nu}]$ . Indeed, by Lemma 1 in a finite number of steps  $N$ ,  $c_{1,2}(x^N)$  reaches the lower end of the support. Finally, the scale is recovered by integrating  $\hat{f}(\nu)$  over its support:

$$\Lambda_2 = \Lambda_2 \int_0^{\bar{\nu}} f(\nu) d\nu = \int_0^{\bar{\nu}} \hat{f}(\nu) d\nu.$$

Therefore  $f(\cdot)$  is identified, as required. Note that  $\mathcal{O}(d_1; \emptyset) = \Lambda_1 = \Pr(d = d_1|x^{\bar{\nu}})$ . Hence  $\mathcal{O}(d_1; \emptyset)$  is identified, and so are  $\mathcal{O}(d_1, d_2; \emptyset) = \Lambda_2$  and  $\mathcal{O}(\{d_1, d_j\}; \{d_2, \dots, d_{j-1}\}) = \Lambda_j$ .

□

*Proof of Theorem 5.* The second condition of Theorem 4 implies that the cutoffs are ordered:  $c_{1,j}(x) < c_{1,j+1}(x)$  for all  $x \in \mathcal{X}$ . Hence,

$$\frac{d\Pr(d = d_1|x)}{dx} = \sum_{j=2}^D \Lambda_j(c_{1,j}(x)) f(c_{1,j}(x)) \frac{dc_{1,j}(x)}{dx},$$



where

$$\Lambda_j(\nu) = \begin{cases} \underline{\Lambda}_j \equiv \sum_{\substack{\mathcal{K} \subset \mathcal{D}: \\ 1, j \in \mathcal{K}, \\ 2, \dots, j-1 \notin \mathcal{K}}} \underline{\mathcal{Q}}(\mathcal{K}) & \text{if } \nu < \nu^* \\ \bar{\Lambda}_j \equiv \sum_{\substack{\mathcal{K} \subset \mathcal{D}: \\ 1, j \in \mathcal{K}, \\ 2, \dots, j-1 \notin \mathcal{K}}} \bar{\mathcal{Q}}(\mathcal{K}) & \text{if } \nu \geq \nu^* \end{cases}.$$

By Assumption I1, we can set  $x^u = x^{\bar{\nu}}$  s.t.  $c_{1,2}(x^{\bar{\nu}}) = \bar{\nu}$  and similarly  $x^l = x^0$  s.t.  $c_{1,2}(x^0) = 0$ . It may be the case that  $c_{1,\hat{D}}(x^0) < \bar{\nu}$  and  $c_{1,\hat{D}+1}(x^0) > \bar{\nu}$  for some  $\hat{D} \geq 2$ . Then,  $\forall j > \hat{D}$ ,  $\Lambda_j$  does not enter the expression for the derivative of  $\Pr(d = d_1|x) \forall x \in [x^0, x^{\bar{\nu}}]$ , because  $f(c_{1,j}(x)) = 0$ . Henceforth, we only consider the relevant alternatives for the derivative of  $\Pr(d = d_1|x)$ , namely  $j \leq \hat{D}$ .

We start at  $x^{\bar{\nu}}$  and hence  $c_{1,2}(x^{\bar{\nu}}) = \bar{\nu}$ . As we lower  $x$  we trace whether  $\frac{d\Pr(d=d_1|x)}{dx}$  jumps. If it does not, identification of  $f(\cdot)$  attains by the proof of Theorem 4.

Suppose there is a point of discontinuity. It arises when a cutoff  $c_{1,j}(x)$  crosses the breakpoint  $\nu^*$ . The identity of the cutoff and hence  $\nu^* = c_{1,j}(x)$  is identified by the fact there is a unique  $\frac{d\Pr(d=d_j|x)}{dx}$  that jumps also. Equipped with the identity of  $\nu^*$  the proof proceeds similarly to that of Theorem 4. Indeed, all  $\bar{\Omega}_j \equiv \bar{\Lambda}_j/\bar{\Lambda}_2$  are identified and so is  $\hat{f}(\nu) \equiv \bar{\Lambda}_2 f(\nu)$  for all  $\nu > \nu^*$ .

The additional step is how to identify  $\underline{\Lambda}_j$  and  $\underline{\Omega}_j \equiv \underline{\Lambda}_j/\underline{\Lambda}_2$ . Start with  $\underline{\Lambda}_2$ . Consider an  $x^*$  s.t.  $c_{1,2}(x^*) = \nu^*$ . The derivatives  $\frac{d\Pr(d=d_1|x)}{dx}$  from the left and from the right of  $x^*$  identify  $\underline{\Lambda}_2 f(\nu^*)$  and  $\bar{\Lambda}_2 f(\nu^*)$ . Hence, the ratio  $\underline{\Lambda}_2/\bar{\Lambda}_2$  is identified. Exactly the same logic applies to all other  $\underline{\Lambda}_j$ 's whenever  $c_{1,j}(x)$  crosses  $\nu^*$ . We can then rewrite

$$\frac{d\Pr(d = d_1|x)}{dx} = \sum_{j=2}^{\hat{D}} (\bar{\Omega}_j)^{\mathbf{1}_j(x)} \left( \frac{\underline{\Omega}_j \underline{\Lambda}_2}{\bar{\Lambda}_2} \right)^{1-\mathbf{1}_j(x)} \hat{f}(c_{1,j}(x)) \frac{dc_{1,j}(x)}{dx},$$

where  $\mathbf{1}_j(x)$  is the indicator function for the event  $\{c_{1,j}(x) > \nu^*\}$ . Now all coefficients of  $\hat{f}(c_{1,j}(x))$  are identified, and identification of  $\hat{f}(\cdot)$  proceeds to the left of  $\nu^*$ . Once it is identified, we integrate it over the support to recover  $\bar{\Lambda}_2$ . Hence  $\underline{\Lambda}_2$  and  $f(\cdot)$  are identified.  $\square$

*Proof of Theorem 6.* Let  $\nu, \tilde{x}, \mathcal{N}_\epsilon(\tilde{x}) \equiv \{x : \|x - \tilde{x}\| < \epsilon\}$  satisfy Condition 2 in the theorem. Then,  $\nu = c_{j,k}(\tilde{x})$  for all  $j, k$ . Consider any pair of alternatives  $(d_j, d_k)$  and set  $\mathcal{L} \subseteq \mathcal{D} \setminus \{j, k\}$ .

Since utility is strictly monotone in  $x_j$  and continuous, we can find  $x \in \mathcal{N}_\epsilon(\tilde{x})$  such that

$$U_\nu(d_j, x) = U_\nu(d_k, x); \quad (\text{A.1})$$

$$U_\nu(d_l, x) > U_\nu(d_j, x) \quad \forall l \in \mathcal{L}; \quad (\text{A.2})$$

$$U_\nu(d_j, x) > U_\nu(d_l, x) \quad \forall l \in \mathcal{D} \setminus \{\mathcal{L} \cup \{j, k\}\}; \quad (\text{A.3})$$

and  $c_{j,k}(x) = \nu$ . The remainder of the proof proceeds in two steps.

*Step 1: Identification of  $f(\nu)\mathcal{Q}_\nu(\mathcal{K})$ :* The singleton sets occur with zero probability by Condition 1 in the theorem, so it remains to show identification for consideration sets larger than one. Consider any two alternatives  $(d_j, d_k)$ . We claim that the following statement holds for  $n = 0, \dots, D$ :

$P(n)$ : For all  $\mathcal{K} \subset \mathcal{D} \setminus \{j, k\}$  satisfying  $|\mathcal{K}| \leq n$ , the quantity  $f(\nu)\mathcal{Q}_\nu(\{j, k\} \cup \mathcal{K})$  is identified.

To show this for  $P(0)$ , set  $\mathcal{L} = \mathcal{D} \setminus \{j, k\}$ . In this case  $\mathcal{K} = \emptyset$ . Let  $x$  satisfy Equations (A.1)-(A.3). Then, all alternatives  $d_l$ ,  $l \neq j, k$ , are preferred to  $d_j$  and  $d_k$  at  $\nu$  and  $c_{j,k}(x) = \nu$ . Hence:

$$\frac{\frac{\partial \Pr(d=d_j|x)}{\partial x_k}}{\frac{\partial c_{j,k}(x)}{\partial x_k}} = f(\nu)\mathcal{Q}_\nu(\{j, k\}).$$

It follows that  $f(\nu)\mathcal{Q}_\nu(\{j, k\})$  is identified.

Next, suppose  $P(n-1)$  is true. Consider any  $\mathcal{K} \subset \mathcal{D} \setminus \{j, k\}$  such that  $|\mathcal{K}| = n$ . Let  $\mathcal{L} = \mathcal{D} \setminus (\mathcal{K} \cup \{j, k\})$ . Let  $x$  satisfy Equations (A.1)-(A.3). Then,

$$\begin{aligned} \frac{\frac{\partial \Pr(d=d_j|x)}{\partial x_k}}{\frac{\partial c_{j,k}(x)}{\partial x_k}} &= f(\nu) \sum_{\mathcal{C} \subset \mathcal{K}} \mathcal{Q}_\nu(\{j, k\} \cup \mathcal{C}) \\ &= f(\nu)\mathcal{Q}_\nu(\{j, k\} \cup \mathcal{K}) + \sum_{\mathcal{C} \subset \mathcal{K}: |\mathcal{C}| < n} f(\nu)\mathcal{Q}_\nu(\{j, k\} \cup \mathcal{C}). \end{aligned}$$

The LHS of this expression is known, and the second term on the RHS is identified by the induction step. Therefore  $P(n)$  holds.

Since  $d_j$  and  $d_k$  were chosen arbitrarily, it follows that  $f(\nu)\mathcal{Q}_\nu(\mathcal{K})$  is identified for all  $\mathcal{K} \subset \mathcal{D}$ .

Step 2: Identification of  $f(\nu)$  and  $\mathcal{Q}_\nu(\mathcal{K})$ . Since

$$\sum_{\mathcal{K} \subset \mathcal{D}} f(\nu) \mathcal{Q}_\nu(\mathcal{K}) = f(\nu) \sum_{\mathcal{K} \subset \mathcal{D}} \mathcal{Q}_\nu(\mathcal{K}) = f(\nu),$$

$f(\nu)$  and  $\mathcal{Q}_\nu(\mathcal{K})$  are identified. □

*Proof of Corollary 1.* The proof follows the same steps as the proof of Theorem 6, but with the following two modifications:

First modification: In Step 1 in the proof of Theorem 6, we start with  $d_j = d_1$  and loop over  $d_k \in \{d_2, \dots, d_D\}$ . This ensures that we only take derivatives with respect to  $x_k$ ,  $k > 1$ . Hence,  $f(\nu) \mathcal{Q}_\nu^{x_1}(\mathcal{K})$  is identified for all sets  $\mathcal{K} \subset \mathcal{D} : |\mathcal{K}| > 1$ .

