

# 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 CWP08/19

# Discrete Choice under Risk with Limited Consideration\*

Levon Barseghyan<sup>†</sup>

Francesca Molinari<sup>‡</sup>

Matthew Thirkettle<sup>§</sup>

February 15, 2019

## Abstract

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. We propose a discrete choice model with unobserved heterogeneity in consideration sets (the collection of alternatives considered by DMs) and unobserved heterogeneity in standard risk aversion. In this framework, stochastic choice is driven both by different rankings of alternatives induced by unobserved heterogeneity in risk preferences and by different sets of alternatives considered. We obtain sufficient conditions for semi-nonparametric point identification of both the distribution of unobserved heterogeneity in preferences and the distribution of consideration sets. Our method yields an estimator that is easy to compute and that can be used in markets with a large number of alternatives. We apply our method to a dataset on property insurance purchases. We find that although households are on average strongly risk averse, they consider lower coverages more frequently than higher coverages. Finally, we estimate the monetary losses associated with limited consideration in our application.

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

---

\*We thank Abi Adams, Jose Apesteguia, Miguel Ballester, Liran Einav, Arthur Lewbel, Chuck Manski, Jack Porter, and the participants of 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 Fall 2018 Duke University Econometrics Seminar, the 2019 ASSA meetings, and the Spring 2019 Stanford IO Lunch for useful comments and suggestions. Part of this research was carried out while Barseghyan and Molinari were visiting 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.<sup>1</sup> Standard EUT assumes that the DM assesses a risky alternative by computing its expected utility; evaluates all available alternatives; and chooses the alternative yielding highest expected utility. The DM's risk aversion is determined by the concavity of her underlying 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 risk aversion and unobserved heterogeneity in *consideration sets*. Each DM evaluates only the alternatives in her consideration set, which is a subset of the choice set. Hence, stochastic choice is driven both by different rankings of alternatives induced by unobserved heterogeneity in risk preferences and by different sets of alternatives considered. Our first contribution is to establish that the requirements of standard economic theory, coupled with a slight strengthening of the classic conditions for semi-nonparametric identification of discrete choice models with full consideration and identical choice sets (see, e.g., Matzkin, 2007),<sup>2</sup> yield semi-nonparametric identification of both the distribution of unobserved heterogeneity in risk aversion and the distribution of consideration sets.<sup>3</sup>

Our second contribution is to provide a simple method to compute our likelihood-based estimator, whose computational complexity grows polynomially in the number of alternatives in the choice set.<sup>4</sup> In particular, 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.

Our third contribution is to elucidate the applicability and the advantages of our framework over the standard application of random utility models (RUMs) with additively separable

---

<sup>1</sup>For a non-exhaustive list of papers in this literature see Starmer (2000) and Barseghyan, Molinari, O'Donoghue, & Teitelbaum (2018). We discuss an important class of non-expected utility theory models and how our analysis applies to these models later in the paper.

<sup>2</sup>In fact, with a binary consideration set, the former and the latter coincide.

<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.

<sup>4</sup>The function evaluation time of the log-likelihood objective function grows linearly with the number of alternatives. 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).

unobserved heterogeneity (e.g., Mixed Logit models) and full consideration. First, our model can generate zero shares for non-dominated alternatives. Second, the model has no difficulty explaining relatively large shares of dominated alternatives. Third, in markets with many insurance 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 recently raised 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.

In general, distinguishing heterogeneous preferences from heterogeneous consideration using discrete choice data is a formidable task. When a DM chooses an alternative, this can be either because that alternative yields the highest expected utility among those in her entire choice set or because the DM does not consider some better available alternatives and the chosen one is the best in her consideration set, implying different distributions of preferences. We show that this seemingly inescapable identification problem can be resolved under certain conditions by leveraging standard requirements of economic theory. Specifically, our random preference models satisfy the classic Single Crossing Property (SCP) of Mirrlees (1971); Spence (1974): the preference order of any two alternatives switches only once on the support of the preference coefficient.<sup>5</sup> The SCP is central to important studies of decision making under risk, as well as those in other fields of Economics.<sup>6</sup> More so, as we make clear, the SCP is necessary for nonparametric identification of the preference parameter distribution in the standard model with full consideration and homogeneous observed choice sets. Coupled with three additional requirements (imposed in the literature on point identification of limited consideration models), we show that the SCP delivers nonparametric identification of the preference-parameter distribution even in the presence of unobserved heterogeneity in consideration sets. The first two requirements are: (1) specification of a consideration set formation model and (2) independence between unobserved heterogeneity in consideration and in risk preferences, conditional on observable characteristics. The second requirement is part of the standard framework: when all DMs consider the entire (non-stochastic) choice set, consideration is independent of underlying preferences

---

<sup>5</sup>The EUT framework with concave Bernoulli utility satisfies the SCP. The SCP 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. As we discuss in Section 8, many non-EU models, when they feature unidimensional preference heterogeneity, also satisfy SCP.

<sup>6</sup>E.g., Athey (2001); Apesteguia, Ballester, & Lu (2017); Chiappori, Salanié, Salanié, & Gandhi (2018).

by definition. Requirements (1) and (2) are motivated by Barseghyan, Coughlin, Molinari, & Teitelbaum (2019), who establish that in the absence of restrictions on the consideration set formation and its relation with risk preferences, one can partially but not point identify the distribution of risk preferences (even parametrically). The final requirement is that there exists a DM-characteristic with large support that shifts preferences over alternatives, but does not affect consideration. In RUMs, requiring existence of a regressor with large support (or an equivalent assumption) is necessary for nonparametric identification even without probabilistic consideration (e.g., Matzkin, 2007). The additional restriction in our framework, implicit in the full-consideration literature, is that the large-support regressor does not affect the probability of considering any of the alternatives in the choice set. We do, however, allow for the case that the large support regressor is not alternative specific, that is, it may only vary across DMs. Moreover, the consideration distribution may depend on the other characteristics of the DMs and of all alternatives.

With this structure in place, our identification result leverages a simple intuition: as the large-support regressor takes values sufficiently large or small, the alternatives in the choice set are unambiguously ranked for all possible realizations of the unobserved risk-preference coefficient. Hence, the choice frequency observed in the data is a function of only the consideration probabilities and, under weak restrictions, this function admits a unique solution for the consideration probabilities. The SCP then allows us to trace out the distribution of preferences given variation in the large-support regressor.

We describe our identification approach in detail for two probabilistic consideration models, each having up to as many parameters as the size of the choice set. These two models are different in nature and can be used as a blueprint to study the empirical content of many others, as we explain in the paper. The first model, termed the Alternative Specific Random Consideration (ARC) model, is inspired by Manski (1977) and Manzini & Mariotti (2014). In this model, alternative  $j$  appears in the DM's consideration set with an alternative-specific probability  $\varphi_j$  and each alternative enters the consideration set independent of all other alternatives. The second model, termed the Random Consideration Level (RCL) model, posits that the DM first draws the size of her consideration set,  $l$  (her consideration level, possibly determined by her cognitive ability), and then randomly picks  $l$  alternatives to consider, with each alternative having the same probability of being picked.

Of course, random preference models like the ones we consider are random utility models as originally envisioned by McFadden (1974) (for a textbook treatment see Manski, 2007). We

show that our random preference models with probabilistic consideration can be written as RUMs with unobserved heterogeneity in risk aversion and with an additive error that has a discrete distribution with support  $\{-\infty, 0\}$ . It is then natural to draw parallels with the mixed (random coefficient) logit model (e.g., McFadden & Train, 2000). In our setting, the Mixed Logit model 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>7</sup> In other contexts, unobservable characteristics may affect choice mostly via consideration – as we model – rather than via “additive noise”.<sup>8</sup>

As in the Mixed Logit model, our models assume independence of the additive error with the observable payoff-relevant characteristics and the unobservable heterogeneity in preferences. However, in the ARC model, the additive error is independent across alternatives but is not identically distributed; in the RCL model, the additive error is identically distributed but is not independent across alternatives; and in the Mixed Logit, the additive error is i.i.d. across alternatives. These differences generate contrasting implications in several respects.

First, the RCL model and the Mixed Logit model generally imply 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 RCL model and the Mixed Logit model satisfy a *Generalized Dominance Property* that we derive: if for any degree of risk aversion alternative  $j$  is dominated by either alternative  $k$  or  $m$ , then the probability of choosing  $j$  must be no larger than the probability of choosing  $k$  or  $m$ . The ARC model does not abide Generalized Dominance. Third, in the ARC and the RCL models, choice probabilities depend on the ordinal expected utility rankings of the alternatives, while, in the Mixed Logit, choice probabilities depend on the cardinal expected utility rankings. As we explain in Section 5, this difference implies that choice probabilities are monotone in risk preferences in our models, while in the Mixed Logit model they are not (Apesteguia & Ballester, 2018). Armed with the identification results obtained for the ARC and RCL models, we show in Section 6 that our approach easily extends to the case where consideration sets form based on liquidity constraints or behavioral phenomena such as extremeness aversion.

---

<sup>7</sup>E.g., employer provided health insurance, auto, or home insurance offered by a single company.

<sup>8</sup>E.g., a DM may only consider those supplemental prescription drug plans that cover specific medications.

Our empirical application is a study of households’ deductible choices across three lines of insurance coverage: auto collision, auto comprehensive, and home (all perils). The aim of our exercise is to estimate the underlying 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 both the RCL model and the EUT model with additive extreme value type I (Gumbel) error. 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 losses resulting from limited consideration are \$49.

## 2 Related Literature

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 et al. (2019). They study models of decision making under risk, where unobserved heterogeneity in preferences as well as in choice and/or consideration sets is allowed for. They additionally allow for arbitrary dependence between consideration sets and preferences, and impose no restrictions on the consideration set formation process. They show that such unrestricted forms of heterogeneity yield, in general, partial but not point identification of the model, even when a parametric distribution for preference heterogeneity is specified. They obtain bounds on the distribution of consideration sets’ size, but no other features of the distribution of consideration sets can be learned.

In this paper we take a conceptually different approach. As in the entire related literature on point identification of limited consideration models, we maintain independence of consideration sets and preferences and we focus on specific consideration sets’ formation processes. The latter are grounded in a sizable literature spanning experimental economics, microeconomics, behavioral economics, psychology, and marketing which aims to formalize the cognitive process underlying the formation of consideration sets.<sup>9</sup>

---

<sup>9</sup>See, e.g., Simon (1959); Tversky (1972); Howard (1977); Manski (1977); Treisman & Gelade (1980); Hauser & Wernerfelt (1990); Shocker, Ben-Akiva, Boccara, & Nedungadi (1991); Roberts & Lattin (1991); Ben-Akiva & Boccara (1995); Eliaz & Spiegel (2011); Masatlioglu, Nakajima, & Ozbay (2012); Manzini & Mariotti (2014); Caplin, Dean, & Leahy (2018). Even when DMs pay full attention, they may face unobserved constraints on what alternatives they can choose (e.g., Gaynor, Propper, & Seiler, 2016).

For the remainder of this section, we discuss the literature on identification of limited consideration models that are closely related to the ARC and RCL models. To identify parametric models of demand, previous contributions in this area have typically relied on auxiliary data revealing the consideration set composition (e.g., Draganska & Klapper, 2011; Conlon & Mortimer, 2013; Honka & Chintagunta, 2017), or on the existence of regressor(s) that impact utility but not consideration (or *vice versa*) (e.g., Goeree, 2008; Gaynor et al., 2016; Heiss, McFadden, Winter, Wuppermann, & Zhou, 2016; Hortaçsu, Madanizadeh, & Puller, 2017).<sup>10</sup> In contrast, we establish semi-nonparametric identification of the distributions of unobserved heterogeneity in preferences and in consideration sets through a combination of the SCP with an exclusion restriction and a large support assumption.

A recent related literature, closest to our own work, studies various departures from the tight parametric structure of the earlier analysis of limited consideration models. Dardanoni, Manzini, Mariotti, & Tyson (2017) consider a stochastic choice model with homogeneous preferences and heterogeneous cognitive types.<sup>11</sup> The cognitive types are implemented through the RCL model and a variant of the ARC model. In the RCL model, the cognitive type is the number of alternatives the DM is able to consider. In the ARC model, the type is the probability with which the DM considers an alternative (which is assumed constant across alternatives). The authors show how one can learn the moments of the distribution of types from a single cross section of aggregate choice shares. A key assumption for identification is that there exists a default alternative and the researcher observes the frequency with which the default alternative is chosen. In our paper, we do not require that the default option is observed and we are flexible on its existence.

Cattaneo, Ma, Masatlioglu, & Suleymanov (2017) propose a general random attention model where the probability of drawing a consideration set decreases as the choice set enlarges. Their model, however, does not allow for unobserved heterogeneity in preferences, and yields partial identification results while requiring rich observable variation in the choice set.

Abaluck & Adams (2018) study identification of an additive error random utility model with consideration formation similar to that in our ARC model, a variant of the RCL model (also

---

<sup>10</sup>Crawford, Grithz, & Iariax (2017) estimate a Fixed-Effect Logit type model that (1) utilizes observed purchase decisions (in a panel or a group-homogeneous cross-section) to construct “sufficient sets” of alternatives that lie within DMs’ feasible sets; (2) given the “sufficient sets”, uses classic techniques on estimating logit models on subsets of feasible sets.

<sup>11</sup>Heterogeneous tastes are also explored. To obtain identification, however, one of two strong assumptions are imposed. Either the taste distribution is known or preferences are linear in observable alternative characteristics and there is an additive error term with extreme value type 1 distribution.



considered in Ho, Hogan, & Scott Morton, 2017; Hortaçsu et al., 2017; Heiss et al., 2016), and a mix of the two. Abaluck & Adams (2018) method and ours are distinct and complementary. They require a default option, and the existence of a regressor (e.g., price) that is alternative specific and enters the indirect utility function linearly (or additively separably with shape restrictions). The price of each alternative is required to have large support, to exhibit cross-alternative variation (i.e., independent variation for each alternative), and to be excluded from the consideration probability of all other alternatives. When modeling choice under risk, concave utility yields that price enters neither linearly nor additively separably. More importantly, our work aims at providing a method to learn DMs’ risk preferences and consideration probabilities from their choices of insurance products. Many important recent empirical contributions in this area use data from a single company – either a firm selling insurance or a firm offering health insurance to its employees.<sup>12</sup> In their data, observable characteristics with large support are typically DM specific and not alternative specific, so that the Abaluck & Adams (2018) method does not apply.<sup>13</sup> In contrast, our framework does not require a default option and we only assume that the large support regressor is independent of consideration set formation. This regressor may or may not be alternative specific.

## 3 Models

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

To set the stage we consider the following insurance market. There is an underlying risk of a loss with probability equal to  $\mu$  that varies across DMs. A finite number of insurance alternatives are available against this loss. Each alternative is a pair  $(d_j, p_j), j \in \{1, \dots, D\}$ . The first element is a deductible, which is the DM’s out of pocket expense in case a loss occurs. Deductibles are decreasing with index  $j$ . All deductibles are less than the lowest realization of the loss. The second element is a price, 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 a multiplication rule,  $p_j = g_j \cdot \bar{p} + \delta$ , where  $\delta$  is a small positive amount and  $g_j$  increases with  $j$ : lower deductibles provide more coverage, and hence cost more. Both  $g_j$

---

<sup>12</sup>See, e.g., Cohen & Einav (2007); Einav, Finkelstein, Pascu, & Cullen (2012); Barseghyan, Molinari, O’Donoghue, & Teitelbaum (2013); Handel (2013); Bhargava, Loewenstein, & Sydnor (2017).

<sup>13</sup>For example, in the context of health insurance there is a fixed price for each insurance plan offered to all employees and there is large variation in risk across employees.

and  $\delta$  are invariant across DMs. The lotteries that the DM faces can be written as follows:

$$L_j(x) \equiv (-p_j, 1 - \mu; -p_j - d_j, \mu),$$

where  $x \equiv (\bar{p}, \mu)$ . DMs are expected utility maximizers. Given initial wealth  $w$ , the expected utility of deductible lottery  $L_j(x)$  is given by

$$EU(L_j(x)) = (1 - \mu) u(w - p_j) + \mu u(w - p_j - d_j),$$

where  $u(\cdot)$  is a Bernoulli utility function defined over final wealth states. We assume that  $u(\cdot)$  belongs to a certain family of utility functions that are fully characterized by a single coefficient  $\nu$  (e.g. CARA, CRRA or NTD).<sup>14</sup> This coefficient of risk aversion is randomly distributed across DMs and has bounded support.

The relationship between risk aversion, underlying prices, and loss probabilities is standard. At sufficiently high  $\bar{p}$  or low  $\mu$ , less coverage is always preferred to more coverage:  $L_1 \succ L_2 \succ L_3 \succ \dots \succ L_D$  for all  $\nu$  on the support. At sufficiently low  $\bar{p}$  or high  $\mu$ , we have the opposite ordering:  $L_D \succ L_{D-1} \succ L_{D-2} \succ \dots \succ L_1$ . For moderate levels of prices and loss probabilities things are more interesting: for each pair of deductible lotteries  $j < k$  there is a cutoff value  $c_{j,k}(x)$  in the interior of the risk-preference coefficient support. 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  $L_j$  and  $L_k$ . Those with lower  $\nu$  choose the riskier alternative  $L_j$ , while those with higher  $\nu$  choose the safer alternative  $L_k$ . Note that  $c_{j,k}(x)$  is a continuous function, since the expected utility from each deductible lottery is continuous in  $x$  as well as in  $\nu$ .

