Discrete Choice under Risk with Limited Consideration

By Levon Barseghyan, Francesca Molinari, and Matthew Thirkettle[*]

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

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

We depart from this standard approach by proposing a discrete choice model with unobserved heterogeneity in preferences and unobserved heterogeneity in consideration sets. Specifically, preferences satisfy the classic Single Crossing Property (SCP) of Mirrlees (1971) and Spence (1974), central to important studies of decision making under risk.[1] That is, the preference order of any two

[1] E.g., Apesteguia, Ballester and Lu (2017); Chiappori et al. (2019). While our focus is on decision making under risk, the SCP property is satisfied in many contexts, ranging from single agent models with goods that can be unambiguously ordered based on quality, to multiple agents models (e.g., Athey (2001)).

alternatives switches only at one value of the preference parameter.[2] Given her
unobserved preference parameter, each DM evaluates only the alternatives in her
unobserved consideration set, which is a subset of the choice set.

Our first contribution is to provide a general framework for point identification
of these models. Our analysis relies on two types of observed data variation. In the
first case, we assume that the data include a single (common) excluded regressor
affecting the utility of each alternative. In the second case, we assume that each
alternative has its own excluded regressor. In both cases, the excluded regres-
sor(s) is independent of unobserved preference heterogeneity. When the excluded
regressor(s) also has large support it becomes a "special regressor" (Lewbel, 2000,
2014). For reasons we explain, the case of the single common excluded regressor is
the most demanding from an identification standpoint. Nonetheless, under clas-
sic conditions for identification of full-consideration discrete choice models (see,
e.g., Lewbel, 2000; Matzkin, 2007) and the SCP, we obtain semi-nonparametric
identification of the preference distribution given basically any consideration set
formation mechanism (henceforth, consideration mechanism).[3] We also prove
identification of the consideration mechanism for the widely used Alternative-
specific Random Consideration (ARC) model of Manski (1977) and Manzini and
Mariotti (2014). The identification argument is constructive and applicable be-
yond the ARC model. We establish identification results for preferences that do
not require large support of the excluded regressor(s). We also show that identi-
fication of both preferences and the consideration mechanism is attainable when
consideration depends on preferences. In particular, we introduce (i) binary con-
sideration types, and (ii) proportionally shifting consideration, both of which can
capture the notion that the DM's attention probabilistically shifts from riskier
to safer alternatives as her risk aversion increases. In these cases, identification
requires that the distribution of the preference parameter admits a continuous
density function.

We can significantly expand our results with alternative-specific excluded re-
gressors. First, we can allow for essentially unrestricted dependence of consid-
eration on preferences without assuming that the excluded regressors have large
support. Second, we show that consideration can depend both on preferences
and on some excluded regressors. We show this for two cases. In the first case,
there is one alternative (the default) that is always considered. The probability of
considering other alternatives can depend on the default-specific excluded regres-
sor. This is a generalization of the models in Heiss et al. (2016); Ho, Hogan and
Scott Morton (2017); Abaluck and Adams (2020), where the consideration mech-
anism only allows for the possibility that either the default or the entire choice set

---

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

[3]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.

is considered. We, however, allow for each subset of the choice set containing the default to have its own probability of being drawn and this probability can vary with the DM's preferences. In the second case, we allow the consideration of each alternative to depend on its own excluded regressor, but not on the regressors of other alternatives (Goeree, 2008; Abaluck and Adams, 2020; Kawaguchi, Uetake and Watanabe, 2020). In addition, consideration may depend on preferences – a feature unique to our paper.

Our second contribution is to provide a simple method to compute our likelihood-based estimator. Its computational complexity grows polynomially in the number of parameters governing the consideration mechanism. Because the SCP generates a natural ordering of alternatives akin to vertical product differentiation, our method does not require enumerating all possible subsets of the choice set. If it did, the computational complexity would grow exponentially with the size of the choice set. Moreover, we compute the utility of each alternative only once for a given value of the preference parameter, gaining enormous computational advantage similar to that of importance-sampling methods.

Our third contribution is to elucidate the applicability and the advantages of our framework over the standard application of full consideration random utility models (RUMs) with additively separable unobserved heterogeneity (e.g., Mixed Logit). First, our model can generate zero shares for non-dominated alternatives. Second, the model has no difficulty explaining relatively large shares of dominated alternatives. Third, in markets with many choice domains, our model can match not only the marginal but also the joint distribution of choices across domains. Forth, our framework is immune to an important criticism by Apesteguia and 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.

Random preference models like the ones we consider are random utility models as envisioned by McFadden (1974) (for a textbook treatment see Manski, 2009). We show that our random preference models can be written as RUMs with unobserved heterogeneity in risk aversion and with an additive error that has a discrete distribution with support $\{-\infty, 0\}$. Then, it is natural to draw parallels with the Mixed (random coefficient) Logit model (e.g., McFadden and Train, 2000). In our setting, the Mixed Logit boils down to assuming that, given the DM's risk aversion, her evaluation of an alternative equals its expected utility summed with an unobserved heterogeneity term capturing the DM's idiosyncratic taste for unobserved characteristics of that alternative. However, in some markets it is hard to envision such characteristics.[4] We show that limited consideration models and

---

[4]Many insurance contracts are identical in *all* aspects *except* for the coverage level and price, e.g., employer provided health insurance, auto, or home insurance offered by a single company. In other contexts, unobservable characteristics may affect choice mostly via consideration – as we model – rather than via "additive noise". E.g., a DM may only consider those supplemental prescription drug plans that cover specific medications.

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

We illustrate our method in a study of households' deductible choices across three lines of insurance: auto collision, auto comprehensive, and home (all perils). We aim to estimate the distribution of risk preferences and the consideration parameters and to assess the resulting fit of the models. We find that the ARC model does a remarkable job at matching the distribution of observed choices, and because of its aforementioned properties, outperforms the Mixed Logit. Under the ARC model, we find that although households are on average strongly risk averse, they consider lower coverages more often than higher coverages. We also find support for proportionally shifting consideration. In particular, risk-neutral DMs consider each of the safer alternatives 15% (11%) less often than do extremely risk averse DMs (DMs with median risk aversion).

The rest of the paper is organized as follows. We describe the model of DMs' preferences in Section I, and study identification in Section II. In Section III we describe the computational advantages of our approach. Section IV compares our model to the Mixed Logit. Section V presents our empirical application. Section VI contextualizes our contribution relative to the extant literature and offers concluding remarks.

I. Preferences

A. *Decision Making under Risk in a Market Setting: An Example*

Consider as an example the following insurance market, which mimics the setting of our empirical application. There is an underlying risk of a loss that occurs with probability μ that may vary across DMs. A finite number of alternatives are available to insure against this loss. Conditional on risk type, i.e., given μ, each alternative $j \in \mathcal{D} \equiv \{1, \ldots, D\}$ is fully characterized by the pair (d_j, p_j). The first element is the insurance deductible, which is the DM's out of pocket expense in the case a loss occurs. Deductibles are decreasing with index j, and all deductibles are less than the lowest realization of the loss. The second element is

the price (insurance premium), which also varies across DMs. For each DM there is a baseline price $\bar{p}$ that determines prices for all alternatives faced by the DM according to the multiplication rule $p_j = g_j \cdot \bar{p} + \delta$. Lower deductibles provide more coverage and cost more, so g_j is increasing with j. Both g_j and δ are invariant across DMs. The lotteries that DMs face are $L_j(x) \equiv (-p_j, 1 - \mu; -p_j - d_j, \mu)$, where $x \equiv \bar{p}$. DMs are expected utility maximizers. Given initial wealth w, the expected utility of deductible lottery $L_j(x)$ is

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

where $u_\nu(\cdot)$ is a Bernoulli utility function defined over final wealth states. We assume that $u_\nu(\cdot)$ belongs to a family of utility functions that are fully characterized by a scalar ν (e.g. Constant Absolute Risk Aversion (CARA), Constant Relative Risk Aversion (CRRA), or Negligible Third Derivative (NTD)), which varies across DMs.[5]

Given the risk type, the relationship between risk aversion and prices is standard. At sufficiently high $\bar{p}$, less coverage is always preferred to more coverage for all ν on the support: $U_\nu(L_1(x)) > U_\nu(L_2(x)) > \cdots > U_\nu(L_D(x))$. At sufficiently low $\bar{p}$, we have the opposite ordering for all ν on the support: $U_\nu(L_D(x)) > U_\nu(L_{D-1}(x)) > \cdots > U_\nu(L_1(x))$. At moderate prices, for each pair of deductible lotteries $j < k$ there is a cutoff value $c_{j,k}(x)$ in the interior of ν's support, found by solving $U_\nu(L_j(x)) = U_\nu(L_k(x))$ for ν. On the left of this cutoff the higher deductible is preferred and on the right the lower deductible is preferred. In other words, $c_{j,k}(x)$ is the unique coefficient of risk aversion that makes the DM indifferent between $L_j(x)$ and $L_k(x)$, known to the researcher at any given x. Those with lower ν choose the riskier alternative $L_j(x)$, while those with higher ν choose the safer alternative $L_k(x)$. Provided $U_\nu(\cdot)$ is smooth in ν, $c_{j,k}(x)$ is smooth in x. In fact, under CARA, CRRA, or NTD, $c_{j,k}(x)$ is a continuously differentiable monotone function. The prices are such that, under CARA, CRRA, or NTD, whenever $U_\nu(L_1(x)) > U_\nu(L_j(x))$ it is also the case that $U_\nu(L_1(x)) > U_\nu(L_{j+1}(x))$.[6] As we show below, this can be stated as $c_{1,j}(x) < c_{1,j+1}(x)$. That is, if the DM's risk aversion is so low that she prefers the riskiest lottery to a safer one, then she also prefers it to an even safer one. Finally, there are no three-way ties. That is, for a given x there are no alternatives $\{j, k, l\}$ such that $U_\nu(L_j(x)) = U_\nu(L_k(x)) = U_\nu(L_l(x))$.[7]

B. Preferences with Single Crossing Property

There is a continuum of DMs. Each of them faces a choice among a finite number of alternatives, i.e., a choice set, which is denoted $\mathcal{D} = \{1, \ldots, D\}$. The

---

[5]Under CRRA, it is implied that DMs' initial wealth is known to the researcher. NTD utility is defined in Cohen and Einav (2007) and in Barseghyan et al. (2013).

[6]We analytically verify this claim for our application in Online Appendix B, but it can also be checked numerically for any given dataset.

[7]It is straightforward to very this condition, and we do so in our application.

number of alternatives is invariant across DMs. Alternatives vary by their utility-relevant characteristics and are distinguished by (at least) one characteristic, $d_j \in \mathbb{R}$, $j \in \mathcal{D}$, which is DM invariant. This characteristic reflects the quality of alternative j (e.g., insurance deductible). When it is unambiguous, we may write d_j instead of "alternative j". Other characteristics may vary across DMs or across alternatives. Our analysis rests on the excluded regressor(s) x. To keep the notation as lean as possible, we state our assumptions and results implicitly conditioning on all remaining characteristics. Hence, alternative j is fully characterized by (d_j, x_j). We consider two cases. In one case, all x_j's are perfectly correlated with a single (common) excluded regressor, x (e.g., $\bar{p}$ in our insurance example). In the other case, each x_j has its own variation conditional on all other x_k, $k \neq j$ (e.g., each alternative on the market exhibits locally independent price variation).

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

Each DM's valuation of the alternatives is defined by a utility function $U_\nu(d_j, x)$, which depends on a DM-specific index ν distributed according to $F(\cdot)$ over a bounded support.[8]

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

The DMs' draws of ν are not observed by the researcher. We require that DMs' preferences satisfy the Single Crossing Property (SCP).

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

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

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

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

Definition 1 (Dominated Alternatives). *Given x, alternative d_j is dominated if*

---

[8]We assume that while ν has bounded support, the utility function is well defined for any real valued ν.

there exists an alternative d_k such that $\forall \nu \in \mathbb{R}$, $U_\nu(d_k, x) > U_\nu(d_j, x)$.

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

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

We assume that alternatives are enumerated according to the Natural Ordering of Alternatives.[9] As the next fact shows, for high values of ν the preference over the Natural Ordering of Alternatives is reversed.

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

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

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

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

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

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

(i) $c_{1,2}(x) < c_{1,3}(x) < c_{2,3}(x)$ *and d_2 is the first best in $\{d_1, d_2, d_3\}$, $\forall \nu \in (c_{1,2}(x), c_{1,3}(x))$;*

(ii) $c_{1,2}(x) > c_{1,3}(x) > c_{2,3}(x)$ *and d_2 is Never-the-First-Best in $\{d_1, d_2, d_3\}$;*

(iii) $c_{1,2}(x) = c_{1,3}(x) = c_{2,3}(x)$ *and d_2 is strictly worse than either d_1 or d_3 for all ν except for $\nu = c_{1,2}(x)$ where there is a three-way tie among these alternatives.*

---

[9]Under this enumeration, d_j will be ordered in either ascending or descending order. In our example from the previous section, since d_j refers to the deductible and ν is the risk aversion coefficient, the natural ordering implies $d_1 > d_2 > \cdots > d_D$.

Fact 4 is a convenient way to distill and exploit the SCP. In particular, for any x, the complete preference order of the alternatives is known for all DMs as well as the identity (of the preference parameter) of the DM indifferent between any two alternatives d_j and d_k.

II. Identification

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

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

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

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

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

The last two conditions are vacuous in the standard full consideration model, while the first two are typically stated as data requirements.

We first discuss how to obtain identification and the role of Assumptions I0-I3 in the simplest case of two alternatives (Section II.A). We then consider the general model with D alternatives. Table 1 organizes our results by assumptions imposed, the consideration mechanism assumed, data availability, and the theorems' conclusions. Theorems 1-3 in Section II.B demonstrate that the preference distribution and some features of the consideration mechanism are identified with a single excluded regressor. Next, we show that alternative-specific variation allows for identification of both the preference distribution and the consideration mechanism when consideration depends on preferences and one of the excluded regressors (Theorem 4 and Corollary 1 in Section II.C). We discuss testing for limited consideration in Section II.D. We then turn to the ARC model in Section II.E. We show that the full model is identified with a single excluded regressor (Theorem 5). Moreover, identification attains for a particular case where consideration depends on preferences (Theorem 6). Finally, Theorem 7 shows that with alternative-specific variation, identification attains when consideration of each alternative depends both on preferences and its own regressor, without requiring full support.

A. The Role of the Canonical Assumptions

Let the choice set be binary and suppose that the DM considers both alternatives. In addition, let x be a scalar so that there is a single excluded regressor. Under Assumptions T0-T2 and I0-I3, any realization of x is associated with a

TABLE 1—IDENTIFICATION THEOREMS

	Assumptions				Consideration Mechanism	Excluded Regressor	Identification Result	
	I0	I1	I2	I3			Preferences	Consideration
Theorem 1	✓	✓	✓	✓	Generic	Single	✓	✓[1]
Theorem 2	✓	✓		✓	Generic	Single	✓	✓[1]
Theorem 3	✓		✓	✓	Loosely Ordered	Single	✓	
Theorem 4	✓			✓	Generic	Alternative Specific	✓	✓
Corollary 1	✓				Generic	Alternative Specific	✓	✓
Theorem 5	✓	✓	✓	✓	ARC	Single	✓	✓
Theorem 6	✓	✓		✓	ARC	Single	✓	✓
Theorem 7	✓				ARC	Alternative Specific	✓	✓

Note: In Theorems 1 and 2 we identify features of the consideration mechanism.

single conditional moment in the data:

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

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

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

Turning to limited consideration, suppose that d_1 is considered with probability $0 < \varphi_1 \leq 1$, and whenever it is considered so is d_2.[10] Then, d_1 is chosen when it is considered and it is preferred to d_2, yielding:

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

At first glance, it appears that the distribution of preferences is identified up to a constant. Yet, at the boundary of the support $\Pr(d = d_1|x^1) = \varphi_1 F(c_{1,2}(x^1)) = \varphi_1$, so that φ_1 is identified. Once φ_1 is known, the distribution $F(\cdot)$ is identified by varying $c_{1,2}(x)$ over the support of ν, similar to the full consideration case. We now explore what happens to identification if Assumptions I0-I3 are not satisfied.

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

Assumption I1 fails: the variation in x is such that $c_{1,2}(x)$ only covers an interval $[\nu^l, \nu^u] \subset [0, \bar{\nu}]$. Then the data provide no information about preferences

[10]With two alternatives this implies that d_2 is always considered.

outside of the interval $[\nu^l, \nu^u]$. Inside the interval, the conditional distribution $F(\nu | \nu \in [\nu^l, \nu^u]) = \frac{F(\nu) - F(\nu^l)}{F(\nu^u) - F(\nu^l)}$ is identified under both limited and full consideration. The consideration probability (and hence the scale of $F(\cdot)$) is partially identified and satisfies the bounds $\Pr(d = d_1 | x^u) \leq \varphi_1 \leq 1$, where x^u is such that $c_{1,2}(x^u) = \nu^u$. Point identification can be attained if an additional assumption is maintained to pin down the scale of $F(\cdot)$. For example, one can simply assume full consideration and set $\varphi_1 = 1$.

Assumption I2 fails: φ_1 depends on preferences and this dependence is arbitrary. Then identification breaks down completely as there is one data moment to identify two unknown objects. However, since we assume – as it is common in the econometrics literature – that the density function of ν is continuous and strictly positive, identification is possible for some types of dependence between consideration and preferences. Suppose there are two consideration types:

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

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

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

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

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

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

To summarize the case of the binary choice set, the only seemingly real difference in identification is that without large support the scale of the preference distribution $F(\cdot)$ is partially identified under limited consideration, while it is assumed to be known under full consideration. The key to identification is a one-to-one mapping from a data moment, $\Pr(d = d_1 | x)$, and the preference dis-

tribution $F(\cdot)$ at a single point on the support, $c_{1,2}(x)$. As we will show next, even with just a single excluded regressor, Assumptions I0-I3 allow for such a mapping to be constructed for a generic consideration mechanism and a choice set of arbitrary size.