Second modification: In Step 2 we obtain

$$f(\nu) \mathcal{Q}^{x_1}(\{1\}) + \sum_{\mathcal{K} \subset \mathcal{D}: |\mathcal{K}| > 1} f(\nu) \mathcal{Q}_\nu^{x_1}(\mathcal{K}) = f(\nu) \sum_{\mathcal{K} \subset \mathcal{D}} \mathcal{Q}_\nu^{x_1}(\mathcal{K}) = f(\nu).$$

Since the first term on the LHS is unknown,  $f(\nu)(1 - \mathcal{Q}^{x_1}(\{1\}))$  is identified for all  $\nu \in [0, \bar{\nu}]$ . The scale is identified, because

$$(1 - \mathcal{Q}^{x_1}(\{1\})) = \int_0^{\bar{\nu}} f(\nu)(1 - \mathcal{Q}^{x_1}(\{1\})) d\nu.$$

Once the scale is identified,  $f(\nu)$  is identified and so are  $\mathcal{Q}_\nu^{x_1}(\mathcal{K})$ ,  $\forall \mathcal{K} \subset \mathcal{D}$ . □

**Lemma 2.** Consider the Basic ARC model. For any  $\mathcal{K} \subset \mathcal{D}$ ,  $\sum_{j \in \mathcal{K}} \Pr(d = d_j | x)$  is increasing in  $\varphi_j$ ,  $\forall j \in \mathcal{K}$ , and decreasing in  $\varphi_j$ ,  $\forall j \notin \mathcal{K}$ .

*Proof.* Fix  $\mathcal{K}$  and consider any  $j \in \mathcal{K}$ . For each  $\nu$  and  $l \in \mathcal{K}$ ,  $l \neq j$ , either  $j \in \mathcal{B}_\nu(d_l, x)$  or not. If  $j \notin \mathcal{B}_\nu(d_l, x)$ , then  $\Pr(d = d_l | x, \nu)$  does not depend on  $\varphi_j$ . Hence,

$$\sum_{l \in \mathcal{K}} \Pr(d = d_l | x, \nu) = A + \Pr(d = d_j | x, \nu) + \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \Pr(d = d_l | x, \nu),$$

where  $A$  is a constant that collects terms that do not depend on  $\varphi_j$ . Continuing,

$$\begin{aligned}
\sum_{l \in \mathcal{K}} \Pr(d = d_l | x, \nu) &= A + \varphi_j \prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) + \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \varphi_l \prod_{k \in \mathcal{B}_\nu(d_l, x)} (1 - \varphi_k) \\
&= A + \varphi_j \prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) + \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \varphi_l (1 - \varphi_j) \prod_{k \in \mathcal{B}_\nu(d_l, x) \setminus \{j\}} (1 - \varphi_k) \\
&= A + \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \left( \varphi_l \prod_{k \in \mathcal{B}_\nu(d_l, x) \setminus \{j\}} (1 - \varphi_k) \right) \\
&\quad + \left( \prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) - \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \varphi_l \prod_{k \in \mathcal{B}_\nu(d_l, x) \setminus \{j\}} (1 - \varphi_k) \right) \varphi_j \\
&\equiv \tilde{A} + B\varphi_j.
\end{aligned}$$

Since  $\mathcal{B}_\nu(j, x) \subset \mathcal{B}_\nu(d_l, x)$  whenever  $j \in \mathcal{B}_\nu(d_l, x)$ ,

$$B = \prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) \left( 1 - \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \varphi_l \prod_{k \in \mathcal{B}_\nu(d_l, x) \setminus \{\mathcal{B}_\nu(d_j, x), j\}} (1 - \varphi_k) \right) \geq 0.$$

Therefore,  $\sum_{j \in \mathcal{K}} \Pr(d = d_j | x) = \int \sum_{j \in \mathcal{K}} \Pr(d = d_j | x, \nu) dF$  is increasing in  $\varphi_j$ .

Finally, for any  $j \notin \mathcal{K}$ ,  $\varphi_j$  only ever enters as  $(1 - \varphi_j)$  in the sum; hence it is decreasing in  $\varphi_j$  for those alternatives.  $\square$

**Lemma 3.** *Consider the Basic ARC model. For any  $\mathcal{K} \subset \mathcal{D}$ ,  $\sum_{j \in \mathcal{K}} \Pr(d = d_j | x)$  is strictly increasing in  $\varphi_j$ ,  $j \in \mathcal{K}$ , whenever there is a  $\nu$  at which alternative  $d_j$  is preferred to all of the always-considered alternatives. It is strictly decreasing in  $\varphi_j$ ,  $j \notin \mathcal{K}$ , whenever there is a  $\nu$  and  $l \in \mathcal{K}$  such that at this  $\nu$  alternative  $d_j$  is preferred to  $d_l$  and  $d_l$  is preferred to all of the always-considered alternatives.*

*Proof.* To show the first claim, notice that  $B = 0$  in the proof of Lemma 2 if and only if  $\varphi_k = 1$  for some  $k \in \mathcal{B}_\nu(d_j, x)$ .

To show the second claim, consider any  $l \in \mathcal{K}$ . Then,

$$\Pr(d = d_l | x, \nu) = \varphi_l \prod_{k \in \mathcal{B}_\nu(d_l, x)} (1 - \varphi_k).$$

For this to be strictly decreasing in  $\varphi_j$ , it must be the case that  $j \in \mathcal{B}_\nu(d_l, x)$  and  $\varphi_k < 1$  for

all  $k \in \mathcal{B}_\nu(d_l, x) \setminus \{j\}$ . □

*Proof of Theorem 7.* By the proof of Theorem 4,  $f(\cdot)$  and  $\Lambda_2 = \varphi_1\varphi_2$  are identified. The consideration parameter  $\varphi_1$  is identified by  $\Pr(d = d_1|x^{\bar{\nu}}) = \varphi_1$ , where  $x^{\bar{\nu}}$  is s.t.  $c_{1,2}(x^{\bar{\nu}}) = \bar{\nu}$ . Since  $\Lambda_2$  is known,  $\varphi_2$  is also identified. The rest of the proof is about identification of the remaining consideration parameters.

To identify  $\varphi_j$  take an  $x$  such that  $\Pr(d = d_j|x) \neq 0$ . Denote  $\mathcal{E} = \{k : \Pr(d = d_k|x) \neq 0\}$ . We claim that  $\Pr(d = d_k|x), \forall k \in \mathcal{E}$ , does not depend on  $\varphi_l$  s.t.  $l \notin \mathcal{E}$ . Suppose otherwise. That is, suppose there exists  $d_l$  such that  $\Pr(d = d_l|x) = 0$  and  $\Pr(d = d_k|x)$  depends on  $\varphi_l$  for some  $k \in \mathcal{E}$ . Then, for each  $\nu$  there is an always-considered alternative that is preferred to  $d_l$ . Since  $\Pr(d = d_k|x)$  depends on  $\varphi_l$ , there exists  $\nu \in [0, \bar{\nu}]$  such that  $d_l$  is preferred to  $d_k$ . However, the always-considered alternative that is preferred to  $d_l$  at  $\nu$  is also preferred to  $d_k$  by transitivity. This leads to a contradiction, because a DM with such preferences will never choose  $d_k$  in the first place. Therefore,  $\Pr(d = d_k|x)$  does not on  $\varphi_l$  for any  $l \notin \mathcal{E}$ .

Since  $F(\cdot)$  is already identified,  $\{\Pr(d = d_k|x)\}_{k \in \mathcal{E}}$  defines a system of  $|\mathcal{E}|$  non-linear equations, where the only unknowns are  $\varphi_k, k \in \mathcal{E}$ . This system has a unique solution. Suppose to the contrary that two sets of consideration parameters  $\{\varphi_k\}_{k \in \mathcal{E}}$  and  $\{\varphi'_k\}_{k \in \mathcal{E}}$  solve this system and they are distinct. Denote  $\mathcal{E}_+ = \{k : \varphi_k > \varphi'_k\}$ . By Lemma 3,  $\sum_{k \in \mathcal{E}_+} \Pr(d = d_k|x)$  is strictly larger at  $\{\varphi_k\}_{k \in \mathcal{E}}$  than at  $\{\varphi'_k\}_{k \in \mathcal{E}}$ . Hence, only one of these sets could satisfy data. Therefore there is a unique set of  $\{\varphi_k\}_{k \in \mathcal{E}}$  that solves this system of equations, and  $\varphi_j$  is identified as claimed. □

*Proof of Theorem 8. Part 1.*

The breakpoint  $\nu^*$  is identified by the argument in the proof of Theorem 5 and so is the preference distribution. We identify  $d^*$  at  $x^{\bar{\nu}}$ . The smallest  $j$  such that  $\Pr(d = d_j|x^{\bar{\nu}}) = 0$  yields  $d^* = j - 1$ . If such a  $j$  does not exist,  $d^* = D$ .

Case 1. Suppose  $d^* = 1$ . Then all  $\underline{\Lambda}_j$ 's and  $\bar{\Lambda}_j$ 's are identified by the argument in the proof of Theorem 5, as  $\hat{D}$  of that proof equals  $D$ . One then recovers  $\underline{\varphi}_j$ 's and  $\bar{\varphi}_j$ 's recursively from  $\underline{\Lambda}_j$ 's and  $\bar{\Lambda}_j$ 's.

Case 2. Suppose  $d^* = D$ . Then identification attains as in Step 2, observing that  $\underline{\varphi}_1 = \sum_{j>1} \underline{\Lambda}_j$  and  $\bar{\varphi}_1 = \sum_{j>1} \bar{\Lambda}_j$ .

Case 3. Suppose  $1 < d^* < D$ . Then  $\underline{\varphi}_j$ 's and  $\bar{\varphi}_j$ 's are identified for all  $j < d^*$  by the same argument as in Step 3. In addition,  $\bar{\varphi}_j$ 's for  $j > d^*$  are identified by the argument

in the proof of Theorem 7. It remains to identify  $\varphi_j$ ,  $j > d^*$ . Take  $x^0$  such that all  $c_{j,k}(x^0) \leq 0$ . Note that  $\Pr(d = d_D|x^0) = \underline{\varphi}_D F(\nu^*) + \overline{\varphi}_D(1 - F(\nu^*))$ , which identifies  $\underline{\varphi}_D$ . Next,  $\Pr(d = d_{D-1}|x^0) = \underline{\varphi}_{D-1}(1 - \underline{\varphi}_D)F(\nu^*) + \overline{\varphi}_{D-1}(1 - \overline{\varphi}_D)(1 - F(\nu^*))$ , which identifies  $\underline{\varphi}_{D-1}$ . This iteration is repeated for  $j = D - 2, \dots, d^* + 1$ .

**Part 2.**

Step 1. Suppose  $D = 3$ . Then proof is complete by the argument developed in Section 3.2 of the paper.