### 3.2 The Model under Full Consideration

There is a continuum of DMs who face a choice among a finite number of alternatives, i.e., the choice set, which is denoted  $\mathcal{D} = \{1, \dots, D\}$ . Alternatives vary by their utility-relevant characteristics. One characteristic,  $d_j \in \mathbb{R}, j \in \mathcal{D}$ , is DM invariant. When it is unambiguous, we may write  $d_j$  instead of “alternative  $j$ ”. Other characteristics, denoted by  $x_j \in \mathcal{X}_j \subset \mathbb{R}^q$ , may vary across DMs as well as across alternatives. That is, alternative  $j$  is fully characterized by  $(d_j, x_j)$ . We denote  $x = (x_1, \dots, x_D)$  and  $\mathcal{X} = \mathcal{X}_1 \otimes \dots \otimes \mathcal{X}_D$ .

---

<sup>14</sup>Under CRRA, it is implied that DMs’ initial wealth is known to the researcher. Negligible Third Derivative (NTD) utility is defined in Cohen & Einav (2007) and in Barseghyan et al. (2013).

Given these characteristics, each DM's preferences over the alternatives are defined by a utility function  $U_\nu(d, x)$ . The latter is fully described by a DM-specific index  $\nu$  assumed to be distributed according to  $F(\cdot)$  over a bounded support  $\Gamma = [0, \bar{\nu}]$ .<sup>15</sup> We assume that the random variables  $\nu$  and  $x$  are independent (and that demographic variables, if available, have been conditioned on). The DM's draw of  $\nu$  is not observed by the researcher and  $F(\cdot)$  is assumed to be continuous and otherwise left completely unspecified. Going forward, we only consider models satisfying a basic Single Crossing Property, as defined below.

**Definition 1** (Single Crossing Property). *The DM's preference relation over alternatives satisfies the Single Crossing Property iff the following condition holds: For all  $j, k, j \neq k$  there exists a continuous function  $c_{jk} : \mathcal{X} \rightarrow \mathbb{R}_{[-\infty, \infty]}$  (or  $c_{kj} : \mathcal{X} \rightarrow \mathbb{R}_{[-\infty, \infty]}$ ) such that*

$$\begin{aligned} U_\nu(d_j, x) &> U_\nu(d_k, x) && \forall \nu \in (-\infty, c_{jk}(x)) \\ U_\nu(d_j, x) &= U_\nu(d_k, x) && \nu = c_{jk}(x) \\ U_\nu(d_j, x) &< U_\nu(d_k, x) && \forall \nu \in (c_{jk}(x), \infty). \end{aligned}$$

That is, we require that the DM's ranking of alternatives is monotone in  $\nu$ : if a DM with a certain degree of risk aversion prefers a safer (riskier) asset to a riskier (safer) one, then all DMs with higher (lower) risk aversion also prefer the safer (riskier) asset.<sup>16</sup>

Full consideration is maintained in this subsection: each DM considers *all* alternatives in the choice set and chooses the one with highest utility (that is, consideration and choice sets coincide). Assumption 1 is a data requirement which guarantees this model's identification: There must be sufficient variation in a utility-relevant characteristic(s) to move the cutoffs (the single crossing points in Definition 1) through the support for the preference coefficient.

**Assumption 1** (Large Support). *For all  $\nu \in \Gamma$  there exists  $x \in \mathcal{X}$  and alternative  $j$  such that either: (1)  $c_{j,k}(x)$  exists for all  $k \neq j$  and  $\nu = \min_{k \neq j} c_{j,k}(x)$ ; or (2)  $c_{k,j}(x)$  exists for all  $k \neq j$  and  $\nu = \max_{k \neq j} c_{k,j}(x)$ .*

**Theorem 1.** *Suppose Assumption 1 holds. Then  $F(\cdot)$  is identified.*

<sup>15</sup>We assume that while  $\nu$  has bounded support, the utility function is well defined for any real valued  $\nu$ .

<sup>16</sup>Since we allow the cutoffs to be infinite, our regularity condition does not exclude strongly dominated choices, i.e. situations in which  $d_k$  is preferred to  $d_j$  for all values of  $\nu$ . In the context of risk preferences this definition of strong dominance is equivalent to first order stochastic dominance. When  $u(\cdot)$  is restricted to the class of concave utility functions strong dominance is equivalent to second order stochastic dominance.

*Proof.* Fix any  $\nu \in \Gamma$ . Find  $x$  and alternative  $j$  such that  $\nu = \min_{k \neq j} c_{j,k}(x)$  or  $\nu = \max_{k \neq j} c_{k,j}(x)$ . Then one of the following is true:

$$F(\nu) = F\left(\min_{k \neq j} c_{j,k}(x)\right) = Pr(d = d_j|x)$$

$$F(\nu) = F\left(\max_{k \neq j} c_{k,j}(x)\right) = 1 - Pr(d = d_j|x).$$

Since  $Pr(d = d_j|x)$  is identified by the data,  $F(\nu)$  is identified. □

Theorem 1 is akin to the “full-support” identification result (Chamberlain, 1986; Heckman, 1990; Lewbel, 2016) and the intuition is straightforward: there must be sufficient variation in the underlying exogenous characteristics to trace out the distribution of  $\nu$  over its entire support. Some variant of Assumption 1 is also necessary for identification. If, for example, there exists an interval  $[\nu_*, \nu^*] \subset \Gamma$  such that for all  $k, j$  there is no  $x$  with  $c_{k,j}(x) \in [\nu_*, \nu^*]$ , then  $F(\cdot)$  will not be identified in this interval. Simply put, the data does not provide any information about the distribution of the preference coefficient in this region.

Next, we present two models of limited consideration. Each of them has the same underlying primitives as the benchmark model, except the consideration set formation is stochastic.

### 3.3 Alternative Specific Random Consideration Model

In 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. These probabilities are assumed to be the same across DMs. We note that without loss of generality  $\varphi_j$  can be interpreted as a function of exogenous characteristics (such as advertisement) that are not utility relevant. In such a case, all of the results below should be interpreted as conditional on a given value of these characteristics.<sup>17</sup> Once the consideration set is drawn, the DM chooses the best alternative according to her preferences. Given that each alternative is considered probabilistically, it is possible that none of the alternatives enter the consideration set. In particular, with probability  $\prod_{k=1}^D (1 - \varphi_k)$  the consideration set is empty. Hence, to close the model, we require a completion rule specifying the behavior of the DM in the case of non-consideration. We

---

<sup>17</sup>More so, these characteristics may include a strict subset of  $x$ . As explained later, we only need one element of  $x$  to have certain properties and be consideration irrelevant.

offer four possible completion rules, each suited for application in different market settings.

**Coin Toss:** Assume  $\varphi_j < 1$  for all  $j$ . Then there is positive probability that the DM does not consider any alternative. In such a case, the DM randomly uniformly picks one alternative from the choice set, i.e. each alternative has probability  $\frac{1}{D}$  of being chosen. Coin Toss is consistent with scenarios in which DMs must choose an alternative (e.g. a deductible when buying home insurance), but lack the desire or ability to meaningfully evaluate them.<sup>18</sup>

**Default Option:** Assume  $\varphi_j < 1$  for all  $j$ . If no alternative is considered, the DM chooses a preset alternative. This completion rule is applicable to scenarios where, without the DM's active choice, she is assigned a pre-specified alternative from the choice set (e.g. employer provided benefits such as 401k allocations and medical insurance).

**Preferred Options** (Manski, 1977): Some alternative(s) is (are) always considered, i.e.  $\varphi_j = 1$  for some  $j$ . The identity of these alternatives does not have to be known to the researcher. However, if there exist multiple  $j$ 's such that  $\varphi_j = 1$ , then these alternatives should be adjacent to each other in the following sense. If there exists an  $x$  such that for all  $\nu \in \Gamma$  some non-preferred alternative dominates a preferred alternative, then it also dominates all other preferred alternatives. This completion rule captures market scenarios in which some alternatives are always discussed or emphasized by the sellers.

**Outside Option** (Manzini & Mariotti, 2014): Assume  $\varphi_j < 1$  for all  $j$ . The first interpretation of this rule is that all DMs who draw the empty set exit the market and are not part of the data. A second interpretation of this rule is as follows. If the empty consideration set is drawn, then the DM redraws a consideration set according to Equation (1) below. The DM continues to draw consideration sets until a non-empty set is obtained.

For all completion rules, the probability that the consideration set takes realization  $\mathcal{K}$  is

$$p(\mathcal{K}) \equiv \prod_{d_k \in \mathcal{K}} \varphi_k \prod_{d_k \in \mathcal{D} - \mathcal{K}} (1 - \varphi_k), \quad \forall \mathcal{K} \subset \mathcal{D}. \quad (1)$$

The differences in completion rules appear in the formulation of the likelihood function. A computationally appealing way to write the likelihood function is to determine the probability that a DM with preference coefficient  $\nu$  chooses alternative  $d_j$  conditional on characteristics  $x$ . Suppose the consideration set is not empty. Then, if  $d_j$  is chosen, it is in the

---

<sup>18</sup>In a classical IO setting, this type of completion rule is consistent with, for example, a shopper randomly choosing a chip packet from the shelf without carefully evaluating the utility derived from consuming various flavors of chips.

consideration set and every alternative that dominates it is not. Denote  $\mathcal{B}_\nu(d_j, x)$  the set of alternatives that dominate  $d_j$  for a DM with preference coefficient  $\nu$  and characteristics  $x$ :

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

It follows that for the first three completion rules

$$Pr(d_j|x, \nu) = \varphi_j \prod_{d_k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) + r_j,$$

where  $r_j$  is the term that accounts for the possibility of an empty consideration set. Under Coin Toss,  $r_j = \frac{1}{D} \prod_{d_k \in \mathcal{D}} (1 - \varphi_k)$ . Under Default Option,  $r_j = \prod_{d_k \in \mathcal{D}} (1 - \varphi_k)$  if  $j$  is the default alternative and is zero otherwise. Finally, under Preferred Options,  $r_j = 0$ . Integrating over  $\nu$  we have that

$$Pr(d_j|x) = \int Pr(d_j|x, \nu) dF = \varphi_j \int \prod_{d_k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) dF + r_j.$$

Similarly, under Outside Option, we have that

$$Pr(d_j|x) = \int Pr(d_j|x, \nu) dF = \frac{1}{1-r} \varphi_j \int \prod_{d_k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) dF,$$

where  $r \equiv \prod_{d_k \in \mathcal{D}} (1 - \varphi_k)$ .

We emphasize that these expressions for  $Pr(d_j|x)$  do not require enumerating all possible consideration sets, which for large choice sets can be hard if not infeasible. Computation of  $Pr(d_j|x)$  simply comes down to evaluating

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

Given  $\varphi$ , the integrand  $\prod_{d_k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k)$  is piecewise constant in  $\nu$  with at most  $D - 1$  breakpoints, corresponding to indifference points between alternatives  $j$  and  $k$  (i.e.  $c_{j,k}(x)$  or  $c_{k,j}(x)$ ). 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  $\mathcal{I}(d_j|x)$  as a weighted sum:

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

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 consideration probabilities, they are computed only once. 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  $\mathcal{I}(d_j|x)$  using Riemann approximation:

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

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  $\nu_m$ ,  $m = 1, \dots, M$ .<sup>19</sup> These rankings determine  $\mathcal{B}_{\nu_m}(d_j, x)$ . The likelihood maximization is now a standard search routine over  $\{\varphi_k\}_{k=1}^D$  and density  $f(\cdot)$ . Our theory restricts  $f(\cdot)$  to the class of continuous functions. In practice, the search is over a class of non-parametric estimators for  $f(\cdot)$ .<sup>20</sup> If the density is parameterized, i.e.  $f(\nu_m) \equiv f(\nu_m, \theta)$ , then the maximization is over  $\{\varphi_k\}_{k=1}^D$  and  $\theta \in \Theta$ . Note also that  $\nu_m$ 's are the same across all DMs, further reducing computational burden.<sup>21</sup>

### 3.4 Random Consideration Level Model

In the Random Consideration Level (RCL) Model, the consideration set forms in two steps. In the first step, each DM draws a consideration level,  $l$ , that is independent of the preference coefficient  $\nu$ . The consideration level determines the size of the consideration set and it takes discrete values in  $\{1, \dots, D\}$  with probability  $\phi_l$  such that  $\sum_{l=1}^D \phi_l = 1$ . In the second step, the consideration set is formed by drawing alternatives uniformly without replacement from

<sup>19</sup>The resulting computational gains are exploited in importance sampling (e.g., Ackerberg, 2009).

<sup>20</sup>One could, for example, use normalized B-splines or a mixture of flexible distributions.

<sup>21</sup>Depending on the class of  $f(\cdot)$ , it may be more accurate to compute  $\mathcal{I}(d_j|x)$  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 choice set until the DM obtains a set with cardinality equal to the consideration level  $l$ . The probability that the consideration set takes realization  $\mathcal{K}$  of size  $l$  is  $p(\mathcal{K}|l) = \binom{D}{l}^{-1}$ .

A computationally appealing way to write the likelihood function is as follows. First, consider a DM with a preference coefficient  $\nu$  and characteristics  $x$  and suppose that  $d_j$  is her  $m^{\text{th}}$ -best alternative. The probability that  $d_j$  is chosen is given by

$$Q_{l,m} = \begin{cases} \frac{\binom{D-m}{l-1}}{\binom{D}{l}} & \text{if } 1 \leq m \leq D - l + 1 \\ 0 & \text{otherwise.} \end{cases}$$

Denote  $\mathbf{b}_\nu(d_j, x)$  the number of alternatives that dominate  $d_j$  for an individual with preference coefficient  $\nu$  and characteristics  $x$ :  $\mathbf{b}_\nu(d_j, x) \equiv \text{card}(\mathcal{B}_\nu(d_j, x))$ .<sup>22</sup> We can write

$$Pr(d_j|x) = \int Pr(d_j|x, \nu) dF = \int \sum_{l=1}^D \phi_l Q_{l, 1+\mathbf{b}_\nu(d_j, x)} dF. \quad (2)$$

We employ similar techniques as those in Section 3.3 to compute the integral in Equation 2.

## 4 Identification

### 4.1 Identification: An Example

Recall our example in Section 3.1. Suppose there are only two alternatives:  $d_1$  is the high deductible and  $d_2$  is the low deductible. From Theorem 1 it is clear that to identify the model under full consideration we need enough variation in  $\bar{p}$  (and/or  $\mu$ ) such that the cutoff  $c_{1,2}(x)$  covers the entire support of the preference coefficient. Here, this variation is sufficient to identify *both* the consideration parameters and the distribution of the risk preferences.

We start with the ARC Model under one of the first three completion rules.<sup>23</sup> For each value of  $x$  we have a single moment identified by the data:

---

<sup>22</sup>In the ARC model the identity of the alternatives dominating  $j$  matters, while in this model only the number of dominating alternatives matter. The reason is that here all alternatives have equal probability of being considered. E.g., suppose that  $D$  is 5 and  $j$  is the second best alternative, so that  $\mathbf{b}(j|x, \nu) = 1$ . The second-best alternative is never chosen under full consideration, so that  $Q_{5,1} = 0$ . Under consideration level 4, the second best is chosen when the first best is not considered, which happens with probability  $\frac{1}{5}$ , that is  $Q_{4,1} = \frac{1}{5}$ . Under consideration level 1 an alternative is chosen iff it is in the consideration set, i.e.  $Q_{1,1} = \frac{1}{5}$ .

<sup>23</sup>Identification under the Outside Default follows similar reasoning.



$$Pr(d = d_1|x) = \varphi_1\varphi_2F(c_{1,2}(x)) + \varphi_1(1 - \varphi_2) + r_1. \quad (3)$$

The first term on the RHS of the Equation 3 captures the case where both alternatives are considered, and hence  $d_1$  is chosen only if the preference coefficient is below the cutoff. The second term captures the case where only  $d_1$  is considered, and thus it is chosen for all values of  $\nu$ . The third term is zero if at least one option is always considered, and otherwise captures the possibility that the consideration set is empty, and, depending on the completion rule,  $r_1$  is equal to either  $\frac{1}{2}(1 - \varphi_1)(1 - \varphi_2)$  or  $(1 - \varphi_1)(1 - \varphi_2)$ . Hence, rather than having a one-to-one mapping between  $Pr(d = d_1|x)$  and  $F(c_{1,2}(x))$  that would identify the latter as in Theorem 1, we have two additional unknown parameters. However, when Assumption 1 holds, we can find  $x^0$  and  $x^1$  such that  $c_{1,2}(x^0) = 0$  and  $c_{1,2}(x^1) = \bar{\nu}$ . This implies that the consideration parameters are the solution to the following system of equations:

$$\begin{aligned} Pr(d = d_1|x^0) &= \varphi_1(1 - \varphi_2) + r_1 \\ Pr(d = d_1|x^1) &= \varphi_1 + r_1. \end{aligned}$$

It is straightforward to show that this system has a unique solution. Hence, identification relies on the assumption that variation in  $x$  is sufficient to generate values for the cutoff  $c_{1,2}(x)$  at the extremes of the support for the preference coefficient, which is also needed for identification in the model with full consideration as discussed in Section 3.2.<sup>24</sup> Once the consideration parameters are known, identification of  $F(\cdot)$  follows from Equation (3), as long as both  $\varphi_1$  and  $\varphi_2$  are strictly positive.<sup>25</sup>

Similarly, under the RCL Model, for each value of  $x$  we have a single moment identified by the data:

$$Pr(d = d_1|x) = \phi_2F(c_{1,2}(x)) + \frac{1}{2}\phi_1 = (1 - \phi_1)F(c_{1,2}(x)) + \frac{1}{2}\phi_1. \quad (4)$$

Again, under Assumption 1, we can drive  $c_{1,2}(x)$  either to zero or to  $\bar{\nu}$ , which turns the expression above into an equation with one unknown, namely  $\phi_1$ . Hence the consideration

---

<sup>24</sup>Since observed variation in characteristics  $x$  identifies the distribution of a latent variable,  $x$  is referred to as the Lewbel special regressor (Lewbel, 2000, 2014).