B.   Single Common Excluded Regressor

We start by introducing general notation for consideration probabilities.

Definition 3. *Let $\mathcal{Q}^x_\nu(\mathcal{K})$ be the probability that, given x, the DM with preference parameter ν draws consideration set $\mathcal{K} \subset \mathcal{D}$ conditional on x. Let $\mathcal{O}^x_\nu(\mathcal{A};\mathcal{B})$ be the probability that, given x, every alternative in set $\mathcal{A}$ is in the consideration set and every alternative in set $\mathcal{B}$ is not for the DM with preference parameter ν:*

$$\mathcal{O}^x_\nu(\mathcal{A};\mathcal{B}) \equiv \sum_{\mathcal{K}:\ \mathcal{A}\subset\mathcal{K},\ \mathcal{B}\cap\mathcal{K}=\emptyset} \mathcal{Q}^x_\nu(\mathcal{K}).$$

The subscript is suppressed when consideration does not depend on preferences, and the superscript is suppressed when it does not depend on the excluded regressor(s).

To ease exposition, we build our discussion around a choice set with three alternatives, $\mathcal{D} = \{d_1, d_2, d_3\}$, such that $c_{1,2}(x) < c_{1,3}(x) < c_{2,3}(x)$ for all x. That is, by Fact 3, if $U_\nu(d_1, x) > U_\nu(d_2, x)$ then $U_\nu(d_1, x) > U_\nu(d_3, x)$ for all x. Suppose consideration is independent of preferences and of the excluded regressor. Then the choice frequencies of d_1 and d_3 are

$$\Pr(d = d_1|x) = \mathcal{O}(\{d_1, d_2\}; \emptyset)F(c_{1,2}(x)) + \mathcal{O}(\{d_1, d_3\}; d_2)F(c_{1,3}(x))$$
$$+ \mathcal{O}(d_1; \{d_2, d_3\});$$

$$\Pr(d = d_3|x) = \mathcal{O}(\{d_1, d_3\}; d_2)(1 - F(c_{1,3}(x))) + \mathcal{O}(\{d_2, d_3\}; \emptyset)(1 - F(c_{2,3}(x)))$$
$$+ \mathcal{O}(d_3; \{d_1, d_2\}).$$

Consider the expression for $\Pr(d = d_1|x)$. Its RHS has three terms. The first term captures the case when d_1 is considered along with d_2, which happens with probability $\mathcal{O}(\{d_1, d_2\}; \emptyset)$. Given the relative position of the cutoffs, whether d_3 is considered or not is irrelevant. The DM will choose d_1 over d_2 if and only if her preference parameter is below $c_{1,2}(x)$. The second term captures the case when d_1 is considered along with d_3, but d_2 is not considered, which happens with probability $\mathcal{O}(\{d_1, d_3\}; d_2)$. Then the relevant cutoff for choosing d_1 is $c_{1,3}(x)$. Third, when d_1 is the only alternative considered, it is chosen regardless of the DM's risk aversion. This event occurs with probability $\mathcal{O}(d_1; \{d_1, d_2\})$.

Since there are two cutoffs, $c_{1,2}(x)$ and $c_{1,3}(x)$, that enter the moment $\Pr(d = d_1|x)$, there is not, without additional assumptions, a one-to-one mapping between the moment and the preference distribution at one point on the support, as it was

the case in Section II.A. That is, as x changes, the observed choice frequency of d_1 may change because of two types of marginal DMs: those indifferent between d_1 and d_2, and those indifferent between d_1 and d_3. This is apparent in the following derivative:

$$\text{(3)} \qquad \frac{d\Pr(d = d_1|x)}{dx} = \mathcal{O}(\{d_1, d_2\}; \emptyset) f(c_{1,2}(x)) \frac{dc_{1,2}(x)}{dx}$$
$$+ \mathcal{O}(\{d_1, d_3\}; d_2) f(c_{1,3}(x)) \frac{dc_{1,3}(x)}{dx}.$$

The corresponding equation for $\Pr(d = d_3|x)$ does not immediately help, as it brings about $f(\cdot)$ evaluated at yet another cutoff, $c_{2,3}(x)$:

$$\text{(4)} \qquad \frac{d\Pr(d = d_3|x)}{dx} = - \mathcal{O}(\{d_1, d_3\}; d_2) f(c_{1,3}(x)) \frac{dc_{1,3}(x)}{dx}$$
$$- \mathcal{O}(\{d_2, d_3\}; \emptyset) f(c_{2,3}(x)) \frac{dc_{2,3}(x)}{dx}.$$

Identification with Large Support

When Assumption I1 holds, we can construct a one-to-one mapping sequentially. The algorithm for doing so consists of four steps. First, we rewrite Equation (3) as

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

where $\phi \equiv \frac{\mathcal{O}(\{d_1, d_3\}; d_2)}{\mathcal{O}(\{d_1, d_2\}; \emptyset)}$ and $\hat{f}(\nu) \equiv \mathcal{O}(\{d_1, d_2\}; \emptyset) f(\nu)$. Second, for ν's near the far end of the support, we can find x and x' such that $\nu = c_{1,2}(x) < \bar{\nu} < c_{1,3}(x)$ and $\nu = c_{1,3}(x') < \bar{\nu} < c_{2,3}(x')$. For any such pair, $f(c_{1,3}(x)) = f(c_{2,3}(x')) = 0$, and, hence, by Equations (3) and (4):

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

The first equation identifies $\hat{f}(\nu)$, while the ratio of the two equations identifies ϕ. Third, whenever $\hat{f}(c_{1,3}(x))$ is known, $\hat{f}(c_{1,2}(x))$ is uniquely pinned down by Equation (5). Because $c_{1,2}(x) < c_{1,3}(x)$, $\forall x$, we can learn $\hat{f}(\cdot)$ sequentially:

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

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

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

For this approach to work, $c_{1,3}(x)$ cannot "catch up" to $c_{1,2}(x)$ (i.e., as assumed, $c_{1,2}(x) < c_{1,3}(x)$ whenever $c_{1,2}(x)$ is on the support). This requires that DMs with preference coefficients on the support are never indifferent between d_1 and two other alternatives – i.e. there are no three way ties involving d_1. Fourth, integration of $\hat{f}(\nu)$ over the entire support recovers the scale and the true density. Indeed,

$$\int_0^{\bar{\nu}} \hat{f}(\nu)d\nu = \mathcal{O}(\{d_1, d_2\}; \emptyset) \int_0^{\bar{\nu}} f(\nu)d\nu = \mathcal{O}(\{d_1, d_2\}; \emptyset)$$

pins down $\mathcal{O}(\{d_1, d_2\}; \emptyset)$, and hence $f(\cdot)$ is identified. A generalization of this strategy yields our first formal result.

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

1) *The consideration mechanism is s.t. with positive probability d_1 and d_2 are considered together;*

2) *Assumption I1 holds for $\mathcal{X} \subset \mathcal{S}$ s.t. $\forall x \in \mathcal{X}$*

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

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

The first assumption of the theorem ensures that a generalized version of Equation (5) is informative. The second assumption implies that the cutoffs for alternative d_1 are ordered: $c_{1,j}(x) < c_{1,j+1}(x)$. While Theorem 1 requires large support for the excluded regressor, it does not generally require it to exhibit variation that forces alternative d_1 to go from being the first best to the least preferred. Rather, the theorem requires that at one extreme of the support alternative d_1 dominates all others. However, at the other extreme we only require that d_2 is preferred to d_1 for all DMs. Identification is attained for any consideration mechanism that allows d_1 and d_2 to be considered together with positive probability. Moreover, if the probability of being considered together is zero for d_1 and d_2, but positive for d_1 and d_3, the theorem still holds as long as the assumptions of the theorem hold for d_3 instead of d_2. Theorem 1 identifies some features of the consideration mechanism. These features may be sufficient for identifying the entire mechanism. In particular, as shown in Section II.E, Theorem 1 yields identification of the ARC model, including the consideration mechanism.

Dependence between consideration and preferences.

We next generalize the example in Section II.A by allowing for high/low consideration types.

Assumption I2.BCT (Binary Consideration Types). *For some unknown $\nu^* \in$*

$(0, \bar{\nu})$:

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

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

Theorem 2. *Suppose Assumptions I0, I2.BCT, I3, T0-T2, and Condition 2 of Theorem 1 hold. Suppose Condition 1 of Theorem 1 holds for all ν. Then $f(\cdot)$ is identified and so is $\mathcal{O}_\nu(\{d_1, d_2\}; \emptyset)$. Suppose $\frac{d \Pr(d = d_1 | x)}{dx}$ is discontinuous. Then ν^* is identified. If, in addition, $c_{1,j}(x) < \nu^*$ for some $x \in \mathcal{X}$ and $j > 2$, then $\mathcal{O}_\nu(\{d_1, d_j\}; \{d_2, \ldots, d_{j-1}\})$ is also identified.*

A discontinuity in $\frac{d \Pr(d = d_1 | x)}{dx}$ may occur when a cutoff $c_{1,j}(x)$ crosses ν^*. In some cases it may not happen despite binary consideration. For example, the probability of considering d_1 and d_2 may jump but in a way that $\mathcal{O}_\nu(\{d_1, d_2\}; \emptyset)$ remains constant. In such a case, $f(\cdot)$ is identified but not necessarily the breakpoint ν^*.

The theorem holds if Assumption I2.BCT is replaced with

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

for some unknown $x^* \in \mathcal{S}$. In sum, preferences can be identified even when there are threshold effects affecting consideration. Assumption I2.BCT is one instance where Assumption I2 does not hold but identification attains. Another instance, which we establish for the ARC model in Section II.E, is proportionally shifting consideration.

IDENTIFICATION WITHOUT LARGE SUPPORT

Returning to our example with three alternatives, it is immediate to see that if whenever d_1 is considered so is d_2, i.e. $\mathcal{O}(\{d_1, d_3\}; d_2) = 0$, the one-to-one mapping is restored. Indeed, the second term on the RHS of Equation (3) disappears and we are back to Equation (1).

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

1) *The consideration mechanism is such that d_1 is considered with positive probability and whenever it is considered so is d_2;*

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

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

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

The proposition above uses (the derivative of) $\Pr(d = d_1|x)$ to create the one-to-one mapping from data to the preference density function. Depending on the consideration mechanism, the same can be achieved using the derivative of $\Pr(d \in \{d_1, d_2, \ldots, d_j\}|x)$.

Definition 4 (Loosely Ordered Consideration)**.** *The consideration mechanism is loosely ordered around j, $j < D$, if whenever alternatives d_k and d_l, $k \leq j < l$, are both considered, so are d_j and d_{j+1}. In addition, d_j and d_{j+1} have a positive probability of being considered together.*

Theorem 3. *Suppose Assumptions I0, I2, I3, T0-T2 hold, and*

1) *The consideration mechanism is loosely ordered around j.*

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

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

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

Condition 1 in Theorem 3 – a loosely ordered consideration mechanism – splits the choice set into "low quality" and "high quality" sets. Any subset of the low quality set can form the consideration set and so can any subset of the high quality set. However, if a consideration set contains both high and low quality alternatives, then it must also contain the "bridging" alternatives $\{d_j, d_{j+1}\}$. The following mechanisms can generate loosely ordered consideration:

I. **Bottom-Up consideration:** Alternative d_k is considered only if d_{k-1} is considered;

II. **Top-Down consideration:** Alternative d_k is considered only if d_{k+1} is considered;

III. **Center-to-edges consideration:** Alternative d_j, $1 < j < D$, is always considered. Alternative d_k, $k > j$, is considered only if d_{k-1} is considered. Alternative d_k, $k < j$, is considered only if d_{k+1} is considered;

IV. **Trimmed-from-the-edges consideration:** Only consideration sets of the form $\mathcal{K} = \{d_k, d_{k+1}, \ldots, d_{k+l}\}$ can occur with positive probability.

The identification result in Theorem 3 extends to mixtures of these mechanisms. They cover a wide array of models including versions of threshold models (Kimya, 2018), (partial) elimination-by-aspects (Tversky, 1972), extremeness aversion (Simonson and Tversky, 1992), and edge aversion (Teigen, 1983; Christenfeld, 1995; Rubinstein, Tversky and Heller, 1997; Attali and Bar-Hillel, 2003), as well as models that embed budget or liquidity constraints.

Condition 2 in Theorem 3 requires that whenever a DM prefers d_j to d_{j+1}, she also prefers d_j to all high quality alternatives; and whenever a DM prefers d_{j+1} to d_j, she also prefers d_{j+1} to all low quality alternatives. This condition can be tested in any given dataset and is automatically satisfied if no alternative is never-the-first-best.

The fundamental difference between Theorems 1 and 3 is that the former imposes the large support requirement, while the latter does not. On the other hand, Theorem 1 imposes less restrictions on the consideration mechanism than Theorem 3.

C. Alternative-specific Excluded Regressors

With alternative-specific excluded regressors we can allow for consideration to depend on preferences. To illustrate, we continue to assume that the choice set is $\{d_1, d_2, d_3\}$. However, now each alternative has its own regressor x_j that only affects the utility of alternative j: $x = (x_1, x_2, x_3)$. In addition, these regressors vary independently of one another and each consideration set contains at least two alternatives.

Identification is built on the following insight. Consider the change in the choice frequency of alternative d_1 in response to an incremental change in x_2 (e.g., a price increase for alternative d_2). The DMs who may switch to d_1 are those indifferent between d_1 and d_2 and consider them both. If these DMs prefer d_1 and d_2 to d_3, whether d_3 is considered is irrelevant; otherwise, for the response to occur, d_3 should not be considered. These two cases translate to the following statements: (i) $c_{1,2}(x) < c_{1,3}(x) < c_{2,3}(x)$; and (ii) $c_{2,3}(x) < c_{1,3}(x) < c_{1,2}(x)$ and d_3 is not considered. No other ordering of cutoffs can occur by Fact 4. With alternative specific variation we can construct two vectors of regressors, x^i and x^{ii}, such that $\nu = c_{1,2}(x^i) < c_{1,3}(x^i) < c_{2,3}(x^i)$ and $c_{2,3}(x^{ii}) < c_{1,3}(x^{ii}) < c_{1,2}(x^{ii}) = \nu$.[11] The derivative of the choice frequency of d_1 with respect to x_2 for these cases are, respectively:

$$(i) \quad \frac{\partial \Pr(d = d_1 | x)}{\partial x_2} = \left[\mathcal{Q}_\nu(\{d_1, d_2\}) + \mathcal{Q}_\nu(\{d_1, d_2, d_3\}) \right] f(\nu) \frac{\partial c_{1,2}(x)}{\partial x_2};$$

$$(ii) \quad \frac{\partial \Pr(d = d_1 | x)}{\partial x_2} = \left[\mathcal{Q}_\nu(\{d_1, d_2\}) \phantom{+ \mathcal{Q}_\nu(\{d_1, d_2, d_3\})} \right] f(\nu) \frac{\partial c_{1,2}(x)}{\partial x_2}.$$

It follows that $\mathcal{Q}_\nu(\{d_1, d_2\}) f(\nu)$ and $\mathcal{Q}_\nu(\{d_1, d_2, d_3\}) f(\nu)$ are identified. In a similar fashion, $Q_\nu(\{d_1, d_3\}) f(\nu)$ and $Q_\nu(\{d_2, d_3\}) f(\nu)$ are identified. Hence, the consideration probability of each non-singleton set is identified up to the same

---

[11]First, we can construct an x such that $U_\nu(d_1, x) = U_\nu(d_2, x) = U_\nu(d_3, x)$. To do so, we fix the price of the first alternative, x_1, and find a price for the second alternative that makes the DM with preference parameter ν indifferent between d_1 and d_2. Since x_3 does not affect the utility of d_1 nor d_2, we can find an x_3 so that the DM is indifferent between d_3 and d_1, and hence she is indifferent between all three alternatives. The two cases are then constructed by taking a small perturbation of x_3. Taken in a direction that reduces $U_\nu(d_3, x)$ generates Case (i); and in the opposite direction Case (ii).

scale. The scale, however, is identified because consideration probabilities must sum to one: $f(\nu) = \sum_{\mathcal{K}} \mathcal{Q}_\nu(\mathcal{K}) f(\nu)$. Hence, $\mathcal{Q}_\nu(\mathcal{K})$ is also identified for each $\mathcal{K}$. The following theorem generalizes this idea.

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

Theorem 4. *Suppose Assumptions I0, I3, T0-T2 hold, there is alternative-specific variation, and the choice set contains at least three alternatives. Suppose*

1) *Each consideration set contains at least two alternatives and $\mathcal{Q}_\nu(\cdot)$ is measurable;*

2) *For a given value of ν, there exists an x with an open neighborhood around it in $\mathcal{S}$ s.t.*

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

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

The assumptions of the theorem above rule out singleton (and empty) consideration sets: identification is impossible with singleton consideration sets and arbitrary dependence on preferences, because any empirical choice frequency can be explained by such consideration sets. An alternative approach is to have one alternative – the "default" – that is always considered as the following corollary demonstrates. The identification argument exploits the response of $\Pr(d = d_j | x)$ to changes in x_k, but not the response of $\Pr(d = d_k | x)$ to changes in x_j. Hence, $D - 1$ excluded regressors are sufficient for identification, allowing for arbitrary dependence of consideration on one (the default's) excluded regressor.