Step 2. Let  $D > 3$ . We identify  $d^*$  at  $x^{\bar{\nu}}$ . The smallest  $j$  such that  $\Pr(d = d_j|x^{\bar{\nu}}) = 0$  yields  $d^* = j - 1$ . We return to the case that  $d^* = 1$  or  $d^* = D$  at the end of this proof. Since  $D > 3$  w.l.o.g. we assume that  $d^* > 2$ . Otherwise, we just relabel alternatives from ascending to descending order and proceed with the analysis starting from the lower part of the support.

Step 3. Using large support we establish that  $\varphi_D$  is a decreasing function of  $\varphi_1$ . We have

$$\Pr(d = d_1|x^{\bar{\nu}}) = \varphi_1 \int_0^{\bar{\nu}} (1 - \alpha(\nu))dF(\nu) = \varphi_1(1 - E\alpha).$$

Similarly,

$$\Pr(d = d_D|x^0) = \varphi_D \int_0^{\bar{\nu}} (1 + \alpha(\nu))dF(\nu) = \varphi_D(1 + E\alpha).$$

Hence

$$\varphi_1 = \frac{\Pr(d = d_1|x^1)}{2 - \frac{\Pr(d=d_D|x^0)}{\varphi_D}}.$$

Step 4. This is an intermediate step, which we use later in the proof. By Fact 3,

$$c_{1,j}(x) < c_j^*(x) \equiv \min_k \{ \{c_{k,j}(x)\}_{1 < k < j}, \{c_{j,k}(x)\}_{j < k \leq D} \}, \quad \forall j.$$

Moreover any sequence  $\{x^s\}_{s=1}^\infty$  such that  $c_j^*(x^s) = c_{1,j}(x^{s+1})$  will reach the lower bound of the support in finite number of steps. Otherwise, by the argument in the proof of Lemma 1,  $c_j^*(x^s)$  and  $c_{1,j}(x^s)$  converge to the same point in the interior of the support, which contradicts the assumptions of the theorem.

Step 5. Identification of  $\{\varphi_j\}_{1 < j < d^*}$ . For each  $j$ , there is an  $x^j$  such that  $c_{1,j}(x^j) = \bar{\nu}$ . It follows by Step 4 that the following equations hold:

$$\begin{aligned}
& - \lim_{x \nearrow x^2} \frac{d \Pr(d = d_2 | x)}{dx} / \frac{dc_{1,2}(x)}{dx} = \varphi_1(\bar{v}) \varphi_2(\bar{v}) f(\bar{v}) \\
& - \lim_{x \nearrow x^3} \frac{d \Pr(d = d_3 | x)}{dx} / \frac{dc_{1,3}(x)}{dx} = \varphi_1(\bar{v}) (1 - \varphi_2(\bar{v})) \varphi_3(\bar{v}) f(\bar{v}) \\
& - \lim_{x \nearrow x^4} \frac{d \Pr(d = d_4 | x)}{dx} / \frac{dc_{1,4}(x)}{dx} = \varphi_1(\bar{v}) (1 - \varphi_2(\bar{v})) (1 - \varphi_3(\bar{v})) \varphi_4(\bar{v}) f(\bar{v}) \\
& \quad \quad \quad \vdots \\
& - \lim_{x \nearrow x^{d^*}} \frac{d \Pr(d = d_{d^*} | x)}{dx} / \frac{dc_{1,d^*}(x)}{dx} = \varphi_1(\bar{v}) (1 - \varphi_2(\bar{v})) (1 - \varphi_3(\bar{v})) \dots (1 - \varphi_{d^*-1}(\bar{v})) f(\bar{v})
\end{aligned}$$

The summation of these expressions recovers the quantity  $\varphi_1(\bar{v})f(\bar{v})$ . Next, substitute  $\varphi_1(\bar{v})f(\bar{v})$  back to the expressions above to sequentially recover  $\{\varphi_j(\bar{v})\}_{2 \leq j \leq d^*}$ . Since  $\varphi_j(\bar{v}) = \varphi_j(1 - \alpha(\bar{v})) = \varphi_j$ ,  $\{\varphi_j\}_{2 \leq j \leq d^*}$  are identified.

Step 6: Identification of  $\varphi_1$  and  $\{\varphi_j\}_{d^* < j \leq D}$ . We now continue with identification of  $\varphi_j$  for  $j > d^*$  and the consideration parameter for the first alternative. The cutoffs are monotone in  $x^t$  and all cutoffs are on the right of  $\bar{v}$  at  $x^{\bar{v}}$ . Consequently,  $\Pr(d = d_j | x^{\bar{v}}) = 0$  for all  $j > d^*$ . Continuously decrease  $t$  until  $\Pr(d = d_{j_1} | x^t) > 0$  for some  $j_1 \in \mathcal{J} \equiv \{d^* + 1, \dots, D\}$  and  $\Pr(d = d_k | x^t) = 0$  for all  $k \in \mathcal{J} - \{j_1\}$ . This will happen when  $c_{d^*,j_1}(x^t)$  crosses  $\bar{v}$ , yielding

$$\lim_{x \nearrow x^t} \frac{d \Pr(d = d_{j_1} | x^t)}{dx} / \frac{dc_{d^*,j_1}(x)}{dx} = -\varphi_{j_1} f(\bar{v}) M_1,$$

where

$$M_1 \equiv \prod_{k \in \{2, \dots, d^*-1\} : c_{k,j_1}(x^t) > \bar{v}} (1 - \varphi_k).$$

Note that  $M_1$  is known, since all relevant  $\varphi_k$ 's are known. Importantly,  $M_1$  does not depend on  $\varphi_1$ , since  $c_{1,j_1}(x^t) < c_{d^*,j_1}(x^t) = \bar{v}$ .

Next, continuously decrease  $t$  further until  $\Pr(d = d_{j_2} | x^t) > 0$  for some  $j_2 \in \mathcal{J} - \{j_1\}$  and  $\Pr(d = d_k | x^t) = 0$  for all  $k \in \mathcal{J} - \{j_1, j_2\}$ . Again, this will happen when  $c_{d^*,j_2}(x)$  crosses  $\bar{v}$ . Hence

$$\lim_{x \nearrow x^t} \frac{d \Pr(d = d_{j_2} | x^t)}{dx} / \frac{dc_{d^*,j_2}(x)}{dx} = -\varphi_{j_2} f(\bar{v}) M_2,$$

$$M_2 \equiv \prod_{k \in \{2, \dots, d^*-1, j_1\} : c_{k,j_2}(x^t) > \bar{v}} (1 - \varphi_k).$$

The term  $M_2$  is known, except possibly for the term  $(1 - \varphi_{j_1})$ , since all other relevant  $\varphi_k$ 's

are known.

The expression above defines  $\varphi_{j_2}$  as a strictly increasing function of  $\varphi_{j_1}$  regardless of whether  $M_2$  depends on  $\varphi_{j_1}$  or not. Indeed, for the case where  $j_1 < j_2$  we have

$$\varphi_{j_2} \propto \begin{cases} \varphi_{j_1} & \text{if } c_{j_1, j_2}(x^t) < \bar{\nu} \\ \frac{\varphi_{j_1}}{1 - \varphi_{j_1}} & \text{if } c_{j_1, j_2}(x^t) \geq \bar{\nu} \end{cases},$$

where the coefficients of proportionality are known. A similar expression holds when  $j_1 > j_2$ . This argument immediately extends to all  $j \in \mathcal{J}$ . In particular, we have for the case where  $j_2 < j_3$

$$\varphi_{j_3} \propto \begin{cases} \varphi_{j_2} & \text{if } c_{j_2, j_3}(x^t) < \bar{\nu} \\ \frac{\varphi_{j_2}}{1 - \varphi_{j_2}} & \text{if } c_{j_2, j_3}(x^t) \geq \bar{\nu} \end{cases}.$$

By assumption,  $Pr(d = d_D | x^0) \neq 0$ ; hence,  $\varphi_D$  is an increasing function of  $\varphi_{j_1}$ . In turn, recall that  $\varphi_1 f(\bar{\nu})$  is known. The ratio of  $\varphi_1 f(\bar{\nu})$  and  $\varphi_{j_1} f(\bar{\nu})$ , which is also known, yields  $\varphi_D$  as an increasing function of  $\varphi_1$ . Hence, taken with the result in Step 3, the quantity  $\varphi_1$  is uniquely pinned down. Identification of all other  $\varphi_j$ 's immediately follow.

Step 7: Identification of  $\alpha(\nu)$  and  $f(\nu)$ . The identification argument is iterative. For each alternative  $j$ , define

$$\Gamma_j^0 \equiv \{\nu \in [0, \bar{\nu}] : \exists x \in \mathcal{X} \text{ s.t. } \nu = c_{1,j}(x) \text{ and } c_j^*(x) \geq \bar{\nu}\}.$$

The set  $\Gamma_j^0$  includes all preference parameters  $\nu$  covered by cutoff,  $c_{1,j}(\cdot)$ , before any other relevant cutoffs for  $d_j$  enter the support. Let  $\Gamma^0 \equiv \bigcap_{j=1}^D \Gamma_j^0$ . By Step 4,  $\Gamma^0 \neq \emptyset$  and  $\bar{\nu} \in \Gamma^0$ . For each  $\nu \in \Gamma^0$  and each  $d_j$ , there is an  $x^j \in \mathcal{X}$  such that  $c_{1,j}(x^j) = \nu$ . As a result, the following system of equations hold for each  $\nu \in \Gamma^0$ :



$$\begin{aligned}
-\frac{d \Pr(d = d_2|x^2)}{dx} / \frac{dc_{1,2}(x^2)}{dx} &= \varphi_1(\nu)\varphi_2(\nu)f(\nu) \\
-\frac{d \Pr(d = d_3|x^3)}{dx} / \frac{dc_{1,3}(x^3)}{dx} &= \varphi_1(\nu)(1 - \varphi_2(\nu))\varphi_3(\nu)f(\nu) \\
-\frac{d \Pr(d = d_4|x^4)}{dx} / \frac{dc_{1,4}(x^4)}{dx} &= \varphi_1(\nu)(1 - \varphi_2(\nu))(1 - \varphi_3(\nu))\varphi_4(\nu)f(\nu) \\
&\vdots \\
-\frac{d \Pr(d = d_{d^*}|x^{d^*})}{dx} / \frac{dc_{1,d^*}(x^{d^*})}{dx} &= \varphi_1(\nu)(1 - \varphi_2(\nu))(1 - \varphi_3(\nu)) \dots (1 - \varphi_{d^*-1}(\nu))f(\nu),
\end{aligned} \tag{A.4}$$

The summation of these expressions recovers the quantity  $\varphi_1(\nu)f(\nu)$ . Substitute this into the first equation to obtain  $\varphi_2(\nu) = \varphi_2(1 - \alpha(\nu))$ . But  $\varphi_2$  is already known, so  $\alpha(\nu)$  is identified on  $\Gamma^0$ . Finally, since  $\varphi_1(\nu) = \varphi_1(1 - \alpha(\nu))$  is now identified, so is  $f(\nu)$  on  $\Gamma^0$ .