<sup>25</sup>If either  $\varphi_1 = 0$  or  $\varphi_2 = 0$  then choice frequencies do not depend on  $x$  and nothing can be learned about the distribution of  $\nu$  as  $F(c_{1,2}(x))$  drops out of Equation (3).

parameters are identified. As long as  $\phi_1 < 1$ , identification of  $F(\cdot)$  follows.

The notable difference between Equations (3) and (4) is that the latter contains only one consideration parameter, while the former contains two. The reason follows from the restriction that the  $\phi$ 's sum to one, while no restriction is imposed on the sum of  $\varphi$ 's. To compensate for this missing moment condition in the ARC Model, the identification argument requires the use of an additional moment. This is the reason why identification of the ARC model will require somewhat stronger conditions.

In sum, using values of  $x$  that put the cutoff at the extremes of the preference-coefficient space allows for identification of the consideration parameters. Once the consideration parameters are known, variation in  $x$  pins down the preference-coefficient distribution. In the next two sections we proceed with formal arguments for identification of limited consideration parameters in both models. The conditions for identification of the preference-coefficient distribution are described in Section 4.3.

## 4.2 Identification of Consideration Parameters

We begin with the ARC model. We relegate all proofs to Appendix A.

**Theorem 2.** *Consider the Coin Toss or Outside Option completion rule. Suppose that there exist  $x^0, x^1$ , and a non-identity permutation  $\{o_1, o_2, \dots, o_D\}$  of the choice set such that  $\forall \nu \in [0, \bar{\nu}]$*

$$\begin{aligned} L_1(x^0) \succ L_2(x^0) \succ \dots \succ L_D(x^0), \\ L_{o_1}(x^1) \succ L_{o_2}(x^1) \succ \dots \succ L_{o_D}(x^1). \end{aligned}$$

*Then the consideration parameters  $\{\varphi_1, \varphi_2, \dots, \varphi_D\}$  are identified.*

While it appears that Theorem 2 (and the other results in this section) make use of the particular ordering of alternatives at  $x^0$ , the theorem can be stated with respect to any ordering of the available alternatives. The intuition for Theorem 2 is as follows. We need to identify  $D$  parameters. Since the preference ordering is deterministic at  $x^0$ , the observed choice frequencies provide  $D - 1$  distinct moments. The last *distinct* moment is obtained from the choice frequency evaluated at  $x^1$  for an alternative that moved position in the preference order (guaranteed by the permutation). Note that the conditions of the theorem allow for the presence of both dominated and dominating choices (choices that are better

or worse than another alternative(s) regardless of the value of  $x$ ). For example, in both preference orderings the best (or the worst)  $D - 2$  choices may be the same at  $x^0$  and  $x^1$ , but the remaining two alternatives switch places.

**Theorem 3.** *Consider the Default Option completion rule. Denote  $d_n$  the default option. Suppose that: (1) There exist  $x^0, x^1$  and a non-identity permutation  $\{o_1, o_2, \dots, o_D\}$  of the choice set such that  $\forall \nu \in [0, \bar{\nu}]$*

$$\begin{aligned} L_1(x^0) \succ L_2(x^0) \succ \dots \succ L_D(x^0), \\ L_{o_1}(x^1) \succ L_{o_2}(x^1) \succ \dots \succ L_{o_D}(x^1). \end{aligned}$$

(2) *There exists an alternative  $d_j \neq d_n$  such that  $L_n(x^0) \succ L_j(x^0)$ ,  $L_j(x^1) \succ L_n(x^1)$ , and  $Pr(d_j|x^1) > 0$ .<sup>26</sup> Then the consideration parameters  $\{\varphi_1, \varphi_2, \dots, \varphi_D\}$  are identified.*

The only difference between the conditions in Theorems 2 and 3 is in the latter case we additionally require an alternative, which is considered with positive probability, that switches rankings with the default option between  $x^0$  and  $x^1$ . This is necessary to obtain information about the consideration parameter for the default option. To see why this must be the case, suppose that the default option is  $d_D$ . Given the ranking of options at  $x = x^0$ , it is immediate to see how  $\varphi_1, \dots, \varphi_{D-1}$  are identified sequentially since  $Pr(d = d_j|x^0) = \varphi_j \prod_{k < j} (1 - \varphi_k)$ .<sup>27</sup> However, we cannot learn  $\varphi_D$  since the last moment at  $x^0$  is redundant and in particular does not reveal any information about  $\varphi_D$ :

$$Pr(d = d_D|x^0) = \varphi_D \prod_{k < D} (1 - \varphi_k) + r = \varphi_D \prod_{k < D} (1 - \varphi_k) + \prod_{k \leq D} (1 - \varphi_k) = \prod_{k < D} (1 - \varphi_k).$$

Now suppose  $d_D$  is dominated by all other alternatives at  $x^1$  (so that there does not exist a  $d_j \neq d_n$  satisfying the assumption in Theorem 3). By the same logic as above,  $Pr(d = d_D|x^1)$  does not reveal  $\varphi_D$ .<sup>28</sup>

**Theorem 4.** *Consider the Preferred Options completion rule. Denote  $d_{\underline{n}}, d_{\underline{n}+1}, \dots, d_{\bar{n}}$  to be the preferred options for some  $\underline{n}$  and  $\bar{n}$ . Suppose that: (1) There exist  $x^0, x^1$  and a*

<sup>26</sup>The condition  $Pr(d_j|x^1) > 0$  (or  $Pr(d_j|x^0) > 0$ ) is equivalent to assuming  $\varphi_j > 0$ . A restriction on the data is testable, so the assumption  $Pr(d_j|x^1) > 0$  is more appealing.

<sup>27</sup>We have that  $\varphi_1 = Pr(d = d_1|x^0)$ ,  $\varphi_2 = \frac{Pr(d = d_2|x^0)}{1 - \varphi_1}$ , etc.

<sup>28</sup>Of course, in this example, if  $d_D$  is always dominated, one may not care about learning  $\varphi_D$ , since it does not affect the probability of any other alternative being chosen.

non-identity permutation  $\{o_1, o_2, \dots, o_D\}$  of the choice set such that  $\forall \nu \in [0, \bar{\nu}]$

$$\begin{aligned} L_1(x^0) \succ L_2(x^0) \succ \dots \succ L_D(x^0), \\ L_{o_1}(x^1) \succ L_{o_2}(x^1) \succ \dots \succ L_{o_D}(x^1). \end{aligned}$$

(2) That  $\forall j > \bar{n}$  we have  $j' < \min_{n' \in \{\underline{n}, \dots, \bar{n}\}} n'$  where  $o_{j'} = j, o_{\underline{n}} = \underline{n}, \dots, o_{\bar{n}} = \bar{n}$ . Then the consideration parameters  $\{\varphi_1, \varphi_2, \dots, \varphi_D\}$  are identified.

The additional restriction for identification under Preferred Options completion rule vis-à-vis Coin Toss is that all alternatives must dominate the preferred options at either  $x^0$  or  $x^1$ . The second condition in Theorem 4 identifies the highest ranked preferred option at  $x^0$ , namely  $d_{\underline{n}}$ . Since  $\varphi_{\underline{n}}$  is equal to one, for  $j > \underline{n}$  the choice frequency is zero at  $x^0$ .<sup>29</sup> If, contrary to the second condition,  $d_j$  is also dominated by a preferred option at  $x^1$ , then by the same logic its choice frequency is zero and hence the consideration parameter  $\varphi_j$  is not identified.

We close this section by making two remarks. First, Theorems 2–4 yield the following condition that guarantees identification under any completion rule:

**Corollary 1.** *If there exist  $x^0$  and  $x^1$  such that  $\forall \nu \in [0, \bar{\nu}]$*

$$\begin{aligned} L_1(x^0) \succ L_2(x^0) \succ \dots \succ L_D(x^0), \\ L_D(x^1) \succ L_{D-1}(x^1) \succ \dots \succ L_1(x^1), \end{aligned}$$

*then the consideration parameters  $\{\varphi_1, \varphi_2, \dots, \varphi_D\}$  are identified under all completion rules.*

Second, Theorems 2–4 are indeed only sufficient: depending on the completion rule and  $x$ 's, there are other conditions that yield identification. For example, the following theorem follows from the proofs of Theorems 2 and 3:

**Theorem 5.** *Consider the Coin Toss, Default Option or Outside Option completion rule. If there exist  $x^0$  and  $x^1$  such that  $\forall \nu \in [0, \bar{\nu}]$*

$$\begin{aligned} L_1(x^0) \succ L_2(x^0) \succ \dots \succ L_D(x^0), \\ L_j(x^1) \succ L_1(x^1) \quad \forall j \neq 1, \quad (\text{or } L_D(x^1) \succ L_j(x^1) \quad \forall j \neq D), \end{aligned}$$

*then the consideration parameters  $\{\varphi_1, \varphi_2, \dots, \varphi_D\}$  are identified.*

---

<sup>29</sup>This follows because  $Pr(d = d_j | x^0) = \varphi_j \prod_{k < j} (1 - \varphi_k) = \varphi_j \times 0 = 0$ .

More generally, the identification of consideration parameters comes down to the following:

**Theorem 6.** *Suppose there exist  $\{x^0, x^1, \dots, x^M\}$  and  $S_j^m \subset \{1, 2, \dots, D\}$  such that  $\forall j \in \{1, \dots, D\}$  and  $\forall \nu \in [0, \bar{\nu}]$*

$$L_i(x^m) \succ L_j(x^m) \succ L_k(x^m), \quad \forall i \in S_j^m \ \& \ \forall k \in D \setminus S_j^m,$$

*then the consideration parameters  $\{\varphi_1, \varphi_2, \dots, \varphi_D\}$  form a system of equations in  $D$  unknowns. If this system admits a unique solution, then consideration parameters are identified.*

The theorem requires that for each alternative  $d_j$  there is some  $x^m$  that preserves the ranking of  $d_j$  relative to all other alternatives regardless of the value of the preference parameter. When this is the case, the observed choice frequency of alternative  $d_j$  conditional on  $x^m$  is a function of the consideration parameters, but not the preference distribution. Hence, for each  $d_j$  we obtain an equation(s) in consideration parameters. If the system of these equations has a unique solution, identification follows. Finally, each set of assumptions in Theorems 2, 3, and 4 guarantee that the aforementioned system of equations exists and that it has a unique solution.

The identifying conditions for the RCL model are similar to those of the ARC model. The conditions are, however, weaker as the RCL model imposes the additional restriction that the consideration parameters must sum to one:  $\sum_{j=1}^D \phi_j = 1$ . For example, the following theorem is the analog of Theorem 2:

**Theorem 7.** *If there exist  $x^0$  such that  $\forall \nu \in [0, \bar{\nu}]$*

$$L_1(x^0) \succ L_2(x^0) \succ \dots \succ L_D(x^0),$$

*then the consideration parameters  $\{\phi_1, \phi_2, \dots, \phi_D\}$  are identified.*

### 4.3 Identification of the preference-coefficient distribution

To set the stage, it is useful to extend our example in Section 4.1 to the case of three alternatives:  $d_1$  is the high deductible,  $d_2$  is the medium deductible, and  $d_3$  is the low deductible. We have that

$$Pr(d = d_1|x) = \varphi_1\varphi_2\varphi_3F(\min\{c_{1,2}(x), c_{1,3}(x)\}) + \varphi_1\varphi_2(1 - \varphi_3)F(c_{1,2}(x)) + \varphi_1\varphi_3(1 - \varphi_2)F(c_{1,3}(x)) + \varphi_1(1 - \varphi_2)(1 - \varphi_3) + r_1. \quad (5)$$

The first term in the sum above captures the case where all three alternatives are considered, and hence alternative  $d_1$  is chosen only if the preference coefficient is below both  $c_{1,2}(x)$  and  $c_{1,3}(x)$ . The second term is the case that alternatives  $d_1$  and  $d_2$  are considered, but alternative  $d_3$  is not considered, so that alternative  $d_1$  is chosen only if the preference coefficient is below the cutoff between alternatives  $d_1$  and  $d_2$ . Only alternatives  $d_1$  and  $d_3$  are considered in the third term, only  $d_1$  is considered in the fourth term, and no alternative are considered in the last term. Note that even though the consideration parameters are point identified (and hence  $\varphi_j$  can be treated as data), one moment of the data,  $Pr(d = d_1|x)$ , is associated with  $F(\cdot)$  evaluated at two different points,  $c_{1,2}(x)$  and  $c_{1,3}(x)$ .<sup>30</sup> If we had variation in  $x$  that effectively shut downs one of the cutoffs (e.g. it drives  $c_{1,3}(x)$  to either zero or to  $\bar{\nu}$ ) without affecting the other cutoff, then we would restore a one-to-one mapping between a moment in the data and a value of  $F(\cdot)$  at a single cutoff. In certain markets this type of variation is possible: For example, the price of the lowest deductible alternative is sufficiently large so that the alternative is strictly dominated. In the insurance context, however, it is rare to observe this type of variation, as the prices for all alternatives tend to move together. We show in Theorem 8 that  $F(\cdot)$  is identified under much weaker conditions, that do not rely on independent variation in characteristics of single alternatives. The intuition for our result can be gleaned from Equation (5). Suppose we start with a value for the characteristics  $\tilde{x}^0$  such that  $c_{1,2}(\tilde{x}^0)$  is close to the boundary, with  $c_{1,2}(\tilde{x}^0) < \bar{\nu}$  but  $c_{1,3}(\tilde{x}^0) > \bar{\nu}$ . Then, since  $F(c_{1,3}(\tilde{x}^0)) = 1$ ,  $Pr(d = d_1|\tilde{x}^0)$  pins down  $F(c_{1,2}(\tilde{x}^0))$ . Next take  $\tilde{x}^1$  such that  $c_{1,3}(\tilde{x}^1) = c_{1,2}(\tilde{x}^1)$ . Since  $F(c_{1,3}(\tilde{x}^1))$  is known,  $Pr(d = d_1|\tilde{x}^1)$  identifies  $F(c_{1,2}(\tilde{x}^1))$ . Repeat these steps to construct a sequence  $\{\tilde{x}^n\}_{n=1}^N$  such that  $c_{1,2}(\tilde{x}^N) \leq 0$ . For this approach to work, in addition to having sufficient variation in  $x$  to cover the support of  $\nu$ , we must also require that  $c_{1,3}(x)$  does not “catch up” to  $c_{1,2}(x)$  (i.e.  $c_{1,2}(x) < c_{1,3}(x)$  whenever  $c_{1,2}(x) \in \Gamma$ ),<sup>31</sup> so that our iteration reaches the other extreme of the support.

In sum, our strategy for identifying the preference-coefficient distribution is to (1) identify

---

<sup>30</sup>Bringing into the analysis another moment, e.g.  $Pr(d = d_2|x)$ , does not help as that brings with itself evaluation of  $F(\cdot)$  at another point,  $c_{2,3}(x)$ .

<sup>31</sup>That is, a DM with a preference coefficient in the interior of the parameter space cannot be indifferent between more than two alternatives.

the distribution of the preference coefficient close to one of the extremes of the support and then (2) move iteratively towards the other extreme. We summarize the variation in  $c_{j,k}(\cdot)$  induced by  $x$  that we need in the following assumption:

**Assumption 2** (Large Support for the Cutoff). *Let  $\underline{c}_j(x) \equiv \min_{k \neq j} c_{j,k}(x)$  and  $\bar{c}_j(x) \equiv \max_{k \neq j} c_{j,k}(x)$ . There exist  $x^0, x^1$ , and a continuous function  $x(t), x(t) \in \mathcal{X}, x(0) = x^0, x(1) = x^1, t \in [0, 1]$ , and an alternative  $j$  such that*

- |  |    |  |
|--|----|--|
| <p>A1. <math>c_{j,k}(x(t))</math> exists <math>\forall k</math>;</p> <p>A2. <math>\underline{c}_j(x^0) = 0</math> and <math>\underline{c}_j(x^1) = \bar{\nu}</math>;</p> <p>A3. <math>\arg \min_{k \neq j} c_{j,k}(x(t))</math> is<br/>unique <math>\forall t \in [0, 1]</math>.</p> | OR | <p>B1. <math>c_{k,j}(x(t))</math> exists <math>\forall k</math>;</p> <p>B2. <math>\bar{c}_j(x^0) = 0</math> and <math>\bar{c}_j(x^1) = \bar{\nu}</math>;</p> <p>B3. <math>\arg \max_{k \neq j} c_{j,k}(x(t))</math> is<br/>unique <math>\forall t \in [0, 1]</math>.</p> |
|--|----|--|

**Theorem 8** (Identification under Limited Consideration). *Suppose in the ARC model or in the RCL model the limited consideration parameters are identified and let Assumption 2 hold for some  $j$ . Furthermore, for the ARC model suppose for some  $k \neq j$  that  $\varphi_j > 0$  and  $\varphi_k > 0$ . For the RCL model suppose  $\phi_1 < 1$ . Then  $F(\cdot)$  is identified.*

Theorem 8 relies on variation in the choice probability of one particular alternative to identify  $F(\cdot)$ . Hence, for  $D > 2$ , both ARC and RCL models are over-identified and therefore testable: in either one  $Pr(d = d_m | x(t))$  is pinned down by given consideration parameters and  $F(\cdot)$ . If the model is correctly specified, it must then be that the predicted  $Pr(d = d_m | x(t))$  coincides with the data.

