Risk Preference Types, Limited Consideration, and Welfare*

Levon Barseghyan
Department of Economics
Cornell University
lb247@cornell.edu

Francesca Molinari
Department of Economics
Cornell University
fm72@cornell.edu

March 20, 2023

Abstract

We provide sufficient conditions for semi-nonparametric point identification of a mixture model of decision making under risk, where agents behave either according to standard expected utility theory with CARA Bernoulli utility function, or according to the dual theory of choice under risk (Yaari 1987) combined with a one-parameter family distortion function. The coefficient of absolute risk aversion and the parameter of the probability distortion function are heterogeneous in the population, and the distribution function of each is left completely unspecified. In addition, the model allows for unobserved heterogeneity in consideration sets within each preference type. The point identification result rests on observing the agents make choices in two distinct contexts, and on sufficient variation in covariates across contexts (without requiring any independent variation across alternatives within a single context). We estimate the model on data on households' deductible choices in two lines of property insurance, and use the results to assess the welfare implications of a hypothetical market intervention where the two lines of insurance are combined into a single one. We study the role of limited consideration in mediating the welfare effects of such intervention.

*We thank Matias Cattaneo, Xinwei Ma, Yusufcan Masatlioglu, Julie Mortimer, Deborah Doukas, Roberta Olivieri, and conference participants at FUR22 and at the JBES session at the ESWM23 for helpful comments. Financial support from NSF grants SES-1824448 and SES-2149374 is gratefully acknowledged.

1 Introduction

This paper is concerned with identification and estimation of risk preferences from agents' choices in property insurance markets, and with assessing the welfare impact of policy interventions in these markets. The key feature of the paper is that it allows for unobserved heterogeneity in preference types, unobserved heterogeneity within each type, and unobserved heterogeneity in the set of alternatives that agents' consider before making a choice (their *consideration set*). Such rich heterogeneity is needed to explain agents' choices, but makes identification analysis challenging.

In property insurance markets, insurance is sold as a collection of coverages, i.e, as a bundle (e.g., for automobiles: collision, comprehensive, liability, etc.). Each part of the bundle is a stand-alone typically vertically differentiated product with a finite menu of options, and can be purchased independently from the other parts of the bundle. Typically, if in a given context an agent faces a larger price than another agent for one alternative, they face a (proportionally) larger price for all other alternatives offered in that context.[1]

There is ample evidence that consumers' choices in insurance markets exhibit patterns that a priori cannot be rationalized by a model of unconstrained optimizing behavior. Handel (2013) and Bhargava et al. (2017) find, in the context of health insurance, that people routinely make or stick to sub-optimal choices. Barseghyan et al. (2016) document, in a setting similar to ours, that more than a quarter of households in their sample make choices that are inconsistent with optimal behavior under any commonly used model of decision making under risk. Some of these households simply pick sub-optimal choices, while others make choices across different insurance lines that imply incompatible levels of risk aversion; i.e., preferences that are not stable, as documented by Barseghyan et al. (2011) and Einav et al. (2012). The traditional additive error random utility model (Luce-McFadden model), or a "trembling hand" alternative (reviewed in Wilcox 2008) that is sometimes used to study insurance demand, often do not remedy the problem, as the model implied choice probabilities can be incompatible with their empirical counterpart. Finally, these incompatibilities are not specific to a particular utility model, but to an entire class of models that satisfy properties that are typically viewed as desirable.[2] This in turn implies that replacing the standard expected utility theory model with its non-expected utility alternatives, such as, e.g., Yaari's (1987) dual theory, Kahneman and Tversky's (1979) prospect theory, and Gul's (1991) disappointment aversion, may be insufficient (see, e.g., Barseghyan et al. 2016, for an analysis of these model's relative (de)merits).

[1] From the perspective of the researcher, this means that there is limited independent variation in covariates within and across coverages.

[2] See Barseghyan et al. (2021b) for a formal discussion and Section 4.3 below for further details.

To resolve these challenges, we propose a rich model suited to study agents' choices in insurance markets with bundled products. We allow for a finite mixture of preference types, with unobserved heterogeneity (a random coefficient) within each type. We focus our empirical analysis on two fundamentally distinct models of decision making under risk: expected utility theory (EU) and Yaari's (1987) dual theory (DT). In the EU model, aversion to risk results from curvature in the Bernoulli utility function through which the expected utility of an alternative is evaluated. The expected utility itself, however, is linear in probabilities. In the DT model, utility is linear in wealth, and aversion to risk results from nonlinearities in how agents evaluate risk (distorting/over-weighting the probability of adverse events). Our approach can accommodate many other preference types, as long as they are *distinct* from one another (as defined in Section 3), and within each context of choice satisfy the standard *single crossing property* of Mirrlees (1971) and Spence (1974), central to important studies of decision making under risk (e.g., Apesteguia et al. 2017; Chiappori et al. 2019).

In addition, we allow for unobserved heterogeneity in consideration sets across agents and for the consideration set formation mechanism to depend on the agent's utility type, but not on the agent's random coefficient conditional on type. The distribution of consideration sets is also required to be independent of at least one regressor $\mathbf{x}$ that enters the utility calculation and to be independent of the random coefficients. The key novelty in this paper relative to our previous work on semi-nonparametric point identification of discrete choice models under risk with limited consideration (Barseghyan et al. 2021b), is that we allow consideration to operate at the bundle level. In particular, whether an alternative offered in one context is considered can depend in unrestricted ways on whether another alternative offered in a distinct context is also considered. As we explain in Section 4.3, this is critical for the model's ability to rationalize the observed data, but makes identification significantly more challenging and the results in Barseghyan et al. (2021b) not directly applicable.

To obtain our semi-nonparametric point identification results (where we have a parametric model for the utility function, with random coefficients whose distribution is not parametrically specified), we assume that the researcher observes agents making choices in two contexts, i.e., the bundle contains two insurance coverages. In keeping with the structure of insurance markets and the limitations of data coming from a single insurance company, we require the covariates $\mathbf{x}$ characterizing products in each context to vary across contexts, but do not require that they vary across alternatives within a context. This very limited form of variation in covariates is among the main challenges that we overcome in the paper.

Our strategy relies on exploiting the single crossing property that each utility model displays within a context, to isolate the response to changes in $\mathbf{x}$ of agents endowed with a specific preference type and random coefficient. Doing so requires working with the elasticity with respect to $\mathbf{x}$ of the choice probability of the cheapest (or the most expensive) bundle.

Under mild restrictions on the consideration set formation mechanism, we show that for both preference types the density function of the random coefficient is identified up to scale without relying on identification-at-infinity arguments nor on large support conditions. When $\mathbf{x}$ does have large support, the entire density function is identified, and the share of each preference type as well as (features of) the distribution of the consideration sets can also be learned.

To exploit the information provided by observing the same agent making choices in distinct contexts, we assume preferences are stable across contexts (a single agent-specific parameterization of the model governs the household's choices in each context). Given the similarity of the two insurance contexts that we study in our empirical application, we consider this assumption an aspect of rationality.[3] Whenever the researcher's interest centers on the analysis of counterfactuals that inherently alter the structure of the decision context(s) studied, stability of preferences across contexts is a necessary ingredient to make progress. As we explain below, we indeed endeavor to estimate the impact of a hypothetical policy intervention that alters the insurance market by restructuring the contexts offered.