In the next step of the iteration, let  $\bar{\nu}^1 = \min_{\nu \in \Gamma^0} \Gamma^0$  be the smallest value of  $\nu$  where  $\alpha(\nu)$  and  $f(\nu)$  are identified. Define

$$\Gamma_j^1 \equiv \{\nu \in [0, \bar{\nu}] : \exists x \in \mathcal{X} \text{ s.t. } \nu = c_{1,j}(x) \text{ and } c_j^*(x) \geq \bar{\nu}^1\} \quad \text{and} \quad \Gamma^1 \equiv \bigcap_{j=1}^D \Gamma_j^1.$$

Then, a similar system to (A.4) holds  $\forall \nu \in \Gamma^1$ , but may include additional terms. These terms are known, because they are functions of  $f(\cdot)$  and  $\alpha(\cdot)$  evaluated at  $\nu \in \Gamma^0$  (and also of  $\{\varphi_j\}_{j=1}^D$ ). We can therefore repeat the argument from the base case to establish that  $\alpha(\nu)$  and  $f(\nu)$  are identified on  $\Gamma^1$ . We repeat this iterative procedure. After a finite number of steps  $N$ , we obtain  $\Gamma^N = [0, \bar{\nu}]$  by Step 4; hence,  $f(\cdot)$  and  $\alpha(\cdot)$  are identified.

Edge Case ( $d^* = 1$  or  $d^* = D$ ): Suppose that  $d^* = D$  (the case when  $d^* = 1$  is symmetric). From Step 5 we obtain  $\varphi_j$  for  $j : 1 < j < D$ . Next, we show how to identify  $\varphi_1$ . We can find

$x^j$  such that  $c_{j,D}(x^j) = 0$ , and so the following system of equations holds:

$$\begin{aligned}
- \lim_{x \searrow x^{D-1}} \frac{d \Pr(d = d_{D-1}|x)}{dx} / \frac{dc_{D-1,D}(x)}{dx} &= \varphi_{D-1}(0) f(0) \\
- \lim_{x \searrow x^{D-3}} \frac{d \Pr(d = d_{D-2}|x)}{dx} / \frac{dc_{D-2,D}(x)}{dx} &= (1 - \varphi_{D-1}(0)) \varphi_{D-2}(0) f(0) \\
- \lim_{x \searrow x^{D-4}} \frac{d \Pr(d = d_{D-3}|x)}{dx} / \frac{dc_{D-3,D}(x)}{dx} &= (1 - \varphi_{D-1}(0))(1 - \varphi_{D-2}(0)) \varphi_{D-3}(0) f(0) \\
&\vdots \\
- \lim_{x \searrow x^1} \frac{d \Pr(d = d_1|x)}{dx} / \frac{dc_{1,D}(x)}{dx} &= (1 - \varphi_{D-1}(0))(1 - \varphi_{D-2}(0)) \dots (1 - \varphi_2(0)) \varphi_1(0) f(0).
\end{aligned}$$

The ratio of the first two questions yield

$$A = \frac{\varphi_{D-2}}{\varphi_{D-1}} \cdot (1 - \varphi_{D-1}(1 + \alpha(0))),$$

where  $A$ ,  $\varphi_{D-2}$ , and  $\varphi_{D-1}$  are known terms; hence,  $\alpha(0)$  is identified. Once  $\alpha(0)$  is identified, the term  $\varphi_1$  is identified from the ratio of the first and last equations in the above system. Finally,  $f(\nu)$  and  $\alpha(\nu)$  are identified by Step 7.  $\square$

*Proof of Theorem 9.* Let  $\nu$ ,  $x$ ,  $\mathcal{N}_\epsilon(x) = \{x' : \|x' - x\| < \epsilon\}$  satisfy the conditions in the theorem. Then  $\nu = c_{j,k}(x)$  for all  $j, k$ . For any pair of alternatives  $(d_j, d_k)$  we can perturb  $x_k$ ,  $k \notin \{j, d^*\}$ , and  $x_l$ ,  $\forall l \notin \{j, d^*, k\}$ , so that the resulting  $x' \in \mathcal{N}_\epsilon(x)$  is such that

$$\begin{aligned}
U_\nu(d_k, x'_k) &> U_\nu(d_j, x_j); \\
U_\nu(d_j, x_j) &> U_\nu(d_l, x'_l) \quad \forall l \in \mathcal{D} \setminus \{j, k, d^*\}.
\end{aligned}$$

And we can do another perturbation of  $x_l$ ,  $\forall l \notin \{j, d^*\}$ , so that the resulting  $x'' \in \mathcal{N}_\epsilon(\tilde{x})$  is such that

$$U_\nu(d_j, x_j) > U_\nu(d_l, x''_l) \quad \forall l \in \mathcal{D} \setminus \{j, d^*\}.$$

We then have

$$\begin{aligned}
\frac{\partial \Pr(d = d_j|x')}{\partial x_{d^*}} &= \varphi_j(x_j, \nu)(1 - \varphi_k(x'_k, \nu)) f(\nu) \frac{\partial c_{j,d^*}(x)}{\partial x_{d^*}}; \\
\frac{\partial \Pr(d = d_j|x'')}{\partial x_{d^*}} &= \varphi_j(x_j, \nu) f(\nu) \frac{\partial c_{j,d^*}(x)}{\partial x_{d^*}}.
\end{aligned}$$

Taking the ratio of the expressions above identifies  $(1 - \varphi_k(x'_k, \nu))$ . By continuity we have  $\varphi_k(x_k, \nu)$ . Identical steps identify  $\varphi_j(x_j, \nu)$ ,  $\forall j \neq d^*$ , and hence  $f(\nu)$ . □

*Proof of Proposition 10.* Take any non-empty consideration set  $\mathcal{K}$ . For a given preference coefficient  $\nu$ , let  $j_{\mathcal{K}}(x, \nu)$  denote the identity of the best alternative in this consideration set. By the natural ordering,  $j_{\mathcal{K}}(x, \nu)$  is an increasing step function in  $\nu$ . Hence,  $I(j_{\mathcal{K}}(x, \nu) \leq J)$  is a decreasing step function. The term  $\Pr\left(\bigcup_{j=1}^J d_j \mid x, \nu\right)$  is a non-negatively weighted sum of  $I(j_{\mathcal{K}}(x, \nu) \leq J)$ . Hence it is decreasing in  $\nu$ . □

*Proof of Proposition 11.* DMs' choices are driven by their preference parameter  $\nu$ , their preferences  $U_{\nu}(\cdot)$ , and the consideration set that they face. Obviously, the consideration process and the preference parameter is unaffected by the scaling of  $U_{\nu}(\cdot)$ . Scaling  $U_{\nu}(\cdot)$  by a common positive factor also leaves the preference order unchanged. Therefore, choice frequencies and other predictions of the model are invariant to the scale of  $U_{\nu}(\cdot)$ . □

*Proof of Proposition 12.* Consider the basic ARC model with preferences  $U_{\nu}(d_j, x)$ . The optimal choice from  $\mathcal{D}$  conditional on the DM facing the consideration set  $\mathcal{K} \neq \emptyset$  is the alternative with the largest value of  $U_{\nu}(d_j, x)$  subject to  $j \in \mathcal{K}$ . This is the same solution as the one that maximizes  $V_{\nu}(d_j, x, \epsilon_j)$  where  $\epsilon_j = 0$  for all  $j \in \mathcal{K}$  and  $\epsilon_j = -\infty$  for all  $j \in \mathcal{D} \setminus \mathcal{K}$ . Finally, since the consideration set  $\mathcal{K}$  has the same distribution as the set of alternatives with  $\epsilon_j = 0$  (this is by construction), the basic ARC model and this ORUM model yield the same model predictions, and hence they are equivalent. □

*Proof of Proposition 13.* As a counter example, suppose  $U_{\nu}(d_j, x) > U_{\nu}(d_k, x)$ , but  $\varphi_k = 1$  and  $\varphi_j = 0$ . Then  $\Pr(d = d_k \mid x, \nu) \geq \Pr(d = d_j \mid x, \nu)$ , with equality if and only if  $d_k$  is dominated by a set of always-considered alternatives. □

## B Application: Verifying Cutoff Order

We start by recalling that CARA and CRRA utility functions satisfy the following basic property (see, e.g., Pratt, 1964; Barseghyan, Molinari, et al., 2018).<sup>44</sup>

---

<sup>44</sup>This property is equivalent to condition (e) in Pratt (1964, Theorem 1). As shown there, it is equivalent to assuming that an increase in  $\nu$  corresponds to an increase in the coefficient of absolute risk aversion.

**Property C.1.** For any  $y_0 > y_1 > y_2 > 0$ , the ratio  $R(y_0, y_1, y_2) \equiv \frac{u_\nu(y_1) - u_\nu(y_2)}{u_\nu(y_0) - u_\nu(y_1)}$  is strictly increasing in  $\nu$ .

It follows that CARA and CRRA utility functions also satisfy a slightly extended version of the property above:

**Property C.2.** For any  $y_0 > y_1 > y_2 > y_3 > 0$ , the ratio  $M_\nu(y_0, y_1, y_2, y_3) \equiv \frac{u_\nu(y_2) - u_\nu(y_3)}{u_\nu(y_0) - u_\nu(y_1)}$  is strictly increasing in  $\nu$ .

*Proof.*

$$\begin{aligned} M_\nu(y_0, y_1, y_2, y_3) &= \frac{u_\nu(y_2) - u_\nu(y_3)}{u_\nu(y_0) - u_\nu(y_1)} = \frac{u_\nu(y_2) - u_\nu(y_3)}{u_\nu(y_1) - u_\nu(y_2)} \times \frac{u_\nu(y_1) - u_\nu(y_2)}{u_\nu(y_0) - u_\nu(y_1)} \\ &= R_\nu(y_1, y_2, y_3) R_\nu(y_0, y_1, y_2) \end{aligned}$$

□

For our application, we show that  $c_{1,j}(\bar{p}, \mu) < c_{1,j+1}(\bar{p}, \mu)$  for any  $j \geq 2$  under both CARA and CRRA preferences.

**Theorem C.1.** Suppose deductibles and prices are such that

$$\frac{p_1 - p_j}{p_1 - p_{j+1}} < \frac{d_1 - d_j}{d_1 - d_{j+1}}.$$

Under either CARA or CRRA expected utility preferences, the cutoff mapping is unique and satisfies  $c_{1,j}(\bar{p}, \mu) < c_{1,j+1}(\bar{p}, \mu)$  for all  $j > 1$ .