## 5 Models' Properties

### 5.1 Parallels with the RUM

We focus on a *standard* application of the RUM with full consideration in the context of our example in Section 3.1. The final evaluation of the utility that the DM derives from alternative  $j$  now includes a separately additive error term:

$$V_\nu(L_j(x)) = EU_\nu(L_j(x)) + \varepsilon_j, \quad (6)$$

where, as before,  $\nu$  captures unobserved heterogeneity in preferences. We emphasize that in the standard RUM  $\varepsilon_j$  is assumed independent of the random coefficients (in this application, the DM's risk-preference coefficient  $\nu$ ) as well as of the observable covariates (in this application,  $x = (\bar{p}, \mu)$ ).

Typical implementations of this model further specify that  $\varepsilon_j$  is identically and independently distributed across alternatives (and DMs) with a Type 1 Extreme Value distribution, following the seminal work of McFadden (1974). This yields a Mixed Logit model that differs from, for example, McFadden & Train (2000) because in the latter the random coefficient(s) enter the utility function linearly, while in the context of expected utility models the random preference coefficient(s) enter nonlinearly. We now discuss two properties of Model (6) that hinder its applicability to the analysis of random expected utility models, and then illustrate how models ARC and RCL are immune from these problems.

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 of RUM-predicted choice probabilities in the coefficient of risk aversion). *In Model (6) with  $\varepsilon_j$  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>32</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 the risk aversion increases, the DM eventually becomes indifferent between  $d_j$  and  $d_k$  and chooses either of these alternatives with equal probability (with probability equal to 0.5 when there are only two alternatives). As the risk aversion increases further, she prefers  $d_k$  to  $d_j$  and hence chooses the latter with lower probability. However, as the 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 each other, again 0.5 with two alternatives.<sup>33</sup> Hence, to “climb back” to 0.5, at some point the probability of choosing  $d_j$  becomes increasing in risk aversion. A careful anatomy of this phenomenon reveals that it originates with the variance of the additive error term  $\varepsilon_j$  being independent of  $\nu$ , a feature that is inescapable in Mixed Logit models.

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

---

<sup>32</sup>See also Wilcox (2008).

<sup>33</sup>Recall that in the Mixed Logit the magnitude of the utility differences is tied to differences in (log) choice probabilities,  $EU_\nu(L_k(x)) - EU_\nu(L_j(x)) = \log(\Pr(d = d_k|x, \nu)) - \log(\Pr(d = d_j|x, \nu))$ , so that as  $\nu \rightarrow \infty$  the choice probabilities are predicted to be all equal.



(1975) that is conditional on  $\nu$ :

**Definition 2.** (*Conditional Rank Order of Choice Probabilities*) *The model yields conditional rank order of the choice probabilities if for given  $\nu$  and for any DM and alternatives  $j, k \in \mathcal{D}$ ,*

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

The standard Mixed Logit model yields conditional rank ordering of the choice probabilities given  $\nu$ .<sup>34</sup> In turn, we show that the conditional rank order property implies the following upper bound on the probability that suboptimal alternatives are chosen:

**Property 2.** (*Generalized Dominance*) *Consider any characteristics  $x$ , alternative  $s$ , and set  $J \subset \mathcal{D} \setminus \{s\}$  satisfying:  $\forall \nu, \exists j_\nu \in J$  s.t.  $EU_\nu(L_s(x)) < EU_\nu(L_{j_\nu}(x))$ . Then*

$$Pr(d = d_s|x) < \sum_{k \in J} Pr(d = d_k|x).$$

Hence, in the standard Mixed Logit model, where the conditional rank order property holds, if for all preference coefficients an alternative  $s$  is dominated either by alternative  $j$  or by alternative  $k$ , then the probability of observing  $s$  is predicted to be less than the sum of the probabilities of observing  $j$  or  $k$ . We remark that neither  $j$  nor  $k$  is required to be optimal in  $\mathcal{D}$ , hence the upper bound in Property 2 is non-trivial.

## 5.2 Monotonicity in Models ARC and RCL

We now formally prove that both the ARC model and the RCL model yield predicted choice probabilities that are monotone in the coefficient of risk aversion. We begin by defining monotonicity for situations in which there are more than two alternatives in the choice set.

**Property 3.** (*Generalized Monotone Preference Property*) *Consider any  $x$  and suppose that  $c_{j,k}(x)$  exists for all  $1 \leq j < k \leq D$ . Then, for any  $\nu_1 < \nu_2$  and  $J \in \{1, 2, \dots, D\}$ :*

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

---

<sup>34</sup>Manski (1975) establishes the rank order property for additive error random utility models (without 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.

The property above states that when alternatives are ordered so that those with lower index are more risky (that is,  $c_{j,k}(x)$  exists for all  $1 \leq j < k \leq D$ ), the probability of choosing one of the  $J$  riskiest alternatives declines as the preference coefficient increases.

**Fact 1.** *The ARC and RCL Models satisfy the Generalized Monotone Preference Property.*

The proof of this fact (and of the ones stated in the next section) is given in Appendix B.

### 5.3 Ordinal Properties of Models ARC and RCL

In the Mixed Logit model, 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 whose variance cannot be a function of  $\nu$ . We now show that both of our models can be recast as an Ordinal Random Utility Model (ORUM) in which only the ordinal and not the cardinal ranking of alternatives based on their expected utility affects DMs' choices. In contrast to the Mixed Logit, we have:

**Fact 2.** *The ARC and RCL Models exhibit the following type of scale invariance: any multiplication of  $U_\nu(\cdot)$  by an arbitrary non-negative function of  $\nu$  leaves the model's predictions unchanged.*

Hence, to turn these models into models with additive error, the errors must have a very particular structure.

**Fact 3.** *(ARC Model as ORUM) The ARC Model is equivalent to an additive error random utility model with unobserved preference heterogeneity where all alternatives are considered, the DM's utility associated with 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 such that:

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

The error terms are independent of  $(x, \nu)$  and across alternatives. Ties, in case  $\varepsilon_j$  takes on  $-\infty$  value  $\forall j$ , are broken according to the completion rule as specified in Section 3.3.

**Table 1** Model Comparisons

	Mixed Logit	ARC	RCL
<b>Error Distribution</b>			
Support	$\mathbb{R}$	$\{-\infty, 0\}$	$\{-\infty, 0\}$
Independent of $x$	Yes	Yes	Yes
Independent of $\nu$	Yes	Yes	Yes
Independent across alternatives	Yes	Yes	No
Identical across alternatives	Yes	No	Yes
<b>Properties</b>			
Monotonicity	No	Yes	Yes
Conditional Rank Order Property	Yes	No	Yes
Generalized Dominance	Yes	No	Yes

**Fact 4.** (*RCL Model as ORUM*) The RCL Model is equivalent to an additive error random utility model with unobserved preference heterogeneity where all alternatives are considered, the DM's utility associated with 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 that takes two values: 0 and  $-\infty$ . The joint distribution of  $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_D)$  is as follows. For every realization  $e$  that has at least one zero element:

$$p(e) = \frac{\phi_l}{\binom{D}{l}}, \text{ where } l = \sum_k \mathbb{1}(e_k = 0).$$

and for  $e = \{-\infty, -\infty, \dots, -\infty\}$ :  $p(e) = 0$ .

The structure of the additive errors derived in Fact 3 and 4, respectively, allow us to learn which of these models satisfy the conditional rank order property, and hence the Generalized Dominance Property.

**Fact 5.** *The ARC Model does not (always) satisfy the Conditional Rank Order Property and, hence, the Generalized Dominance.*

**Fact 6.** *The RCL Model satisfies the Conditional Rank Order Property and, hence, Generalized Dominance.*

We summarize this section with Table 1, that lists the differences across the Mixed Logit,

ARC, and RCL models. The first panel lists the differences in the assumptions and the second panel lists the differences in implied properties.

## 6 Beyond the ARC and RCL Models

As shown above, the ARC and RCL models are complimentary in terms of consideration set formation if viewed through the lens of an error structure corresponding to an ORUM. Each model relaxes an assumption imposed on the error structure (either independence or identically distributed). Yet identification of these models rests on a similar logic. We conjecture that other consideration models that relax both the independence and identically distributed assumptions can also be identified. While it is beyond the scope of this paper to argue which consideration set formation is the right one in a given context, we offer two additional examples of consideration set formation based on well established economic/behavioral phenomena. The first example captures economic situations in which a DM will consider alternatives with an attribute that is below a certain DM specific threshold (Kimya, 2018). Within the insurance context, a threshold on the deductible level attribute will arise immediately if there are DM specific (unobserved) liquidity constraints: Anticipating that her liquidity constraint might bind if a loss occurs, a forward looking DM discards high deductible alternatives from the consideration set.<sup>35</sup> The second mechanism builds on the notion of extremeness aversion, one way Behavioral Economics and Marketing literature addresses the context dependency of preferences. Simonson & Tversky (1992) define extremeness aversion as the situation when “the attractiveness of an option is enhanced if it is an intermediate option in the choice set and is diminished if it is an extreme option”. In our framework, the relative location of the alternative in the choice set will determine its likelihood of being considered.

### 6.1 The Threshold Model

We return to our example in Section 3.1. Suppose that DMs find it prohibitively costly to have out-of pocket expenses above certain DM-specific limit. Hence there is an upper bound on what deductibles they consider. Formally, we assume that, in addition to the preference parameter  $\nu$ , each DM has an unobserved threshold parameter  $\underline{d}$ . The DM’s draw of the threshold parameter defines her consideration set. In particular, alternative  $d_k$

---

<sup>35</sup>For a discussion on how liquidity constraints affect households’ risk aversion see Chetty & Szeidl (2007).

is considered if and only if  $d_k < \underline{d}$ .<sup>36</sup> We assume that  $\nu$  and  $\underline{d}$  are independent conditional on observables (e.g., wealth, credit score, and age), and that the threshold parameter is continuously distributed  $G(d)$  with support  $(d_D, \infty)$ .

Let  $\xi_1 = 1 - G(d_1)$ ,  $\xi_2 = G(d_1) - G(d_2)$ ,  $\dots$ ,  $\xi_{D-2} = G(d_{D-3}) - G(d_{D-2})$ , and  $\xi_{D-1} = G(d_{D-2})$ . Then the fraction of DMs considering all deductibles is  $\xi_1$ , considering all but the highest deductible is  $\xi_2$ , and considering only the lowest deductible is  $\xi_D$ .  $F(\nu)$  and  $\{\xi_1, \xi_2, \dots, \xi_{D-1}\}$  are identified provided sufficient variation in  $p$  and/or  $\mu$ , as discussed above, exists. Indeed, consider DMs with  $x^0 = (p^0, \mu^0)$  such that the deductibles are ranked from the highest to the lowest for all  $\nu$  on the support. Then  $\xi_k$  is equal to  $Pr(d = d_k | x^0)$ , which is identified by the data. Identification of  $F(\nu)$  follows from similar arguments made in Theorem 8.

## 6.2 Extremeness Aversion Model

Consider again example in Section 3.1. To ease notation we assume that the number of alternatives in the choice set,  $D$ , is odd and we let  $m = \frac{D+1}{2}$ . All DMs consider the median alternative  $d_m$  with probability equal to one. The remaining alternatives are considered with probability that is decreasing in the distance from the median alternative. In particular, the consideration set is formed according to the following product rule:

$$Pr(d_k \text{ is considered}) = \begin{cases} \prod_{j=m+1}^k \xi_j & \text{if } k > m \\ \prod_{j=k}^{m-1} \xi_j & \text{if } k < m \end{cases}.$$

As before, consideration set formation is independent  $\nu$  conditional on observables. Suppose that there is sufficient variation in  $p$  and/or  $\mu$  so that there exists DMs with  $x^0 = (p^0, \mu^0)$  such that the deductibles are ranked from the highest to the lowest for all  $\nu$  on the support. Then  $\xi_k$  is equal to the ratio  $\frac{Pr(d_k | x^0)}{Pr(d_{k-1} | x^0)}$ , which is identified by the data.

# 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

---

<sup>36</sup>Recall that deductibles are decreasing in  $k$ , and hence  $d_D$  is the smallest deductible providing maximum coverage.

**Table 2** Premiums 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

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. We only consider their *first* purchases.<sup>37</sup> Table E.1 provides descriptive statistics for households’ observable characteristics, which we use later to estimate households’ preference coefficients.<sup>38</sup> For each household and each coverage we observe the exact menu of alternatives available at the time of the purchase. The deductible alternatives vary across coverages but not across households. Table E.2 presents the frequency of chosen deductibles in our data.

Premiums are determined coverage-by-coverage as in the example from Section 3.1. For each household, the company determines a baseline price  $\bar{p}$  using a coverage-specific rating function, which takes into account the household’s coverage-relevant characteristics and any applicable discounts. Given  $\bar{p}$ , the premium for alternative  $j$  is determined based on a coverage-specific rule,  $p_j = g_j \cdot \bar{p} + \delta$ . Table E.5 reports the average premium by context and deductible, and Table 2 summarizes the premium distributions for the \$500 deductible. As the latter table shows, premiums vary dramatically. In each coverage, the 99th percentile of the \$500 deductible is more than ten times the corresponding 1st percentile.

The underlying loss probabilities are derived from expected claim rates that are estimated using coverage-by-coverage Poisson-Gamma Bayesian credibility models applied to a large auxiliary panel. The unbalanced panel contains over 400,000 households from 1998 to 2007, yielding more than 1.3 million household-year observations for each coverage. We assume that household  $i$ ’s claims under coverage  $j$  in year  $t$  follow a Poisson distribution with arrival rate  $\lambda_{ijt}$ . We treat  $\lambda_{ijt}$  as latent random variables and assume that  $\ln \lambda_{ijt} = \mathbf{W}'_{ijt} \zeta_j + \epsilon_{ij}$ , where  $\mathbf{W}_{ijt}$  is a vector of observables,  $\epsilon_{ij}$  is an unobserved i.i.d. error term, and  $\exp(\epsilon_{ij})$

<sup>37</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>38</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 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

follows a Gamma distribution with unit mean and variance  $\eta_j$ .<sup>39</sup> Poisson panel regressions with random effects yield estimates of  $\zeta_j$  and  $\eta_j$  for each coverage  $j$ . For each household  $i$ , we use the regression estimates to generate a predicted claim rate  $\hat{\lambda}_{ij}$  for each coverage  $j$ , conditional on the household's *ex ante* characteristics  $\mathbf{W}_{ij}$  and *ex post* claims experience. In the model, we assume that households expect no more than one claim.<sup>40</sup> Hence, we transform  $\hat{\lambda}_{ij}$  into a predicted claim probability  $\hat{\mu}_{ij} = 1 - \exp(-\hat{\lambda}_{ij})$ . 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>41</sup>

## 7.2 The Model

The model is identical to the one in Section 3.1, augmented with either the ARC or the RCL Model. As in the example, the DM's problem amounts to choice over deductible lotteries of the form  $L_k(x) \equiv (-p_k, 1 - \mu; -p_k - d_k, \mu)$ , where  $x = (\bar{p}, \mu)$ . The utility function is assumed to be CARA. For all  $\nu > 0$ :

$$EU_\nu(L_k(x)) = -(1 - \mu)e^{-\nu(w-p_k)} - \mu e^{-\nu(w-p_k-d_k)} = -e^{-\nu w} [(1 - \mu)e^{\nu p_k} + \mu e^{\nu(p_k+d_k)}],$$

where  $w$  denotes the DM's initial wealth.<sup>42</sup> Note that  $e^{-\nu w}$  enters multiplicatively in the expression above and hence does not affect the rankings of the alternatives.

<sup>39</sup>We refer to this model as a Bayesian credibility model because  $\hat{\lambda}$  corresponds to the Bayesian credibility premium in the actuarial literature (Denuit, Maréchal, Pitrebois, & Walhin, 2007, Ch 3).

<sup>40</sup>The claim rates are small and, consequently, the likelihood of two or more claims is small. For home insurance 86.2% of predicted claim rates in the core sample are less than 0.1 and 97.4% percent are less than 0.15. For collision the frequencies are 79.8% and 98.6%, respectively. For comprehensive – 99.95% and 100%.

<sup>41</sup>See Barseghyan et al. (2013) (and Cohen & Einav (2007) in the context of Israeli auto insurance) for a detailed discussion of where such independent variation comes from.

<sup>42</sup>When  $\nu$  is zero, expected utility is simply  $-p_k - \mu d_k$ .

We now establish the assumptions required for the identification results in Corollary 1 and Theorem 8.<sup>43</sup> It is immediate to see that for any  $\nu$  on a compact support, when  $\bar{p}$  is sufficiently large and/or  $\mu$  is close to zero, the preference ordering is sequential:  $EU_\nu(L_1) > EU_\nu(L_2) > \dots > EU_\nu(L_D)$ . Alternatively, when  $\bar{p}$  is small (and  $\mu > 0$ ) or  $\mu$  is close to one (and  $\bar{p}$  is not very large) we have that  $EU_\nu(L_D) > EU_\nu(L_{D-1}) > \dots > EU_\nu(L_1)$ .

Turning to the identification of the preference-coefficient distribution, note that all cutoffs exist and are continuous functions of  $x = (\bar{p}, \mu)$ . It remains to show that  $\arg \min_{k \neq 1} c_{1,k}(x)$  is unique. To establish this, we show in Appendix C that  $c_{1,2}(x) < c_{1,m}(x)$  for any  $m > 2$ .

## 7.3 Estimation Results

### 7.3.1 The 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. We do so to illustrate the key features of our method and to ascertain that multiple contexts and demographic variables play no particular role in identification.

In this market there are no preset defaults for deductibles, which implies that Default Option is not a proper completion rule for these data. We assume the Coin Toss completion rule. The estimation under Coin Toss naturally encompasses Preferred Options – if estimated values of one or more  $\varphi_j$ 's turn out to be one, then we have Preferred Options.<sup>44</sup> To execute our estimation procedure we need to choose the upper bound of the preference-coefficient support. We set it to  $\bar{\nu} = 0.02$ , 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 AIC/BIC criteria indicate that a single component is sufficient for our analysis.