Corollary 1. *Suppose Assumptions I0, T0-T2 hold, there is alternative-specific variation, the choice set contains at least three alternatives, all consideration sets contain d_1, and*

1) *Consideration is independent of $x_{-1} \equiv (x_2, \ldots, x_D)$: $\mathcal{Q}_\nu^x(\mathcal{K}) = \mathcal{Q}_\nu^{x_1}(\mathcal{K})$, and $\mathcal{Q}_\nu^{x_1}(\cdot)$ are measurable functions, continuous in x_1;*

2) *The consideration of $\mathcal{K} = \{d_1\}$ is independent of ν: $\mathcal{Q}_\nu^{x_1}(d_1) = \mathcal{Q}^{x_1}(d_1) < 1$, $\forall \nu$;*

3) *For a given value of x_1 and each value of $\nu \in [0, \bar{\nu}]$, there exists an x_{-1} and an open neighborhood around $x = (x_1, x_{-1})$ in $\mathcal{S}$ s.t.*

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

Then $f(\nu)$ is identified and so are $\mathcal{Q}_\nu^{x_1}(\mathcal{K})$, $\forall \mathcal{K} \subset \mathcal{D}$, for all ν on the support.

Corollary 1 generalizes the model of Heiss et al. (2016); Ho, Hogan and Scott Morton (2017) in two dimensions. First, here each subset of the choice set containing the default has its own probability of being drawn. Second, this probability can vary with the DM's preferences as well as with the excluded regressor of the default alternative.

D.   Testing for Limited Consideration

Since full consideration is a special case of limited consideration, it follows from the identification results above that under the SCP one can test for full consideration. The proposition below states one way of doing so *without*: (1) relying on large support; (2) specifying a consideration mechanism; or (3) invoking the independence assumptions I2 and I3.

Proposition 2. *Suppose Assumptions I0, T0-T2 hold. Suppose there exist* $x, x' \in \mathcal{S}$, *and sets* $\mathcal{L}, \mathcal{L}' \subset \mathcal{D}$ *s.t. for some* $\nu^* \in [0, \bar{\nu}]$

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

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

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

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

E.   The ARC Model

We now introduce a specific consideration mechanism, while maintaining the preference structure, including the SCP, from Section I.B. We refer to this model as the Alternative-specific Random Consideration (ARC) model (Manski, 1977; Manzini and Mariotti, 2014). Each alternative d_j appears in the consideration set with probability φ_j independently of other alternatives. For now, we assume that these probabilities do not depend on DMs' preferences or the excluded regressor. Once the consideration set is drawn, the DM chooses the best alternative according to her preferences. To avoid empty consideration sets, following Manski (1977), we assume that at least one alternative whose identity is unknown to the researcher is always considered.[12]

---

[12]In the previous version of this paper (Barseghyan, Molinari and Thirkettle, 2019) this completion rule is called Preferred Option(s). There we also provide identification results for other completion

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

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

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

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

Under the assumptions of Theorem 1, identification attains. Notably, each consideration parameter φ_j is identified (as long as d_j is chosen with positive probability at some x).

Theorem 5. *Suppose Assumptions I0, I2-I3, T0-T2, ARC hold, and Assumption I1 holds for $\mathcal{X} \subset \mathcal{S}$ s.t. $\forall x \in \mathcal{X}$*

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

Then $f(\cdot)$ is identified and so are φ_1 and φ_2. In addition, if $\Pr(d = d_j | x) \neq 0$ for some x, then φ_j is identified.

PREFERENCE-DEPENDENT CONSIDERATION

Returning to our example with three alternatives, recall that we have an additional moment $\Pr(d = d_3 | x)$. The information it provides allows us to identify some forms of dependence between consideration and preferences, i.e., to relax Assumption I2. To see how, suppose d_2 is always considered. Then, with preference dependence, the choice frequencies become:

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

The ratio of the derivatives of these two moments yields $\frac{\varphi_1(\nu)}{\varphi_3(\nu)}$. More assumptions are required to obtain point identification of the $\varphi_j(\nu)$'s. In Section II.B we provided identification results for Binary Consideration Types. Here, leveraging the additional structure provided by the ARC model, we can allow for more flexible dependence between consideration and preferences. We do so through

---

rules, including Coin Toss (if the empty consideration set is drawn, the DM randomly uniformly picks one alternative from the choice set, i.e. each alternative has probability $1/D$ of being chosen), Default Option (there is a preset alternative that is chosen if the empty set is drawn), and Outside Option (the DM exits the market if the empty set is drawn).

a *proportionally shifting* consideration mechanism, formally defined below. This mechanism may arise when there is a cost to evaluate each alternative. In such a case, the DMs may consider alternatives that they ex-ante deem more aligned with their preferences (e.g., the DM's consideration shifts away from riskier to safer alternatives as her risk aversion increases).

Assumption ARC.P (ARC with Proportional Consideration). *The consideration mechanism follows the ARC model with $\{D \geq 4 \ \& \ 1 \leq d^* \leq D\}$ or $\{D = 3 \ \& \ d^* = 2\}$, and*

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

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

In the case with three alternatives, $\frac{\varphi_1(\nu)}{\varphi_3(\nu)} = \frac{\varphi_1(1-\alpha(\nu))}{\varphi_3(1+\alpha(\nu))}$. From this, $\frac{\varphi_1}{\varphi_3}$ is identified when x and x' are chosen such that $c_{1,2}(x) = c_{2,3}(x') = \bar{\nu}$. Once $\frac{\varphi_1}{\varphi_3}$ is identified, $\frac{1-\alpha(\nu)}{1+\alpha(\nu)}$ is known for all ν; hence, $\alpha(\nu)$ can be solved for. Identification of $f(\nu)$ follows from substituting $\alpha(\nu)$ into the expression for $\frac{\Pr(d=d_1|x)}{dx}$. The theorem below generalizes this argument.

Definition 6. *(No Three Way Ties) For a given x, there are no-three way ties if $\nexists \nu \in [0, \bar{\nu}]$ and $\{j, k, l\}$ s.t. $U(d_j, x) = U(d_k, x) = U(d_l, x)$.*

Theorem 6. *Suppose Assumptions I0, I3, T0-T2, ARC.P hold, and Assumption I1 holds for $\mathcal{X}$ s.t. $\forall x \in \mathcal{X}$ there are no three-way ties and*

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

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

The conditions of the theorem are stronger than in Theorem 5, as they impose relative order of the cutoffs not only for alternative d_1 but also for d_D. In many cases, the relative order of $c_{1,j}(x)$'s alone is sufficient, for example when d_1 is always considered.

IDENTIFICATION WITH ALTERNATIVE-SPECIFIC EXCLUDED REGRESSORS

By leveraging features of the ARC model, the identification results in Section II.C can be extended to the case where the consideration of d_j is a function both of x_j and preferences. This differs from Corollary 1, which restricted the

consideration of alternative d_j to depend only on the default alternative's excluded regressor. We continue to assume that the choice set is $\{d_1, d_2, d_3\}$ and that d_2 is always considered. Let the consideration of d_j be a measurable function of x_j and ν, continuous in its first argument: $\varphi_j = \varphi_j(x_j, \nu)$. Similar to the example in Section II.C, we construct two vectors, x^i and x^{ii}, such that: *(i)* $\nu = c_{1,2}(x^i) < c_{1,3}(x^i) < c_{2,3}(x^i)$; and *(ii)* $c_{2,3}(x^{ii}) < c_{1,3}(x^{ii}) < c_{1,2}(x^{ii}) = \nu$. The derivative of the choice frequency of d_1 with respect to x_2 for these cases are, respectively:

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

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

The ratio of the expressions in Equation (6) identifies $\varphi_3(x_3, \nu)$. Using a similar logic, we can identify $\varphi_1(x_1, \nu)$. Plugging these consideration probabilities into Equation (6) identifies $f(\nu)$. In sum, alternative-specific variation yields identification without large support and without the independence Assumptions I2 and I3. It is also possible to allow consideration of d_1 (and d_3) to depend on x_1, ν, as well as x_3. The key exclusion restriction in this case is that the consideration of d_2 is independent of all components of x. Our last identification result generalizes this example.

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

Theorem 7. *Suppose Assumptions I0, T0-T2, ARC.AS hold. Suppose there is alternative-specific variation and the choice sets contain at least three alternatives. Suppose for a given value of ν there exists an $x = (x_1, x_2, \ldots, x_D)$, and an open neighborhood around it in $\mathcal{S}$, s.t.*

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

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

Existing identification results that rely on alternate-specific variation (Goeree, 2008; Abaluck and Adams, 2020; Kawaguchi, Uetake and Watanabe, 2020) allow for consideration dependence on its own regressor, but not preferences. Theorem 7 states identification for a general version of the ARC model where the alternative-specific consideration probability can depend on both its own regressor and DMs' preferences.

III.   Likelihood and Tractability

We now turn to the computational aspects of limited consideration models under the SCP and, in particular, of their likelihood function. Consider a generic consideration mechanism.

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

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

Then,

$$(7) \qquad \Pr(d_j|x) = \int \Pr(d_j|x, \nu)dF = \int \mathcal{O}_\nu^x(d_j; \mathcal{B}_\nu(d_j, x))dF.$$

The object on the RHS does not require evaluating the utility of each alternative within each possible consideration set. In fact, $U_\nu(d_j, x)$ needs to be computed only once for each ν, d_j, and x to create $\mathcal{B}_\nu(d_j, x)$, which does not vary with the consideration set. Hence the computational complexity lies in the mapping from $\mathcal{O}_\nu^x(\cdot)$ to the parameters governing the consideration mechanism. This, however, may not even require enumerating all possible consideration sets. To demonstrate this with a concrete example, we proceed with the basic ARC model. In this case, the RHS of Equation (7) is:

$$(8) \qquad I(d_j|x) \equiv \varphi_j \int \prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k)dF.$$

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

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

where ν_h's are the sequentially ordered breakpoints augmented by the integration endpoints: $\nu_0 = 0$ and $\nu_D = \bar{\nu}$. This expression is trivial to evaluate given $F(\cdot)$ and breakpoints $\{\nu_h\}_{h=0}^D$. More importantly, since the breakpoints are invariant with respect to the consideration probabilities, they are computed only once for

each x. This simplifies the likelihood maximization routine by orders of magnitude, as each evaluation of the objective function involves a summation over products with at most D terms. A second approach is to compute $I(d_j|x)$ using Riemann approximation:

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

where M is the number of intervals in the approximating sum, $\frac{\bar{\nu}}{M}$ is the intervals' length, ν_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 ν_m, $m = 1, \ldots, M$. These rankings determine $\mathcal{B}_{\nu_m}(d_j, x)$. The likelihood maximization is now a standard search routine over $\{\varphi_j\}_{j=1}^{D}$ and $f(\cdot)$. Our theory restricts $f(\cdot)$ to the class of continuous and strictly positive functions. In practice, the search is over a class of non-parametric estimators for $f(\cdot)$.[13] If the density is parameterized, i.e., $f(\nu_m) \equiv f(\nu_m; \theta^f)$, then the maximization is over $\{\varphi_j\}_{j=1}^{D}$ and θ^f. Finally, the interval midpoints are the same across all DMs as they do not depend on x, further reducing computational burden.[14]

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

As a final remark, if alternative d_j is never chosen, then one can conduct estimation as if d_j were not in the choice set. Indeed, per Equation (8), φ_j contributes positively to the likelihood if and only if alternative d_j is chosen. When it is never chosen, it may only enter via the term $(1 - \varphi_j)$; hence, the likelihood will be maximized by setting $\varphi_j = 0$. Therefore, setting $\varphi_j = 0$ for all zero-share alternatives, regardless of why they were not chosen, has no impact on estimation. This too may speed up estimation.

IV. Limited Consideration and RUM: A Comparison

We focus on a *standard* application of the RUM with full consideration in the context of our example in Section I.A. The final evaluation of the utility that the

---

[13]One could use a mixture of Beta distributions (Ghosal, 2001), as we do in Section V.

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

DM derives from alternative j now includes an additively separable error term:

$$(9) \qquad V_\nu(L_j(x)) = U_\nu(L_j(x)) + \varepsilon_j,$$

where, as before, ν captures unobserved heterogeneity in preferences, and ε_j is assumed independent of the random coefficients (in this application, ν).

Typical implementations of this model further specify that ε_j is i.i.d. across alternatives (and DMs) with a Type 1 Extreme Value distribution, following the seminal work of McFadden (1974). This yields a Mixed Logit that is distinct from the commonly used one in McFadden and Train (2000). In their model, random coefficient(s) enter the utility function linearly, while in the context of expected utility they enter nonlinearly. We now discuss two properties of the Mixed Logit that hinder its applicability in our context.

A. Monotonicity

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:

Proposition 3. *(Non-monotonicity in RUM, Apesteguia and Ballester, 2018; Wilcox, 2008) In Model (9) with HARA preferences and ε_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.*

To see why, consider two non-dominated alternatives d_j and d_k such that d_j is riskier than d_k. A risk neutral DM prefers d_j to d_k, and hence will choose the former with higher probability. As risk aversion increases, the DM eventually becomes indifferent between d_j and d_k and chooses either of these alternatives with equal probability. As risk aversion increases further, she prefers d_k to d_j and chooses the latter with lower probability. However, as risk aversion gets even larger, the expected utility under HARA of any lottery with finite stakes converges to zero. Consequently, the choice probabilities of all alternatives, regardless of their riskiness, converge to a common value.[15] Hence, at some point the probability of choosing d_j is increasing in risk aversion.

To the contrary, our model with a limited consideration mechanism that is independent of preferences yields choice probabilities that are monotone in the preference parameter.

Property 1 (Generalized Preference Monotonicity). *A model satisfies generalized*

---

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

preference monotonicity if for any $\nu_1 < \nu_2$ and $J \in \mathcal{D}$:

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

In the context of risk preferences, Property 1 states that the probability of choosing one of the J riskiest alternatives declines as ν increases. Since Property 1 is satisfied for any choice set under the SCP and full consideration, it is also satisfied under limited consideration:

Proposition 4. *A model that satisfies the SCP (i.e., Assumption T2) and Assumption I2 satisfies Generalized Preference Monotonicity.*

B. Generalized Dominance

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

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

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

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) A model satisfies Generalized Dominance if for any x, d_j, and set $\mathcal{K} \subset \mathcal{D} \setminus \{d_j\}$ s.t. alternative d_j is never-the-first-best in $\mathcal{K} \cup \{d_j\}$*

$$\Pr(d = d_j | x) < \sum_{k \in \mathcal{K}} \Pr(d = d_k | x).$$

Generalized Dominance holds in the Mixed Logit model and, more broadly, in models that satisfy the conditional rank order property. However, it may not hold in some limited consideration models. For example, Generalized Dominance is violated if d_j is never-the-first best among $\{d_j, d_k, d_l\}$, is almost always considered, and alternatives d_k and d_l are rarely considered.

C. Limited Consideration as Ordinal RUM

In the Mixed Logit, the cardinality of the differences in the (random) expected utility of alternatives plays a crucial role in the determination of choice probabili-

ties, as it interacts with the realization of the additive error. In contrast, in models that satisfy the SCP, the DMs' choices are determined by the ordinal expected utility ranking of the alternatives. Hence, limited consideration models can be recast as Ordinal Random Utility models (ORUM), where the key departure from standard RUMs is the distribution of the additive error term.

Proposition 5. *(Limited Consideration as ORUM) A Limited Consideration Model is equivalent to an additive error random utility model with unobserved preference heterogeneity where all alternatives are considered, the DM's utility of each alternative $d_j \in \mathcal{D}$ is given by*

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

and $\vec{\varepsilon}(\nu, x) = \{\varepsilon_j(\nu, x)\}_{j=1}^{D}$ is distributed on $\{-\infty, 0\}^D$ according to $\Pr(\vec{\varepsilon}(\nu, x)) = \mathcal{Q}_\nu^x(\mathcal{K})$ for $\mathcal{K}$ s.t. $d_j \in \mathcal{K}$ if $\varepsilon_j(\nu, x) = 0$ and $d_k \in \mathcal{D} \setminus \mathcal{K}$ if $\varepsilon_j(\nu, x) = -\infty$.

Casting our limited consideration model as an ORUM clearly demonstrates its flexibility. In particular, our results show how to obtain identification when the errors are correlated with the excluded regressors, the preference parameter, and across alternatives.

TABLE 2—MODEL COMPARISONS

Error Structure	Mixed Logit	Basic ARC	Binary Types	Prop. Shifting	Generic Consideration
Support	$\mathbb{R}$	$\{-\infty, 0\}$	$\{-\infty, 0\}$	$\{-\infty, 0\}$	$\{-\infty, 0\}$
Indep. of x	Yes	Yes	Yes[1]	Yes	No
Indep. of ν	Yes	Yes	No[1]	No	No
Indep. across alternatives	Yes	Yes	Yes	Yes	No
Iden. across alternatives	Yes	No	No	No	No

Note: Binary Types can also be dependent on x but would require independence with ν in that case.

We conclude this section with Table 2, listing the differences across the Mixed Logit and limited consideration models. The first two columns summarize the differences between the basic ARC model and the Mixed Logit. The third column and fourth column remind the reader our two models with consideration depending on preferences. Finally, the last column highlights the fact that with alternative-specific variation we may also have dependence of the error term on the excluded regressor(s) as well as on the preference parameter.

V. Application

We offer an empirical analysis of households' decisions under risk. This analysis aims to illustrate how our method works and its ability to fit the data.

A. Data

We study households' deductible choices across three lines of property insurance: auto collision, auto comprehensive, and home all perils. The data come from a U.S. insurance company (Anonymous Firm, 2018). 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.[16] Online Table D.1 provides descriptive statistics for households' observable characteristics, which we use later to estimate households' preferences.[17] We observe the exact menu of alternatives available at the time of the purchase for each household and each line of coverage. The deductible alternatives vary across lines of coverage but not across households. Online Table D.2 presents the frequency of chosen deductibles in our data.

TABLE 3—PREMIUM QUANTILES FOR THE $500 DEDUCTIBLE

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

Premiums are set coverage-by-coverage as in the example from Section I.A. Online Table D.4 reports the average premium by context and deductible, and Table 3 summarizes the premium distributions for the $500 deductible. Premiums vary dramatically. The 99th percentile of the $500 deductible is more than ten times the corresponding 1st percentile in each line of coverage.