*Proof.* We start with CARA preferences. The existence and the uniqueness of  $c_{j,k}(x)$  for all  $j < k$  follows directly from the Property C.2. Indeed note that  $p_j < p_k < p_k + d_k < p_j + d_j$ .<sup>45</sup> At the cutoff the DM is indifferent between lotteries  $j$  and  $k$ . Equating two expected utilities and rearranging we have that

$$\frac{e^{-\nu(w-p_k-d_k)} - e^{-\nu(w-p_j-d_j)}}{e^{-\nu(w-p_j)} - e^{-\nu(w-p_k)}} = \frac{1 - \mu}{\mu}, \quad (\text{B.1})$$

where  $w$  is the DM's initial wealth. By Property C.2, the L.H.S. of Equation B.1 is strictly monotone in  $\nu$ , and it tends to  $+\infty$  when  $\nu$  goes to  $+\infty$  and to zero when  $\nu$  goes to  $-\infty$ . It follows that there exists a unique  $\nu$ , i.e the cutoff  $c_{j,k}(x)$ , that solves the Equation B.1.

---

<sup>45</sup>If  $p_k + d_k > p_j + d_j$ , then alternative  $j$  first order stochastically dominates  $k$  and hence the cutoff is  $+\infty$ .

Moreover, since the L.H.S. is strictly monotone in  $\nu$  it follows from the Implicit Function Theorem that  $c_{j,k}(x)$  is continuous in  $\mu$  and  $\bar{p}$ .

The next step is to establish  $c_{1,j}(\bar{p}, \mu) < c_{1,j+1}(\bar{p}, \mu)$ ,  $j > 1$ . For the purpose of obtaining a contradiction, suppose that there exists  $(\bar{p}, \mu)$  and an  $j$  such that  $c_{1,j}(\bar{p}, \mu) = c_{1,j+1}(\bar{p}, \mu)$ . Since the expected utility of lottery  $k$  is proportional to

$$EU_\nu(L_k) \propto -e^{\nu p_k} (1 - \mu + \mu e^{\nu d_k}),$$

there exists  $\nu = c_{1,j}(\bar{p}, \mu) = c_{1,j+1}(\bar{p}, \mu)$  such that

$$\frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_j}} e^{\nu(g_1 - g_j)\bar{p}} = 1 = \frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_{j+1}}} e^{\nu(g_1 - g_{j+1})\bar{p}}$$

Taking logs yields

$$\begin{aligned} \log\left(\frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_j}}\right) &= -\nu(g_1 - g_j)\bar{p} \\ \log\left(\frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_{j+1}}}\right) &= -\nu(g_1 - g_{j+1})\bar{p}. \end{aligned}$$

Dividing through yields

$$\frac{\log\left(\frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_j}}\right)}{\log\left(\frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_{j+1}}}\right)} = \frac{g_1 - g_j}{g_1 - g_{j+1}}.$$

The R.H.S. is less than one. The L.H.S. is monotonically decreasing in  $\mu < 1$ . Indeed, denote  $\hat{\mu} = \frac{1 - \mu}{\mu}$ ,  $\Delta_1 = e^{\nu d_1}$ ,  $\Delta_j = e^{\nu d_j}$ , and  $\Delta_{j+1} = e^{\nu d_{j+1}}$  to rewrite the L.H.S. as follows

$$f(\hat{\mu}) \equiv \frac{\log(\Delta_1 + \hat{\mu}) - \log(\Delta_j + \hat{\mu})}{\log(\Delta_1 + \hat{\mu}) - \log(\Delta_{j+1} + \hat{\mu})}.$$

We claim that the expression above is monotonically increasing in  $\hat{\mu}$ . Indeed, we have

$$\frac{f'(\hat{\mu})}{f(\hat{\mu})} = \left(\frac{1}{\Delta_1 + \hat{\mu}} - \frac{1}{\Delta_j + \hat{\mu}}\right) \frac{1}{\log(\Delta_1 + \hat{\mu}) - \log(\Delta_j + \hat{\mu})} - \left(\frac{1}{\Delta_1 + \hat{\mu}} - \frac{1}{\Delta_{j+1} + \hat{\mu}}\right) \frac{1}{\log(\Delta_1 + \hat{\mu}) - \log(\Delta_{j+1} + \hat{\mu})}$$

After relabeling  $\Lambda_1 = -\log(\Delta_1 + \hat{\mu})$ ,  $\Lambda_j = -\log(\Delta_j + \hat{\mu})$  and  $\Lambda_{j+1} = -\log(\Delta_{j+1} + \hat{\mu})$  we

obtain

$$\frac{f'(\hat{\mu})}{f(\hat{\mu})} = \frac{e^{\Lambda_1} - e^{\Lambda_{j+1}}}{\Lambda_1 - \Lambda_{j+1}} - \frac{e^{\Lambda_1} - e^{\Lambda_j}}{\Lambda_1 - \Lambda_j}.$$

Since  $\Lambda_1 < \Lambda_j < \Lambda_{j+1}$  and exponential function is convex, the expression above is positive. Hence the derivative of  $f\left(\frac{1-\mu}{\mu}\right)$  W.R.T.  $\mu$  is negative, and hence it achieves its lowest value at  $\mu = 1$ . When  $\mu = 1$ , the L.H.S. is equal to  $\frac{d_1 - d_j}{d_1 - d_{j+1}}$ . Hence, the question is whether the following equality may hold

$$\frac{d_1 - d_j}{d_1 - d_{j+1}} = \frac{g_1 - g_j}{g_1 - g_{j+1}}.$$

It naturally would hold in perfectly competitive markets where additional coverage is simply proportional to its price. In practice, however, one might expect that with some market power the prices increase faster than then coverage, which is exactly what we find in our data (as well as for a larger number of firms appearing in Barseghyan et al. (2011)). Hence  $c_{1,j}(\bar{p}, \mu) \neq c_{1,j+1}(\bar{p}, \mu)$ . Since the cutoffs are continuous, it follows that  $c_{1,j}(\bar{p}, \mu) < c_{1,j+1}(\bar{p}, \mu)$ .

Under CRRA,  $c_{j,k}(\bar{p}, \mu)$  exist and are continuous exactly for the same reasons as under CARA. It remains to establish that  $c_{1,j}(\bar{p}, \mu) < c_{1,j+1}(\bar{p}, \mu)$ . Consider the following Taylor expansion for the CRRA Bernoulli utility function  $u(w)$  about point  $w - p_k$ :

$$\begin{aligned} u_\nu(w) &\equiv \frac{w^{1-\nu}}{1-\nu} \\ &= \frac{(w - p_k)^{1-\nu}}{1-\nu} + \frac{w^{-\nu}}{1!} p_k - \nu \frac{w^{-\nu-1}}{2!} p_k^2 + \nu(\nu+1) \frac{w^{-\nu-2}}{3!} p_k^3 - \nu(\nu+1)(\nu+2) \frac{w^{-\nu-3}}{4!} p_k^4 + \dots \end{aligned}$$

This can be written as follows

$$\frac{(w - p_k)^{1-\nu} - w^{1-\nu}}{w^{1-\nu}} = -\frac{1-\nu}{w} p_k + \frac{(1-\nu)(-\nu)}{2!w^2} p_k^2 - \frac{(1-\nu)(-\nu)(-\nu-1)}{3!w^3} p_k^3 + \dots$$

Hence, we can write

$$EU_\nu(L_k) \propto (1-\mu) \sum_{t=1}^{\infty} \omega_t p_k^t + \mu \sum_{t=1}^{\infty} \omega_t (p_k + d_k)^t.$$

The coefficients  $\omega_t \equiv (t!w^t)^{-1} \prod_{t'=0}^{t-1} (1-\nu-t')(-1)^t$  are negative for all  $t$ , so the two power series above are absolutely convergent. The element-wise difference between  $EU_\nu(L_j)$  and

$EU_\nu(L_k)$  is therefore well-defined with:

$$\begin{aligned}
EU_\nu(L_j) - EU_\nu(L_k) &\propto (1 - \mu) \sum_{t=1}^{\infty} \omega_t (p_j^t - p_k^t) + \mu \sum_{t=1}^{\infty} \omega_t ((p_j + d_j)^t - (p_k + d_k)^t) \\
&= (p_j - p_k) (1 - \mu) \sum_{t=1}^{\infty} \omega_t \sum_{h=0}^t p_j^h p_k^{t-h} + \\
&\quad + ((p_j - p_k) + (d_j - d_k)) \mu \sum_{t=1}^{\infty} \omega_t \sum_{h=0}^t (p_j + d_j)^h (p_k + d_k)^{t-h}
\end{aligned}$$

This implies that if  $\nu = c_{1,j}(\bar{p}, \mu) = c_{1,j+1}(\bar{p}, \mu)$ , then

$$\begin{aligned}
\frac{p_1 - p_j}{p_1 - p_{j+1}} &= \frac{p_1 - p_j + d_1 - d_j}{p_1 - p_{j+1} + d_1 - d_{j+1}} \times \\
&\quad \frac{\sum_{t=1}^{\infty} \omega_t \sum_{h=0}^t (p_1 + d_1)^h (p_j + d_j)^{t-h}}{\sum_{t=1}^{\infty} \omega_t \sum_{h=0}^t (p_1 + d_1)^h (p_{j+1} + d_{j+1})^{t-h}} \frac{\sum_{t=1}^{\infty} \omega_t \sum_{h=0}^t p_1^h p_{j+1}^{t-h}}{\sum_{t=1}^{\infty} \omega_t \sum_{h=0}^t p_1^h p_j^{t-h}}
\end{aligned}$$

Note that  $p_{j+1} > p_j$ . Moreover, when  $\nu = c_{1,j}(\bar{p}, \mu) = c_{1,j+1}(\bar{p}, \mu)$  it is also the case that  $\nu = c_{1,j}(\bar{p}, \mu) = c_{1,j+1}(\bar{p}, \mu) = c_{j,j+1}(\bar{p}, \mu)$ .

For the cutoff  $c_{j,j+1}(\bar{p}, \mu)$  to be on the support it must be the case that  $p_j + d_j > p_{j+1} + d_{j+1}$ . Indeed otherwise  $p_{j+1} - p_j > d_j - d_{j+1}$ , which is a violation of the first order stochastic dominance. Hence if we can show that

$$\frac{p_1 - p_j}{p_1 - p_{j+1}} < \frac{p_1 - p_j + d_1 - d_j}{p_1 - p_{j+1} + d_1 - d_{j+1}},$$

we would arrive to a contradiction, since it would be mean that the LHS of the equation is smaller than the RHS. Re-arranging:

$$\begin{aligned}
\frac{p_1 - p_{j+1} + d_1 - d_{j+1}}{p_1 - p_{j+1}} &< \frac{p_1 - p_j + d_1 - d_j}{p_1 - p_j} \\
\frac{d_1 - d_{j+1}}{p_1 - p_{j+1}} &< \frac{d_1 - d_j}{p_1 - p_j} \\
\frac{p_1 - p_j}{p_1 - p_{j+1}} &< \frac{d_1 - d_j}{d_1 - d_{j+1}}.
\end{aligned}$$

The latter inequality holds in the data, as discussed in the case of CARA.