We illustrate the relevance of allowing for multiple preference types, unobserved heterogeneity within type, and limited consideration, by estimating risk preferences from data on household's choices in two lines of property insurance, auto collision and auto comprehensive. While currently U.S. property insurance companies offer these two lines of coverage as two separate products, we investigate the implications of offering a combined auto insurance product at a price that equals the sum of the prices for the two separate coverages. Such pricing arises if firms operate under perfect competition or if they use a constant markup rule. This counterfactual exercise is of substantive interest as combined lines of coverage already exist elsewhere (e.g., in Israel; see Cohen and Einav (2007)) and even in the U.S. auto insurance industry: for example, property damage and bodily injury coverage can be offered both as separate lines of coverage, as well as combined in the form of single limit liability coverage. The exercise has the virtue of illustrating the potentially different predictions of the EU model and of the DT model. Moreover, it informs the debate on the need to simplify insurance choice, and it clarifies how limited consideration interacts with the nudging effects associated with this type of market intervention. Under EU, which is linear in probabilities, such an intervention has a weakly negative impact on consumer welfare. This is because imposing single coverage effectively forces households to choose the same deductible vis-à-vis both types of losses. Under DT, however, aversion to risk is generated by concave overweighting of loss probabilities. Hence, the probability weighting function is sub-additive and consumers' welfare may increase. This is because the cumulative weight on the loss state under the combined coverage is less than the sum of the weights under separate coverages. The extent to which either of these effects prevails is crucially impacted

[3]This view of stability as rationality exists elsewhere in the literature; see, e.g., Kahneman (2003).

by whether consideration increases or decreases after the intervention. The richness of our model allows us to account for all these effects.

The rest of the paper is organized as follows. Section 2 lays out the model. Section 3 presents our sufficient conditions for its semi-nonparametric identificationl. Section 4 describes our empirical model and the data. Section 5 reports the results of our estimation exercise. Section 6 reports the results of the welfare exercise. Section 7 concludes by contextualizing our work in the broader literature.

2 Discrete Choice Under Risk in Multiple Contexts

Our starting point is the random utility model developed by McFadden (1974), applied to study choices over risky alternatives with monetary outcomes. We further adapt the model to analyze the behavior of agents who make choices under risk in multiple distinct contexts.

2.1 Lotteries as objects of choice

Following Barseghyan et al. (2021b) and mimicking the setting of our empirical application in Section 4, we focus on risky contexts with two states of the world. We index the finite collection of contexts in which the agents make choices by $j = \mathtt{I}, \mathtt{II}, \ldots, \mathtt{J}$. In each context j, agent i faces an underlying risk of a loss that occurs with probability μ_i^j (and no loss with probability $1 - \mu_i^j$). A finite, context-dependent universe of feasible alternatives $\mathcal{D}^j \equiv \{1^j, \ldots, M^j\}$ is available to insure against this loss. Conditional on risk level, i.e., given μ_i^j, each alternative $\ell^j \in \mathcal{D}^j$ is fully characterized by the pair $(\mathtt{d}^{\ell j}, p_i^{\ell j})$. The first element is the insurance deductible, which is the agent's out of pocket expense if a loss occurs. All deductibles are less than the lowest realization of the loss and $\mathtt{d}^{1j} > \mathtt{d}^{2j} > \cdots > \mathtt{d}^{M^j j}$. The second element is the price (insurance premium), and varies across agents. Each agent is assigned a base price $\bar{p}_i^j$. The agent-specific price of each alternative in $\mathcal{D}^j$ is determined through an agent-invariant function, so that $p_i^{1j} < p_i^{2j} < \cdots < p_i^{M^j j}$ (lower deductibles provide more coverage and cost more). Hence, the lotteries that the agents face are

$$\mathcal{L}(\mathtt{d}^{\ell j}, \bar{p}_i^j, \mu_i^j) \equiv \left(-p_i^{\ell j}, 1 - \mu_i^j; -p_i^{\ell j} - \mathtt{d}^{\ell j}, \mu_i^j\right) \tag{2.1}$$

where $(\bar{p}_i^j, \mu_i^j)$ is observed by the researcher for each agent i and context j. We remark that in this framework there is *no* independent variation in covariates across alternatives. Throughout, we implicitly condition on μ_i^j, and let $\mathbf{x}_i^j \equiv \bar{p}_i^j$ and $\mathbf{x}_i = (\mathbf{x}_i^{\mathtt{I}}, \ldots, \mathbf{x}_i^{\mathtt{J}})$; we let

$\mathcal{L}(\mathtt{d}^{\ell j}, \mathbf{x}_i^j) \equiv \mathcal{L}(\mathtt{d}^{\ell j}, \bar{p}_i^j, \mu_i^j)$. We do not use variation in μ_i^j to establish our identification results, although doing so is potentially useful and the subject of ongoing research.

2.2 Preference types with unobserved heterogeneity within type

We allow the population of agents to be a *mixture of preference types*. The literature has put forward many models of decision making under risk which can generate demand for insurance at actuarially unfair prices, including the workhorse expected utility theory model and a host of non-expected utility theory models. Each of these models has relative (de)merits in rationalizing observed choices, and may deliver different predictions for counterfactual policies (see Barseghyan et al. 2018a, for a review). We hence think it important to provide identification results for a model where multiple preference types are allowed for, and where unobserved heterogeneity within type is also present.

For notational simplicity, we detail here the case with two preference types. The results extend to more than two types (even when one observes choices only in two contexts). Let each agent i draw a preference type t_i as follows:

$$t_i = \begin{cases} 1 & \text{with probability } \alpha, \\ 0 & \text{with probability } 1-\alpha, \end{cases} \tag{2.2}$$

with $\alpha \in (0, 1)$ the unknown mixing probability.

Each realization of t_i is associated with a family of utility functions with *distinct functional forms*, denoted $\mathcal{U}^1 = \{U_\nu,\ \nu \in [0, \bar{\nu}]\}$ for $t_i = 1$, and $\mathcal{U}^0 = \{U_\omega,\ \omega \in [0, \bar{\omega}]\}$ for $t_i = 0$. Functions in each family are known up to a scalar random coefficient that depends on type, denoted ν_i (with support $[0, \bar{\nu}]$) for agents with $t_i = 1$, and ω_i (with support $[0, \bar{\omega}]$) for agents with $t_i = 0$. For example, in our empirical application $\mathcal{U}^1$ is the collection of expected utility functions associated with preferences that exhibit constant absolute risk aversion (CARA) with agent-specific Arrow-Pratt coefficient ν_i,[4] and $\mathcal{U}^0$ is a family of non-expected utility functions that do not nest expected utility as a special case and are parametrized by ω_i (see Eqs. (4.1)-(4.2) and Assumptions 4.2 & 4.3 below). As the preference types are distinct, each agent either receives a draw of ν_i or a draw of ω_i, hence by construction the two random coefficients are independent. We do not impose *any* parametric restrictions on their distributions. Rather, in Section 3 we provide nonparametric point identification results for the two marginal distributions of preferences and for the share of each type.

[4]Other preferences that are characterized by a scalar parameter include ones exhibiting constant relative risk aversion (CRRA), or negligible third derivative (NTD; see, e.g., Cohen and Einav 2007; Barseghyan et al. 2013). Under CRRA, it is required that agents' initial wealth is known to the researcher.

ASSUMPTION 2.1 (Restrictions on distribution of random coefficients): *The random coefficient ν_i (respectively, ω_i) is distributed according to a cumulative distribution function F (respectively, G) that satisfies the properties of CDFs, and admits a density function f that is continuous and strictly positive on $\mathcal{V} \equiv [0, \bar{\nu}]$ (respectively, g strictly positive on $\mathcal{W} \equiv [0, \bar{\omega}]$). Both ν_i and ω_i are independent of $\mathbf{x}_i$.*[5]

We make three fundamental assumptions about utility functions in both families. First, we assume that households' preferences are *stable* across contexts, which allows us to leverage variation in observed choices and covariates across contexts (recall that we have no covariate variation within each context).

ASSUMPTION 2.2 (Stability): *The utility function U_{ν_i} of each agent i with $t_i = 1$ (respectively, U_{ω_i} for agents with $t_i = 0$) is* context-invariant.