The estimated distribution and consideration parameters are reported in Table F.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 alternative (i.e., the distribution of optimal choices under full consideration). Predicted choices are

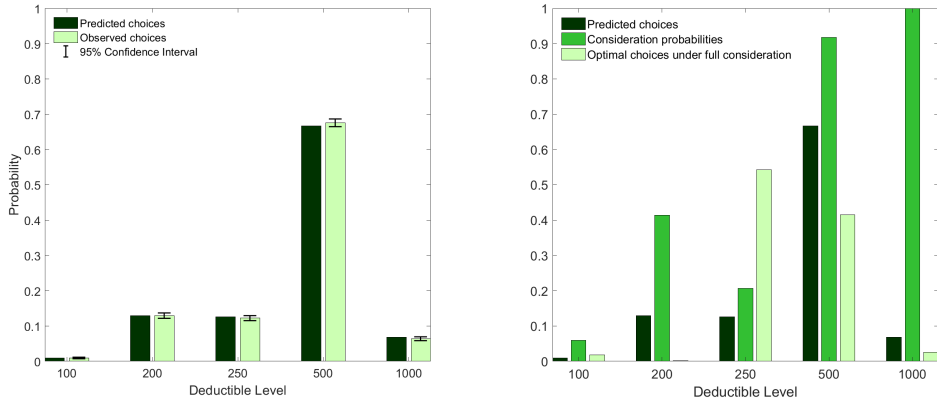
---

<sup>43</sup>In Appendix C, we show that these assumptions are satisfied also under CRRA.

<sup>44</sup>We could have also assumed Outside Default (under the second interpretation). The collision only results under this completion rule are nearly identical to those presented in the paper.



**Figure 1: The ARC Model**



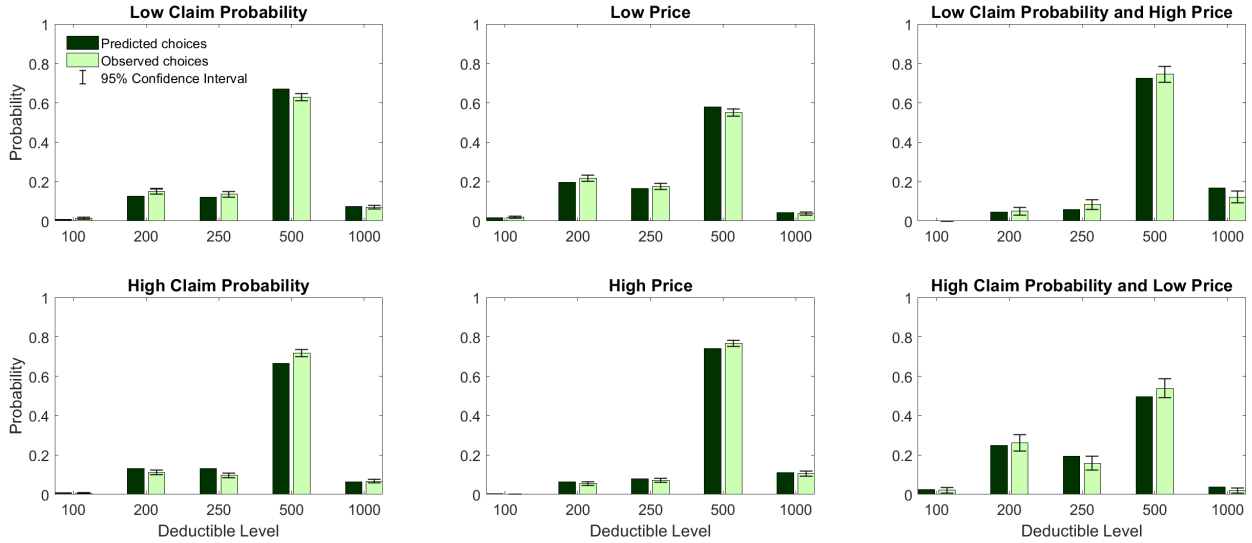
The first panel reports the distribution of predicted and observed choices. The second panel displays consideration probabilities and the distribution of optimal choices under full consideration.

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. It is noteworthy that the two highest deductibles (\$1,000 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 relatively 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 to compensate for this shift and to “push back up” the likelihood function. 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 0.999.<sup>45</sup> To put these numbers into context, a DM with risk aversion equal to 0.0037 is willing to pay \$424 to avoid a \$1,000 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 model's ability to match data extends also to conditional moments. The first two panels

<sup>45</sup>We cannot assume that all consideration probabilities are equal to one, since the \$200 deductible is dominated under full consideration and is chosen with positive probability.

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



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>46</sup> Finally, the last two panels capture 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 DMs 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, the \$200 collision deductible is always dominated either by the \$100 deductible or the \$250 deductible. This happens because of the particular pricing schedule in collision. It costs the same to get an additional \$50 of coverage by lowering 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>47</sup> Yet, overall the \$200 deductible is chosen roughly as often as the \$100 and \$250

<sup>46</sup>Low/high groups here are defined as households whose claim rate (or baseline price) are in the bottom/top third of the distribution.

<sup>47</sup>This pattern is at odds not only with EUT but also many non-EU models (Barseghyan et al., 2016).

deductibles combined. More so, for certain sub-groups the \$200 deductible is chosen *much more often* than the \$100 and \$250 deductible combined.<sup>48</sup> It follows that a model satisfying Generalized Dominance cannot rationalize these choices.

In the next step of our estimation analysis we relax the assumption that demographic variables do not influence risk preferences. 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>49</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>50</sup> The results of this estimation are in line with our first estimation. (See Column 2 in Table F.1, as well as Figures F.1 and F.2 in Appendix F.) 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.

### 7.3.2 The RCL Model and the RUM

For completeness, we now discuss estimation results for the RCL Model and the RUM with unobserved heterogeneity. In both cases, we assume that risk-preference coefficient is Beta distributed with support  $[0, \bar{\nu}]$ , where, as before,  $\bar{\nu} = 0.02$ . *A priori*, neither of these models should do well in matching the distribution of observed choices. Both of them satisfy

---

<sup>48</sup>For example, the fraction of households facing low  $\bar{\nu}$  and high  $\mu$  that choose the \$200 deductible is 0.26, while the fraction that choose the \$100 deductibles or the \$250 deductibles is 0.18.

<sup>49</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>50</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  $\tilde{\gamma} = -\gamma$ .

the Conditional Rank Order Property and have no ability to direct households’ choices in a particular direction. Instead, they smoothly spread households’ choices around their respective first bests: the closer the expected utility of a given alternative is to the expected utility of the first best, the higher the frequency at which it will be chosen.

Consequently, these models cannot match the observed distribution and, in particular, are unable to explain the relatively high observed share of the \$200 deductible. Table F.2 reports the estimation results for the RCL model and the RUM. Figure F.3 compares the observed distribution of choices and the predicted ones under both models. The predicted distributions are similar to each other, but are a much poorer fit to the data than that of the ARC Model.

To formally assess how well these models fit the data relative to the ARC, we rely on the Vuong test. The latter takes into account both the fact that the models are not nested and that they can have different number of parameters. The test soundly (at 1% level) rejects both the RCL and the RUM in favor of the ARC Model.

### 7.3.3 The ARC Model: All Coverages

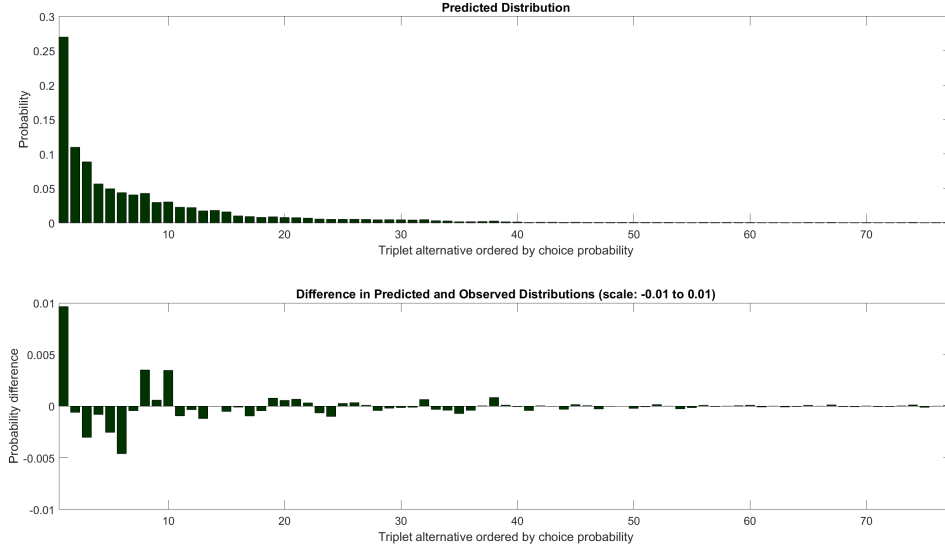
We now proceed with estimation of the full model. We consider two cases. In the first case households’ risk preferences are invariant across coverages, but consideration sets form independently within each coverage. There are three sets of consideration parameters  $\{\varphi^{coll}, \varphi^{comp}, \varphi^{home}\}$  and the probability that alternative  $k$  is considered in one coverage (e.g. collision) is independent of the probability that alternative  $j$  is considered in another coverage (comprehensive or home). Hence, within each coverage, the households’ problem is identical to that from the previous section.<sup>51</sup> The estimation results are presented in Figure F.4 and Table F.3. Just as in the case of collision coverage only, the model matches well the choice distributions within each coverage. However, the independence of consideration sets across coverages 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 deductibles and that 12% of households choose an alternative with a larger comprehensive deductible than collision deductible. In the data 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

---

<sup>51</sup>Effectively this scenario amounts to assuming “narrow bracketing” (Read, Loewenstein, & Rabin, 1999), a common approach in the literature. Note that under full consideration there is no loss of generality in assuming narrow bracketing. As it is well known, with CARA preferences the decision in one context is independent of the decisions in other contexts as long as loss events are mutually exclusive.

**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.

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 F.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 distributions is equal to 0.96 percentage points, which occurs at the  $(\$500, \$500, \$500)$  triplet that is chosen by 26% of the households. The *integrated* absolute error across all triplets is 4.61%. We note that in our data 43 out of 120 triplets are never chosen (these are omitted from Figure 3). It is straightforward to show analytically that likelihood maximization implies that the consideration probabilities for these triplets must be zero, so that their predicted shares are *de facto* zero.<sup>52</sup> Consequently, the likelihood maximization routine is faster and more reliable as we

<sup>52</sup>Since we are estimating the model with the Coin Toss completion rule, these options still can be chosen if the consideration set is empty and  $\varphi_j < 1$  for all  $j$ . In our estimation, the probability that the consideration set is empty is 0.0015, which implies that an alternative with zero consideration probability is chosen with probability  $0.0015/120=0.000013$ .

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 et al., 2012; Barseghyan et al., 2016). On the one hand, households' risk aversion relative to their peers is correlated across contexts, implying that households preferences have a stable component. On the other hand, analyses based on revealed preference reject the standard models of risk aversion: under full consideration, for the vast majority of households one cannot find a (household-specific) risk aversion parameter that can justify their choices simultaneously across all contexts. Relaxing full consideration allows to match the observed joint distribution of choices, and hence their rank correlations.

Estimated risk preferences are similar to those estimated with collision only data, although the variance is slightly smaller. Turning to consideration, the triplet considered far more frequently than any other alternative is the cheapest one: (\$1,000, \$1,000, \$1,000).<sup>53</sup> Its consideration probability is 0.81, while the next two most considered triplets are (\$500, \$500, \$1,000) 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 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, since our likelihood calculation does not require summation of probabilities over all possible consideration sets containing each household's choice.<sup>54</sup> Second, the model's parameters grow linearly with the choice set – one parameter per an additional alternative – which keeps the computations in check. 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 alternatives or hundred twenty would not make any difference neither from an identification nor 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 monetary cost of limited consideration. In our data, it is \$49 (see Appendix D).

---

<sup>53</sup>The first entry is for collision, the second is for comprehensive, and the third is for home.

<sup>54</sup>This is contrast to Goeree's (2008) method, which utilizes the logit structure and hence must keep track of all consideration sets containing the household's choice. In our setting, it is feasible to estimate an additive error RUM assuming the DMs consider each deductible triplet as a separate alternative (Figure F.5 and Table F.5). As the figure shows, the failure to match data is evident. The Vuong test formally rejects it in favor of the ARC model.

## 8 Beyond Expected Utility

Our framework can also be applied to non-EU models as long as risk attitudes are determined by a unidimensional index. Consider, e.g., the probability distortions model in Barseghyan et al. (2013).<sup>55</sup> Under this model, the expected utility of a lottery is evaluated using a distorted claim probability  $\Omega(\mu)$  instead of  $\mu$ :

$$EU(L_j(x)) = (1 - \Omega(\mu)) u(w - p_j) + \Omega(\mu) u(w - p_j - d_j).$$

Let  $u(\cdot)$  be linear across all DMs, and hence  $\Omega(\cdot)$  is the only source of risk aversion. Assume that for a given  $\mu$ ,  $\Omega(\mu)$  is randomly distributed across DMs with support  $[\underline{\Omega}, \overline{\Omega}] \subset [0, 1]$ . Since the SCP is trivially satisfied, identification of both the ARC and the RCL models follows under the same conditions as in Section 4. If, however,  $u(\cdot)$  is concave and varies across DMs, then there are two distinct sources of aversion to risk,  $\nu$  and  $\Omega(\mu)$ . While parametric identification of the joint distribution of  $\{\nu, \Omega(\cdot)\}$  under full consideration is straightforward, non-parametric identification is an open question (Barseghyan et al., 2018). If conditions can be derived for non-parametric identification under full consideration, then our identification strategy may be used to obtain identification under limited consideration as well.

## 9 Conclusion

In this paper we built a framework where DMs consider only a subset of the available alternatives. We offered two models with different consideration formation mechanisms and established their identification. There are many ways – such as liquidity constraints or extremeness aversion that we discussed – in which limited consideration may arise. While much effort in applied theory has been towards constructing non-EU models that can generate rankings of alternatives that are different from those in EUT, a promising and complementary avenue is to build and test theories that allow for limited consideration in the decision making process. There are many open questions. First, what economic forces determine the formation of consideration sets and how does consideration change with the market setting? Second, once we allow for limited consideration, how do our conclusions about the underlying models of risk change? Would we still need non-EU models to explain DMs’ behavior in real market situations, as it is commonly argued in the literature, and if yes which ones?

---

<sup>55</sup>In the context of binary lotteries this model incorporates many leading alternatives to EUT. See Barseghyan et al. (2018) for a discussion.

## References

- Abaluck, J., & Adams, A. (2018). What do consumers consider before they choose? Identification from asymmetric demand responses. *Working Paper*.
- 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.
- Apesteguia, J., & Ballester, M. A. (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy*, 126(1), 74–106.
- Apesteguia, 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.
- Barseghyan, L., Coughlin, M., Molinari, F., & Teitelbaum, J. (2019). Heterogeneous choice sets and preferences. *Working paper, Cornell University*.
- 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), 501–64.
- Barseghyan, L., Molinari, F., & Teitelbaum, J. C. (2016). Inference under stability of risk preferences. *Quantitative Economics*, 7(2), 367–409.
- 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.
- Ben-Akiva, M., & Boccara, B. (1995). Discrete choice models with latent choice sets. *International Journal of Research in Marketing*, 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.
- Caplin, A., Dean, M., & Leahy, J. (2018). Rational inattention, optimal consideration sets, and stochastic choice. *The Review of Economic Studies*.
- Cattaneo, M. D., Ma, X., Masatlioglu, Y., & Suleymanov, E. (2017). A random attention model. *arXiv preprint arXiv:1712.03448*.
- Chamberlain, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics*, 32(2), 189–218.
- Chetty, R., & Szeidl, A. (2007, 05). Consumption Commitments and Risk Preferences. *The Quarterly Journal of Economics*, 122(2), 831–877.
- Chiappori, P.-A., Salanié, B., Salanié, F., & Gandhi, A. (2018). From aggregate betting data to individual risk preferences. *Forthcoming in Econometrica*.
- Cohen, A., & Einav, L. (2007). Estimating risk preferences from deductible choice. *American Economic Review*, 97(3), 745–788.
- Conlon, C. T., & Mortimer, J. H. (2013). Demand estimation under incomplete product



- availability. *American Economic Journal: Microeconomics*, 5(4).
- Crawford, G., Grithz, R., & Iariax, A. (2017). Demand estimation with unobserved choice set heterogeneity. *Work in progress*.
- Dardanoni, V., Manzini, P., Mariotti, M., & Tyson, C. J. (2017). Inferring cognitive heterogeneity from aggregate choices. *Work in progress, University of St Andrews*.
- Denuit, M., Maréchal, X., Pitrebois, S., & Walhin, J.-F. (2007). John Wiley & Sons.
- 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.
- Eliasz, K., & Spiegler, R. (2011). On the strategic use of attention grabbers. *Theoretical Economics*, 6(1), 127–155.
- 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.
- Goeree, M. S. (2008). Limited information and advertising in the US 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.
- Hauser, J. R., & Wernerfelt, B. (1990). An evaluation cost model of consideration sets. *Journal of Consumer Research*, 16(4), 393-408.
- Heckman, J. (1990). Varieties of selection bias. *The American Economic Review*, 80(2), 313–318.
- Heiss, F., McFadden, D., Winter, J., Wuppermann, A., & Zhou, B. (2016). Inattention and switching costs as sources of inertia in medicare part d. *NBER WP 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.
- 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.
- Howard, J. A. (1977). *Consumer behavior: Application of theory*. New York: McGraw-Hill.
- Kimya, M. (2018). Choice, consideration sets and attribute filters. *American Economic Journal: Microeconomics*.
- 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. *Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, 38 – 62.
- Lewbel, A. (2016). The identification zoo—meanings of identification in econometrics. *Forthcoming*

- on *Journal of Economic Literature*, 7.
- 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. (2007). *Identification for prediction and decision*. Harvard University Press.
- Manzini, P., & Mariotti, M. (2014). Stochastic choice and consideration sets. *Econometrica*, 82(3), 1153-1176.
- Masatlioglu, Y., Nakajima, D., & Ozbay, E. Y. (2012). Revealed attention. *American Economic Review*, 102(5), 2183-2205.
- 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. *Frontiers in Econometrics*, 105–142.
- McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15, 447–470.
- Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2), 175–208.
- Read, D., Loewenstein, G., & Rabin, M. (1999). Choice bracketing. *Journal of Risk and Uncertainty*, 19(1–3), 171–197.
- 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.
- Shocker, A. D., Ben-Akiva, M., Boccara, B., & Nedungadi, P. (1991). Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions. *Marketing Letters*, 2(3), 181–197.
- Simon, H. A. (1959). Theories of decision-making in economics and behavioral science. *The American Economic Review*, 49(3), 253–283.
- 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.
- Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, 38(2), 332-382.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97 - 136.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological review*, 79(4), 281.
- Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. In *Risk aversion in experiments* (pp. 197–292). EGPL.