□

## C Monetary Cost of Limited Consideration

We view limited consideration as a process that constrains households from achieving their first-best alternative either because the market setting forces some alternatives to become more salient than others (e.g. agent effects) or because of time or psychological costs that prevent the household from evaluating all alternatives in the choice set. Regardless of the underlying mechanism(s) of limited consideration, we can quantify its *monetary* cost within our framework. We ask, *ceteris paribus*, how much money the households “leave on the table” when choosing deductibles in property insurance under limited consideration rather than under full consideration. This is likely to be a lower bound on actual monetary losses arising from limited consideration, because insurance companies might be exploiting sub-optimality of households choices when setting prices or choosing menus.

We measure the monetary costs of limited consideration as follows. For each household we compute (the expected value of) the certainty equivalent of the lottery associated with the households’ optimal choice, as well as of the one associated with their choice under limited consideration.<sup>46</sup> We then take the difference between these certainty equivalent values and average them across all households in the sample. On average, we find that households lose \$49 dollars across the three deductibles because of limited consideration. See Table E.7 for variation conditional on demographic characteristics and insurance score. We also find wide dispersion in loss across households (see Figure E.7). In particular, the 10<sup>th</sup> percentile of losses is \$30 and the 90<sup>th</sup> is \$72.

---

<sup>46</sup>Certainty equivalent of the lottery is defined as the minimum amount they are willing to accept in lieu of the lottery. In our case, for alternative  $j$ , it is simply  $ce_j \equiv \frac{1}{\nu} \ln[(1 - \mu) \exp(\nu p_j) + \mu \exp(\nu(p_j + d_j))]$ .



## D Data

**Table D.1** Descriptive Statistics

Variable	Mean	Std. Dev.	1st %	99th %
Age	53.3	15.7	25.4	84.3
Female	0.40			
Single	0.22			
Married	0.55			
Second Driver	0.43			
Insurance Score	767	112	532	985

**Table D.2** Frequency of Deductible Choices Across Contexts

Deductible	1000	500	250	200	100	50
Collision	0.064	0.676	0.122	0.129	0.009	
Comprehensive	0.037	0.430	0.121	0.329	0.039	0.044
Home	0.176	0.559	0.262		0.002	

**Table D.3** Deductible Rank Correlations Across Contexts

	Collision	Comprehensive	Home
Collision	1		
Comprehensive	0.61	1	
Home	0.37	0.35	1

**Table D.4** Joint Distribution of Auto Deductibles

Collision	Comprehensive					
	1000	500	250	200	100	50
1000	3.71	1.93	0.18	0.44	0.05	0.04
500	0	40.99	6.46	17.84	1.27	1.00
250	0	0.04	5.42	4.55	1.28	0.94
200	0.01	0.05	0.03	9.99	1.07	1.78
100	0	0	0	0.04	0.23	0.66

The distribution is reported in percent.

**Table D.5** Average Premiums Across Coverages

Deductible	1,000	500	250	200	100	50
Collision	145	187	243	285	321	
Comprehensive	94	117	147	155	178	224
Home	594	666	720		885	

## E Empirical Results: Figures and Tables

### E.1 The ARC Model with Observable Demographics

While it is ideal to control for households' observable characteristics non-parametrically, it is data demanding. In practice, it is commonly assumed that household characteristics shift the expected value of the preference-coefficient distribution.<sup>47</sup> We adopt the same strategy here by assuming that for each household  $i$ ,  $\log \frac{\beta_{1,i}}{\beta_2} = \mathbf{Z}_i \gamma$ , where  $\gamma$  is an unknown vector to be estimated. The terms  $\beta_{1,i}$  and  $\beta_2$  denote the parameters of the Beta distribution, where  $\beta_{1,i}$  is household specific and  $\beta_2$  is common across households. The preference coefficients are random draws from a distribution with an expected value that is a function of the observable characteristics given by  $E(\nu_i) = \frac{\beta_{1,i}}{\beta_{1,i} + \beta_2} \bar{\nu} = \frac{e^{\mathbf{Z}_i \gamma}}{1 + e^{\mathbf{Z}_i \gamma}} \bar{\nu}$ .<sup>48</sup> The results of this estimation are in line with our first estimation. (See Column 2 in Table E.1, as well as Figures E.1 and E.2 in Appendix E.) The new observation here is that the model closely matches the distribution of choices across various sub-populations in the sample including gender, age, credit worthiness, and contracts with multiple drivers. The model's ability to match these conditional distributions can be attributed, in part, to the dependence of risk preferences on household characteristics. The model is, however, fairly parsimonious as the consideration parameters are restricted to be the same across all households. Finally, estimated consideration probabilities are close in magnitude to those estimated above. In particular, the highest deductibles (\$1,000 and \$500) are most likely to be considered, with respective frequencies of 0.95 and 0.91. The remaining alternatives are considered at much lower frequencies.

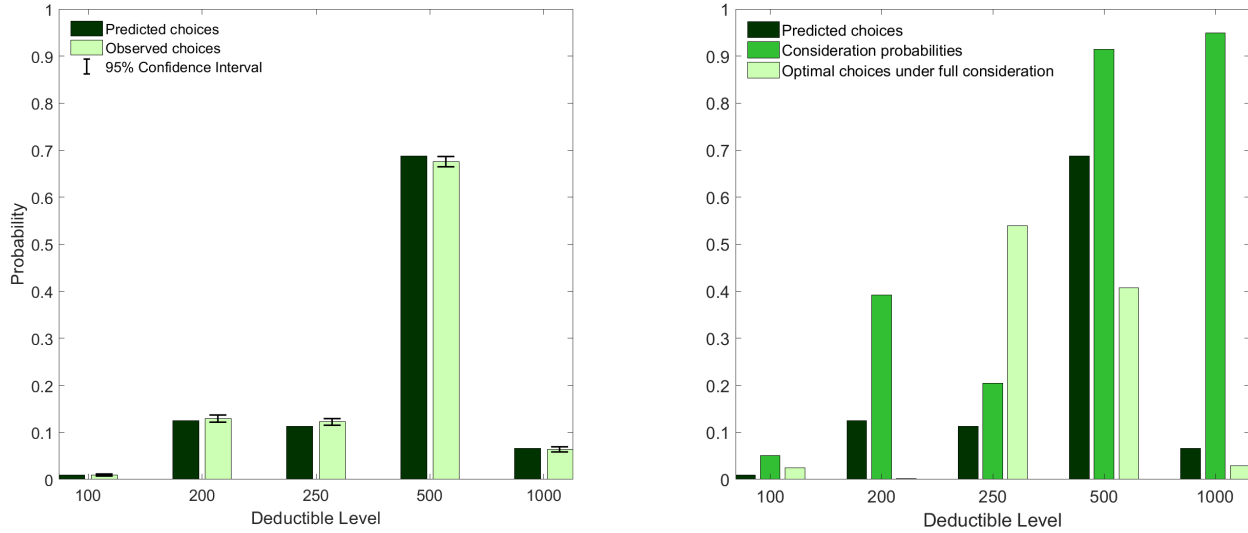
---

<sup>47</sup>For example, Cohen & Einav (2007) assume that  $\log \nu_i = \mathbf{Z}_i \gamma + \varepsilon_i$ , where  $\mathbf{Z}_i$  are the observables for household  $i$  and  $\varepsilon_i$  is i.i.d.  $N(0, \sigma^2)$ . Hence,  $E(\nu_i) = e^{\mathbf{Z}_i \gamma + \sigma^2/2}$ .

<sup>48</sup>If, instead, we assume  $\log \frac{\beta_{2,i}}{\beta_1} = \mathbf{Z}_i \tilde{\gamma}$ , then we arrive to the same expression for the expected value with the exception that  $\tilde{\gamma} = -\gamma$ .

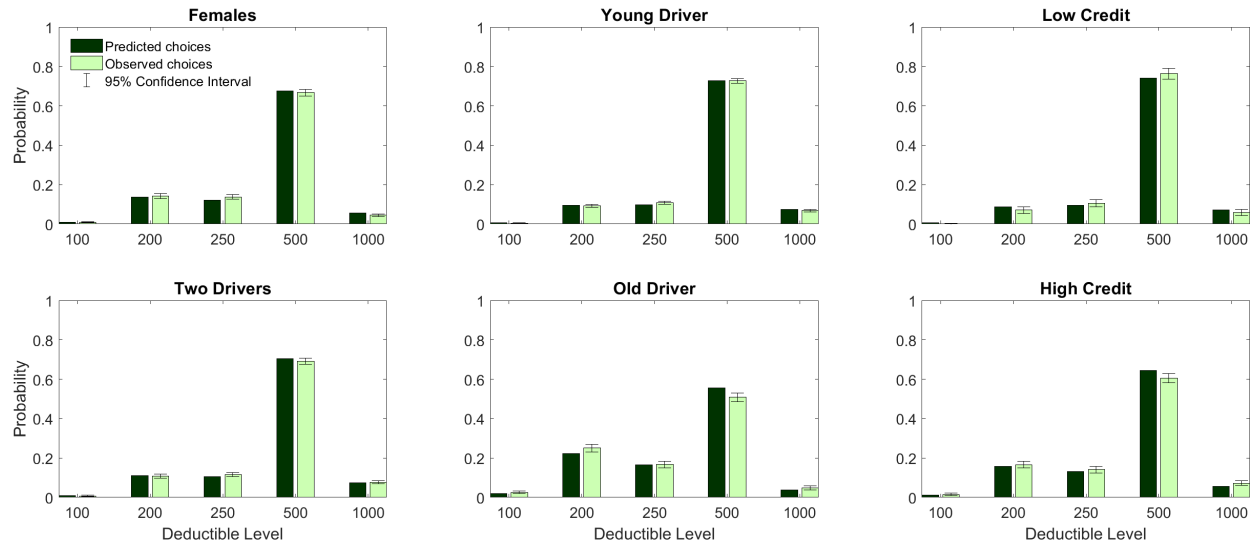
## E.2 Figures

**Figure E.1:** The ARC Model with Observable Demographics

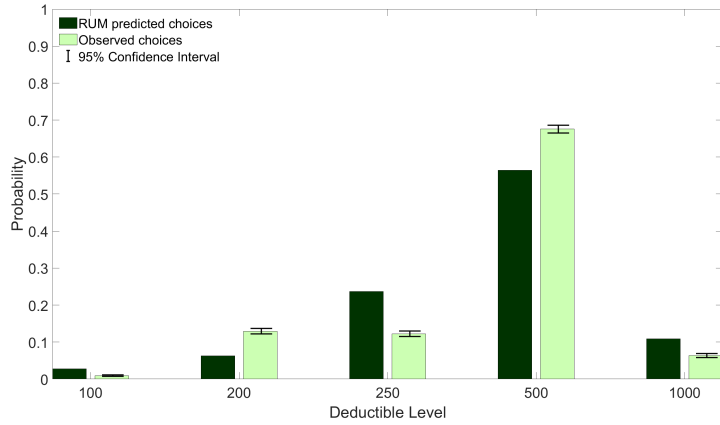


The first panel reports the distribution of predicted and observed choices. The second panel displays consideration probabilities and the distribution of optimal choices under full consideration.