Second, we need to take a stand on how agents make choices in multiple contexts. To this end, it is important to introduce notation for *bundles* of alternatives. Denote bundles as $\mathcal{I}_{\mathtt{I},\mathtt{II},\ldots,\mathtt{J}}$, where the first index refers to the alternative in context $\mathtt{I}$, the second one to that in context $\mathtt{II}$, etc. Let $CE_{\nu_i}(\mathcal{L}(\mathtt{d}^{\ell j}, \mathbf{x}^j))$ (respectively, $CE_{\omega_i}(\mathcal{L}(\mathtt{d}^{\ell j}, \mathbf{x}^j))$) denote the *certainty equivalent* of lottery $\mathcal{L}(\mathtt{d}^{\ell j}, \mathbf{x}^j)$ (see, e.g., Mas-Colell et al. 1995, Definition 6.C.2) in a specific context j for an agent of type $t_i = 1$ (respectively, $t_i = 0$). We impose a standard, albeit sometimes implicit, assumption in the literature,[6] according to which agents' choices are made without taking into account any background risk (e.g., Read et al. 1999).

ASSUMPTION 2.3 (Narrow Bracketing): *Agent i's certainty equivalent for the lottery associated with bundle $\mathcal{I}_{I,II,\ldots,J}$ is equal to the sum of that agents' certainty equivalents for lotteries $\mathcal{L}(\mathtt{d}^{\ell j}, \mathbf{x}^j)$, $j = \mathtt{I}, \mathtt{II}, \ldots, \mathtt{J}$.*

Third, we assume that each preference type satisfies the classic Single Crossing Property (SCP) of Mirrlees (1971) and Spence (1974), central to important studies of decision making under risk (see, for example Apesteguia et al. 2017; Chiappori et al. 2019).[7] Formally,

ASSUMPTION 2.4 (Single Crossing Property): *For a given context j and any two lotteries $\mathcal{L}(\mathtt{d}^{\ell j}, \mathbf{x})$ and $\mathcal{L}(\mathtt{d}^{kj}, \mathbf{x})$, $\ell < k$, there exists a continuously differentiable and strictly monotone*

[5] Recall that our analysis conditions on μ_i^j, hence the distribution of preferences may depend on it.

[6] All papers that estimate risk preferences in the field as reviewed in Barseghyan et al. (2018a) impose it.

[7] The SCP 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).

function $\mathcal{Z}_k^\ell : \mathrm{supp}(\mathbf{x}) \to \mathbb{R}_{[-\infty,\infty]}$ *s.t.*

$$\begin{aligned} U_\zeta(\mathcal{L}(d^{kj}, \mathbf{x})) &< U_\zeta(\mathcal{L}(d^{\ell j}, \mathbf{x})) \quad \forall \zeta \in (-\infty, \mathcal{Z}_k^\ell(\mathbf{x})) \\ U_\zeta(\mathcal{L}(d^{kj}, \mathbf{x})) &= U_\zeta(\mathcal{L}(d^{\ell j}, \mathbf{x})) \quad \zeta = \mathcal{Z}_k^\ell(\mathbf{x}) \\ U_\zeta(\mathcal{L}(d^{kj}, \mathbf{x})) &> U_\zeta(\mathcal{L}(d^{\ell j}, \mathbf{x})) \quad \forall \zeta \in (\mathcal{Z}_k^\ell(\mathbf{x}), \infty). \end{aligned}$$

where $\zeta = \nu_i$ *for agents of type* $t_i = 1$*,* $\zeta = \omega_i$ *for type* $t_i = 0$*, and* $\mathrm{supp}(\mathbf{x})$ *is the union of the supports of* $\mathbf{x}^j$ *across contexts* j*. We refer to* $\mathcal{Z}_k^\ell(\cdot)$ *as the cutoff between* $\mathcal{L}(d^{\ell j}, \mathbf{x})$ *and* $\mathcal{L}(d^{kj}, \mathbf{x})$*, and denote it* $\mathcal{V}_k^\ell(\cdot)$ *for* $t_i = 1$ *and* $\mathcal{W}_k^\ell(\cdot)$ *for* $t_i = 0$*.*[8]

Within a single context, the expected utility theory framework generally satisfies the SCP, which requires that if an agent with a certain degree of risk aversion (the random coefficient ν_i) prefers a safer lottery to a riskier one, then all agents with higher risk aversion also prefer the safer lottery.[9] The same is true for the non-expected utility theory model that we use in our empirical analysis in Section 4. The SCP implies that *within a single context*, the household's ranking of alternatives is monotone in ν_i for $t_i = 1$ and in ω_i for $t_i = 0$, yielding vertical differentiation of alternatives within each preference type.

2.3 Unobserved heterogeneity in consideration sets

Across contexts, agents face an overall large and potentially overwhelming universe of feasible alternatives, leading to choice overload, cognitive ability constraints, etc. The subset of alternatives actually available to each agent is unknown to the researcher, due, e.g., to unobserved budget constraints, liquidity constraints, etc. Hence, we allow for *unobserved heterogeneity in consideration sets*, i.e., in the collection of alternatives that the agents evaluate when making their choices. We denote the overall universe of alternatives across contexts as $\mathcal{D} \equiv \times_{j=\mathrm{I}}^{\mathrm{J}} \mathcal{D}^j$, with $\mathcal{I}_{\mathrm{I},\mathrm{II},\dots,\mathrm{J}}$ denoting each of the *bundles* in $\mathcal{D}$.

Assumption 2.5 (Consideration set formation mechanism): *Conditional on* t_i*, agent* i *draws a consideration set* $C_i \subseteq \mathcal{D}$ *independently from its random coefficient and from* $\mathbf{x}_i$ *s.t.*

$$\begin{aligned} \mathcal{Q}_1(\mathcal{K}) &\equiv \Pr(C_i = \mathcal{K} | t_i = 1) = \Pr(C_i = \mathcal{K} | \mathbf{x}_i, \nu_i, t_i = 1), \quad \mathcal{K} \subseteq \mathcal{D} \\ \mathcal{Q}_0(\mathcal{K}) &\equiv \Pr(C_i = \mathcal{K} | t_i = 0) = \Pr(C_i = \mathcal{K} | \mathbf{x}_i, \omega_i, t_i = 0), \quad \mathcal{K} \subseteq \mathcal{D} \end{aligned}$$

[8] We assume that while ν and ω have bounded support, the utility functions in $\mathcal{U}^1$ and $\mathcal{U}^0$ are well defined for any real valued ν and ω, respectively.

[9] For a discussion of possible failures of SCP, see Apesteguia and Ballester (2018).

The fundamental restrictions imposed in Assumption 2.5 are that conditional on preference type, consideration is independent of the agent's random coefficient and of the observed covariate $\mathbf{x}$.[10] However, the distribution of consideration sets may depend on preference type. Importantly, we allow consideration to be *broad*, as it is determined at the bundle level instead of within context. A significantly more restrictive approach would posit that consideration is *narrow*: agent i draws a tuple of consideration sets $C_i^j \in \mathcal{D}^j$, $j = \mathtt{I}, \mathtt{II}, \ldots, \mathtt{J}$ independently across contexts, and forms $C_i = \times_{j=\mathtt{I}}^{\mathtt{J}} C_i^j$. As we further discuss in Section 2.4, allowing consideration sets to be drawn at the bundle level substantially complicates the identification analysis, but delivers a more realistic model.

2.4 Optimal choice within the consideration set

Once the consideration set is drawn, each agent chooses the best alternative in each context according to their preferences.

$$\mathcal{I}^* \equiv [\ell^{*\mathtt{I}}, \ell^{*\mathtt{II}}, \ldots, \ell^{*\mathtt{J}}] = \arg \max_{[\ell^{\mathtt{I}}, \ell^{\mathtt{II}}, \ldots, \ell^{\mathtt{J}}] \in C} \sum_{j=\mathtt{I}}^{\mathtt{J}} CE_\zeta(\mathcal{L}(\mathtt{d}^{\ell^j}, \mathbf{x}^j)) \tag{2.3}$$

where $\zeta = \nu$ if $t = 1$ and $\zeta = \omega$ if $t = 0$. The bundle choice $\mathcal{I}^*$ depends on the agent's preference type, random coefficient, consideration set, associated premium-deductible tuples, and claim probabilities μ^j, $j = \mathtt{I}, \ldots, \mathtt{J}$.