Claim probabilities stem from (Barseghyan, Teitelbaum and Xu, 2018), who derived them using coverage-by-coverage Poisson-Gamma Bayesian credibility models applied to a large auxiliary panel. Predicted claim probabilities (summarized in Table 4) 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.

---

[16]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.

[17]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 4—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

B.   Estimation Results

THE BASIC ARC MODEL: COLLISION

We start by presenting estimation results in a simple setting where the only choice is the collision deductible and observable demographics do not affect preferences. To execute our estimation procedure we set $\bar{\nu} = 0.02$, which is conservative (see Barseghyan, Molinari and 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, resulting in a total of seven parameters to be estimated. We let the data speak to the identity of the always-considered alternative.[18]

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

---

[18]In fact, the estimation is run under the Coin Toss completion rule that nests the possibility that any alternative can be always considered. The data chooses $\varphi_{1000} = 1$.
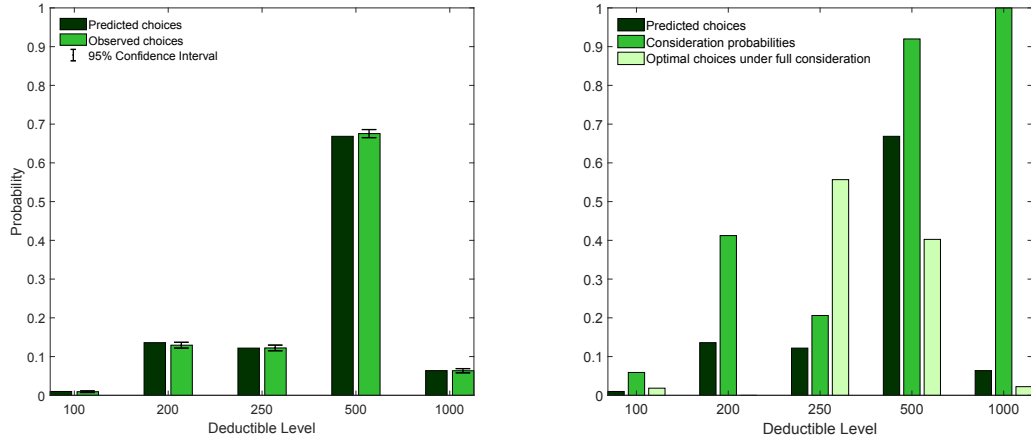
FIGURE 1. THE ARC MODEL

Note: 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.

increases and therefore the higher deductibles are chosen with lower probability. Average risk aversion must decline to compensate for this shift. 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 32% from 0.0037 to 0.0025 when all consideration parameters equal 0.9999.[19] To put these numbers into context, a DM with risk aversion equal to 0.0037 is willing to pay \$431 to avoid a \$1000 loss with probability 0.1, while a DM with risk aversion equal to 0.0025 is only willing to pay \$300 to avoid the loss.

The basic ARC model's ability to match the data extends also to conditional moments. The first two panels of Figure 2 show observed and predicted choices for the fraction of households facing low and high premiums, respectively, and the next two panels are for households facing low and high claim probabilities.[20] Finally, the last two panels display households who face both low claim probabilities and high prices and *vice versa*. It is transparent from Figure 2 that the model matches closely the observed frequency of choices across different subgroups of households facing a variety of prices and claim probabilities, even though some of these frequencies are quite different from the aggregate ones.

The ARC model's ability to violate Generalized Dominance is key in matching the data. In our dataset, because of the pricing schedule in collision, the \$200 is never-the-first best among {\$100, \$200, \$250} for 99.84% of all households and 100% of households who have chosen the \$200 deductible. It costs the same

---

[19]We cannot assume that all consideration probabilities are equal to one, since the \$200 deductible is never-the-first-best under full consideration and is chosen with positive probability.

[20]Low and high groups here are defined as households whose claim rate (or baseline price) are in the first and third terciles, respectively.
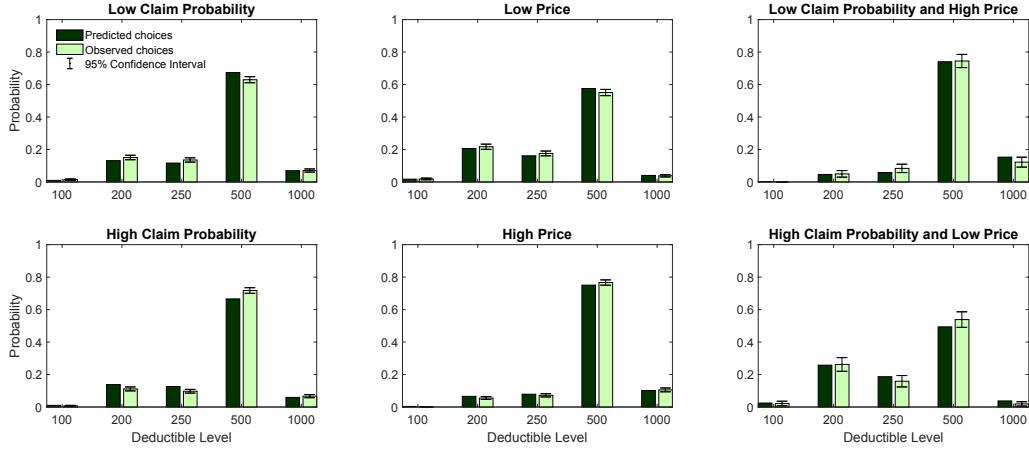
FIGURE 2. THE ARC MODEL: CONDITIONAL DISTRIBUTIONS

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.[21] Yet, overall the \$200 deductible is chosen roughly as often as the \$100 and \$250 deductibles combined. More so, for certain sub-groups the \$200 deductible is chosen *much more often* than the \$100 and \$250 deductible combined. It follows that a model satisfying Generalized Dominance cannot rationalize these choices.

Next we relax the assumption that demographic variables, $\mathbf{Z}$, do not influence risk preferences. In particular, conditional on demographics, preferences are distributed $\text{Beta}(\beta_1(\mathbf{Z}), \beta_2)$, where $\log \frac{\beta_1(\mathbf{Z})}{\beta_2} = \mathbf{Z}\gamma$, yielding a conditional mean preference value $E(\nu|\mathbf{Z}) = \frac{e^{\mathbf{Z}\gamma}}{1+e^{\mathbf{Z}\gamma}}\bar{\nu}$. The details of this step and the results are reported in Online Appendix E.1. Both consideration and preference estimates remain close to those reported above.

PROPORTIONALLY SHIFTING CONSIDERATION

We estimate the model with proportionally shifting consideration where the preference distribution and function $\alpha(\nu)$ may depend on demographic variables (Online Table E.2). Motivated by the findings of the previous section, we assume

---

[21]This pattern is at odds not only with EUT but also many non-EU models (Barseghyan, Molinari and Teitelbaum, 2016).

that the cheapest/riskiest alternative is always considered. The consideration probability of the remaining alternatives is equal to $\varphi_j(x, \nu) = \varphi_j(1 - \alpha(\nu|\mathbf{Z}))$, where $\alpha(\nu|\mathbf{Z}) = \xi_1(\mathbf{Z})\left(1 - \frac{\nu}{\bar{\nu}}\right)^{\xi_2}$, $\xi_1(\mathbf{Z}) = \frac{e^{\mathbf{Z}\rho}}{1 + e^{\mathbf{Z}\rho}}$, and ξ_2 is positive. We continue to assume that preferences are distributed $\text{Beta}(\beta_1(\mathbf{Z}), \beta_2)$.

The estimated average value of $\xi_1(\mathbf{Z})$ is 0.15, with the 95% CI of $[0.02, 0.21]$. When $\xi_1(\mathbf{Z}) = 0.15$, a risk-neutral DM considers each of the safer alternatives $\{\$100, \$200, \$250, \$500\}$ 15% less often than does an extremely risk averse DM. The estimated value of ξ_2 is 7.14. This implies that as the risk aversion parameter increases from 0 to its estimated average median value of 0.0035, consideration probability of the safer alternatives increases by 11%, and it is essentially flat after that rising by an additional 4% as risk aversion reaches its upper bound.

The Mixed Logit Random Utility Model

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

The ARC Model: All Coverages

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

The largest difference between the predicted and observed shares equals 0.38 percentage points, which is for the $(\$250, \$250, \$500)$ triplet that is chosen by 2.9% of the households. The integrated absolute error across all triplets is 3.46%. In our data, 43 out of 120 triplets are never chosen (these are omitted from Figure 3). As discussed in Section III, the likelihood maximization implies that the consideration probabilities for these triplets must be zero, so their predicted
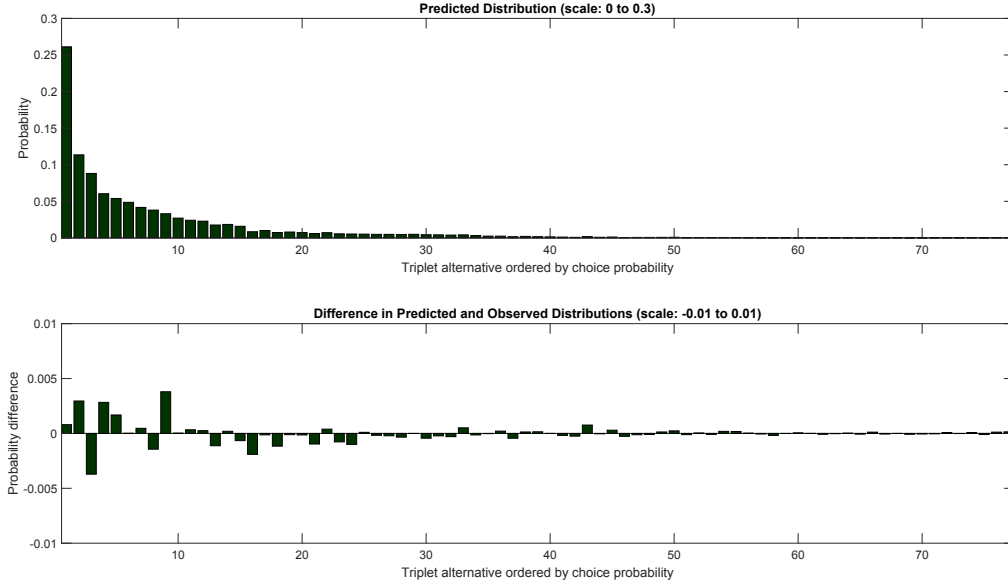
FIGURE 3. THE ARC MODEL, THREE COVERAGES

Note: 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. The scale of the second panel is magnified by a factor of 30 to make the probability difference visible.

shares are zero. Hence, the likelihood maximization routine is faster and more reliable as we do not need to search for φ_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 and Teitelbaum, 2011; Einav et al., 2012; Barseghyan, Molinari and Teitelbaum, 2016). On the one hand, households' risk aversion relative to their peers is correlated across lines of coverage, implying that households preferences have a stable component. On the other hand, analyses based on revealed preference reject the standard models: under full consideration, for the vast majority of households one cannot find a level of (household-specific) risk aversion that justifies their choices simultaneously across all contexts. Limited consideration allows the model to match the observed joint distribution of choices, and hence their rank correlations.

The estimated risk preferences are similar to those estimated with collision only data, although the variance is slightly smaller. The triplet considered most frequently is the cheapest one: ($1000, $1000, $1000). Its consideration probability is 0.76, while the next two most considered triplets are ($500, $500, $1000) and ($500, $500, $500). These are considered with probability 0.47 and 0.43, respectively. Overall, there is a strong positive correlation (0.53) between the consideration probability and the sum of the deductibles in a given alternative.

We summarize once more the computational advantages of our procedure. First,

estimation of our model remains feasible for a large choice set.[22] Second, the model's parameters grow linearly with the size of the choice set – one parameter per an additional alternative. Third, enlarging the choice set does not call for new independent sources of data variation. For example, in our model whether there are five or one hundred twenty deductible alternatives does not make any difference either from an identification or an estimation stand point: with sufficient variation in $\bar{p}$ and/or μ, the model is identified and can be estimated. As a final remark, once the model is estimated, one can compute the average monetary cost of limited consideration. In our data it is \$50 (see Online Appendix C).

VI. Discussion

The literature concerned with the formulation, identification, and estimation of discrete choice models with limited consideration is vast. However, to our knowledge, there is no previous work applying such models to the study of decision making under risk, except for the contemporaneous work of Barseghyan et al. (2019). In particular, this paper is the first to exploit the SCP for identification purposes. As a result, several fundamental differences emerge between our work and existing papers. First, we achieve identification in the most challenging case where there is a single excluded regressor that affects the utility of all alternatives.[23] Second, we allow for consideration to depend on preferences. Third, with alternative-specific excluded regressors, this dependence can be essentially unrestricted and can be combined with dependence of consideration on (some of) the excluded regressors. Fourth, we scrutinize the large support assumption, show why it may be necessary, and when and how it is possible to make progress when it is not satisfied. Fifth, our approach comes with an easy to implement and computationally fast estimation strategy. Finally, we make a contribution specific to the study of decision making under risk by proposing a model that is immune from Apesteguia and Ballester (2018) criticism and features two sources of unobserved heterogeneity – risk aversion and limited consideration – whose distributions are identified. More generally, the paper establishes that, as long as the DMs' preferences satisfy the SCP, allowing for limited consideration does not hinder the model's identifiability or applicability. Hence, we view our framework as a stepping stone for studies of consumer behavior in markets where limited consideration may be present (one example is Coughlin, 2020, who builds on our framework to study consumer choice in Medicare Part D markets).

Papers that allow for limited consideration or more broadly for choice set heterogeneity can be classified in four groups. The first relies on auxiliary informa-

---

[22]In our setting, it is feasible to estimate an additive error RUM assuming the DMs consider each deductible triplet as a separate alternative (Online Figure E.4 and Online Table E.5). As the figure shows, the failure to match the data is evident. The Vuong test formally rejects it in favor of the ARC model.

[23]This setting is common in insurance markets, see, e.g.,Cohen and Einav (2007); Einav et al. (2012); Sydnor (2010); Barseghyan, Prince and Teitelbaum (2011); Barseghyan et al. (2013); Handel (2013); Bhargava, Loewenstein and Sydnor (2017).

tion about the composition or distribution of DMs' choice sets, such as brand awareness (e.g., Draganska and Klapper, 2011; Honka and Chintagunta, 2017) or search activity (e.g., Honka and Chintagunta, 2017; De los Santos, Hortaçsu and Wildenbeest, 2012; Kim, Albuquerque and Bronnenberg, 2010; Honka, Hortaçsu and Vitorino, 2017).[24] We do not require such information.

The second group attains identification via two-way exclusion restrictions, i.e., by assuming that some variables impact consideration but not utility and *vice versa*. A well-known example of this approach is Goeree (2008), who posits that advertising intensity affects the likelihood of considering a computer, but does not impact consumer preferences, while computer attributes such as CPU speed affect preferences but not consideration (see also van Nierop et al. (2010) and Gaynor, Propper and Seiler (2016)). Hortaçsu, Madanizadeh and Puller (2017) create an exclusion restriction by exploiting the dynamic aspect of consumer choice.[25] The consumer's decision to consider alternatives to her current service provider is a function of (her experiences with) the last period provider but not her next period provider (see also Heiss et al. (2016)). In contrast, we achieve identification with as little as one common excluded regressor and a single cross section.

The third group relies on restricting the consideration mechanism to a specific class of models. Abaluck and Adams (2020) consider two such models (and their hybrid): a variant of the ARC and a "default specific" model (as in, e.g., Ho, Hogan and Scott Morton, 2017; Heiss et al., 2016) in which each DM's consideration set comprises either a single default alternative or the entire feasible set. They assume that consideration and preferences are independent, and that each alternative has a characteristic with large support that is additively separable in utility and may only affect its own consideration but not the consideration of other alternatives.[26] They exploit violations of symmetry in the Slutsky matrix (i.e., in cross-alternative demand responses to prices) to detect limited consideration. Kawaguchi, Uetake and Watanabe (2020) study beverage purchases from vending machines, allowing advertisement to be a driver of consideration, but also to affect utility. Their approach is close to that of Goeree (2008), though they provide a formal argument for identification with large support and exclusion restrictions even when there is no choice set variation. A key assumption is that all beverages are considered with probability equal to one as the advertising intensity of each beverage becomes very large.

The methods we propose relate to the papers in the third group in two aspects. First, we too *sometimes* require large support as a "fail safe" assumption, but only

---

[24]For canonical cites see, e.g., Roberts and Lattin (1991) and Ben-Akiva and Boccara (1995).

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

[26]The exception is the "default" alternative, whose characteristic may trigger the consideration of the entire choice set.

in the most challenging case of a single common excluded regressor. Second, we too rely on exclusion restrictions. The reliance on these assumptions is inescapable given the econometrics literature on point identification of discrete choice models. Our approach elucidates the identifying power of a single excluded regressor in models that satisfy the SCP and, in particular, the relative ranking of alternatives encapsulated in Facts 3 and 4 (see Lewbel and Yang, 2016, for related results for average treatment effects in ordered discrete choice models). We further exploit this structure to establish identification in models with substantially richer levels of unobserved heterogeneity, by allowing for dependence between consideration and preferences.

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

REFERENCES

Abaluck, Jason, and Abi Adams. 2020. "What do consumers consider before they choose? Identification from asymmetric demand responses." *Quarterly Journal of Economics*, forthcoming.

Anonymous Firm. 2018. "Property Insurance Purchases." *Accessed via Cornell Restricted Access Data Center (CRADC)*, last accessed on May 4 2018.

Apesteguia, Jose, and Miguel A. Ballester. 2018. "Monotone Stochastic Choice Models: The Case of Risk and Time Preferences." *Journal of Political Economy*, 126(1): 74–106.

Apesteguia, Jose, Miguel A Ballester, and Jay Lu. 2017. "Single-Crossing Random Utility Models." *Econometrica*, 85(2): 661–674.