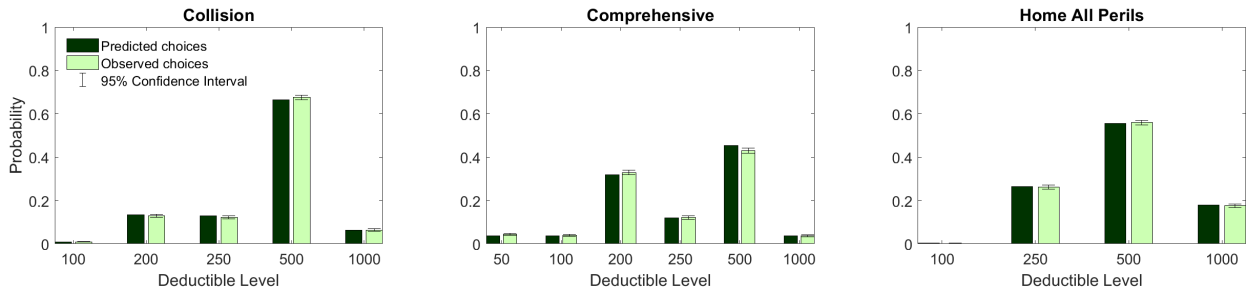
**Figure E.2:** The ARC Model with Observable Demographics: Conditional Distributions



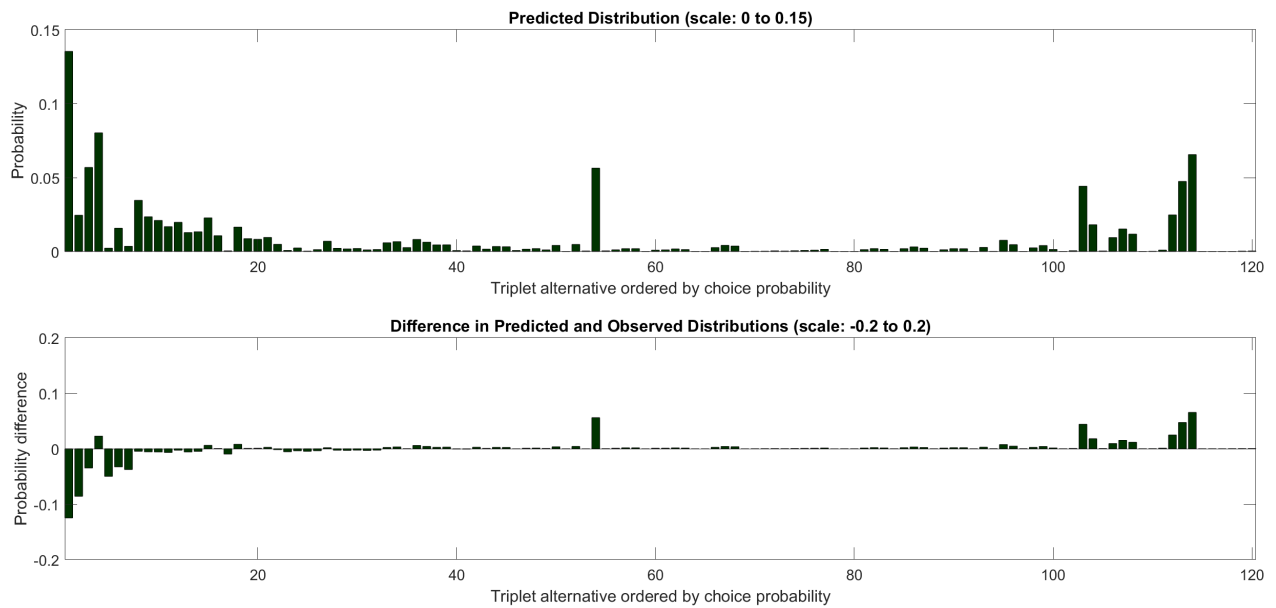
**Figure E.3: The Mixed Logit**



**Figure E.4: The ARC Model, Three Coverages, “Narrow” Consideration**

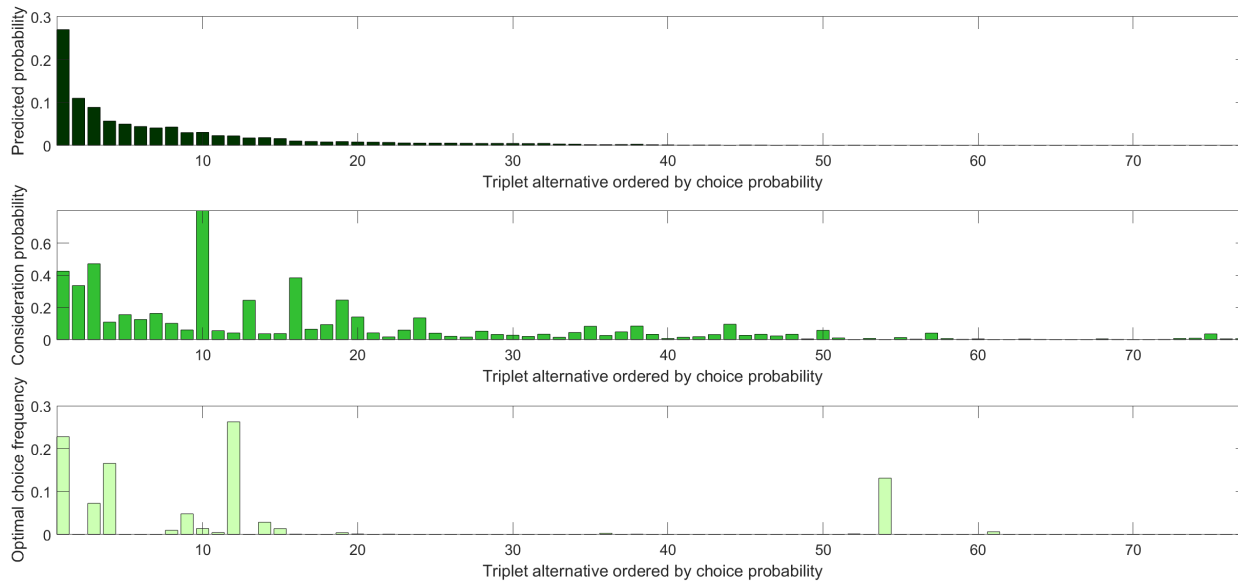


**Figure E.5: The Mixed Logit, Three Coverages**



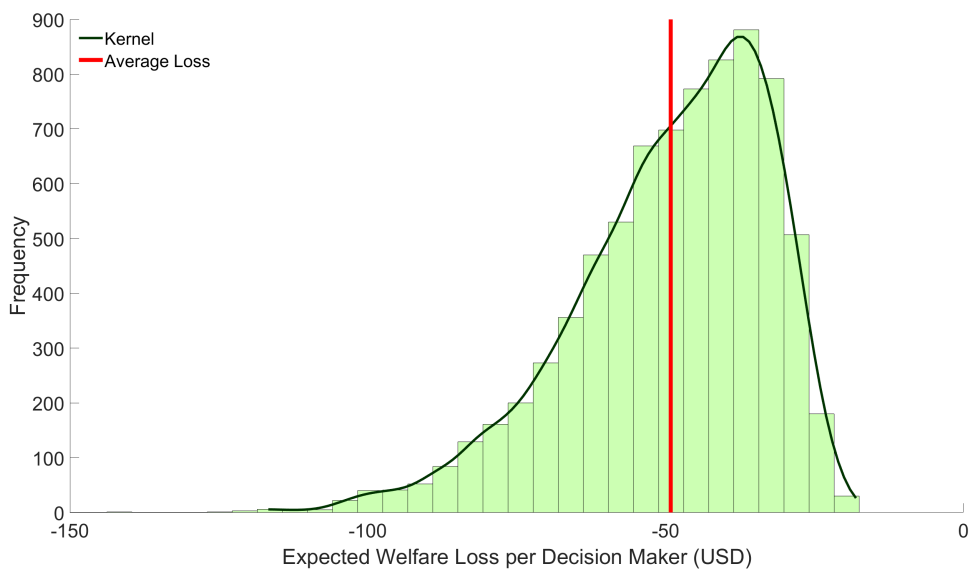
Triplets are sorted by observed frequency at which they are chosen. The first panel reports the predicted choice frequency and the second panel reports the difference in predicted and observed choice frequencies.

**Figure E.6:** The ARC Model, Three Coverages: Consideration and Optimal Choice Distribution



Triplets are sorted by observed frequency at which they are chosen.

**Figure E.7:** The ARC Model with Three Coverages: Monetary Loss From Limited Consideration



### E.3 Tables

**Table E.1** MLE Estimation Results for the ARC Model: Collision Only

	ARC Model		ARC Model with Observables	
$\beta_1$	1.621	[1.378, 1.948]	2.090	[1.556, 2.816]
$\beta_2$	7.319	[5.946, 9.177]	8.855	[6.934, 11.758]
<b>Mean of <math>\nu</math></b>	0.004	[0.003, 0.004]	0.004	[0.003, 0.004]
<b>SD of <math>\nu</math></b>	0.002	[0.002, 0.003]	0.002	[0.002, 0.002]
<b>Intercept</b>	-	-	-1.432	[-1.600, -1.302]
<b>Age</b>	-	-	0.211	[0.149, 0.298]
<b>Age<sup>2</sup></b>	-	-	0.047	[-0.002, 0.106]
<b>Female Driver</b>	-	-	0.075	[0.019, 0.145]
<b>Single Driver</b>	-	-	0.050	[-0.011, 0.114]
<b>Married Driver</b>	-	-	0.102	[0.022, 0.196]
<b>Credit Score</b>	-	-	0.137	[0.078, 0.199]
<b>2+ Drivers</b>	-	-	-0.310	[-0.479, -0.155]
<b>Collision \$100</b>	0.059	[0.041, 0.081]	0.051	[0.033, 0.071]
<b>Collision \$200</b>	0.414	[0.371, 0.465]	0.392	[0.344, 0.453]
<b>Collision \$250</b>	0.207	[0.190, 0.224]	0.205	[0.188, 0.227]
<b>Collision \$500</b>	0.918	[0.904, 0.931]	0.915	[0.896, 0.927]
<b>Collision \$1000</b>	1.000	[0.972, 1.000]	0.949	[0.690, 1.000]