The flexible model of consideration set formation that we allow for has important implications for the choice problem in Eq. (2.3). If we were to assume narrow consideration, hence restrict agents to draw consideration sets *independently* across contexts, the choice problems would break into independent decisions, with

$$\ell^{*j} = \arg \max_{\ell^j \in C^j} CE_\zeta(\mathcal{L}(\mathtt{d}^{\ell^j}, \mathbf{x}^j))$$

Each context-specific choice problem satisfies the SCP in Assumption 2.4. Barseghyan et al. (2021b) offer a comprehensive analysis of the implications of the single crossing property for semi-nonparametric identification of a model of discrete choice under risk that features a single preference type and unobserved heterogeneity in consideration sets. As we allow for multiple preference types, our analysis extends theirs even in the simplified framework where consideration is narrow. More importantly, a narrow model of consideration implies that very similar alternatives in different contexts (e.g., a \$500 deductible at price $p^{\mathtt{I}}$ in collision insurance and a \$500 deductible at price $p^{\mathtt{II}}$ in comprehensive insurance) enter

[10] Recall that our analysis conditions on μ_i^j, hence the distribution of consideration sets may depend on it.

the consideration set independently. This assumption is unpalatable, particularly when analyzing demand for bundled products. We therefore allow for broad consideration. In doing so, we overcome a substantial hurdle relative to Barseghyan et al. (2021b). When consideration is broad and C_i is formed at the bundle level, the SCP may not necessarily hold across tuples of alternatives, because alternatives may not be monotonically ranked (with respect to ν_i or ω_i) against each other. Hence, here we develop a new approach to obtain point identification of the distribution of preferences, $f(\cdot)$ for agents of type $t_i = 1$ and $g(\cdot)$ for agents of type $t_i = 0$; of the shares α and $(1-\alpha)$ of preferences types; and of features of the distribution of consideration sets given type, $\mathcal{Q}_{t_i}(\cdot)$, $t_i = 0, 1$.

3 Identification Results

We derive our formal identification results for the case of two contexts of choice, $j = \mathtt{I}, \mathtt{II}$ (the results extend easily to more contexts, at the cost of heavier notation). We begin by describing the conditions under which we can prove our point identification results. We index bundles as $\mathcal{I}_{k,r}$. We let $\mathcal{V}^{\ell,q}_{k,r}(\mathbf{x})$ and $\mathcal{W}^{\ell,q}_{k,r}(\mathbf{x})$ denote, respectively, cutoff levels for ν_i and ω_i at which the agent is indifferent between bundles $\mathcal{I}_{k,r}$ and $\mathcal{I}_{\ell,q}$. Under Assumption 2.3, at cutoff $\mathcal{V}^{\ell,q}_{k,r}(\mathbf{x})$ (and similarly for $\mathcal{W}^{\ell,q}_{k,r}(\mathbf{x})$) we have

$$CE_{\mathcal{V}^{\ell,q}_{k,r}(\mathbf{x})}(\mathcal{L}(\mathtt{d}^{\ell\mathtt{I}}, \mathbf{x}^{\mathtt{I}})) + CE_{\mathcal{V}^{\ell,q}_{k,r}(\mathbf{x})}(\mathcal{L}(\mathtt{d}^{q\mathtt{II}}, \mathbf{x}^{\mathtt{II}})) = CE_{\mathcal{V}^{\ell,q}_{k,r}(\mathbf{x})}(\mathcal{L}(\mathtt{d}^{k\mathtt{I}}, \mathbf{x}^{\mathtt{I}})) + CE_{\mathcal{V}^{\ell,q}_{k,r}(\mathbf{x})}(\mathcal{L}(\mathtt{d}^{r\mathtt{II}}, \mathbf{x}^{\mathtt{II}}))$$

Relative to the cutoffs introduced in Assumption 2.4, which compared alternatives within a single context and we denoted $\mathcal{V}^{\ell}_{k}(\mathbf{x})$ (single superscript and subscript for a single context of choice), we have $\mathcal{V}^{\ell,s}_{k,s}(\mathbf{x}) = \mathcal{V}^{\ell}_{k}(\mathbf{x})$ and $\mathcal{V}^{s,q}_{s,r}(\mathbf{x}) = \mathcal{V}^{q}_{r}(\mathbf{x})$ for all $\mathtt{d}^{s\mathtt{II}} \in \mathcal{D}^{\mathtt{II}}$ and $\mathtt{d}^{s\mathtt{I}} \in \mathcal{D}^{\mathtt{I}}$ (and similarly for $\mathcal{W}^{\ell,s}_{k,s}(\mathbf{x})$). While cutoffs $\mathcal{V}^{\ell,q}_{k,r}(\mathbf{x})$ and $\mathcal{W}^{\ell,q}_{k,r}(\mathbf{x})$ for $\ell \neq k, q \neq r$ depend on both $\mathbf{x}^{\mathtt{I}}$ and $\mathbf{x}^{\mathtt{II}}$, cutoffs $\mathcal{V}^{\ell,s}_{k,s}(\mathbf{x})$ and $\mathcal{W}^{\ell,s}_{k,s}(\mathbf{x})$ depend *only* on $\mathbf{x}^{\mathtt{I}}$, while cutoffs $\mathcal{V}^{s,q}_{s,r}(\mathbf{x})$ and $\mathcal{W}^{s,q}_{s,r}(\mathbf{x})$ depend *only* on $\mathbf{x}^{\mathtt{II}}$. These properties will be used to establish our identification results.

We remark that the cutoffs $\mathcal{V}^{\ell,q}_{k,r}(\mathbf{x})$ and $\mathcal{W}^{\ell,q}_{k,r}(\mathbf{x})$ may not be unique if $\ell > k$ but $q < r$ (or vice versa). However, they are unique whenever $\mathcal{I}_{1,1}$ is compared with *any* other bundle (and similarly whenever $\mathcal{I}_{M^{\mathtt{I}},M^{\mathtt{II}}}$ is compared with any other bundle).

Throughout, we assume that the researcher has access to data that identify the joint distribution of chosen bundles and covariates. The consideration set, however, is not observed.

Assumption 3.1 (Observed data): *A random sample* $\{(\mathcal{I}^*, \mathbf{x}^{I}_i, \mathbf{x}^{II}_i) : i = 1, \ldots, n\}$ *is observed, with* $\mathcal{I}^*_i$*, as defined in Eq. (2.3).*

3.1 Restrictions on variation in $\mathbf{x}^j$ across contexts

Identification of the model's functionals rests on the interplay between the model and the variation in the observed covariates. We only require the covariates $\mathbf{x}_i \equiv (\mathbf{x}_i^{\mathtt{I}}, \mathbf{x}_i^{\mathtt{II}})$ to vary across agents and contexts, as formally stated below. This creates substantial challenges for identification analysis, as we have no independent variation in covariates across the alternatives in a single context. Hence, for identification to be possible one needs sufficient variation across contexts.

Assumption 3.2 (Preferred within a triplet): *In each context $j \in \{I, II\}$, for any $\mathbf{x}$ and triplet $\{d^{1j}, d^{kj}, d^{(k+1)j}\}$, $\forall k \in \{2, ..., M^j - 1\}$, there are values of ν (and ω) at which each alternative in this triplet is strictly preferred to the other two.*

Assumption 3.2 requires that given three coverage levels including the cheapest, each one is preferred by at least some agent. As shown in Barseghyan et al. (2021b), under Assumption 2.4, this condition is satisfied for agents of type $t_i = 1$ within context I if and only if $-\infty < \mathcal{V}_{2,1}^{1,1}(\mathbf{x}^{\mathtt{I}}) < \mathcal{V}_{3,1}^{1,1}(\mathbf{x}^{\mathtt{I}}) < \mathcal{V}_{4,1}^{1,1}(\mathbf{x}^{\mathtt{I}}) \cdots < +\infty$ (and similarly for agents of type $t_i = 0$, and for context II with appropriate modifications in the compared bundles and evaluation at $\mathbf{x}^{\mathtt{II}}$ instead of $\mathbf{x}^{\mathtt{I}}$). Hence, any agent of type $t_i = 1$ who draws $\nu < \mathcal{V}_{2,1}^{1,1}(\mathbf{x}^{\mathtt{I}})$ unambiguously prefers alternative $\ell^{1\mathtt{I}}$ to any other alternative in $\mathcal{D}^{\mathtt{I}}$.