Athey, Susan. 2001. "Single crossing properties and the existence of pure strategy equilibria in games of incomplete information." *Econometrica*, 69(4): 861–889.

Attali, Yigal, and Maya Bar-Hillel. 2003. "Guess where: The position of correct answers in multiple-choice test items as a psychometric variable." *Journal of Educational Measurement*, 40(2): 109–128.

Barseghyan, Levon, Francesca Molinari, and Joshua C. Teitelbaum. 2016. "Inference Under Stability of Risk Preferences." *Quantitative Economics*, 7(2): 367–409.

Barseghyan, Levon, Francesca Molinari, and Matthew Thirkettle. 2019. "Discrete Choice Under Risk With Limited Consideration." *CEMMAP Working Paper*, CWP08/19.

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum. 2013. "The Nature of Risk Preferences: Evidence from Insurance Choices." *American Economic Review*, 103(6): 2499–2529.

Barseghyan, Levon, Jeffrey Prince, and Joshua C. Teitelbaum. 2011. "Are Risk Preferences Stable Across Contexts? Evidence from Insurance Data." *American Economic Review*, 101(2): 591–631.

Barseghyan, Levon, Joshua C Teitelbaum, and Lin Xu. 2018. "Different Contexts, Different Risk Preferences?" *SRRN Working Paper*, No. 3200657.

Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C. Teitelbaum. 2019. "Heterogeneous Choice Sets and Preferences." *CEMMAP Working Paper*, CWP37/19.

Ben-Akiva, Moshe, and Bruno Boccara. 1995. "Discrete Choice Models with Latent Choice Sets." *International Journal of Marketing Research*, 12(1): 9–24.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor. 2017. "Choose to Lose: Health Plan Choices from a Menu with Dominated Options." *Quarterly Journal of Economics*, 132(3): 1319–1372.

Cattaneo, Matias D, Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov. 2020. "A random attention model." *Journal of Political Economy*, 128(7): 2796–2836.

Chiappori, Pierre-André, Bernard Salanié, François Salanié, and Amit Gandhi. 2019. "From aggregate betting data to individual risk preferences." *Econometrica*, 87(1): 1–36.

Christenfeld, Nicholas. 1995. "Choices from identical options." *Psychological Science*, 6(1): 50–55.

Cohen, Alma, and Liran Einav. 2007. "Estimating Risk Preferences from Deductible Choice." *American Economic Review*, 97(3): 745–788.

Coughlin, Maura. 2020. "Insurance Choice with Non-Monetary Plan Attributes: Limited Consideration in Medicare Part D." *Working paper.*

Crawford, Gregory S., Rachel Griffith, and Alessandro Iaria. 2020. "A Survey of Preference Estimation with Unobserved Choice Set Heterogeneity." *Journal of Econometrics*, forthcoming.

Dardanoni, Valentino, Paola Manzini, Marco Mariotti, and Christopher J. Tyson. 2020. "Inferring Cognitive Heterogeneity From Aggregate Choices." *Econometrica*, 88(3): 1269–1296.

De los Santos, Babur, Ali Hortaçsu, and Matthus R. Wildenbeest. 2012. "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior." *American Economic Review*, 102(6): 2955–2980.

Draganska, Michaela, and Daniel Klapper. 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, Liran, Amy Finkelstein, Iuliana Pascu, and Mark R. Cullen. 2012. "How General are Risk Preferences? Choice under Uncertainty in Different Domains." *American Economic Review*, 102(6): 2606–2638.

Gaynor, Martin, Carol Propper, and Stephan Seiler. 2016. "Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service." *American Economic Review*, 106(11): 3521–57.

Ghosal, Subhashis. 2001. "Convergence rates for density estimation with Bernstein polynomials." *Annals of Statistics*, 29(5): 1264–1280.

Goeree, Michelle Sovinsky. 2008. "Limited Information and Advertising in the U.S. Personal Computer Industry." *Econometrica*, 76(5): 1017–1074.

Handel, Benjamin R. 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." *American Economic Review*, 103(7): 2643–2682.

Heiss, Florian, Daniel McFadden, Joachim Winter, Amelie Wuppermann, and Bo Zhou. 2016. "Inattention and Switching Costs as Sources of Inertia in Medicare Part D." *NBER Working Paper*, No. 22765.

Ho, Kate, Joseph Hogan, and Fiona Scott Morton. 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, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino. 2017. "Advertising, Consumer Awareness, and Choice: Evidence from the U.S. Banking Industry." *RAND Journal of Economics*, 48(3): 611–646.

Honka, Elisabeth, and Pradeep Chintagunta. 2017. "Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry." *Marketing Science*, 36(1): 21–42.

Hortaçsu, Ali, Seyed Ali Madanizadeh, and Steven L. Puller. 2017. "Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market." *American Economic Journal: Economic Policy*, 9(4): 192–226.

Kawaguchi, Kohei, Kosuke Uetake, and Yasutora Watanabe. 2020. "Designing Context-Based Marketing: Product Recommendations under Time Pressure." *SSRN Working Paper*, No. 2529294.

Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg. 2010. "Online Demand Under Limited Consumer Search." *Marketing Science*, 29(6): 1001–1023.

Kimya, Mert. 2018. "Choice, Consideration Sets, and Attribute Filters." *American Economic Journal: Microeconomics*, 10(4): 223–47.

Lewbel, Arthur. 2000. "Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables." *Journal of Econometrics*, 97(1): 145 – 177.

Lewbel, Arthur. 2014. "An overview of the special regressor method." In *Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics.* , ed. Jeffrey Racine, Liangjun Su and Aman Ullah, 38 – 62. Oxford University Press.

Lewbel, Arthur, and Thomas Tao Yang. 2016. "Identifying the average treatment effect in ordered treatment models without unconfoundedness." *Journal of Econometrics*, 195(1): 1–22.

Manski, Charles F. 1975. "Maximum score estimation of the stochastic utility model of choice." *Journal of Econometrics*, 3(3): 205 – 228.

Manski, Charles F. 1977. "The structure of random utility models." *Theory and decision*, 8(3): 229–254.

Manski, Charles F. 2009. "Identification for Prediction and Decision." *Harvard University Press.*

Manzini, Paola, and Marco Mariotti. 2014. "Stochastic Choice and Consideration Sets." *Econometrica*, 82(3): 1153–1176.

Matzkin, Rosa L. 2007. "Nonparametric identification." In *Handbook of Econometrics.* Vol. 6, , ed. James J. Heckman and Edward E. Leamer, Chapter 73, 5307 – 5368. Elsevier.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics.* , ed. P. Zarembka, 105–142. New York:Academic Press.

McFadden, Daniel, and Kenneth Train. 2000. "Mixed MNL Models for Discrete Reponse." *Journal of Applied Econometrics*, 15(5): 447–470.

Mirrlees, James A. 1971. "An exploration in the theory of optimum income taxation." *The Review of Economic Studies*, 38(2): 175–208.

Roberts, John H., and James M. Lattin. 1991. "Development and Testing of a Model of Consideration Set Composition." *Journal of Marketing Research*, 28(4): 429–440.

Rubinstein, Ariel, Amos Tversky, and Dana Heller. 1997. "Naive strategies in competitive games." In *Understanding Strategic Interaction.* 394–402. Springer.

Simonson, Itamar, and Amos Tversky. 1992. "Choice in Context: Trade-off Contrast and Extremeness Aversion." *Journal of Marketing Research*, XXIX: 281–95.

Spence, Andrew Michael. 1974. "Market signaling: Informational transfer in hiring and related screening processes." *Harvard University Press*, 143.

Sydnor, Justin. 2010. "(Over)Insuring Modest Risks." *American Economic Journal: Applied Economics*, 2(4): 177–199.

Teigen, Karl Halvor. 1983. "Studies in subjective probability 1: Prediction of random events." *Scandinavian Journal of Psychology*, 24(1): 13–25.

Tversky, Amos. 1972. "Choice by elimination." *Journal of Mathematical Psychology*, 9(4): 341 – 367.

van Nierop, Erjen, Bart Bronnenberg, Richard Paap, Michel Wedel, and Philip Hans Franses. 2010. "Retrieving Unobserved Consideration Sets from Household Panel Data." *Journal of Marketing Research*, 47(1): 63–74.

Vuong, Quang H. 1989. "Likelihood ratio tests for model selection and non-nested hypotheses." *Econometrica*, 57(2): 307–333.

Wilcox, Nathaniel T. 2008. "Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison." *Risk aversion in experiments*, 12: 197–292.

Discrete Choice under Risk with Limited Consideration[*]

Levon Barseghyan[†]   Francesca Molinari[‡]   Matthew Thirkettle[§]

February 18, 2021

For Online Publication

Appendices

A Proofs

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

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

The following lemma is useful for establishing Theorem 1.

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

Proof. The cutoff $\{c_{1,2}(x^t)\}_{t=1}^\infty$ is a strictly declining sequence. Suppose all its elements are non-negative. Then it converges to some $\nu^\infty \geq 0$ such that $\nu^\infty = c_{1,2}(x^\infty) = c_{1,j}(x^\infty)$ for some $x^\infty \in \mathcal{X}$, a contradiction. □

Proof of Theorem 1. The second condition in the theorem implies that the cutoffs are or-

dered: $c_{1,j}(x) < c_{1,j+1}(x)$ for all $x \in \mathcal{X}$. Hence

$$\Pr(d = d_1|x) = \sum_{j=2}^{D} \sum_{\substack{\mathcal{K} \subset \mathcal{D}: \\ 1,j \in \mathcal{K}, \\ 2,\ldots,j-1 \notin \mathcal{K}}} Q(\mathcal{K})F(c_{1,j}(x)) + Q(\{d_1\})$$

$$= \sum_{j=2}^{D} \mathcal{O}(\{d_1, d_j\}; \{d_2, \ldots, d_{j-1}\})F(c_{1,j}(x)) + \mathcal{O}(d_1; \emptyset)$$

$$\equiv \sum_{j=2}^{D} \Lambda_j F(c_{1,j}(x)) + \Lambda_1,$$

so that

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

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

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

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

The ratio of $\lim_{x \nearrow x^j} d\Pr(d = d_j|x)/dx$ and $\lim_{x \nearrow x^2} d\Pr(d = d_2|x)/dx$ identifies $\Omega_j \equiv \frac{\Lambda_j}{\Lambda_2}$, where $\Lambda_2 \neq 0$ by the first assumption in the theorem. Rewrite the derivative of $\Pr(d = d_1|x)$ as

follows:

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

$$= \sum_{j=2}^{\hat{D}} \frac{\Lambda_j}{\Lambda_2} \Lambda_2 f(c_{1,j}(x)) \frac{dc_{1,j}(x)}{dx}$$

$$= \sum_{j=2}^{\hat{D}} \Omega_j [\Lambda_2 f(c_{1,j}(x))] \frac{dc_{1,j}(x)}{dx}$$

$$= \sum_{j=2}^{\hat{D}} \Omega_j \hat{f}(c_{1,j}(x)) \frac{dc_{1,j}(x)}{dx},$$

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

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

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

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

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

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

$\square$

Proof of Theorem 2. The second condition of Theorem 1 implies that the cutoffs are ordered: $c_{1,j}(x) < c_{1,j+1}(x)$ for all $x \in \mathcal{X}$. Hence,

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

3

where

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

Similar to the proof of Theorem 1, we only consider the relevant alternatives for the derivative of $\Pr(d = d_1|x)$, namely $j \leq \hat{D}$.

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

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

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

$$
\frac{d\Pr(d = d_1|x)}{dx} = \sum_{j=2}^{\hat{D}} \left(\overline{\Omega}_j\right)^{\mathbb{1}(c_{1,j}(x)\geq\nu^*)} \left(\underline{\Omega}_j \frac{\underline{\Lambda}_2}{\overline{\Lambda}_2}\right)^{\mathbb{1}(c_{1,j}(x)<\nu^*)} \hat{f}(c_{1,j}(x)) \frac{dc_{1,j}(x)}{dx},
$$

Now all coefficients of $\hat{f}(c_{1,j}(x))$ are identified, and identification of $\hat{f}(\cdot)$ proceeds to the left of ν^*. Once it is identified, we integrate it over the support to recover $\overline{\Lambda}_2$. Hence $\underline{\Lambda}_2$ and $f(\cdot)$ are identified.

When the breakpoint occurs at x^* rather than ν^*, the same proof strategy can be applied.

$\square$

Proof of Proposition 1. The second condition in the theorem implies that $c_{1,2}(x) < c_{1,j}(x)$ for any $j > 2$. The cutoffs $c_{1,j}(x)$'s, $j > 2$, are irrelevant for evaluating $\Pr(d = d_1|x)$ by the

first condition of the theorem. Therefore,

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

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

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

and hence variation in x guarantees that $f(\nu)$ is identified up-to-scale on $[c_{1,2}(x^l), c_{1,2}(x^u)] = [\nu^l, \nu^u]$. It follows that $F(\nu|\nu \in [\nu^l, \nu^u])$ is identified. $\qquad\square$

Proof of Theorem 3. Under Condition *1* of the Theorem there can only be three types of consideration sets. The first type are all possible subsets of $\{d_1, \ldots, d_j\}$; the second type are all possible subsets of $\{d_{j+1}, \ldots, d_D\}$. The third type necessarily contains both d_j and d_{j+1}. The probability of choosing an alternative in $\{d_1, \ldots, d_j\}$ is one for the first type and zero for the second type. Hence,

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

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

Proof of Theorem 4. Let ν, $\tilde{x}$, $\mathcal{N}_\epsilon(\tilde{x}) \equiv \{x : \|x - \tilde{x}\| < \epsilon\}$ satisfy Condition *2* in the theorem. Then, $\nu = c_{j,k}(\tilde{x})$ for all j, k. Consider any pair of alternatives (d_j, d_k). Since utility is strictly monotone in x_j and continuous, for each $\mathcal{L} \subseteq \mathcal{D} \setminus \{j, k\}$ we can find $x \in \mathcal{N}_\epsilon(\tilde{x})$ such that

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

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

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

The remainder of the proof proceeds in two steps.

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

5

holds for $n = 0, \ldots, D$:

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

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

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

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

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

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

$$= f(\nu)\mathcal{Q}_\nu(\{j, k\} \cup \mathcal{K}) + \sum_{\mathcal{C} \subset \mathcal{K}: |\mathcal{C}| < n} f(\nu)\mathcal{Q}_\nu(\{j, k\} \cup \mathcal{C}).$$

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

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

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

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

$f(\nu)$ is identified. Identification of $\mathcal{Q}_\nu(\mathcal{K})$ follows from Step 1. $\square$

Proof of Corollary 1. The proof follows the same steps as the proof of Theorem 4, but with the following two modifications:

First modification: In Step 1 in the proof of Theorem 4, we start with $d_j = d_1$ and loop over $d_k \in \{d_2, \ldots, d_D\}$. This ensures that we only take derivatives with respect to x_k, $k > 1$. Hence, $f(\nu)\mathcal{Q}_\nu^{x_1}(\mathcal{K})$ is identified for all sets $\mathcal{K} \subset \mathcal{D} : |\mathcal{K}| > 1$.

Second modification: In Step 2 we obtain

$$f(\nu)\mathcal{Q}^{x_1}(d_1) + \sum_{\mathcal{K}\subset\mathcal{D}:|\mathcal{K}|>1} f(\nu)\mathcal{Q}^{x_1}_\nu(\mathcal{K}) = f(\nu)\sum_{\mathcal{K}\subset\mathcal{D}}\mathcal{Q}^{x_1}_\nu(\mathcal{K}) = f(\nu).$$

Since the second term on the LHS is known, $f(\nu)(1 - \mathcal{Q}^{x_1}(d_1))$ is identified for all $\nu \in [0,\bar{\nu}]$. The scale is identified, because

$$(1 - \mathcal{Q}^{x_1}(d_1)) = \int_0^{\bar{\nu}} f(\nu)(1 - \mathcal{Q}^{x_1}(d_1))d\nu.$$

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

Proof of Proposition 2. For the purpose of obtaining a contradiction, suppose that there is full consideration. Then

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

This is a contradiction. Therefore there is limited consideration.

□

The following two Lemmas are used in the proof of Theorem 5. The proofs of these Lemmas rest on the following claims.

1. The probability of alternative d_j being chosen can only increase in its consideration probability.

2. The probability of alternative d_j being chosen can only decline in consideration probability of any other alternative d_k.

3. The probability of alternative d_j or d_k being chosen can only increase in the consideration probability of d_j as the positive effect of this change on $Pr(d = d_j|x)$ dominates the negative effect on $Pr(d = d_k|x)$.

4. The probability of an alternative in $\mathcal{K}$ being chosen can only decline in consideration probability of any alternative that does not belong to $\mathcal{K}$.

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

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

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

where A is a constant that collects terms that do not depend on φ_j. Continuing,

$$\sum_{l \in \mathcal{K}} \Pr(d = d_l | x, \nu) = A + \varphi_j \prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) + \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \varphi_l \prod_{k \in \mathcal{B}_\nu(d_l, x)} (1 - \varphi_k)$$

$$= A + \varphi_j \prod_{k \in \mathcal{B}_\nu(d_j, x)} (1 - \varphi_k) + \sum_{l \in \mathcal{K}: j \in \mathcal{B}_\nu(d_l, x)} \varphi_l (1 - \varphi_j) \prod_{k \in \mathcal{B}_\nu(d_l, x) \setminus \{j\}} (1 - \varphi_k)$$

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

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

$$\equiv \tilde{A} + B \varphi_j.$$

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

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

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

Finally, for any $k \notin \mathcal{K}$, φ_k may only appear on the RHS as $(1 - \varphi_k)$. Hence $\sum_{j \in \mathcal{K}} \Pr(d = d_j | x)$ is decreasing in φ_k. $\square$

Lemma A.3. *Consider the Basic ARC model. For any $\mathcal{K} \subset \mathcal{D}$, $\sum_{j \in \mathcal{K}} \Pr(d = d_j | x)$ is strictly increasing in φ_j, $j \in \mathcal{K}$, whenever there is an open interval of ν's at which alternative d_j is preferred to all of the always-considered alternatives. It is strictly decreasing in φ_k, $j \notin \mathcal{K}$,*

whenever there is an open interval of ν's and $l \in \mathcal{K}$ such that at these ν's alternative d_k is preferred to d_l and d_l is preferred to all of the always-considered alternatives.