# Appendices

## A Identification Proofs

**Lemma A.1.** *Consider the ARC Model under the Coin Toss completion rule. If there exists characteristics  $x^L$  and permutation  $\{o_1, \dots, o_D\}$  such that  $\forall \nu \in [0, \bar{\nu}]$*

$$L_{o_1}(x^L) \succ L_{o_2}(x^L) \succ \dots \succ L_{o_D}(x^L),$$

then the following relationship holds for all  $j = 1, \dots, D$ :

$$\varphi_{o_j} = \frac{q_{o_j} - r}{1 + (j-1)r - \sum_{k=1}^{j-1} q_{o_k}},$$

where  $q_{o_k} \equiv Pr(d = d_{o_k} | x^L)$  and  $r = \frac{1}{D} \prod_{k=1}^N (1 - \varphi_{o_k})$ .

*Proof.* Without loss of generality suppose that  $k = o_k$  for all  $k$ . Fix any  $j \in \mathcal{D}$ . We first show that

$$\prod_{k=1}^{j-1} (1 - \varphi_k) = 1 + (j-1)r - \sum_{k=1}^{j-1} q_k.$$

On the one hand, by additivity of probability and the definition of  $q_k$

$$Pr(d \in \{d_1, d_2, \dots, d_{j-1}\} | x^L) = \sum_{k=1}^{j-1} q_k.$$

On the other hand, according to the model, it is the probability that at least one of  $\{d_1, d_2, \dots, d_{j-1}\}$  is considered plus the probability one of them is chosen when the consideration set is empty:

$$Pr(d \in \{d_1, d_2, \dots, d_{j-1}\} | x^L) = 1 - \prod_{k=1}^{j-1} (1 - \varphi_k) + (j-1)r.$$

Finally, due to the assumption of the preference ordering,

$$\begin{aligned} q_j &= \varphi_j \prod_{k=1}^{j-1} (1 - \varphi_k) + r \\ &= \varphi_j \left( 1 + (j-1)r - \sum_{k=1}^{j-1} q_k \right) + r. \end{aligned}$$

□

**Lemma A.2.** Consider the ARC Model under the Default Option completion rule. If there exist characteristics  $x^L$  and permutation  $\{o_1, \dots, o_D\}$  such that  $\forall \nu \in [0, \bar{\nu}]$

$$L_{o_1}(x^L) \succ L_{o_2}(x^L) \succ \dots \succ L_{o_D}(x^L),$$

then the following relationship holds for all  $j = 1, \dots, D$ :

$$\varphi_{o_j} = \begin{cases} \frac{q_{o_j} - r}{1 - \sum_{k=1}^{j-1} q_{o_k}} & \text{if } d_{o_j} \text{ is the default option} \\ \frac{q_{o_j}}{1 + r - \sum_{k=1}^{j-1} q_{o_k}} & \text{if } d_{o_n} \text{ is the default option for some } n < j, \\ \frac{q_{o_j}}{1 - \sum_{k=1}^{j-1} q_{o_k}} & \text{otherwise} \end{cases}$$

where  $q_{o_k} \equiv Pr(d = d_{o_k} | x^L)$  and  $r = \prod_{k=1}^N (1 - \varphi_{o_k})$ .

*Proof.* Let  $d_n$  denote the default option. The proof follows exactly the same steps as the proof of the previous lemma, except with the following two changes:

$$\begin{aligned} Pr(d \in \{d_1, d_2, \dots, d_{j-1}\} | x^L) &= \begin{cases} 1 - \prod_{k=1}^{j-1} (1 - \varphi_k) + r & \text{if } n < j \\ 1 - \prod_{k=1}^{j-1} (1 - \varphi_k) & \text{otherwise.} \end{cases} \\ q_j &= \begin{cases} \varphi_j \prod_{k=1}^{j-1} (1 - \varphi_k) + r & \text{if } n = j \\ \varphi_j \prod_{k=1}^{j-1} (1 - \varphi_k) & \text{otherwise.} \end{cases} \end{aligned}$$

The three cases immediately follow depending on whether  $n < j$ ,  $n = j$ , or  $n > j$ . □

**Lemma A.3.** Consider the ARC Model under the Outside Option completion rule. If there exist characteristics  $x^L$  and permutation  $\{o_1, \dots, o_D\}$  such that  $\forall \nu \in [0, \bar{\nu}]$

$$L_{o_1}(x^L) \succ L_{o_2}(x^L) \succ \dots \succ L_{o_D}(x^L),$$

then the following relationship holds for all  $j = 1, \dots, D$ :

$$\varphi_{o_j} = \frac{(1-r)q_{o_j}}{1 - (1-r)\sum_{k=1}^{j-1} q_{o_k}},$$

where  $q_{o_k} \equiv Pr(d = d_{o_k} | x^L, \mathcal{K} \neq \emptyset)$ ,  $\mathcal{K}$  is the consideration set, and  $r = \prod_{k=1}^N (1 - \varphi_{o_k})$ .

*Proof.* Without loss of generality suppose that  $k = o_k$  for all  $k$ . Fix any  $j \in \mathcal{D}$ . We first show that

$$\prod_{k=1}^{j-1} (1 - \varphi_k) = 1 - (1-r) \sum_{k=1}^{j-1} q_k.$$

On the one hand, by additivity of probability and the definition of  $q_k$

$$Pr(d \in \{d_1, d_2, \dots, d_{j-1}\} | x^L) = Pr(\mathcal{K} \neq \emptyset) Pr(d \in \{d_1, d_2, \dots, d_{j-1}\} | x^L, \mathcal{K} \neq \emptyset) = (1-r) \sum_{k=1}^{j-1} q_k.$$

On the other hand, according to the model, it is the probability that at least one of  $\{d_1, d_2, \dots, d_{j-1}\}$  is considered

$$Pr(d \in \{d_1, d_2, \dots, d_{j-1}\} | x^L) = 1 - \prod_{k=1}^{j-1} (1 - \varphi_k).$$

Finally, due to the assumption of the preference ordering,

$$\begin{aligned} q_j &= \frac{1}{(1-r)} \varphi_j \prod_{k=1}^{j-1} (1 - \varphi_k) \\ &= \frac{1}{(1-r)} \varphi_j \left( 1 - (1-r) \sum_{k=1}^{j-1} q_k \right). \end{aligned}$$

□

## Proof of Theorem 2

*Proof.* Start with Coin Toss. Let  $d_{o_m}$  be the first alternative in the sequence  $L_{o_1}(x^1) \succ L_{o_2}(x^1) \succ \dots$  such that  $o_m \neq m$ . Let  $j$  be the position of alternative  $o_m$  in the sequence  $L_1(x^0) \succ L_2(x^0) \succ \dots$ . Note that  $m < j$ ,  $d_{o_m} = d_j$ , and all lotteries that dominate  $d_{o_m}$  at  $x^1$  also dominate  $d_j$  at  $x^0$ , since, by construction  $o_1 = 1, o_2 = 2, o_3 = 3, \dots, o_{m-1} = m - 1$ .

The assumptions are satisfied in Lemma A.1 for  $d_j$  at  $x^0$  and  $d_{o_m}$  at  $x^1$ . It follows that:

$$\frac{q_j - r}{1 + (j - 1)r - \sum_{k=1}^{j-1} q_k} = \varphi_j = \varphi_{o_m} = \frac{s_{o_m} - r}{1 + (m - 1)r - \sum_{k=1}^{m-1} s_{o_k}}. \quad (\text{A.1})$$

where  $q_k \equiv Pr(d = d_k | x^0)$  and  $s_{o_k} \equiv Pr(d = d_{o_k} | x^1)$ . This is a quadratic equation in  $r$ . Note that

$$s_{o_m} = \varphi_{o_m} \prod_{k=1}^{m-1} (1 - \varphi_{o_k}) + r \geq r.$$

So any admissible solution for  $r$  ought to be in the interval  $[0, s_{o_m}]$ ; we show that Equation (A.1) has a unique solution in  $[0, s_{o_m}]$ .

Collecting terms we can write Equation (A.1) as follows:

$$\begin{aligned} g(r) &\equiv ar^2 + br + c \\ &\equiv (m - j)r^2 + \left( s_{o_m}(j - 1) - q_j(m - 1) + \sum_{k=1}^{j-1} q_k - \sum_{k=1}^{m-1} s_{o_k} \right) r + s_{o_m} \left( 1 - \sum_{k=1}^{j-1} q_k \right) - q_j \left( 1 - \sum_{k=1}^{m-1} s_{o_k} \right) = 0. \end{aligned}$$

We first show the following

1.  $q_j < s_{o_m}$
2.  $\sum_{k=1}^{j-1} q_k > \sum_{k=1}^{m-1} s_{o_k}$
3.  $s_{o_m} \left( 1 - \sum_{k=1}^{j-1} q_k \right) < q_j \left( 1 - \sum_{k=1}^{m-1} s_{o_k} \right)$

from which it follows that the coefficients for the quadratic function satisfy  $a < 0$ ,  $b > 0$ , and  $c < 0$ .

Indeed, we have:

1.  $q_j < s_{o_m}$ :

$$\begin{aligned}
q_j &= \varphi_j \prod_{k=1}^{j-1} (1 - \varphi_k) + r \\
&= \varphi_{o_m} \prod_{k=1}^{m-1} (1 - \varphi_{o_k}) \prod_{k=m}^{j-1} (1 - \varphi_k) + r \quad (\text{since } o_1 = 1, \dots, o_{m-1} = m-1). \\
&< \varphi_{o_m} \prod_{k=1}^{m-1} (1 - \varphi_{o_k}) + r \\
&= s_{o_m}.
\end{aligned}$$

2.  $\sum_{k=1}^{j-1} q_k > \sum_{k=1}^{m-1} s_{o_k}$ :

$$\sum_{k=1}^{j-1} q_k = \sum_{k=1}^{m-1} s_{o_k} + \sum_{k=m}^{j-1} q_k > \sum_{k=1}^{m-1} s_{o_k}.$$

3.  $s_{o_m} \left(1 - \sum_{k=1}^{j-1} q_k\right) < q_j \left(1 - \sum_{k=1}^{m-1} s_{o_k}\right)$ :

$$\begin{aligned}
\left(1 - \sum_{k=1}^{j-1} q_k\right) &= \prod_{k=1}^{j-1} (1 - \varphi_k) - (j-1)r \\
&= \prod_{k=1}^{m-1} (1 - \varphi_{o_k}) \prod_{k=m}^{j-1} (1 - \varphi_k) - (j-1)r \\
&\equiv uv - (j-1)r, \\
\left(1 - \sum_{k=1}^{j-1} s_{o_k}\right) &= \prod_{k=1}^{m-1} (1 - \varphi_{o_k}) - (m-1)r = u - (m-1)r, \\
s_{o_m} &= \varphi_{o_m} \prod_{k=1}^{m-1} (1 - \varphi_{o_k}) + r = \varphi_j u + r, \quad \text{and} \\
q_j &= \varphi_j \prod_{k=1}^{j-1} (1 - \varphi_k) + r = \varphi_j uv + r.
\end{aligned}$$

Putting this together we have:

$$\begin{aligned}
s_{o_m} \left( 1 - \sum_{k=1}^{j-1} q_k \right) - q_j \left( 1 - \sum_{k=1}^{m-1} s_{o_k} \right) &= (\varphi_j u + r)(uv - (j-1)r) - (\varphi_j uv + r)(u - (m-1)r) \\
&= (m-j)r^2 + \varphi_j ur(v(m-1) - (j-1)) + ur(v-1) \\
&< 0
\end{aligned}$$

since  $u, v, r, \varphi_j \in [0, 1]$  and  $m < j$ .

We have shown that the coefficients of the quadratic function  $g(r)$  have the following signs

$$\begin{aligned}
a &= m - j < 0 \\
b &= \left( s_{o_m}(j-1) - q_j(m-1) + \sum_{k=1}^{j-1} q_k - \sum_{k=1}^{m-1} s_{o_k} \right) > 0 \\
c &= s_{o_m} \left( 1 - \sum_{k=1}^{j-1} q_k \right) - q_j \left( 1 - \sum_{k=1}^{m-1} s_{o_k} \right) < 0.
\end{aligned}$$

Thus  $g(r)$  is a concave quadratic function. To understand its behavior we show that  $g(r)$  evaluated at  $r = s_{o_m}$  is positive.

$$\begin{aligned}
g(s_{o_m}) &= (m-j)s_{o_m}^2 + \left( s_{o_m}(j-1) - q_j(m-1) + \sum_{k=1}^{j-1} q_k - \sum_{k=1}^{m-1} s_{o_k} \right) s_{o_m} + s_{o_m} \left( 1 - \sum_{k=1}^{j-1} q_k \right) - q_j \left( 1 - \sum_{k=1}^{m-1} s_{o_k} \right) \\
&= (m-1)s_{o_m}^2 + \left( 1 - q_j(m-1) - \sum_{k=1}^{m-1} s_{o_k} \right) s_{o_m} - q_j \left( 1 - \sum_{k=1}^{m-1} s_{o_k} \right) \\
&= (m-1)s_{o_m}(s_{o_m} - q_j) + (s_{o_m} - q_j) \left( 1 - \sum_{k=1}^{m-1} s_{o_k} \right) \\
&= (s_{o_m} - q_j) \left( ms_{o_m} + 1 - \sum_{k=1}^m s_{o_k} \right) > 0,
\end{aligned}$$

since  $s_{o_m} > q_j$  and  $\sum_{k=1}^m s_{o_k} \in [0, 1]$ . It follows that one root of  $g(r)$  is always contained in the interval  $(-\infty, s_{o_m})$ , say  $r^-$ , and the other root is always contained in the interval  $(s_{o_m}, \infty)$ , say  $r^+$ .

Now since  $c$  is negative both roots are positive so that the unique solution to  $g(r) = 0$  on  $[0, s_{o_m}]$  is  $r^-$  and hence  $r$  is identified. Once  $r$  is known, all  $\varphi_j$ 's are derived according to the expression in Lemma A.1 applied to  $d_1, \dots, d_D$  at  $x^0$ .

We now turn to Default Option. Assume W.L.O.G that all  $\varphi_j$ 's are positive. Indeed,  $\varphi_j = 0$



iff alternative  $j$  is never chosen for any  $x$  and hence is identified. As above, let  $d_{o_m}$  be the first alternative in the sequence  $L_{o_1}(x^1) \succ L_{o_2}(x^1) \succ \dots$  such that  $o_m \neq m$ . Let  $j$  be the position of alternative  $o_m$  in the sequence  $L_1(x^0) \succ L_2(x^0) \succ \dots$ . Note that  $m < j$ ,  $d_{o_m} = d_j$ , and all lotteries that dominate  $d_{o_m}$  at  $x^1$  also dominate  $d_j$  at  $x^0$ , since, by construction  $o_1 = 1, o_2 = 2, o_3 = 3, \dots, o_{m-1} = m - 1$ .

The assumptions are satisfied in Lemma A.3 for  $d_j$  at  $x^0$  and  $d_{o_m}$  at  $x^1$ . It follows that:

$$\frac{(1-r)q_j}{1 + (1-r)\sum_{k=1}^{j-1} q_k} = \varphi_j = \varphi_{o_m} = \frac{(1-r)s_{o_m}}{1 + (1-r)\sum_{k=1}^{m-1} s_{o_k}}. \quad (\text{A.2})$$

where  $q_k \equiv Pr(d = d_k | x^0, d \in \{d_1, d_2, \dots, d_D\})$  and  $s_{o_k} \equiv Pr(d = d_{o_k} | x^1, d \in \{d_1, d_2, \dots, d_D\})$ .

If  $\varphi_j = 0$ , it is immediate that  $r = 0$ . On the other hand, if  $\varphi_j > 0$ , then Equation A.2 implies that

$$r = 1 - \frac{s_{o_m} \sum_{k=1}^{j-1} q_k - q_j \sum_{k=1}^{m-1} s_{o_k}}{s_{o_m} - q_j}.$$

Since  $q_j < s_{o_m}$  and  $\sum_{k=1}^{j-1} q_k > \sum_{k=1}^{m-1} s_{o_k}$ , there is a unique  $r \in [0, 1]$  that solves the Equation A.2. With known  $r$ , we can learn  $\varphi_j$ 's sequentially according to Lemma A.3:  $\varphi_1 = (1-r)q_1$ ,  $\varphi_2 = \frac{(1-r)q_2}{1-(1-r)q_1}$ , and so on.