In what follows, an important role is played by the values of $\mathbf{x} = (\mathbf{x}^{\mathtt{I}}, \mathbf{x}^{\mathtt{II}})$ at which the indifference cutoff for an agent of type t_i between alternatives $\ell^{1\mathtt{I}}$ and $\ell^{2\mathtt{I}}$ (the two cheapest alternatives in context I) is equal to that agent's indifference cutoff between alternatives $\ell^{1\mathtt{II}}$ and $\ell^{2\mathtt{II}}$ (the two cheapest alternatives in context II). We first define these values of $\mathbf{x}$, and then make assumptions on the support of $\mathbf{x}$ to guarantee that it includes them.

Definition 3.1 (Covariate values delivering indifference): *Given t_i, fix a value of $\nu \in [0, \bar{\nu}]$ if $t_i = 1$ and of $\omega \in [0, \bar{\omega}]$ if $t_i = 0$. Let the set of covariate values at which the agent has preference ν (respectively, ω) and is indifferent between bundles $\mathcal{I}_{1,1}$, $\mathcal{I}_{1,2}$, and $\mathcal{I}_{2,1}$, be:*

$$\mathbf{X}^1(\nu) \equiv \{(\mathbf{x}^I, \mathbf{x}^{II}) : \mathcal{V}_{2,1}^{1,1}(\mathbf{x}^I) = \mathcal{V}_{1,2}^{1,1}(\mathbf{x}^{II}) = \nu\}$$
$$\mathbf{X}^0(\omega) \equiv \{(\mathbf{x}^I, \mathbf{x}^{II}) : \mathcal{W}_{2,1}^{1,1}(\mathbf{x}^I) = \mathcal{W}_{1,2}^{1,1}(\mathbf{x}^{II}) = \omega\}$$

The covariate values $\mathbf{X}^1(\nu)$ (respectively, $\mathbf{X}^0(\omega)$) are the values of $\mathbf{x} = (\mathbf{x}^{\mathtt{I}}, \mathbf{x}^{\mathtt{II}})$ at which an agent with preferences ν (respectively, ω) is indifferent between the two cheapest coverage levels in context I and, at the same time, also in context II. In other words, the agent is indifferent between $\mathcal{I}_{1,1}$, $\mathcal{I}_{1,2}$, and $\mathcal{I}_{2,1}$ (and, hence, $\mathcal{I}_{2,2}$). Given the single crossing property in Assumption 2.4, within each context it is immediate to see that both elements of $\mathbf{X}^1(\nu)$ (the covariate value in context I and the covariate value in context II) are strictly monotone in ν (and, similarly, both elements of $\mathbf{X}^0(\omega)$ are monotone in ω). For example, the higher is ν, the higher is the base price in context `I` at which the agent with random coefficient ν is indifferent between $\mathcal{I}_{1,1}$ and $\mathcal{I}_{2,1}$. Hence, we can represent $\mathbf{X}^1(\nu)$ (respectively, $\mathbf{X}^0(\omega)$) as a strictly monotone function on the support of $(\mathbf{x}^{\mathtt{I}}, \mathbf{x}^{\mathtt{II}})$.[11] We assume that these strictly monotone functions intersect on a set of measure zero.

Assumption 3.3 (Distinct contexts): *The contexts are distinct, in the sense that:*

(I) $\mathbf{X}^1(\nu) \neq \mathbf{X}^0(\omega)$ *a.e.*

(II) The following four conditions are satisfied:

$$\mathcal{V}^{1,1}_{\ell,q}(\mathbf{X}^0(\omega)) \neq \mathcal{V}^{1,1}_{k,r}(\mathbf{X}^0(\omega)) \text{ a.e.} \tag{3.1}$$

$$\mathcal{W}^{1,1}_{\ell,q}(\mathbf{X}^1(\nu)) \neq \mathcal{W}^{1,1}_{k,r}(\mathbf{X}^1(\nu)) \text{ a.e.} \tag{3.2}$$

$$\mathcal{V}^{1,1}_{\ell,q}(\mathbf{X}^1(\nu)) \neq \mathcal{V}^{1,1}_{k,r}(\mathbf{X}^1(\nu)) \text{ a.e. } \forall\{\ell,q,k,r\} \text{ s.t. } \{\ell,q,k,r\}\backslash\{1,2\} \neq \varnothing \tag{3.3}$$

$$\mathcal{W}^{1,1}_{\ell,q}(\mathbf{X}^0(\omega)) \neq \mathcal{W}^{1,1}_{k,r}(\mathbf{X}^0(\omega)) \text{ a.e. } \forall\{\ell,q,k,r\} \text{ s.t. } \{\ell,q,k,r\}\backslash\{1,2\} \neq \varnothing \tag{3.4}$$

Assumption 3.3-(II) implies Assumption 3.3-(I), as Eqs. (3.1)-(3.2) for $\ell, q = 2, 1$ and $k, r = 1, 2$ imply $\mathbf{X}^1(\nu) \neq \mathbf{X}^0(\omega)$ *a.e.* Both conditions require that at any value of $(\mathbf{x}^{\mathtt{I}}, \mathbf{x}^{\mathtt{II}})$ at which indifference across $\mathcal{I}_{1,1}$, $\mathcal{I}_{1,2}$, and $\mathcal{I}_{2,1}$ occurs for an agent of type $t_i = 1$, such indifference cannot occur for an agent of type $t_i = 0$. Additionally, Assumption 3.3-(II) requires that at any value of $(\mathbf{x}^{\mathtt{I}}, \mathbf{x}^{\mathtt{II}})$ at which indifference across $\mathcal{I}_{1,1}$, $\mathcal{I}_{1,2}$, and $\mathcal{I}_{2,1}$ occurs, no other triplet of bundles including $\mathcal{I}_{1,1}$ can generate a three-way tie in utility ranking. Given the data and utility models across preference types, one can directly check whether Assumption 3.3 is satisfied. Finally, we require that the support of $\mathbf{x}$ is sufficiently rich.

Assumption 3.4 (Independent variation in $\mathbf{x}$): *Given intervals* $[\nu^*, \nu^{**}] \subseteq [0, \bar{\nu}]$ *and* $[\omega^*, \omega^{**}] \subseteq [0, \bar{\omega}]$*, for some* $\epsilon > 0$*, the random vector* $\mathbf{x} = (\mathbf{x}^I, \mathbf{x}^{II})$ *has strictly positive density on the sets* $\mathcal{S}^1_\epsilon(\nu^*, \nu^{**}) \subset \mathbb{R}^2$ *and* $\mathcal{S}^0_\epsilon(\omega^*, \omega^{**}) \subset \mathbb{R}^2$*, with*

$$\mathcal{S}^1_\epsilon(\nu^*, \nu^{**}) = \left\{\mathsf{B}_\epsilon(\mathbf{X}^1(\nu)),\ \nu \in [\nu^*, \nu^{**}]\right\}$$
$$\mathcal{S}^0_\epsilon(\omega^*, \omega^{**}) = \left\{\mathsf{B}_\epsilon(\mathbf{X}^0(\omega)),\ \omega \in [\omega^*, \omega^{**}]\right\}$$

where $\mathsf{B}_a(c)$ *denotes a ball in* $\mathbb{R}^2$ *of radius* a *centered at* c*.*

Assumption 3.4 guarantees that for each $\nu \in [\nu^*, \nu^{**}]$ there are values of $\mathbf{x}$ such that $\mathbf{X}^1(\nu)$ is non-empty and that there is an ϵ-neighborhood around $\mathbf{X}^1(\nu)$ with positive density (and similarly for $\mathbf{X}^0(\omega)$ and all $\omega \in [\omega^*, \omega^{**}]$). This yields sufficient observed variation in $\mathbf{x}$

[11] See Figure 3.1 and its discussion below.