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

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

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

For this to be strictly decreasing in φ_k, it must be the case that $k \in \mathcal{B}_\nu(d_j, x)$ and $\varphi_l < 1$ for all $l \in \mathcal{B}_\nu(d_j, x) \setminus \{k\}$. $\qquad\square$

Proof of Theorem 5. By the proof of Theorem 1, $f(\cdot)$ and $\Lambda_2 = \varphi_1 \varphi_2$ are identified. The consideration parameter φ_1 is identified by $\Pr(d = d_1 | x^{\bar\nu}) = \varphi_1$, where $x^{\bar\nu}$ is s.t. $c_{1,2}(x^{\bar\nu}) = \bar\nu$. Since Λ_2 is known, φ_2 is also identified. The rest of the proof is about identification of the remaining consideration parameters.

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

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

Proof of Theorem 6. Step 1. Suppose $D = 3$. Then $d^* = 2$ and

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

The ratio of the derivatives of these two moments yields $\frac{\varphi_1(\nu)}{\varphi_3(\nu)}$, where

$$\varphi_1(\nu) = \varphi_1(1 - \alpha(\nu))$$
$$\varphi_3(\nu) = \varphi_3(1 + \alpha(\nu)),$$

First, $\frac{\varphi_1}{\varphi_3}$ is identified when x and x' are chosen such that $c_{1,2}(x) = c_{2,3}(x') = \bar{\nu}$. Once $\frac{\varphi_1}{\varphi_3}$ is identified, $\frac{1-\alpha(\nu)}{1+\alpha(\nu)}$ is known for all ν; hence, $\alpha(\nu)$ can be solved for. Identification of $f(\nu)$ follows from substituting $\alpha(\nu)$ into the expression for $\frac{\Pr(d=d_1|x)}{dx}$.

Step 2. Let $D > 3$. We identify d^* at $x^{\bar{\nu}}$. The smallest j such that $\Pr(d = d_j|x^{\bar{\nu}}) = 0$ yields $d^* = j - 1$ (or $d^* = D$ if no such j exists). We return to the case that $d^* = 1$, $d^* = 2$, or $d^* = D$ at the end of this proof.

Step 3. Using large support we establish that φ_D is a decreasing function of φ_1. We have

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

Similarly,

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

Hence

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

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

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

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

Step 5. Identification of $\{\varphi_j\}_{1<j<d^*}$. For each ν in a sufficiently small neighborhood near $\bar{\nu}$,

say $(\bar{\nu} - \varepsilon, \bar{\nu})$, and for each j, there is an x^j such that $c_{1,j}(x^j) = \nu$ and $c_{k,j}(x^j) > \bar{\nu}$ for all $k \neq 1, j$. It follows by Step 4 that the following equations hold:

$$\frac{d\Pr(d = d_2|x^2)}{dx} \bigg/ \frac{dc_{1,2}(x^2)}{dx} = \varphi_1(\nu)\varphi_2(\nu)f(\nu)$$

$$\frac{d\Pr(d = d_3|x^3)}{dx} \bigg/ \frac{dc_{1,3}(x^3)}{dx} = \varphi_1(\nu)(1 - \varphi_2(\nu))\varphi_3(\nu)f(\nu)$$

$$\frac{d\Pr(d = d_4|x^5)}{dx} \bigg/ \frac{dc_{1,4}(x^4)}{dx} = \varphi_1(\nu)(1 - \varphi_2(\nu))(1 - \varphi_3(\nu))\varphi_4(\nu)f(\nu)$$

$$\vdots$$

$$\frac{d\Pr(d = d_{d^*}|x^{d^*})}{dx} \bigg/ \frac{dc_{1,d^*}(x^{d^*})}{dx} = \varphi_1(\nu)(1 - \varphi_2(\nu))(1 - \varphi_3(\nu))\dots(1 - \varphi_{d^*-1}(\nu))f(\nu)$$

The summation of these expressions recovers the quantity $\varphi_1(\nu)f(\nu)$. Next, substitute $\varphi_1(\nu)f(\nu)$ into the expressions above to sequentially recover $\{\varphi_j(\nu)\}_{2 \leq j \leq d^*}$. Since ν can be made arbitrarily close to $\bar{\nu}$ by selecting a smaller value of ε and since $\alpha(\cdot)$ is continuous, $\lim_{\nu \to \bar{\nu}} \varphi_j(\nu) = \varphi_j(\bar{\nu}) = \varphi_j(1 - \alpha(\bar{\nu})) = \varphi_j$ is also identified. Hence, $\{\varphi_j\}_{j=2}^{d^*}$ are identified.

Step 6: Identification of φ_1 and $\{\varphi_j\}_{d^* < j \leq D}$. The cutoffs are monotone in x^t and all cutoffs are on the right of $\bar{\nu}$ at $x^{\bar{\nu}}$. Consequently, $\Pr(d = d_j|x^{\bar{\nu}}) = 0$ for all $j > d^*$. Continuously decrease t until $\Pr(d = d_{j_1}|x^t) > 0$ for some $j_1 \in \mathcal{J} \equiv \{d^* + 1, \dots, D\}$ and $\Pr(d = d_k|x^t) = 0$ for all $k \in \mathcal{J} \setminus \{j_1\}$. This will happen when $c_{d^*, j_1}(x^t)$ crosses $\bar{\nu}$, yielding

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

where ν is in a small neighborhood near $\bar{\nu}$ s.t. $c_{k,j_1}(x^t) > \bar{\nu}$ for all $k > d^*, k \neq j_1$ and

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

Near the end of this proof-step we will take the limit as $\nu \to \bar{\nu}$ after dividing out $f(\nu)$. Importantly, $M_1 \equiv \lim_{\nu \to \bar{\nu}} M_1(\nu)$ is known, since all relevant φ_k's are known, and M_1 does not depend on φ_1, since $c_{1,j_1}(x^t) < c_{d^*, j_1}(x^t) < \bar{\nu}$.

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

Hence,

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

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

The term $M_2 \equiv \lim_{\nu \to \bar{\nu}} M_2(\nu)$ is known, except possibly for the term $(1-\varphi_{j_1})$, since all other relevant φ_k's are known. The expression above defines $\varphi_{j_2}(\nu)$ as a strictly increasing function of $\varphi_{j_1}(\nu)$ regardless of whether $M_2(\nu)$ depends on $\varphi_{j_1}(\nu)$ or not. Indeed, for the case where $j_1 < j_2$ we have

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

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

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

Since $Pr(d=d_D|x^0) \neq 0$, the above sequential argument yields that $\varphi_D(\nu)$ is an increasing function of $\varphi_{j_1}(\nu)$. In turn, recall that $\varphi_1(\nu)f(\nu)$ is known for ν arbitrary close to $\bar{\nu}$. The limit of the ratio between $\varphi_1(\nu)f(\nu)$ and $\varphi_{j_1}(\nu)f(\nu)$, which is also known, yields φ_D as an increasing functions of φ_1. Hence, taken with the result in Step 3, the quantity φ_1 is uniquely pinned down. Identification of all other φ_j's immediately follow.

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

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

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

$$-\frac{d\Pr(d=d_2|x^2)}{dx}\bigg/\frac{dc_{1,2}(x^2)}{dx}=\varphi_1(\nu)\varphi_2(\nu)f(\nu)$$

$$-\frac{d\Pr(d=d_3|x^3)}{dx}\bigg/\frac{dc_{1,3}(x^3)}{dx}=\varphi_1(\nu)(1-\varphi_2(\nu))\varphi_3(\nu)f(\nu)$$

$$-\frac{d\Pr(d=d_4|x^4)}{dx}\bigg/\frac{dc_{1,4}(x^4)}{dx}=\varphi_1(\nu)(1-\varphi_2(\nu))(1-\varphi_3(\nu))\varphi_4(\nu)f(\nu) \qquad (A.4)$$

$$\vdots$$

$$-\frac{d\Pr(d=d_{d^*}|x^{d^*})}{dx}\bigg/\frac{dc_{1,d^*}(x^{d^*})}{dx}=\varphi_1(\nu)(1-\varphi_2(\nu))(1-\varphi_3(\nu))\ldots(1-\varphi_{d^*-1}(\nu))f(\nu),$$

The summation of these expressions recovers the quantity $\varphi_1(\nu)f(\nu)$. Substitute this into the first equation to obtain $\varphi_2(\nu) = \varphi_2(1-\alpha(\nu))$. But φ_2 is already known, so $\alpha(\nu)$ is identified on Γ^0. Finally, since $\varphi_1(\nu) = \varphi_1(1-\alpha(\nu))$ is now identified, so is $f(\nu)$ on Γ^0.

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

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

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

Edge Case I: Suppose $d^* = d_1$ (the identity is known from Step 2). The following expressions hold for ν close to $\bar{\nu}$:

$$-\frac{d\Pr(d=d_2|x^2)}{dx}\bigg/\frac{dc_{1,2}(x^2)}{dx}=\varphi_2(\nu)f(\nu)$$

$$-\frac{d\Pr(d=d_3|x^3)}{dx}\bigg/\frac{dc_{1,3}(x^3)}{dx}=(1-\varphi_2(\nu))\varphi_3(\nu)f(\nu)$$

$$-\frac{d\Pr(d=d_4|x^4)}{dx}\bigg/\frac{dc_{1,4}(x^4)}{dx}=(1-\varphi_2(\nu))(1-\varphi_3(\nu))\varphi_4(\nu)f(\nu) \qquad (A.5)$$

$$\vdots$$

$$-\frac{d\Pr(d=d_D|x^D)}{dx}\bigg/\frac{dc_{1,D}(x^D)}{dx}=(1-\varphi_2(\nu))(1-\varphi_3(\nu))\ldots(1-\varphi_{D-1}(\nu))\varphi_D(\nu)f(\nu),$$

13

The ratio of the first and second equations in System (A.5) yields:

$$\frac{\varphi_2(\nu)}{\varphi_3(\nu)} = (1 - \varphi_2(\nu))\frac{1}{A_3(\boldsymbol{x}_3(\nu))},$$

where $\boldsymbol{x}(\nu) = (x^2(\nu), x^3(\nu), \ldots, x^D(\nu))$ is a known implicit function of ν satisfying $\nu = c_{1,j}(x^j(\nu))$ for $j = 2, \ldots, D$, and $A_3(\boldsymbol{x}(\nu))$ is data:

$$A_3(\boldsymbol{x}(\nu)) = \frac{\frac{d\Pr(d=d_3|x^3(\nu))}{dx} \big/ \frac{dc_{1,3}(x^3(\nu))}{dx}}{\frac{d\Pr(d=d_2|x^2(\nu))}{dx} \big/ \frac{dc_{1,2}(x^2(\nu))}{dx}}.$$

From this, we obtain

$$\frac{\varphi_2}{\varphi_3} = (1 - \varphi_2(1 - \alpha(\nu)))\frac{1}{A_3(\boldsymbol{x}(\nu))}$$

$$\frac{A_3(\boldsymbol{x}(\nu))}{\varphi_3} = \frac{1}{\varphi_2} - (1 - \alpha(\nu))$$

$$\alpha(\nu) = 1 - \frac{1}{\varphi_2} + \frac{A_3(\boldsymbol{x}(\nu))}{\varphi_3}$$

$$\alpha'(\nu) = \left[\frac{\partial A_3(\boldsymbol{x}(\nu))}{\partial x^2}\frac{dx^2(\nu)}{d\nu} + \frac{\partial A_3(\boldsymbol{x}(\nu))}{\partial x^3}\frac{dx^3(\nu)}{d\nu}\right]\frac{1}{\varphi_3} \equiv B_3(\boldsymbol{x}(\nu))\frac{1}{\varphi_3},$$

where $B_3(\boldsymbol{x}(\nu))$ is a known function of data. A similar idea yields

$$\alpha'(\nu) = B_4(\boldsymbol{x}(\nu))\frac{1}{\varphi_4}$$

$$\alpha'(\nu) = B_5(\boldsymbol{x}(\nu))\frac{1}{\varphi_5}$$

$$\vdots$$

$$\alpha'(\nu) = B_D(\boldsymbol{x}(\nu))\frac{1}{\varphi_D}.$$

Hence, the ratios $\frac{\varphi_3}{\varphi_4}, \frac{\varphi_4}{\varphi_5}, \frac{\varphi_5}{\varphi_6} \ldots$ are identified. The ratio and the limit at the far end of the support of the third and fourth equations in System (A.5) yields

$$0 = 1 - \frac{1}{\varphi_3} + A_4(\boldsymbol{x}(\bar{\nu}))\frac{1}{\varphi_4}$$

$$\varphi_3 = 1 - A_4(\boldsymbol{x}(\bar{\nu}))\frac{\varphi_3}{\varphi_4},$$

so φ_3 is identified and so are $\varphi_4, \ldots, \varphi_D$. The ratio of the first and second equation in System (A.5) identifies φ_2. The proof continues on with Step 7. As such it does not require the

second condition of the theorem.

Edge Case II: Suppose that $d^* = D$. From Step 5 we obtain φ_j for $j : 1 < j < D$. Next, we show how to identify φ_1. We can find x^j such that $\nu = c_{j,D}(x^j)$ is arbitrary close to zero satisfying $c_{j,k}(x^j) < 0$ for all $k \neq j, D$, and so the following system of equations holds:

$$\frac{d\Pr(d = d_{D-1}|x^{D-1})}{dx} \Big/ \frac{dc_{D-1,D}(x^{D-1})}{dx} = \varphi_{D-1}(\nu)f(\nu)$$
$$\frac{d\Pr(d = d_{D-2}|x^{D-2})}{dx} \Big/ \frac{dc_{D-2,D}(x^{D-2})}{dx} = (1 - \varphi_{D-1}(\nu))\varphi_{D-2}(\nu)f(\nu)$$
$$\frac{d\Pr(d = d_{D-3}|x^{D-3})}{dx} \Big/ \frac{dc_{D-3,D}(x^{D-3})}{dx} = (1 - \varphi_{D-1}(\nu))(1 - \varphi_{D-2}(\nu))\varphi_{D-3}(\nu)f(\nu)$$
$$\vdots$$
$$\frac{d\Pr(d = d_1|x^1)}{dx} \Big/ \frac{dc_{1,D}(x^1)}{dx} = (1 - \varphi_{D-1}(\nu))(1 - \varphi_{D-2}(\nu))\ldots(1 - \varphi_2(\nu))\varphi_1(\nu)f(\nu).$$

The ratio of the first two equations yield

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

where A, φ_{D-2}, and φ_{D-1} are known terms; hence, $\alpha(0) = \lim_{\nu \to 0} \alpha(\nu)$ is identified. Once $\alpha(0)$ is identified, the term φ_1 is identified from the ratio of the first and last equations in the above system. Finally, $f(\nu)$ and $\alpha(\nu)$ are identified by Step 7.

Edge Case III: Suppose $d^* = 2$. By Steps 3, 5 and 6 all φ_j's are identified. A modified version of Steps 4 and 7 can be applied. We begin by starting at the lower end of the distribution with d_D taking the role of d_1; d_{D-1} taking the role of d_2, etc. Step 4 can be restated for

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

Finally Step 7 can be repeated starting at the lower end of the support and building on the

following equations

$$\frac{d\Pr(d=d_{D-1}|x^{D-1})}{dx}\bigg/\frac{dc_{D-1,D}(x^{D-1})}{dx}=\varphi_D(\nu)\varphi_{D-1}(\nu)f(\nu)$$

$$\frac{d\Pr(d=d_{D-2}|x^{D-2})}{dx}\bigg/\frac{dc_{D-2,D}(x^{D-2})}{dx}=\varphi_D(\nu)(1-\varphi_{D-1}(\nu))\varphi_{D-2}(\nu)f(\nu)$$

$$\frac{d\Pr(d=d_{D-3}|x^{D-3})}{dx}\bigg/\frac{dc_{D-3,D}(x^{D-3})}{dx}=\varphi_D(\nu)(1-\varphi_{D-1}(\nu))(1-\varphi_{D-2}(\nu))\varphi_{D-3}(\nu)f(\nu)$$

$$\vdots$$

$$\frac{d\Pr(d=d_3|x^3)}{dx}\bigg/\frac{dc_{1,3}(x^3)}{dx}=\varphi_D(\nu)(1-\varphi_{D-1}(\nu))(1-\varphi_{D-2}(\nu))\dots(1-\varphi_3(\nu))f(\nu).$$

$\square$

Proof of Theorem 7. Let ν, x, $\mathcal{N}_\epsilon(x)=\{x':\|x'-x\|<\epsilon\}$ satisfy the conditions in the theorem. Then $\nu=c_{j,k}(x)$ for all j,k. For any pair of alternatives (d_j,d_k) we can perturb x_k, $k\notin\{j,d^*\}$, and x_l, $\forall l\notin\{j,d^*,k\}$, so that the resulting $x'\in\mathcal{N}_\epsilon(x)$ is such that

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

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

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

Then

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

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

$\square$

Proof of Proposition 4. Take any non-empty consideration set $\mathcal{K}$. For a given preference

coefficient ν, let $j_{\mathcal{K}}(x, \nu)$ denote the identity of the best alternative in this consideration set. By the natural ordering, $j_{\mathcal{K}}(x, \nu)$ is an increasing step function in ν. Hence, $I(j_{\mathcal{K}}(x, \nu) \leq J)$ is a decreasing step function. The term $\Pr\left(\bigcup_{j=1}^{J} d_j \middle| x, \nu\right)$ is a non-negatively weighted sum of $I(j_{\mathcal{K}}(x, \nu) \leq J)$. Hence it is decreasing in ν. $\qquad\square$

Proof of Proposition 5. Consider a limited consideration model with preferences $U_\nu(d_j, x)$ and consideration probability $\mathcal{Q}_\nu^x(\mathcal{K})$, $\mathcal{K} \subset \mathcal{D}$. The optimal choice from $\mathcal{D}$ conditional on the DM facing the consideration set $\mathcal{K} \neq \emptyset$ is the alternative with the largest value of $U_\nu(d_j, x)$ subject to $j \in \mathcal{K}$. This is the same solution as the one that maximizes $V_\nu(d_j, x, \epsilon_j)$ where $\epsilon_j = 0$ for all $j \in \mathcal{K}$ and $\epsilon_j = -\infty$ for all $j \in \mathcal{D} \setminus \mathcal{K}$. Finally, since conditional on x the consideration set $\mathcal{K}$ has the same distribution as the set of alternatives with $\epsilon_j = 0$ (this is by construction), the limited consideration model and this ORUM model yield the same model predictions, and hence they are equivalent. $\qquad\square$

B   Application: Verifying Cutoff Order

We start by recalling that CARA and CRRA utility functions satisfy the following basic property (see, e.g., Pratt, 1964; Barseghyan, Molinari, O'Donoghue, & Teitelbaum, 2018).[1]

Lemma B.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 ν.*

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

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

Proof.