□

### Proof of Theorem 3

*Proof.* Let  $d_n$  denote the default alternative so that it is  $n^{\text{th}}$  best at  $x^0$ . Let  $r = \prod_{k=1}^D (1 - \varphi_k)$ . We first show that  $r$  is identified.

Let  $d_j$  and (W.L.O.G) let  $j > n$ , so that  $d_n \succ d_j$  by all DMs with characteristics  $x^0$  regardless of their  $\nu$ . Let  $o_{j'}$  and  $o_{n'}$  index the position of  $d_j$  and  $d_n$  at  $x^1$ . That is,  $d_j = d_{o_{j'}}$  and  $d_n = d_{o_{n'}}$ . By assumption we have  $j' < n'$  so that  $d_j$  is preferred to  $d_n$  by any DM with characteristics  $x^1$ . The conditions for Lemma A.2 hold, so that

$$\frac{q_j}{r + 1 - \sum_{k=1}^{n-1} q_k} = \varphi_j = \varphi_{o_{j'}} = \frac{s_{o_{j'}}}{1 - \sum_{k=1}^{n'-1} s_{o_k}},$$

Solving for  $r$  yields:

$$r = \frac{q_j \left(1 - \sum_{k=1}^{n'-1} s_{o_k}\right) - s_{o_{j'}} \left(1 - \sum_{k=1}^{n-1} q_k\right)}{s_{o_{j'}}}.$$

Note that, by assumption,  $q_j, s_{o_{j'}} > 0$  so  $r$  is well defined. Once  $r$  is known, all  $\varphi_j$ 's are derived according to the expression in Lemma A.2 applied to  $d_1, \dots, d_D$  at  $x^0$ .  $\square$

#### Proof of Theorem 4

*Proof.* Let  $q_j = Pr(d_j|x^0)$  and  $s_{o_k} = Pr(d_{o_k}|x^1)$ . We can learn  $\underline{n}$  and  $\varphi_j$  for all  $j \leq \underline{n}$  as follows. By assumption, there are no preferred options among alternatives  $d_1, \dots, d_{\underline{n}-1}$ . Hence,

1.  $\varphi_1 = q_1$ . If  $\varphi_1 = 1$  then  $\underline{n} = 1$ . Otherwise, set  $j = 2$  and proceed to Step 2.
2.  $\varphi_j = \frac{q_j}{1 - \sum_{k=1}^{j-1} q_k}$ . If  $\varphi_j = 1$  then  $\underline{n} = j$ . Otherwise, set  $j = j + 1$  and repeat Step 2.

Repeating this argument for the moments evaluated at  $x^1$ , we find the first  $n^*$  such that  $\varphi_{o_{n^*}} = 1$  (i.e.  $n^* = \min_{n' \in \{\underline{n}', \dots, \bar{n}'\}} n'$  where  $o_{\underline{n}'} = \underline{n}, \dots, o_{\bar{n}'} = \bar{n}$ ) and  $\varphi_{o_{j'}}$  for all  $j' \leq n^*$ .

To summarize we have identified  $\varphi_j$  for all  $j \leq \underline{n}$  and all  $j \geq \bar{n}$  (since whenever  $j \geq \bar{n}$  it also the case that  $j' \leq n^*$  where  $o_{j'} = j$ ). By assumption, Preferred Options are adjacent so that whenever  $\underline{n} \leq n \leq \bar{n}$ ,  $d_n$  is also Preferred Options and hence  $\varphi_n = 1$ .  $\square$

#### Proof of Theorem 7

*Proof.* We have that

$$\begin{aligned} Pr(d = d_D|x^1) &= \phi_1 Q_{1,D} \\ Pr(d = d_{D-1}|x^1) &= \phi_1 Q_{1,D-1} + \phi_2 Q_{2,D-1} \\ &\vdots \\ Pr(d = d_1|x^1) &= \phi_1 Q_{1,1} + \phi_2 Q_{2,1} + \dots + \phi_D Q_{D,1} \end{aligned}$$

The  $Q$ 's in the equations above are known and are strictly positive. It follows that  $\phi$ 's are identified sequentially.  $\square$

#### Proof of Theorem 8

*Proof.* We start with Assumption 2.A1-A3. Denote  $\mathcal{A} = \{k : \varphi_k > 0\}$  under the ARC Model and  $\mathcal{A} = \mathcal{D}$  under the RCL Model. Fix  $j$  corresponding to Assumption 2.A1-A3 and denote  $\underline{j}(j, t) \equiv \arg \min_{k \in \mathcal{A} - \{j\}} c_{j,k}(x(t))$ . By Assumption 2.A3 and the continuity of  $c(\cdot)$ ,

the alternative corresponding to  $\underline{j}(j, t)$  is the same for all  $t \in [0, 1]$ . Thus, we can write WLOG  $\underline{j} = \underline{j}(j, t)$ . For any  $t \in [0, 1]$  we have

$$Pr(d = d_j | x(t)) = w_{j, \underline{j}} F(c_{\underline{j}}(x(t))) + \sum_{k \in \mathcal{A} - \{j, \underline{j}\}} w_{j, k} F(c_{j, k}(x(t))) + \hat{r}_j,$$

where  $\hat{r}_j \geq 0$ ,  $w_{j, \underline{j}} > 0$  and all  $w_{j, k} > 0$  are known functions of the limited consideration parameters. Since the latter are identified, so are  $\hat{r}_j$ ,  $w_{j, \underline{j}}$  and  $w_{j, k}$ ,  $k \neq j, \underline{j}$ . Next, find the smallest  $t_1$  such that  $c_{j, k}(x(t_1)) \geq \bar{\nu}$  for all  $k \in \mathcal{A} - \{j, \underline{j}\}$ . In other words,  $t_1$  is the smallest value of  $t$  for which only the lowest cutoff is below the upper bound of the support. It follows that for any  $t \in [t_1, 1]$ ,

$$Pr(d = d_j | x(t)) = w_{j, \underline{j}} F(c_{\underline{j}}(x(t))) + \hat{r}_j,$$

which implies that  $F(\cdot)$  is identified for all  $\nu \in [\nu_1, \bar{\nu}]$  where  $\nu_1 \equiv c_{\underline{j}}(x(t_1))$ . It is clear that if  $\mathcal{A} - \{j, \underline{j}\} = \emptyset$  we are done. Otherwise, find the smallest  $t_2$  such that  $c_{j, k}(x(t_2)) \geq \nu_1$  for all  $k \in \mathcal{A} - \{j, \underline{j}\}$ . In other words,  $t_2$  is the smallest value of  $t$  for which only the lowest cutoff is below  $\nu_1$ . Since all other cutoffs lie in the region where  $F(\cdot)$  is known, it follows that  $F(\cdot)$  is identified for all  $\nu \in [\nu_2, \nu_1]$ , and, hence for all  $\nu \in [\nu_2, \bar{\nu}]$ , where  $\nu_2 \equiv c_{\underline{j}}(x(t_2))$ . Proceeding in this way we have that  $F(\cdot)$  is identified over  $[\nu_n, \bar{\nu}]$ .  $\{\nu_n\}$  is a strictly monotonically declining sequence defined recursively as

$$\begin{aligned} \nu_n &\equiv c_{\underline{j}}(x(t_n)) \\ t_0 &= 1 \\ t_n &= \min_{t \in [0, 1]} t \\ &\text{s.t. } c_{j, k}(x(t)) \geq \nu_{n-1} \quad \forall k \in \mathcal{A} - \{j, \underline{j}\}. \end{aligned}$$

This sequence either eventually converges to 0, crosses to the left of 0, or converges to some accumulation point  $\nu^*$  in the interior of  $[0, \bar{\nu}]$ . In the former cases we have identification. We claim that the latter case cannot arise. For the purpose of obtaining a contradiction, suppose that  $\nu^* = \lim_{n \rightarrow \infty} \nu_n > 0$ . By continuity it follows that

$$\lim_{n \rightarrow \infty} c_{\underline{j}}(x(t_n)) = c_{\underline{j}}(x(t^*)),$$

where

$$\begin{aligned}
t^* &\equiv \lim_{n \rightarrow \infty} t_n \\
&= \lim_{n \rightarrow \infty} \left( \min_{t \in [0,1]} t \quad \text{s.t.} \quad c_{j,k}(x(t)) \geq \nu_{n-1} \quad \forall k \in \mathcal{A} - \{j, \underline{j}\} \right) \\
&= \min_{t \in [0,1]} t \quad \text{s.t.} \quad c_{j,k}(x(t)) \geq \lim_{n \rightarrow \infty} \nu_{n-1} \quad \forall k \in \mathcal{A} - \{j, \underline{j}\} \quad (\text{by continuity}) \\
&= \min_{t \in [0,1]} t \quad \text{s.t.} \quad c_{j,k}(x(t)) \geq \nu^* \quad \forall k \in \mathcal{A} - \{j, \underline{j}\}.
\end{aligned}$$

Now since the cutoffs are strictly decreasing in  $t$ , there is a  $k \in \mathcal{A} - \{j, \underline{j}\}$  such that  $c_{j,k}(x(t^*)) = \nu^*$ . Putting this together we yield

$$c_{j,k}(x(t^*)) = \nu^* = \underline{c}_j(x(t^*)),$$

which contradicts Assumption 2.A3.

Under Assumption 2.B1-3 the proof works in the exactly same way, only we start at the lower end of the preference-coefficient support.

□

## B Proofs of Properties

### Proof of Fact 1

*Proof.* Take any non empty consideration set  $\mathcal{K}$ . For a given preference coefficient  $\nu$ , let  $j_{\mathcal{K}}(\nu)$  denote the identity of the best alternative in this consideration set. Because of the way alternatives are ordered,  $j_{\mathcal{K}}(\nu)$  is an increasing step function. Hence,  $I(j_{\mathcal{K}}(\nu) \leq J)$  is a decreasing step function, Note, that  $Pr \left( \bigcup_{j=1}^J d_j \mid x, \nu \right)$  is the sum of  $I(j_{\mathcal{K}}(\nu) \leq J)$  weighted by the probability of  $\mathcal{K}$  being drawn. Hence it is decreasing in  $\nu$ . □

### Proof of Fact 5

*Proof.* Suppose  $U_{\nu}(d_j, x) > U_{\nu}(d_k, x)$ , but  $\varphi_k = 1$  and  $\varphi_j = 0$ . □

**Proof of Fact 6**

*Proof.* Suppose  $U_\nu(d_j, x) > U_\nu(d_k, x)$ . If both  $d_j$  and  $d_k$  are in the consideration set, then  $d_j$  will be chosen. For any consideration set that contains  $d_k$  but not  $d_j$ , there is an equal size consideration set that contains  $d_j$  but not  $d_k$ , namely  $(\mathcal{K} \cup d_j) \setminus d_k$ . These two sets have the same probability of being formed. If  $d_k$  is preferred to all other alternatives in  $\mathcal{K}$ , then the same is true for  $d_j$  in  $\mathcal{K} \cup d_j \setminus d_k$ . Summing over all consideration sets delivers the result.  $\square$

## C Verifying Identification in Our Application

We start by recalling that CARA and CRRA utility functions satisfy the following basic property (see, e.g., Pratt, 1964; Barseghyan, Molinari, O’Donoghue, & Teitelbaum, 2018):<sup>1</sup>

**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  $Q_\nu(y_0, y_1, y_2, y_3) \equiv \frac{u_\nu(y_2) - u_\nu(y_3)}{u_\nu(y_0) - u_\nu(y_1)}$  is strictly increasing in  $\nu$ .*

*Proof.*

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

$\square$

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

**Theorem C.1.** *Under either CARA or CRRA expected utility preferences, the cutoff mappings satisfy  $c_{1,2}(\bar{p}, \mu) < c_{1,m}(\bar{p}, \mu)$  for any  $m > 2$ .*

*Proof.* We start with CARA preferences. The existence and the uniqueness of  $c_{j,k}(x)$  for all

---

<sup>1</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.

$j < k$  follows directly from the Property C.2. Indeed note that  $p_j < p_k < p_k + d_k < p_j + d_j$ .<sup>2</sup> At the cutoff the DM is indifferent between lotteries  $j$  and  $k$ . Equating two expected utilities and rearranging we have that

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

where  $w$  is the DM's initial wealth. By Property C.2, the L.H.S. of Equation C.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 C.1. 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 that  $c_{1,2}(\bar{p}, \mu) < c_{1,m}(\bar{p}, \mu)$ ,  $m > 2$ . First, note that the expected utility of lottery  $k$  is proportional to

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

For the purpose of obtaining a contradiction, suppose that there exists  $(\bar{p}, \mu)$  and an  $m$  such that  $c_{1,2}(\bar{p}, \mu) = c_{1,m}(\bar{p}, \mu)$ . That is, there exists a  $\nu = c_{1,2}(\bar{p}, \mu) = c_{1,m}(\bar{p}, \mu)$  such that

$$\frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_2}} e^{\nu(g_1 - g_2)\bar{p}} = 1 = \frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_m}} e^{\nu(g_1 - g_m)\bar{p}}$$

Taking logs for each side and rearranging we have that

$$\begin{aligned} \log\left(\frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_2}}\right) &= -\nu(g_1 - g_2)\bar{p} \\ \log\left(\frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_m}}\right) &= -\nu(g_1 - g_m)\bar{p}. \end{aligned}$$

Dividing through we have that

$$\frac{\log\left(\frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_2}}\right)}{\log\left(\frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_m}}\right)} = \frac{g_1 - g_2}{g_1 - g_m}.$$

The R.H.S. is less than one. The L.H.S. is monotonically decreasing in  $\mu < 1$ . Indeed, denote

---

<sup>2</sup>If  $p_k + d_k > p_j + d_j$ , then alternative  $j$  first order stochastically dominates  $k$  and hence the cutoff is  $+\infty$ .

$\hat{\mu} = \frac{1-\mu}{\mu}$ ,  $\Delta_1 = e^{\nu d_1}$ ,  $\Delta_2 = e^{\nu d_2}$ , and  $\Delta_m = e^{\nu d_m}$  to rewrite the L.H.S. as follows

$$f(\hat{\mu}) \equiv \frac{\log(\Delta_1 + \hat{\mu}) - \log(\Delta_2 + \hat{\mu})}{\log(\Delta_1 + \hat{\mu}) - \log(\Delta_m + \hat{\mu})}.$$

We claim that the expression above is monotonically increasing in  $\hat{\mu}$ . Its derivative is equal to

$$\frac{f'(\hat{\mu})}{f(\hat{\mu})} = \left( \frac{1}{\Delta_1 + \hat{\mu}} - \frac{1}{\Delta_2 + \hat{\mu}} \right) \frac{1}{\log(\Delta_1 + \hat{\mu}) - \log(\Delta_2 + \hat{\mu})} - \left( \frac{1}{\Delta_1 + \hat{\mu}} - \frac{1}{\Delta_m + \hat{\mu}} \right) \frac{1}{\log(\Delta_1 + \hat{\mu}) - \log(\Delta_m + \hat{\mu})}$$

Once more relabeling  $\Lambda_1 = -\log(\Delta_1 + \hat{\mu})$ ,  $\Lambda_2 = -\log(\Delta_2 + \hat{\mu})$  and  $\Lambda_m = -\log(\Delta_m + \hat{\mu})$  we can write the above as

$$\frac{f'(\hat{\mu})}{f(\hat{\mu})} = \frac{e^{\Lambda_1} - e^{\Lambda_m}}{\Lambda_1 - \Lambda_m} - \frac{e^{\Lambda_1} - e^{\Lambda_2}}{\Lambda_1 - \Lambda_2}.$$

Since  $\Lambda_1 < \Lambda_2 < \Lambda_m$  and exponential function is convex, we have that 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  $\frac{d_1 - d_2}{d_1 - d_m}$ . Hence, the question is whether the following equality may hold

$$\frac{d_1 - d_2}{d_1 - d_m} = \frac{g_1 - g_2}{g_1 - g_m}.$$

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, Prince, & Teitelbaum (2011)). Hence  $c_{1,2}(\bar{p}, \mu) \neq c_{1,m}(\bar{p}, \mu)$ , for  $m > 2$ . Since the cutoffs are continuous, it follows that  $c_{1,2}(\bar{p}, \mu) < c_{1,m}(\bar{p}, \mu)$  for  $m > 2$ .

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  $c_{1,2}(\bar{p}, \mu) < c_{1,m}(\bar{p}, \mu)$  for  $m > 2$ . 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. Hence, we take the element-wise difference between  $EU_\nu(L_j)$  and  $EU_\nu(L_k)$ :

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

This implies that if  $\nu = c_{1,2}(\bar{p}, \mu) = c_{1,m}(\bar{p}, \mu)$ ,  $m > 2$  we must have that

$$\begin{aligned} \frac{p_1 - p_2}{p_1 - p_m} &= \frac{p_1 - p_2 + d_1 - d_2}{p_1 - p_m + d_1 - d_m} \times \\ &\quad \frac{\sum_{t=1}^{\infty} \omega_t \sum_{h=0}^t (p_1 + d_1)^h (p_2 + d_2)^{t-h} \sum_{t=1}^{\infty} \omega_t \sum_{h=0}^t p_1^h p_m^{t-h}}{\sum_{t=1}^{\infty} \omega_t \sum_{h=0}^t (p_1 + d_1)^h (p_m + d_m)^{t-h} \sum_{t=1}^{\infty} \omega_t \sum_{h=0}^t p_1^h p_2^{t-h}} \end{aligned}$$

Note that  $p_m > p_2$ . More over, when  $\nu = c_{1,2}(\bar{p}, \mu) = c_{1,m}(\bar{p}, \mu)$  it is also the case that  $\nu = c_{1,2}(\bar{p}, \mu) = c_{1,m}(\bar{p}, \mu) = c_{2,m}(\bar{p}, \mu)$ .