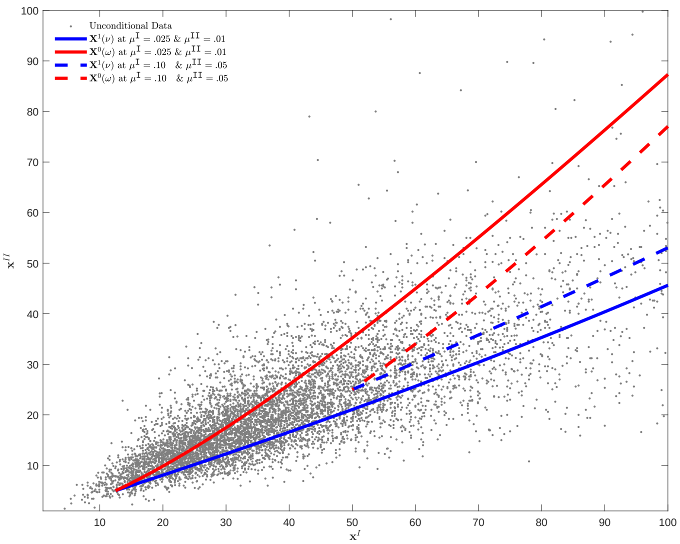

Figure 3.1: $\mathbf{X}^0(\omega)$ and $\mathbf{X}^1(\nu)$ in our application, with data in the background.

to identify the functionals that we are after. We illustrate the notion of distinct contexts and independent variation in $\mathbf{x}$ via Figure 3.1, which depicts $\mathbf{X}^0(\omega)$ and $\mathbf{X}^1(\nu)$ drawn for different pairs of μ's. First, $\mathbf{X}^0(\omega)$ and $\mathbf{X}^1(\nu)$ intersect only at a single point, which corresponds to $\nu = 0$ and $\omega = 1$, i.e., respectively, no risk aversion and no probability distortions. Second, these curves are both monotone. We present them with a scatterplot of unconditional data from our empirical application in the background, to highlight the fact that even when variation in $\mathbf{x}$ does not cover the entire R^2_+, identification is attainable since Assumption 3.4 requires variation in $\mathbf{x}$ only to cover respective neighborhoods of $\mathbf{X}^0(\omega)$ and $\mathbf{X}^1(\nu)$.

We next explain why, under full consideration, our assumptions suffice for identification of the share of preference types and the distributions of the respective random coefficients. Fix a value of $\nu \in [\nu^*, \nu^{**}]$ at which one wants to learn $f(\nu)$. Under Assumption 3.4, $\mathbf{X}^1(\nu)$ is non-empty and there is an ϵ-ball of positive density around it. Along with Assumption 3.3, this implies that there is a vector $(\mathbf{x}^{\mathtt{I}\prime}, \mathbf{x}^{\mathtt{II}\prime}) \in \mathsf{B}_\epsilon(\mathbf{X}^1(\nu))$ such that $\nu = \mathcal{V}^{1,1}_{2,1}(\mathbf{x}^{\mathtt{I}\prime}) < \mathcal{V}^{1,1}_{1,2}(\mathbf{x}^{\mathtt{II}\prime})$ and $\mathcal{W}^{1,1}_{2,1}(\mathbf{x}^{\mathtt{I}\prime}) > \mathcal{W}^{1,1}_{1,2}(\mathbf{x}^{\mathtt{II}\prime})$. Then, as shown in Figure 3.2, under Assumptions 2.4 and 3.2,[12]

$$\Pr(\mathcal{I}^*_i = \mathcal{I}_{1,1}|\mathbf{x}') = \alpha F(\mathcal{V}^{1,1}_{2,1}(\mathbf{x}^{\mathtt{I}\prime})) + (1-\alpha) G(\mathcal{W}^{1,1}_{1,2}(\mathbf{x}^{\mathtt{II}\prime}))$$

[12] Recall that these assumptions, jointly, imply that any agent who draws $\nu < \mathcal{V}^{1,1}_{2,1}(\mathbf{x}^{\mathtt{I}\prime}) < \mathcal{V}^{1,1}_{1,2}(\mathbf{x}^{\mathtt{II}\prime})$ unambiguously prefers alternative $\ell^{1\mathtt{I}}$ to all other alternatives in $\mathcal{D}^{\mathtt{I}}$, unambiguously prefers alternative $\ell^{1\mathtt{II}}$ to all other alternatives in $\mathcal{D}^{\mathtt{II}}$, and therefore unambiguously prefers bundle $\mathcal{I}_{1,1}$ to any other bundle in $\mathcal{D}$.

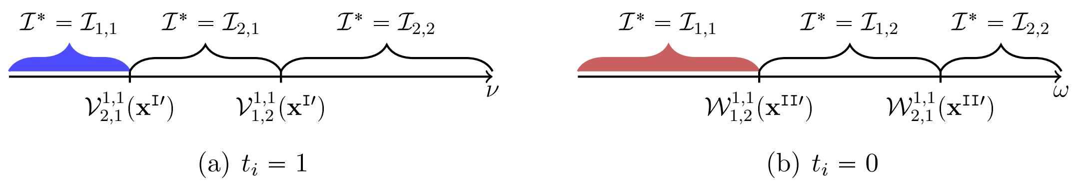

Figure 3.2: Stylized depiction of regions where, under full consideration, $\mathcal{I}^* = \mathcal{I}_{1,1}$.

In turn, owing to the fact that $\mathcal{V}^{1,1}_{2,1}(\mathbf{x}^{\mathrm{I}\prime})$ depends on $\mathbf{x}^{\mathrm{I}}$ but $\mathcal{W}^{1,1}_{1,2}(\mathbf{x}^{\mathrm{II}\prime})$ does not, this yields

$$\frac{\partial \Pr(\mathcal{I}^*_i = \mathcal{I}_{1,1}|\mathbf{x}')}{\partial \mathbf{x}^{\mathrm{I}}} = \alpha f(\nu) \frac{\partial \mathcal{V}^{1,1}_{2,1}(\mathbf{x}^{\mathrm{I}\prime})}{\partial \mathbf{x}^{\mathrm{I}}}$$

where the term $\frac{\partial \mathcal{V}^{1,1}_{2,1}(\mathbf{x}^{\mathrm{I}\prime})}{\partial \mathbf{x}^{\mathrm{I}}}$ is a known function of $\mathbf{x}^{\mathrm{I}}$ and is different from zero due to Assumption 2.4 (where cutoff functions are assumed to be strictly monotone in $\mathbf{x}$). If $[\nu^*, \nu^{**}] = [0, \bar{\nu}]$, one can repeat the above argument for all ν on the support and then use the fact that $f(\nu)$ integrates to one to learn α. One can similarly learn $g(\omega)$, $\omega \in [\omega^*, \omega^{**}]$.

3.2 Restrictions on the consideration set formation mechanism

In the presence of limited consideration, the above argument does not directly apply, as one needs to carefully account for all possible consideration sets in which bundle $\mathcal{I}_{1,1}$ is included. We therefore need to introduce additional notation and some restrictions.

For any $\mathcal{K}_1, \mathcal{K}_2 \subseteq \mathcal{D}$, $\mathcal{K}_1 \cap \mathcal{K}_2 = \varnothing$, denote the probability that all elements of $\mathcal{K}_1$ are included in the consideration set while all elements of $\mathcal{K}_2$ are excluded from it, by

$$\mathcal{O}_1(\mathcal{K}_1; \mathcal{K}_2) \equiv \sum_{\mathcal{K}:\ \mathcal{K}_1 \subset \mathcal{K},\ \mathcal{K}_2 \cap \mathcal{K} = \varnothing} \Pr(C_i = \mathcal{K}|t_i = 1) = \sum_{\mathcal{K}:\ \mathcal{K}_1 \subset \mathcal{K},\ \mathcal{K}_2 \cap \mathcal{K} = \varnothing} \mathcal{Q}_1(\mathcal{K})$$

and define $\mathcal{O}_0(\mathcal{K}_1; \mathcal{K}_2)$ similarly.