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

$\qquad\square$

---

[1]This property is equivalent to condition (e) in Pratt (1964, Theorem 1). As shown there, it is equivalent to assuming that an increase in ν corresponds to an increase in the coefficient of absolute risk aversion.

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

Proposition B.1. *Suppose deductibles and prices are such that*

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

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

Note that in a perfectly competitive markets where additional coverage is simply proportional to its price both ratios will be equal. In practice, however, one might expect that with some market power the prices increase faster than then coverage, and hence

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

This is exactly what we find in our data (as well as for a larger number of firms appearing in Barseghyan, Prince, & Teitelbaum (2011)).

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

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

where w is the DM's initial wealth. By Lemma B.2, the L.H.S. of Equation B.1 is strictly monotone in ν, and it tends to $+\infty$ when ν goes to $+\infty$ and to zero when ν goes to $-\infty$. It follows that there exists a unique ν, i.e the cutoff $c_{j,k}(x)$, that solves the Equation B.1. Moreover, since the L.H.S. is strictly monotone in ν it follows from the Implicit Function Theorem that $c_{j,k}(x)$ is continuous in μ and $\bar{p}$.

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

---

[2]If $p_k + d_k > p_j + d_j$, then alterantive j first order stochastically dominates k and hence the cuttoff is $+\infty$.

Since the expected utility of lottery k is proportional to

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

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

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

Taking logs yields

$$\log\left(\frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_j}}\right) = -\nu(g_1 - g_j)\bar{p}$$

$$\log\left(\frac{1 - \mu + \mu e^{\nu d_1}}{1 - \mu + \mu e^{\nu d_{j+1}}}\right) \geq -\nu(g_1 - g_{j+1})\bar{p}.$$

Dividing through and using the fact that $-\nu(g_1 - g_{j+1})\bar{p} \geq -\nu(g_1 - g_j)\bar{p} \geq 0$ yields

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

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

$$\text{L.H.S} = f\left(\frac{1-\mu}{\mu}\right) = f(\hat{\mu}) \equiv \frac{\log(\Delta_1 + \hat{\mu}) - \log(\Delta_j + \hat{\mu})}{\log(\Delta_1 + \hat{\mu}) - \log(\Delta_{j+1} + \hat{\mu})}.$$

First, we show that the above expression is monotonically increasing in $\hat{\mu}$. Observe that

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

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

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

Since $\Lambda_1 < \Lambda_j < \Lambda_{j+1}$ and exponential function is convex, the expression above is positive. Thus, the derivative of $f\left(\frac{1-\mu}{\mu}\right)$ W.R.T. μ is negative as claimed. That is, $f\left(\frac{1-\mu}{\mu}\right)$ achieves

its lowest value at $\mu = 1$ and is equal to $\frac{d_1 - d_j}{d_1 - d_{j+1}}$. Finally, a contradiction is obtained since

$$\text{L.H.S} \geq \min_{\mu} f\left(\frac{1-\mu}{\mu}\right) = \frac{d_1 - d_j}{d_1 - d_{j+1}} > \frac{g_1 - g_j}{g_1 - g_{j+1}} = \text{R.H.S},$$

where the strict inequality is by assumption. Therefore, $c_{1,j}(\bar{p}, \mu) < c_{1,j+1}(\bar{p}, \mu)$ under CARA as claimed.

Under CRRA, $c_{j,k}(\bar{p}, \mu)$ exist and are continuous exactly for the same reasons as under CARA. It remains to establish that $c_{1,j}(\bar{p}, \mu) < c_{1,j+1}(\bar{p}, \mu)$. For the purpose of obtaining a contradiction, suppose $c_{1,j}(\bar{p}, \mu) \geq c_{1,j+1}(\bar{p}, \mu)$ for some $(\bar{p}, \mu)$. Consider the following Taylor expansion for the CRRA Bernoulli utility function $u_\nu(w) \equiv \frac{w^{1-\nu}}{1-\nu}$:

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

Or, equivalently,

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

Hence,

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

where $\omega_t(\nu) \equiv (t!w^t)^{-1}\prod_{t'=0}^{t-1}(\nu - 1 + t') < 0$ when $\nu \in (0, 1)$. When $\nu > 1$, $\omega_t(\nu) > 0$ but the factor premultiplying $u_\nu(w - p_k)$ above is negative, so we would still come to the same conclusion that $EU_\nu(L_k)$ is proportional to a power series with coefficients $\tau_t(\nu) = -\omega_t(\nu) < 0$. The power series are absolutely convergence provided that $p_k + d_k < w$, so the difference in the power series for $EU_\nu(L_j)$ and $EU_\nu(L_k)$ is equal to the sum of the difference:

$$EU_\nu(L_j) - EU_\nu(L_k) \propto (1-\mu)\sum_{t=1}^{\infty}\omega_t(\nu)\left(p_j^t - p_k^t\right) + \mu\sum_{t=1}^{\infty}\omega_t(\nu)\left((p_j + d_j)^t - (p_k + d_k)^t\right)$$

$$= (p_j - p_k)(1-\mu)\sum_{t=1}^{\infty}\omega_t(\nu)\sum_{h=0}^{t-1}p_j^h p_k^{t-h}+$$

$$+ ((p_j - p_k) + (d_j - d_k))\mu\sum_{t=1}^{\infty}\omega_t(\nu)\sum_{h=0}^{t-1}(p_j + d_j)^h(p_k + d_k)^{t-h}.$$

The condition $\nu = c_{1,j}(\bar{p}, \mu) \geq c_{1,j+1}(\bar{p}, \mu)$ implies

$$\frac{p_1 - p_j}{p_1 - p_{j+1}} \geq \frac{p_1 - p_j + d_1 - d_j}{p_1 - p_{j+1} + d_1 - d_{j+1}} \delta(\nu), \tag{B.2}$$

where

$$\delta(\nu) \equiv \frac{\sum_{t=1}^{\infty} \omega_t(\nu) \sum_{h=0}^{t-1} (p_1 + d_1)^h (p_j + d_j)^{t-h}}{\sum_{t=1}^{\infty} \omega_t(\nu) \sum_{h=0}^{t-1} (p_1 + d_1)^h (p_{j+1} + d_{j+1})^{t-h}} \frac{\sum_{t=1}^{\infty} \omega_t(\nu) \sum_{h=0}^{t-1} p_1^h p_{j+1}^{t-h}}{\sum_{t=1}^{\infty} \omega_t(\nu) \sum_{h=0}^{t-1} p_1^h p_j^{t-h}}.$$

Under the assumption $\nu = c_{1,j}(\bar{p}, \mu) \geq c_{1,j+1}(\bar{p}, \mu)$ it is also the case that $\nu = c_{1,j}(\bar{p}, \mu) \geq c_{1,j+1}(\bar{p}, \mu) \geq c_{j,j+1}(\bar{p}, \mu)$ by Fact 4. Hence, $p_j + d_j > p_{j+1} + d_{j+1}$. Indeed otherwise $p_{j+1} - p_j > d_j - d_{j+1}$ is a violation of the first order stochastic dominance. Taken with $p_{j+1} > p_j$, it follows that $\delta(\nu) > 1$. Finally, a contradiction will be obtained if

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

since then Equation (B.2) will not hold. Re-arranging this expression we obtain:

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

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

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

The latter inequality holds by assumption. It follows that $c_{1,j}(\bar{p}, \mu) < c_{1,j+1}(\bar{p}, \mu)$, $j > 1$.

$\square$

C   Monetary Cost of Limited Consideration

We view limited consideration as a mechanism 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.[3] 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 $50 dollars across the three deductibles because of limited consideration. See Table E.6 for variation conditional on demographic characteristics and insurance score. We also find wide dispersion in loss across households (see Figure E.6). In particular, the 10th percentile of losses is $31 and the 90th is $73.

---

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

D Data

Table D.1 Descriptive Statistics

Variable	Mean	Std. Dev.	1st %	99th %
Age	53.9	15.6	25.4	84.1
Female	0.40			
Single	0.22			
Married	0.55			
Second Driver	0.43			
Insurance Score	767	112	532	985

Table D.2 Frequency of Deductible Choices Across Contexts

Deductible	1000	500	250	200	100	50
Collision	0.063	0.676	0.122	0.129	0.009	
Comprehensive	0.037	0.430	0.121	0.329	0.039	0.044
Home	0.176	0.559	0.262		0.002	

Table D.3 Deductible Rank Correlations Across Contexts

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

Table D.4 Average Premiums Across Coverages

Deductible	1,000	500	250	200	100	50
Collision	145	187	243	285	327	
Comprehensive	94	117	147	155	178	224
Home	594	666	720		885	

E   Empirical Results: Figures and Tables

## E.1   The ARC Model with Observable Demographics

While it is ideal to control for households' observable characteristics non-parametrically, it is data demanding. In practice, it is commonly assumed that household characteristics shift the expected value of the preference-coefficient distribution.[4] 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 γ is an unknown vector to be estimated. The terms $\beta_{1,i}$ and β_2 denote the parameters of the Beta distribution, where $\beta_{1,i}$ is household specific and β_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 $E(\nu_i) = \frac{\beta_{1,i}}{\beta_{1,i}+\beta_2}\bar{\nu} = \frac{e^{\mathbf{Z}_i\gamma}}{1+e^{\mathbf{Z}_i\gamma}}\bar{\nu}$.[5] The results of this estimation are in line with our first estimation. (See Column 2 in Table E.1, as well as Figures E.1 and E.2.) 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 0.94 and 0.92. The remaining alternatives are considered at much lower frequencies.

---

[4]For exmaple, 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 ε_i is i.i.d. N$(0, \sigma^2)$. Hence, $E(\nu_i) = e^{\mathbf{Z}_i\gamma+\sigma^2/2}$.

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

E.2   Figures

Figure E.1: The ARC Model with Observable Demographics
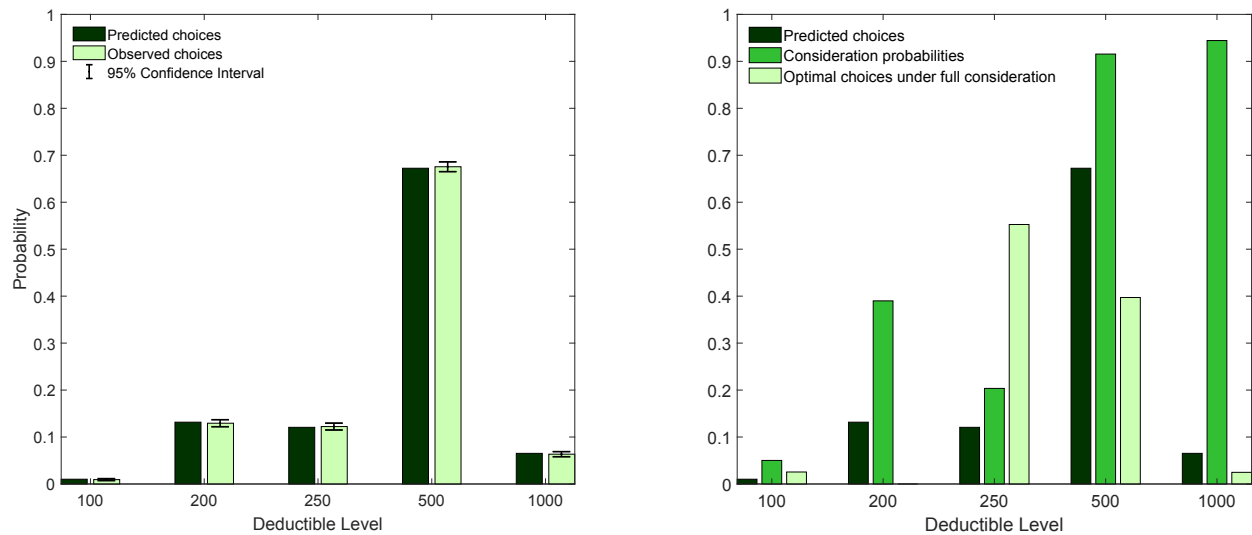
The first panel reports the distribution of predicted and observed choices. The second panel displays consideration probabilities and the distribution of optimal choices under full consideration.

Figure E.2: The ARC Model with Observable Demographics: Conditional Distributions
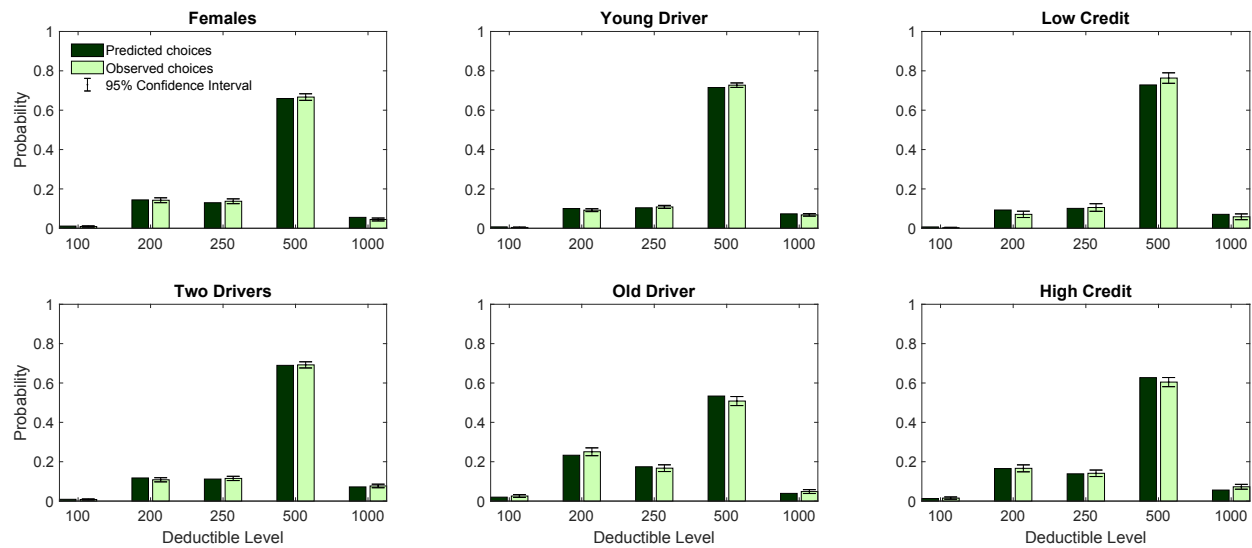
Figure E.3: The Mixed Logit
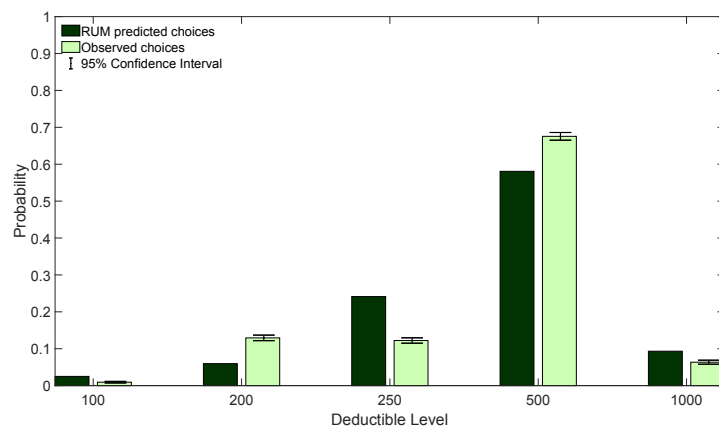
Figure E.4: The Mixed Logit, Three Coverages
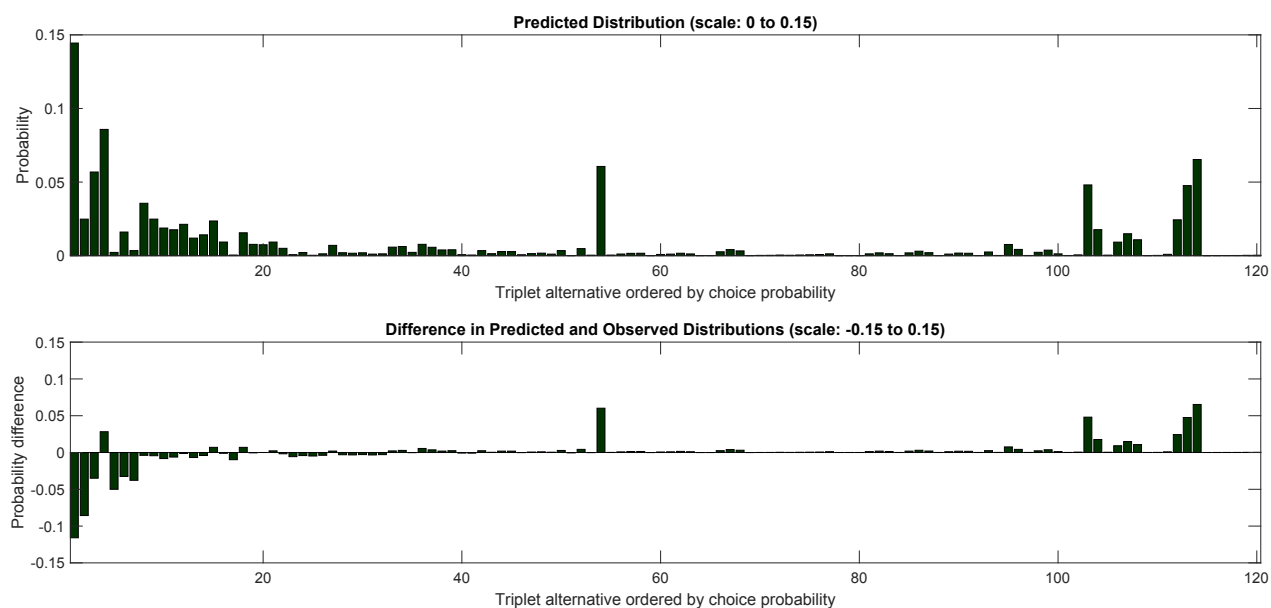
Triplets are sorted by observed frequency at which they are chosen. The first panel reports the predicted choice frequency and the second panel reports the difference in predicted and observed choice frequencies.