For the cutoff  $c_{2,m}(\bar{p}, \mu)$  to be on the support it must be the case that  $p_2 + d_2 > p_m + d_m$ . Indeed otherwise we have that  $p_m - p_2 > d_2 - d_m$ , which is a violation of the first order stochastic dominance. Hence if we can show that

$$\frac{p_1 - p_2}{p_1 - p_m} < \frac{p_1 - p_2 + d_1 - d_2}{p_1 - p_m + d_1 - d_m},$$

we would arrive to a contradiction, since it would be mean that the LHS of the equation is



smaller than the RHS. Re-arranging we have that

$$\begin{aligned} \frac{p_1 - p_m + d_1 - d_m}{p_1 - p_m} &< \frac{p_1 - p_2 + d_1 - d_2}{p_1 - p_2} \\ \frac{d_1 - d_m}{p_1 - p_m} &< \frac{d_1 - d_2}{p_1 - p_2} \\ \frac{p_1 - p_2}{p_1 - p_m} &< \frac{d_1 - d_2}{d_1 - d_m}. \end{aligned}$$

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

□

## D 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>3</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 F.7 for variation conditional on demographic characteristics and insurance score. We also find wide dispersion in loss across households (see Figure F.7). In particular, the 10<sup>th</sup> percentile of losses is \$30 and the 90<sup>th</sup> is \$72.

---

<sup>3</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  $ce_j \equiv \frac{1}{\nu} \ln[(1 - \mu) \exp(\nu p_j) + \mu \exp(\nu(p_j + d_j))]$ .

## E Data

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

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

**Table E.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.

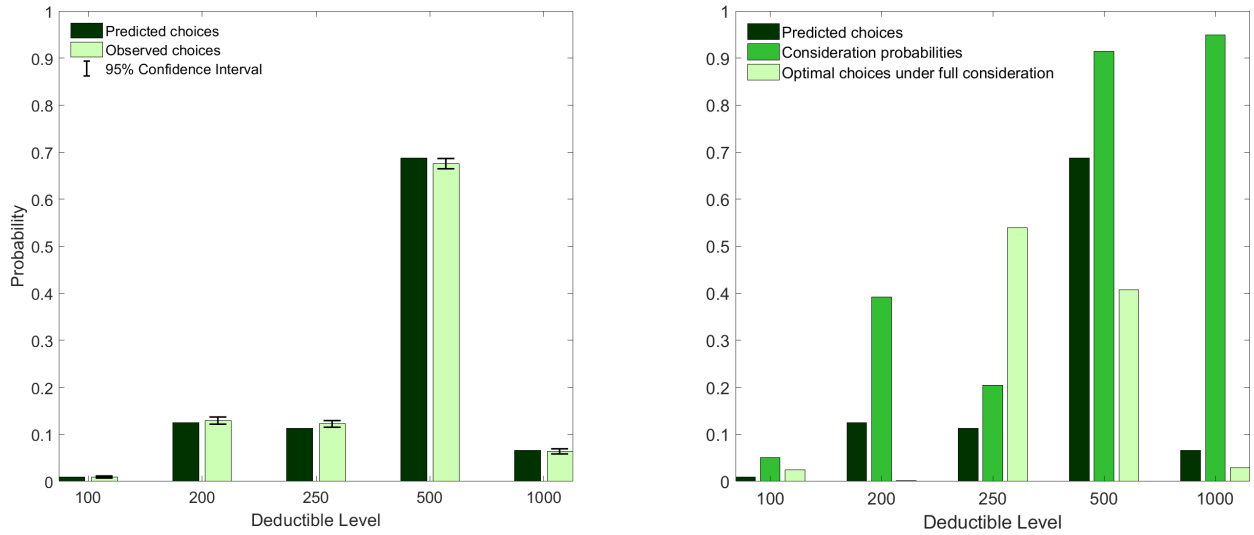
**Table E.5** Average Premiums Across Coverages

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	

# F Empirical Results: Figures and Tables

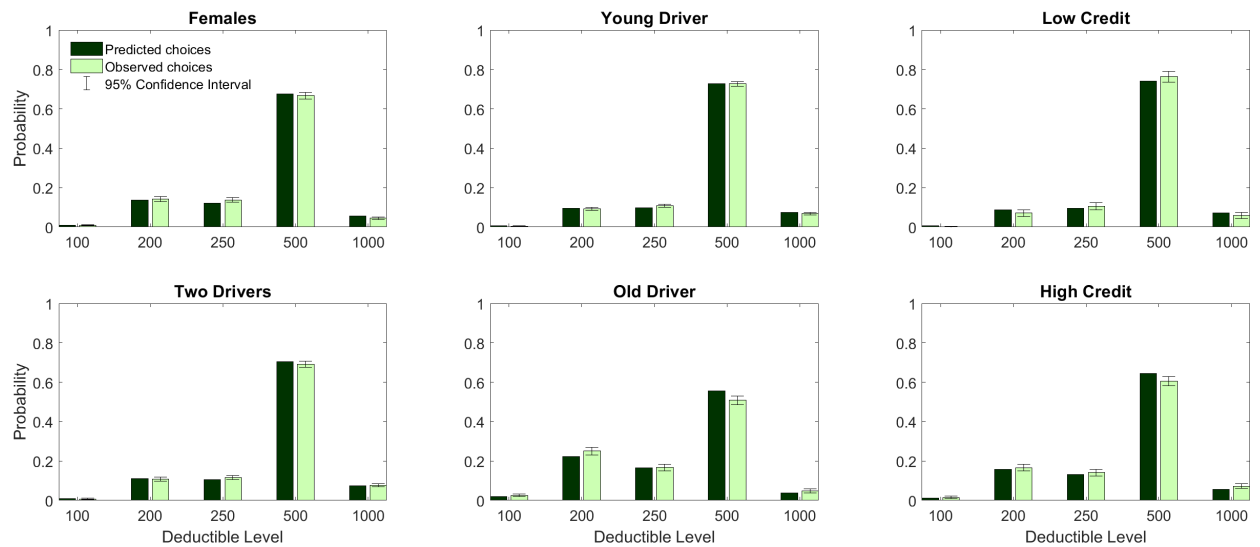
## F.1 Figures

Figure F.1: The ARC Model with Observable Demographics

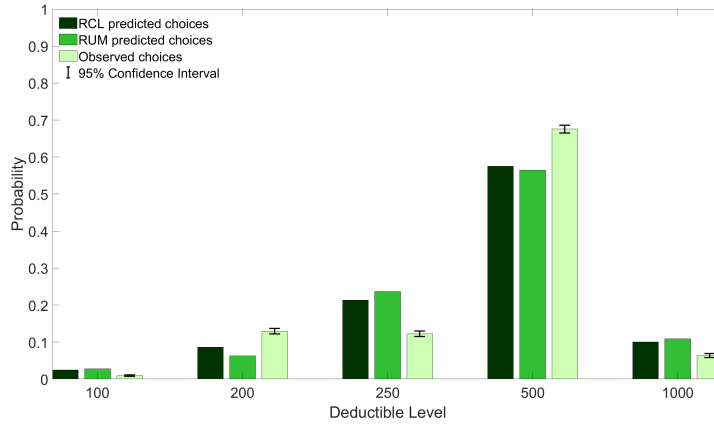


The first panel reports the distribution of predicted and observed choices. The second panel displays consideration probabilities and the distribution of optimal choices under full consideration.

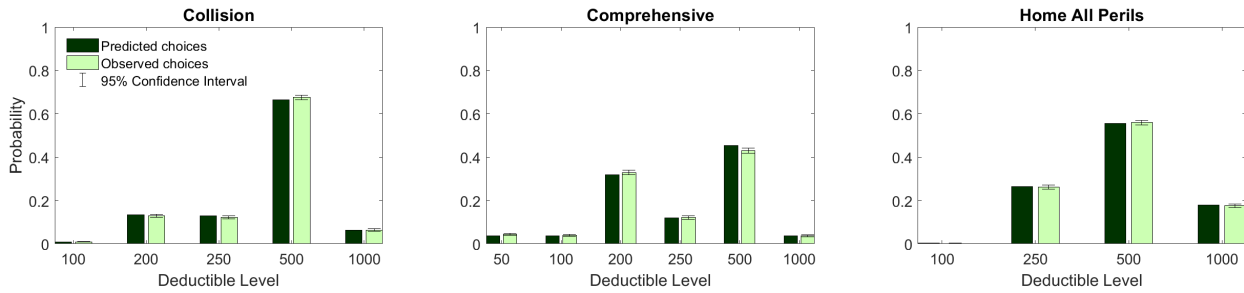
Figure F.2: The ARC Model with Observable Demographics: Conditional Distributions



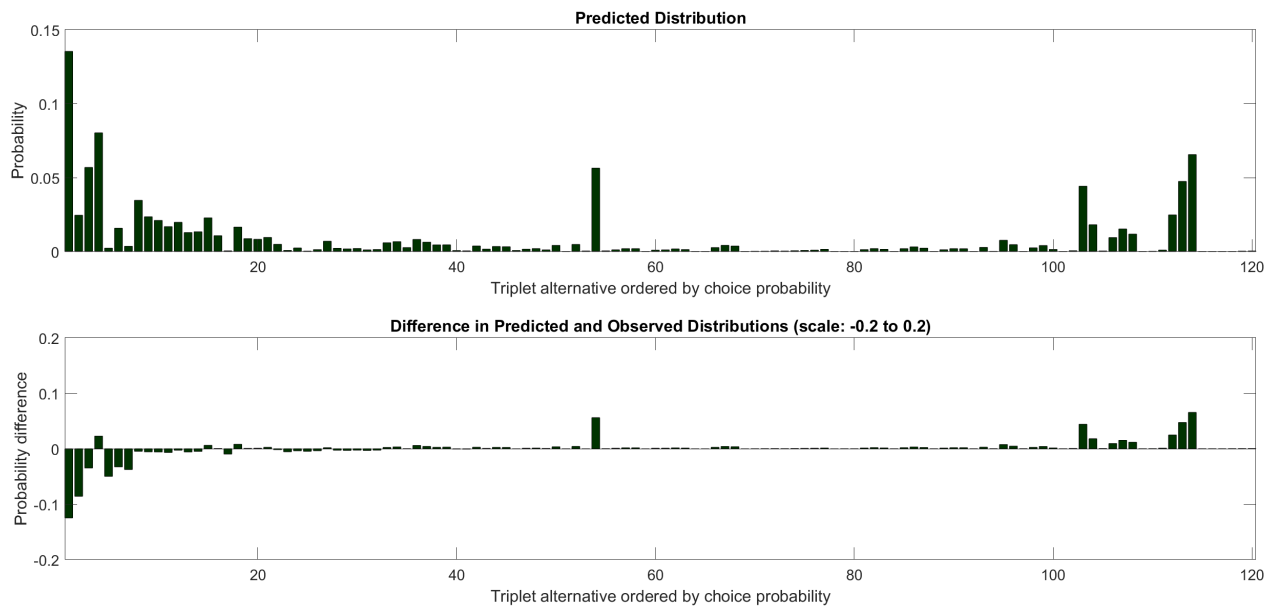
**Figure F.3:** The RCL Model and the RUM



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

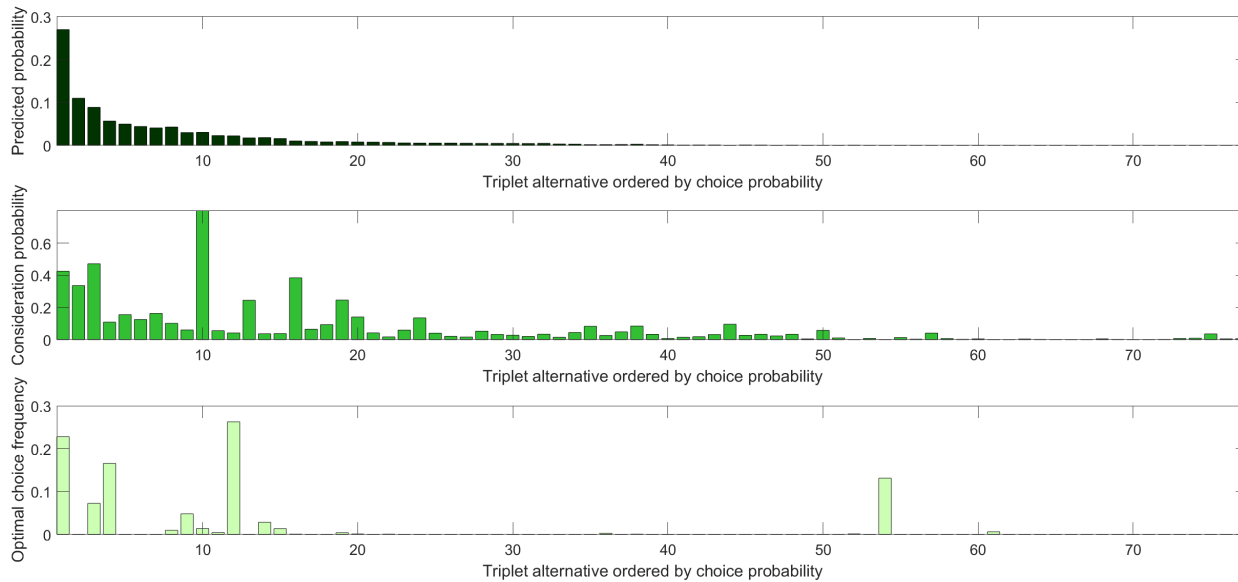


**Figure F.5:** The RUM, Three Coverages



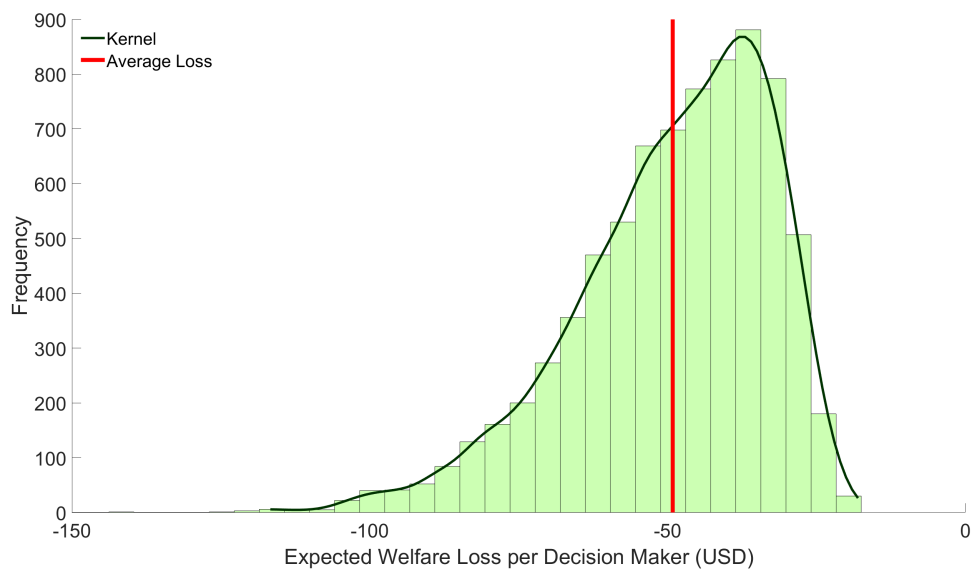
Triplets are sorted by observed frequency at which they are chosen. The first panel reports the predicted choice frequency and the second panel reports the difference in predicted and observed choice frequencies.

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



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

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



## F.2 Tables

**Table F.1** MLE Estimation Results for the ARC Model

	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 F.2** MLE Estimation Results for the RCL and RUM Models

	<b>RCL Model</b>		<b>RUM Model</b>	
$\beta_1$	6.330	[5.085, 8.993]	8.401	[6.794, 10.650]
$\beta_2$	100.232	[80.526, 142.598]	122.603	[102.578, 152.511]
<b>Mean of <math>\nu</math></b>	0.001	[0.001, 0.001]	0.001	[0.001, 0.001]
<b>SD of <math>\nu</math></b>	0.0005	[0.0004, 0.0005]	0.0004	[0.0004, 0.0005]
<b>Intercept</b>	-2.569	[-2.656, -2.479]	-2.647	[-2.713, -2.586]
<b>Age</b>	-0.142	[-0.198, -0.090]	-0.146	[-0.178, -0.118]
<b>Age<sup>2</sup></b>	-0.047	[-0.098, 0.003]	-0.026	[-0.051, -0.002]
<b>Female Driver</b>	0.011	[-0.039, 0.064]	-0.004	[-0.032, 0.025]
<b>Single Driver</b>	0.038	[-0.014, 0.092]	-0.010	[-0.039, 0.020]
<b>Married Driver</b>	0.027	[-0.044, 0.101]	-0.031	[-0.069, 0.009]
<b>Credit Score</b>	0.232	[0.180, 0.288]	0.096	[0.073, 0.124]
<b>2+ Drivers</b>	-0.390	[-0.535, -0.233]	-0.021	[-0.101, 0.061]
<b>Attention Level 1</b>	0.031	[0.019, 0.046]	-	-
<b>Attention Level 2</b>	0.509	[0.447, 0.536]	-	-
<b>Attention Level 3</b>	0.000	[0.000, 0.086]	-	-
<b>Attention Level 4</b>	0.000	[0.000, 0.000]	-	-
<b>Full Attention</b>	0.460	[0.423, 0.484]	-	-
<b>Sigma</b>	-	-	0.040	[0.036, 0.043]

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



**Table F.3** MLE Estimation Results for the ARC Model, Three Coverages: “Narrow” Consideration

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

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

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

<b>RUM Model</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 F.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