Denote by $\mathbb{B}(\mathcal{I}_{\ell,q}, \mathbf{x}; \zeta)$ the collection of bundles that at a given value of ζ strictly dominate bundle $\mathcal{I}_{\ell,q}$, with $\zeta = \nu_i$ for agents of type $t_i = 1$, and $\zeta = \omega_i$ for $t_i = 0$:

$$\mathbb{B}(\mathcal{I}_{\ell,q}, \mathbf{x}; \zeta) \equiv \{\mathcal{I}_{k,r}\ s.t.\ CE_\zeta(\mathcal{I}_{k,r}, \mathbf{x}) > CE_\zeta(\mathcal{I}_{\ell,q}, \mathbf{x})\}$$

Then, for a given value of $\mathbf{x}$, any bundle $\mathcal{I}_{\ell,q} \in \mathcal{D}$ is chosen if and only if it is considered and every bundle that dominates it is not:

$$\Pr(\mathcal{I}^*_i = \mathcal{I}_{\ell,q}|\mathbf{x}) = \alpha \int \mathcal{O}_1(\mathcal{I}_{\ell,q}; \mathbb{B}(\mathcal{I}_{\ell,q}, \mathbf{x}; \nu))dF + (1-\alpha) \int \mathcal{O}_0(\mathcal{I}_{\ell,q}; \mathbb{B}(\mathcal{I}_{\ell,q}, \mathbf{x}; \omega))dG \quad (3.5)$$

As the indifference cutoffs involving bundle $\mathcal{I}_{1,1}$ are unique, differentiating Eq. (3.5) with respect to $\mathbf{x}^{\mathrm{I}}$ we have:

$$\begin{aligned}\frac{\partial \Pr(\mathcal{I}_i^* = \mathcal{I}_{1,1}|\mathbf{x})}{\partial \mathbf{x}^{\mathrm{I}}} = \alpha &\sum_{(k,r)\neq(1,1)} \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{k,r}\}; \mathbb{B}(\mathcal{I}_{1,1}, \mathbf{x}; \mathcal{V}_{k,r}^{1,1})) f(\mathcal{V}_{k,r}^{1,1}) \frac{\partial \mathcal{V}_{k,r}^{1,1}}{\partial \mathbf{x}^{\mathrm{I}}} \\ &+ (1-\alpha) \sum_{(k,r)\neq(1,1)} \mathcal{O}_0(\{\mathcal{I}_{1,1}, \mathcal{I}_{k,r}\}; \mathbb{B}(\mathcal{I}_{1,1}, \mathbf{x}; \mathcal{W}_{k,r}^{1,1})) g(\mathcal{W}_{k,r}^{1,1}) \frac{\partial \mathcal{W}_{k,r}^{1,1}}{\partial \mathbf{x}^{\mathrm{I}}} \end{aligned} \quad (3.6)$$

Eq. (3.5) with $(\ell, q) = (1, 1)$ shows that bundle $\mathcal{I}_{1,1}$ is chosen when it is the bundle in the consideration set yielding the highest certainty equivalent, i.e., no bundle that yields a higher certainty equivalent (any of the bundles in $\mathbb{B}(\mathcal{I}_{1,1}, \mathbf{x}; \cdot)$) is considered.[13] Consequently, an agent choosing bundle $\mathcal{I}_{1,1}$ will switch to or from a different bundle $\mathcal{I}_{k,r}$ if and only if (i) they are indifferent between $\mathcal{I}_{1,1}$ and $\mathcal{I}_{k,r}$; and (ii) they do not consider any bundle in $\mathcal{D}$ that dominates $\mathcal{I}_{1,1}$ and $\mathcal{I}_{k,r}$. The summation in Eq. (3.6) collects all relevant consideration sets across the different preference types and indifference points (cutoffs), weighted by the density function at these indifference points and taking into account how the change in $\mathbf{x}^{\mathrm{I}}$ affects the indifference points themselves.[14]

We impose the following restrictions on the consideration set formation mechanism:

Assumption 3.5 (Minimally informative consideration): *One of the following holds:*

(I) $\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}, \mathcal{I}_{2,1}\}; \varnothing) = \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}, \mathcal{I}_{1,2}\}; \varnothing) > 0.$

(II) $\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}, \mathcal{I}_{2,1}\}; \varnothing) - \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}, \mathcal{I}_{1,2}\}; \varnothing) \neq 0$, *and*

$\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,1}\}; \varnothing) = \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,1}\}; \{\mathcal{I}_{2,2}, \mathcal{I}_{1,2}\})$.[15]

One of these two restrictions also holds with $\mathcal{O}_0$ replacing $\mathcal{O}_1$.

Assumption 3.5-(I) requires symmetry in the probability with which the triplets $(\mathcal{I}_{1,1}, \mathcal{I}_{2,2}, \mathcal{I}_{1,2})$ and $(\mathcal{I}_{1,1}, \mathcal{I}_{2,2}, \mathcal{I}_{2,1})$ are included in the consideration set, and that each probability is strictly

[13] Equivalently, bundle $\mathcal{I}_{\ell,q}$ is chosen if and only if it is the first best among the ones considered:

$$\begin{aligned}\Pr(\mathcal{I}_i^* = \mathcal{I}_{\ell,q}|\mathbf{x}) = \alpha &\sum_{\mathcal{I}_{\ell,q}\in\mathcal{K}} \mathcal{Q}_1(\mathcal{K}) \int \mathbf{1}(CE_\nu(\mathcal{I}_{k,r}, \mathbf{x}) \leqslant CE_\nu(\mathcal{I}_{\ell,q}, \mathbf{x})\ \forall \mathcal{I}_{k,r} \in \mathcal{K}|\mathbf{x}; \nu) dF \\ &+ (1-\alpha) \sum_{\mathcal{I}_{\ell,q}\in\mathcal{K}} \mathcal{Q}_0(\mathcal{K}) \int \mathbf{1}(CE_\omega(\mathcal{I}_{k,r}, \mathbf{x}) \leqslant CE_\omega(\mathcal{I}_{\ell,q}, \mathbf{x})\ \forall \mathcal{I}_{k,r} \in \mathcal{K}|\mathbf{x}; \omega) dG\end{aligned}$$

[14] A similar expression holds for $\partial \Pr(\mathcal{I}_i^* = \mathcal{I}_{1,1}|\mathbf{x})/\partial \mathbf{x}^{\mathrm{II}}$: the right-hand-side of Eq. (3.6) remains as is, except that instead of $\partial\mathcal{V}_{k,r}^{1,1}/\partial\mathbf{x}^{\mathrm{I}}$ and $\partial\mathcal{W}_{k,r}^{1,1}/\partial\mathbf{x}^{\mathrm{I}}$ we have $\partial\mathcal{V}_{k,r}^{1,1}/\partial\mathbf{x}^{\mathrm{II}}$ and $\partial\mathcal{W}_{k,r}^{1,1}/\partial\mathbf{x}^{\mathrm{II}}$, respectively.

[15] Alternatively, $\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{1,2}\}; \varnothing) = \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{1,2}\}; \{\mathcal{I}_{2,2}, \mathcal{I}_{2,1}\})$ can replace the last condition in Assumption 3.5-(II). In our application this alternative restriction is satisfied because bundle $\mathcal{I}_{1,2}$ (which is the deductible bundle $\{\$1000, \$500\}$) is chosen with probability zero, and hence both probabilities are zero.

positive, so that information can be extracted through the differentiation in Eq. (3.6). Assumption 3.5-(II) requires that if such symmetry is absent, then alternatives $\mathcal{I}_{1,1}$ and $\mathcal{I}_{2,1}$ can only be considered together when neither $\mathcal{I}_{1,2}$ nor $\mathcal{I}_{2,2}$ are considered (a trivial case that would guarantee this condition is that $\mathcal{I}_{2,1}$ is never considered when $\mathcal{I}_{1,1}$ is). The conditions in Assumption 3.5 are sufficient (together with the other assumptions listed above) for our identification results. However, they can be replaced by technical yet verifiable assumptions on the behavior of the cutoffs involving comparisons of alternatives $\mathcal{I}_{1,1}, \mathcal{I}_{2,1}, \mathcal{I}_{1,2}, \mathcal{I}_{2,2}$.[16]

3.3 Identification results

We next state our main identification results, whose proofs are in the Appendix.