Figure E.5: The ARC Model, Three Coverages:
Consideration and Optimal Choice Distribution
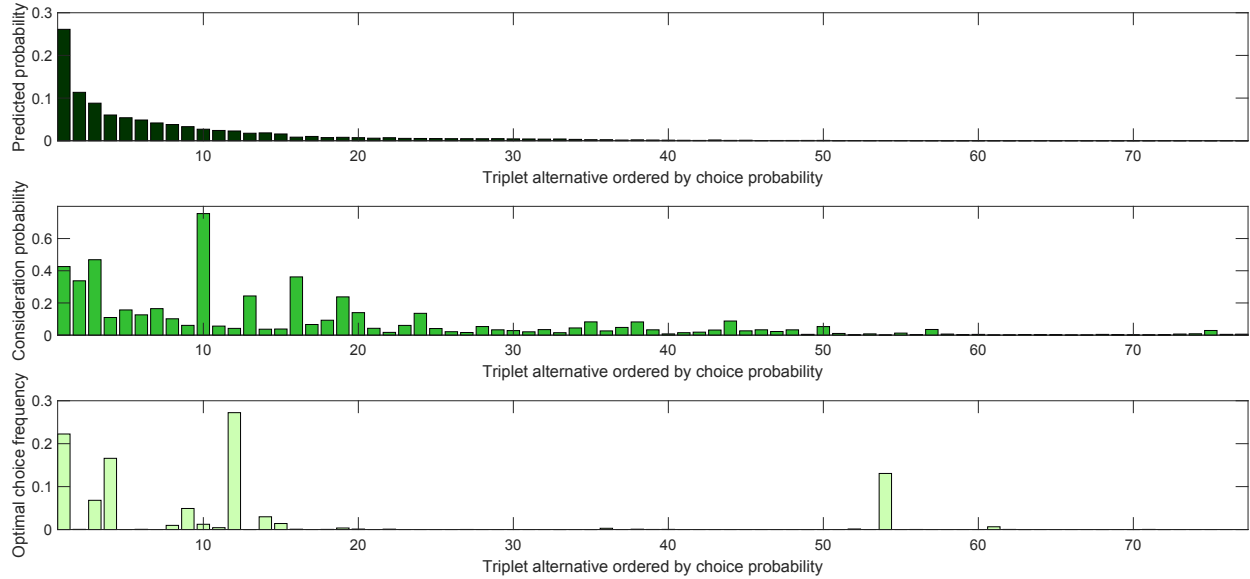
Triplets are sorted by observed frequency at which they are chosen.

Figure E.6: The ARC Model with Three Coverages:
Monetary Loss From Limited Consideration

28
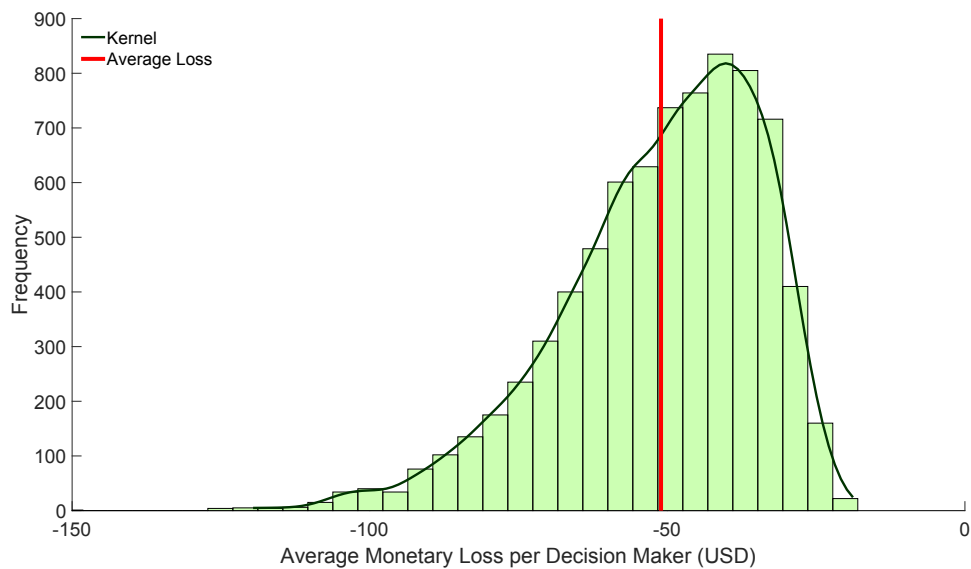
E.3 Tables

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

	ARC Model		ARC Model with Observables	
Average β_{1i}	1.70	[1.56, 1.82]	2.23	[1.93, 2.50]
β_2	7.45	[6.68, 8.08]	9.20	[8.09, 10.1]
Mean of ν	0.0037	[0.0036, 0.0038]	0.0038	[0.0036, 0.0040]
SD of ν	0.0024	[0.0023, 0.0026]	0.0022	[0.0021, 0.0023]
Intercept	-	-	-1.41	[-1.47, -1.33]
Age	-	-	0.207	[0.173, 0.237]
Age^2	-	-	0.048	[0.022, 0.073]
Female Driver	-	-	0.077	[0.051, 0.104]
Single Driver	-	-	0.050	[0.022, 0.079]
Married Driver	-	-	0.103	[0.062, 0.144]
Credit Score	-	-	0.134	[0.107, 0.160]
2+ Drivers	-	-	-0.302	[-0.370, -0.224]
Collision \$100	0.059	[0.050, 0.068]	0.050	[0.042, 0.058]
Collision \$200	0.412	[0.391, 0.433]	0.390	[0.364, 0.413]
Collision \$250	0.206	[0.198, 0.214]	0.204	[0.193, 0.212]
Collision \$500	0.920	[0.913, 0.926]	0.915	[0.909, 0.924]
Collision \$1000	1.000	[1.000, 1.000]	0.944	[0.899, 1.000]

Table E.2 MLE Estimation Results for the Proportionally Shifting Consideration Model

	ARC Model		ARC Model with Observables	
Average β_{1i}	1.44	[1.31, 1.55]	2.11	[1.86, 2.28]
β_2	6.07	[5.35, 6.67]	8.74	[7.73, 9.58]
Mean of ν	0.0038	[0.0037, 0.0040]	0.0038	[0.0036, 0.0040]
SD of ν	0.0027	[0.0026, 0.0028]	0.0023	[0.0021, 0.0024]
Intercept	-	-	-1.40	[-1.47, -1.35]
Age	-	-	0.194	[0.160, 0.222]
Age^2	-	-	0.036	[0.010, 0.059]
Female Driver	-	-	0.070	[0.046, 0.096]
Single Driver	-	-	0.049	[0.021, 0.076]
Married Driver	-	-	0.091	[0.047, 0.130]
Credit Score	-	-	0.135	[0.110, 0.160]
2+ Drivers	-	-	-0.283	[-0.348, -0.200]
Collision \$100	0.061	[0.051, 0.070]	0.055	[0.046, 0.063]
Collision \$200	0.424	[0.401, 0.446]	0.408	[0.382, 0.433]
Collision \$250	0.211	[0.202, 0.220]	0.212	[0.201, 0.222]
Collision \$500	0.985	[0.974, 0.998]	0.961	[0.929, 0.977]
Collision \$1000	1.000	-	1.000	-
Average ξ_{1i}	0.478	[0.277, 0.652]	0.148	[0.021, 0.212]
ξ_2	26.7	[16.1, 37.9]	7.14	[0.939, 10.3]
ξ_1**: Intercept**	-	-	-2.24	[-3.59, 0.0011]
ξ_1**: Age**	-	-	1.24	[-0.736, 1.75]
ξ_1**: Age**^{2}	-	-	-0.382	[-0.701, 0.584]
ξ_1**: Female Driver**	-	-	-0.323	[-1.16, 0.910]
ξ_1**: Single Driver**	-	-	0.382	[-1.51, 0.650]
ξ_1**: Married Driver**	-	-	0.0017	[-2.35, 1.45]
ξ_1**: Credit Score**	-	-	0.405	[-0.688, 0.642]
ξ_1**: 2+ Drivers**	-	-	0.485	[-2.22, 1.98]

Table E.3 MLE Estimation Results for the
Mixed Logit:
Collision Only

	Mixed Logit	
Average β_{1i}	9.07	[7.54, 10.2]
β_2	124.4	[106.0, 137.5]
Mean of ν	0.0014	[0.0013, 0.0014]
SD of ν	0.0004	[0.0004, 0.0005]
Intercept	-2.59	[-2.63, -2.55]
Age	-0.139	[-0.156, -0.122]
Age^2	-0.024	[-0.037, -0.010]
Female Driver	-0.0035	[-0.019, 0.012]
Single Driver	-0.0098	[-0.026, 0.0060]
Married Driver	-0.030	[-0.054, -0.0078]
Credit Score	0.091	[0.076, 0.105]
2+ Drivers	-0.016	[-0.061, 0.029]
Sigma	0.039	[0.037, 0.041]

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

	ARC Model			ARC Model	
Average β_{1i}	4.70	[3.89, 5.30]	**(250,200,250)**	0.037	[0.033, 0.041]
β_2	24.0	[19.7, 27.2]	**(250,200,500)**	0.056	[0.051, 0.061]
Mean of ν	0.0032	[0.0032, 0.0033]	**(250,200,1000)**	0.045	[0.035, 0.055]
SD of ν	0.0013	[0.0012, 0.0014]	**(250,250,100)**	0.0011	[0.0004, 0.0020]
Intercept	-1.68	[-1.72, -1.64]	**(250,250,250)**	0.042	[0.038, 0.046]
Age	0.162	[0.146, 0.180]	**(250,250,500)**	0.061	[0.056, 0.066]
Age^2	0.041	[0.026, 0.054]	**(250,250,1000)**	0.026	[0.019, 0.034]
Female Driver	0.043	[0.027, 0.061]	**(250,500,500)**	0.0007	[0.0004, 0.0013]
Single Driver	0.010	[-0.0075, 0.028]	**(500,50,250)**	0.034	[0.028, 0.042]
Married Driver	0.030	[0.0054, 0.054]	**(500,50,500)**	0.053	[0.044, 0.063]
Credit Score	0.137	[0.121, 0.153]	**(500,50,1000)**	0.033	[0.018, 0.047]
2+ Drivers	-0.097	[-0.141, -0.052]	**(500,100,250)**	0.015	[0.012, 0.019]
(100,50,250)	0.041	[0.033, 0.049]	**(500,100,500)**	0.043	[0.036, 0.049]
(100,50,500)	0.015	[0.0099, 0.021]	**(500,100,1000)**	0.048	[0.034, 0.062]
(100,50,1000)	0.013	[0.0048, 0.023]	**(500,200,100)**	0.0079	[0.0040, 0.012]
(100,100,100)	0.0023	[0.0009, 0.0042]	**(500,200,250)**	0.126	[0.119, 0.133]
(100,100,250)	0.0077	[0.0049, 0.010]	**(500,200,500)**	0.337	[0.322, 0.352]
(100,100,500)	0.0050	[0.0027, 0.0078]	**(500,200,1000)**	0.243	[0.221, 0.264]
(100,100,1000)	0.0047	[0.0019, 0.0086]	**(500,250,100)**	0.0018	[0.0008, 0.0032]
(100,200,250)	0.0005	[0.0002, 0.0010]	**(500,250,250)**	0.038	[0.034, 0.042]
(100,200,500)	0.0008	[0.0004, 0.0015]	**(500,250,500)**	0.102	[0.095, 0.109]
(100,200,1000)	0.0036	[0.0013, 0.0065]	**(500,250,1000)**	0.093	[0.080, 0.106]
(200,50,100)	0.011	[0.0052, 0.016]	**(500,500,100)**	0.0032	[0.0015, 0.0059]
(200,50,250)	0.066	[0.057, 0.075]	**(500,500,250)**	0.110	[0.103, 0.116]
(200,50,500)	0.061	[0.051, 0.071]	**(500,500,500)**	0.426	[0.413, 0.439]
(200,50,1000)	0.033	[0.018, 0.048]	**(500,500,1000)**	0.469	[0.450, 0.486]
(200,100,100)	0.0020	[0.0008, 0.0037]	**(1000,50,250)**	0.0069	[0.0020, 0.013]
(200,100,250)	0.021	[0.017, 0.026]	**(1000,50,500)**	0.0085	[0.0022, 0.015]
(200,100,500)	0.028	[0.023, 0.034]	**(1000,50,1000)**	0.029	[0.0031, 0.053]
(200,100,1000)	0.023	[0.012, 0.033]	**(1000,100,250)**	0.0049	[0.0019, 0.0090]
(200,200,100)	0.0017	[0.0007, 0.0032]	**(1000,100,500)**	0.0060	[0.0021, 0.011]
(200,200,250)	0.157	[0.147, 0.167]	**(1000,100,1000)**	0.035	[0.0080, 0.064]
(200,200,500)	0.165	[0.154, 0.176]	**(1000,200,250)**	0.032	[0.019, 0.045]
(200,200,1000)	0.136	[0.113, 0.158]	**(1000,200,500)**	0.082	[0.061, 0.105]
(200,250,250)	0.0004	[0.0002, 0.0006]	**(1000,200,1000)**	0.088	[0.046, 0.128]
(200,250,500)	0.0005	[0.0002, 0.0009]	**(1000,250,250)**	0.0067	[0.0027, 0.012]
(200,500,250)	0.0015	[0.0008, 0.0023]	**(1000,250,500)**	0.027	[0.015, 0.039]
(200,1000,1000)	0.0047	[0.0017, 0.0085]	**(1000,250,1000)**	0.053	[0.025, 0.083]
(250,50,100)	0.0020	[0.0008, 0.0037]	**(1000,500,250)**	0.033	[0.022, 0.044]
(250,50,250)	0.021	[0.017, 0.025]	**(1000,500,500)**	0.140	[0.119, 0.161]
(250,50,500)	0.033	[0.027, 0.039]	**(1000,500,1000)**	0.362	[0.309, 0.405]
(250,100,250)	0.017	[0.015, 0.020]	**(1000,1000,250)**	0.082	[0.058, 0.107]
(250,100,500)	0.016	[0.014, 0.020]	**(1000,1000,500)**	0.238	[0.199, 0.267]
(250,100,1000)	0.019	[0.011, 0.026]	**(1000,1000,1000)**	0.755	[0.652, 0.829]
(250,200,100)	0.0010	[0.0003, 0.0019]			

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

	Mixed Logit	
Average β_{1i}	4.89	[4.60, 5.16]
β_2	54.2	[51.6, 56.6]
Mean of ν	0.0017	[0.0016, 0.0017]
SD of ν	0.0007	[0.0007, 0.0007]
Intercept	-2.37	[-2.39, -2.34]
Age	-0.077	[-0.088, -0.066]
Age^2	-0.015	[-0.024, -0.0059]
Female Driver	0.0008	[-0.0098, 0.012]
Single Driver	-0.014	[-0.025, -0.0030]
Married Driver	-0.018	[-0.033, -0.0029]
Credit Score	0.034	[0.023, 0.045]
2+ Drivers	-0.048	[-0.075, -0.020]
Sigma	0.224	[0.209, 0.238]

Table E.6 Average Monetary Loss by Group

	Average Monetary Loss	
All	-50.2	[-52.4, -47.3]
Female Driver	-54.3	[-56.8, -51.0]
Single Driver	-45.1	[-47.1, -42.3]
Young	-45.5	[-47.2, -43.2]
Old	-65.4	[-69.3, -59.4]
Low Credit Driver	-47.6	[-49.3, -44.9]
High Credit Driver	-54.3	[-57.3, -50.3]

References

Barseghyan, L., Molinari, F., O'Donoghue, T., & Teitelbaum, J. C. (2018). Estimating risk preferences in the field. *Journal of Economic Literature*, *56*(2), 501-564.

Barseghyan, L., Prince, J., & Teitelbaum, J. C. (2011). Are risk preferences stable across contexts? Evidence from insurance data. *American Economic Review*, *101*(2), 591-631.

Cohen, A., & Einav, L. (2007). Estimating risk preferences from deductible choice. *American Economic Review*, *97*(3), 745-788.

Pratt, J. W. (1964). Risk aversion in the small and in the large. *Econometrica*, *32*(1/2), 122–136.