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table E.2** MLE Estimation Results for the  
Mixed Logit:  
Collision Only

	<b>Mixed Logit</b>	
$\beta_1$	8.401	[6.794, 10.650]
$\beta_2$	122.603	[102.578, 152.511]
<b>Mean of <math>\nu</math></b>	0.001	[0.001, 0.001]
<b>SD of <math>\nu</math></b>	0.0004	[0.0004, 0.0005]
<b>Intercept</b>	-2.647	[-2.713, -2.586]
<b>Age</b>	-0.146	[-0.178, -0.118]
<b>Age<sup>2</sup></b>	-0.026	[-0.051, -0.002]
<b>Female Driver</b>	-0.004	[-0.032, 0.025]
<b>Single Driver</b>	-0.010	[-0.039, 0.020]
<b>Married Driver</b>	-0.031	[-0.069, 0.009]
<b>Credit Score</b>	0.096	[0.073, 0.124]
<b>2+ Drivers</b>	-0.021	[-0.101, 0.061]
<b>Full Attention</b>	0.040	[0.036, 0.043]

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01



**Table E.3** MLE Estimation Results for the ARC Model: Narrow Bracketing

<b>ARC Model</b>		
$\beta_1$	1.152	[1.010, 1.284]
$\beta_2$	3.141	[2.639, 3.694]
Mean of $\nu$	0.005	[0.005, 0.006]
SD of $\nu$	0.004	[0.004, 0.004]
Intercept	-1.127	[-1.225, -1.032]
Age	0.198	[0.164, 0.235]
Age <sup>2</sup>	0.090	[0.059, 0.121]
Female Driver	0.052	[0.018, 0.088]
Single Driver	0.004	[-0.037, 0.047]
Married Driver	0.008	[-0.038, 0.062]
Credit Score	0.110	[0.077, 0.145]
2+ Drivers	-0.089	[-0.186, 0.004]
Collision \$100	0.033	[0.023, 0.043]
Collision \$200	0.324	[0.299, 0.351]
Collision \$250	0.199	[0.185, 0.216]
Collision \$500	0.953	[0.945, 0.960]
Collision \$1000	1.000	[0.870, 1.000]
Comprehensive \$50	1.000	[1.000, 1.000]
Comprehensive \$100	0.337	[0.291, 0.384]
Comprehensive \$200	0.765	[0.744, 0.790]
Comprehensive \$250	0.325	[0.295, 0.357]
Comprehensive \$500	0.892	[0.853, 0.928]
Comprehensive \$1000	0.277	[0.226, 0.316]
Home \$100	0.002	[0.000, 0.010]
Home \$250	0.387	[0.368, 0.409]
Home \$500	0.859	[0.844, 0.877]
Home \$1000	0.824	[0.774, 0.873]

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table E.4** MLE Estimation Results for the ARC Model, Three Coverages

	ARC Model		ARC Model (cont.)		
$\beta_1$	4.515	[3.432, 6.255]	<b>(250,200,250)</b>	0.037	[0.029, 0.045]
$\beta_2$	23.623	[17.528, 33.251]	<b>(250,200,500)</b>	0.056	[0.046, 0.067]
Mean of $\nu$	0.003	[0.003, 0.003]	<b>(250,200,1000)</b>	0.045	[0.025, 0.067]
SD of $\nu$	0.001	[0.001, 0.002]	<b>(250,250,100)</b>	0.001	[0.000, 0.005]
Intercept	-1.706	[-1.792, -1.623]	<b>(250,250,250)</b>	0.042	[0.035, 0.050]
Age	0.166	[0.130, 0.207]	<b>(250,250,500)</b>	0.061	[0.051, 0.070]
Age <sup>2</sup>	0.041	[0.011, 0.073]	<b>(250,250,1000)</b>	0.026	[0.011, 0.044]
Female Driver	0.043	[0.006, 0.079]	<b>(250,500,500)</b>	0.0007	[0.000, 0.002]
Single Driver	0.011	[-0.028, 0.052]	<b>(500,50,250)</b>	0.034	[0.020, 0.049]
Married Driver	0.031	[-0.020, 0.085]	<b>(500,50,500)</b>	0.053	[0.032, 0.077]
Credit Score	0.141	[0.108, 0.175]	<b>(500,50,1000)</b>	0.034	[0.007, 0.074]
2+ Drivers	-0.099	[-0.196, -0.0004]	<b>(500,100,250)</b>	0.015	[0.009, 0.022]
<b>(100,50,250)</b>	0.041	[0.026, 0.059]	<b>(500,100,500)</b>	0.042	[0.029, 0.059]
<b>(100,50,500)</b>	0.015	[0.005, 0.029]	<b>(500,100,1000)</b>	0.049	[0.022, 0.081]
<b>(100,50,1000)</b>	0.013	[0.000, 0.043]	<b>(500,200,100)</b>	0.008	[0.000, 0.019]
<b>(100,100,100)</b>	0.002	[0.000, 0.010]	<b>(500,200,250)</b>	0.125	[0.109, 0.142]
<b>(100,100,250)</b>	0.008	[0.002, 0.014]	<b>(500,200,500)</b>	0.336	[0.305, 0.370]
<b>(100,100,500)</b>	0.005	[0.000, 0.011]	<b>(500,200,1000)</b>	0.245	[0.202, 0.296]
<b>(100,100,1000)</b>	0.005	[0.000, 0.019]	<b>(500,250,100)</b>	0.002	[0.000, 0.008]
<b>(100,200,250)</b>	0.0006	[0.000, 0.003]	<b>(500,250,250)</b>	0.038	[0.030, 0.046]
<b>(100,200,500)</b>	0.0008	[0.000, 0.003]	<b>(500,250,500)</b>	0.101	[0.088, 0.118]
<b>(100,200,1000)</b>	0.004	[0.000, 0.016]	<b>(500,250,1000)</b>	0.094	[0.066, 0.123]
<b>(200,50,100)</b>	0.011	[0.000, 0.025]	<b>(500,500,100)</b>	0.003	[0.000, 0.011]
<b>(200,50,250)</b>	0.065	[0.047, 0.088]	<b>(500,500,250)</b>	0.109	[0.097, 0.122]
<b>(200,50,500)</b>	0.060	[0.039, 0.082]	<b>(500,500,500)</b>	0.426	[0.399, 0.454]
<b>(200,50,1000)</b>	0.034	[0.007, 0.073]	<b>(500,500,1000)</b>	0.472	[0.435, 0.512]
<b>(200,100,100)</b>	0.002	[0.000, 0.009]	<b>(1000,50,250)</b>	0.008	[0.000, 0.033]
<b>(200,100,250)</b>	0.021	[0.013, 0.030]	<b>(1000,50,500)</b>	0.009	[0.000, 0.040]
<b>(200,100,500)</b>	0.028	[0.018, 0.039]	<b>(1000,50,1000)</b>	0.036	[0.000, 0.150]
<b>(200,100,1000)</b>	0.023	[0.005, 0.048]	<b>(1000,100,250)</b>	0.005	[0.000, 0.022]
<b>(200,200,100)</b>	0.002	[0.000, 0.007]	<b>(1000,100,500)</b>	0.006	[0.000, 0.028]
<b>(200,200,250)</b>	0.155	[0.133, 0.178]	<b>(1000,100,1000)</b>	0.041	[0.000, 0.126]
<b>(200,200,500)</b>	0.163	[0.140, 0.189]	<b>(1000,200,250)</b>	0.032	[0.007, 0.060]
<b>(200,200,1000)</b>	0.135	[0.090, 0.188]	<b>(1000,200,500)</b>	0.083	[0.042, 0.135]
<b>(200,250,250)</b>	0.0004	[0.000, 0.001]	<b>(1000,200,1000)</b>	0.096	[0.021, 0.195]
<b>(200,250,500)</b>	0.0005	[0.000, 0.002]	<b>(1000,250,250)</b>	0.007	[0.000, 0.022]
<b>(200,500,250)</b>	0.002	[0.000, 0.004]	<b>(1000,250,500)</b>	0.027	[0.006, 0.057]
<b>(200,1000,1000)</b>	0.005	[0.000, 0.024]	<b>(1000,250,1000)</b>	0.058	[0.000, 0.134]
<b>(250,50,100)</b>	0.002	[0.000, 0.009]	<b>(1000,500,250)</b>	0.033	[0.012, 0.060]
<b>(250,50,250)</b>	0.020	[0.013, 0.030]	<b>(1000,500,500)</b>	0.141	[0.095, 0.188]
<b>(250,50,500)</b>	0.033	[0.021, 0.047]	<b>(1000,500,1000)</b>	0.384	[0.297, 0.492]
<b>(250,100,250)</b>	0.017	[0.012, 0.023]	<b>(1000,1000,250)</b>	0.085	[0.037, 0.143]
<b>(250,100,500)</b>	0.016	[0.010, 0.023]	<b>(1000,1000,500)</b>	0.246	[0.180, 0.324]
<b>(250,100,1000)</b>	0.019	[0.004, 0.037]	<b>(1000,1000,1000)</b>	0.808	[0.627, 1.000]
<b>(250,200,100)</b>	0.001	[0.000, 0.005]			

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table E.5** MLE Estimation Results for RUM,  
Three Coverages

<b>Mixed Logit</b>		
$\beta_1$	4.363	[3.953, 4.840]
$\beta_2$	51.093	[47.265, 55.484]
<b>Mean of <math>\nu</math></b>	0.002	[0.002, 0.002]
<b>SD of <math>\nu</math></b>	0.0007	[0.0007, 0.0007]
<b>Intercept</b>	-2.422	[-2.469, -2.379]
<b>Age</b>	-0.081	[-0.103, -0.059]
<b>Age<sup>2</sup></b>	-0.016	[-0.032, 0.002]
<b>Female Driver</b>	0.0007	[-0.018, 0.018]
<b>Single Driver</b>	-0.015	[-0.034, 0.005]
<b>Married Driver</b>	-0.018	[-0.047, 0.009]
<b>Credit Score</b>	0.037	[0.020, 0.055]
<b>2+ Drivers</b>	-0.049	[-0.100, -0.0001]
<b>Sigma</b>	0.223	[0.201, 0.249]

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table E.6** Expected Monetary Loss by Group

	<b>Expected Monetary Loss</b>	
<b>All</b>	-49.1	[-55.3, -44.7]
<b>Female Driver</b>	-53.2	[-59.9, -48.0]
<b>Single Driver</b>	-44.1	[-49.7, -40.2]
<b>Young</b>	-44.4	[-49.1, -40.9]
<b>Old</b>	-64.6	[-76.8, -56.1]
<b>Low Credit Driver</b>	-46.3	[-51.4, -42.5]
<b>High Credit Driver</b>	-53.6	[-62.0, -47.6]

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01