THEOREM 3.1: *Let Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, 3.1, 3.2, 3.3, 3.4, 3.5 hold. Then*

1. *$f(\cdot)$ is identified up to scale on $[\nu^*, \nu^{**}]$.*
2. *$g(\cdot)$ is identified up to scale on $[\omega^*, \omega^{**}]$.*
3. *If $[\nu^*, \nu^{**}] = [0, \bar{\nu}]$ and $[\omega^*, \omega^{**}] = [0, \bar{\omega}]$, then $f(\cdot)$ and $g(\cdot)$ are identified.*

Theorem 3.1 shows that under limited consideration, despite the lack of independent variation in observed covariates across alternatives (within a single context), it is nonetheless possible to identify the distribution of the random coefficient for each preference type without relying on identification at infinity arguments.[17] While to pin down the entire distribution of preferences large support is required, our approach identifies (up to scale) the density function of each random coefficient conditional on a given interval.

One can identify the shares of preference types under a mild additional restriction that requires that the probability of including one specific pair of bundles in the consideration set and excluding another specific bundle (or pair of bundles) is independent of preference type.

COROLLARY 3.1: *Suppose all Assumptions of Theorem 3.1 hold, and that either:*

(i) Assumption 3.5-(I) holds for both agents with preference types $t_i = 1$ and $t_i = 0$, and

$$\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,1}\}; \varnothing) - \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,1}\}; \{\mathcal{I}_{2,2}, \mathcal{I}_{1,2}\}) = \mathcal{O}_0(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,1}\}; \varnothing) - \mathcal{O}_0(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,1}\}; \{\mathcal{I}_{2,2}, \mathcal{I}_{1,2}\})$$

[16] These conditions are available from the authors upon request, and require that $\partial \mathcal{V}^{1,1}_{1,2}(\mathbf{x})/\partial \mathbf{x}^{\mathrm{II}}$ does not equal a specific linear function of $\partial \mathcal{V}^{1,1}_{2,1}(\mathbf{x})/\partial \mathbf{x}^{\mathrm{I}}$.

[17] If one had variation in $\mathbf{x}^j$ across alternatives and unbounded support, letting the observed covariate (say, price) for a given alternative go to infinity would be akin to assuming that one observes agents repeated choices in context j while facing feasible sets that include/exclude each single alternative.

(ii) Assumption 3.5-(II) holds for both agents with preference types $t_i = 1$ and $t_i = 0$, and

$$\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \mathcal{I}_{2,1}) - \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \mathcal{I}_{1,2}) = \mathcal{O}_0(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \mathcal{I}_{2,1}) - \mathcal{O}_0(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \mathcal{I}_{1,2})$$

Then α is identified.

We conclude by observing that given the distributions of the random coefficients, $F(\cdot)$ and $G(\cdot)$, the system of equations defined in Eq. (3.5) ($L \times M$ equations for a given $\mathbf{x}$) is linear in the consideration probabilities across the two types, weighted by their respective shares α and $1-\alpha$. This in turn implies that we have a continuum of $L \times M$ linear equations to pin down $2^{L\times M+1}$ parameters. In general, with sufficient variation in $\mathbf{x}$, these parameters are over-identified, subject to standard non-redundancy assumptions.[18] However, depending on the specific models of preferences assumed, and on the richness of variation in the data observed, it may not be possible to identify some parts of the distribution of consideration sets. Nevertheless, for a specific model, given the data, one can test whether a full rank system of equations results across observed values of $\mathbf{x}$ (see., e.g., Chen and Fang 2019).

4 Model & Data on Choices in Automobile Insurance

4.1 Empirical model

We model agents' choices in two contexts of insurance coverage, where each coverage provides full insurance against covered losses in excess of a deductible chosen by the agent. As the decision maker in our data is a household, we refer to agents as households. We denote the probability of household i experiencing a claim in context j by μ_i^j. For each coverage $j \in \{\mathtt{I}, \mathtt{II}\}$, household i faces a menu of premium-deductible pairs, $\mathcal{M}_i^j \equiv \{(p_i^{\ell j}, \mathtt{d}^{\ell j}) : \mathtt{d}^{\ell j} \in \mathcal{D}^j\}$, where $p_i^{\ell j}$ is the household-specific premium associated with deductible $\mathtt{d}^{\ell j}$ and $\mathcal{D}^j$ is the set of deductible options offered in context j. As we explain in Section 4.2, for each context $j \in \{\mathtt{I}, \mathtt{II}\}$ the ratio of the price of deductible $\mathtt{d}^{\ell j}$ to the price of deductible $\mathtt{d}^{kj}$ is constant across households for all $\mathtt{d}^{\ell j}, \mathtt{d}^{kj} \in \mathcal{D}^j$. As such, in empirical applications from single company data, as in our case, no cross-product within-context price variation can be exploited for identification, a challenge that we address in the paper.

We make assumptions, that are widespread in the literature on property insurance, related to filing claims and their probabilities:

[18]For example, if for type $t_i = 1$ alternative $\mathcal{I}_{\ell,k}$ dominates alternative $\mathcal{I}_{q,r}$, $\mathcal{Q}_1(\{\mathcal{I}_{\ell,k}, \mathcal{I}_{q,r}\})$ cannot be separately identified from $\mathcal{Q}_1(\{\mathcal{I}_{\ell,k}\})$.

Assumption 4.1 (Restrictions Related to Claim Probabilities):

(I) Households disregard the possibility of experiencing more than one claim during the policy period.

(II) Any claim exceeds the highest available deductible; payment of the deductible is the only cost associated with a claim; the household's deductible choice does not influence its claim probability.

We assume that the two types of preferences described in Section 2.2 result from either Expected Utility Theory (EU) or Yaari's (1987) Dual Theory (DT). Within EU, a single-context lottery is evaluated through

$$U_i(\mathcal{L}(\mathtt{d}^{\ell j}, p_i^{\ell j}, \mu_i^j)) \equiv (1-\mu_i^j)u_i(w_i - p_i^{\ell j}) + \mu_i^j u_i(w_i - p_i^{\ell j} - \mathtt{d}^{\ell j}), \tag{4.1}$$

where w_i is the household's wealth and $u_i(\cdot)$ is its Bernoulli utility function, which under Assumption 2.2 is the same for each context. In the EU model, utility is linear in the probabilities and aversion to risk is driven by the shape of the utility function $u_i(\cdot)$.

Yaari's (1987) DT model aims at decoupling the decision maker's attitude towards risk from her attitude towards wealth. Within DT, a single-context lottery is evaluated through

$$U_i(\mathcal{L}(\mathtt{d}^{\ell j}, p_i^{\ell j}, \mu_i^j)) \equiv (1-\Omega_i(\mu_i^j))(w_i - p_i^{\ell j}) + \Omega_i(\mu_i^j)(w_i - p_i^{\ell j} - \mathtt{d}^{\ell j}), \tag{4.2}$$

where $\Omega_i(\cdot)$ is the household's probability distortion function, which under Assumption 2.2 is the same for each context. In the DT model, utility is linear in the outcomes and aversion to risk is driven by the shape of the probability distortion function $\Omega_i(\cdot)$.[19] We remark that in our setting (as well as in many others where subjective beliefs data are not collected and the analysis relies on an often implicit rational expectations assumption), the DT model is indistinguishable from one in which agents' subjective loss probabilities systematically deviate through the $\Omega_i(\cdot)$ function from the objective ones.

To strike a balance between model generality and its empirical tractability, we impose shape restrictions on $u_i(\cdot)$ and $\Omega_i(\cdot)$, respectively. We assume $u_i(\cdot)$ exhibits constant absolute risk aversion (CARA):

Assumption 4.2 (CARA): *$u_i(y) = \frac{1-\exp(-\nu_i y)}{\nu_i}$ for $\nu_i \neq 0$ and $u_i(y) = y$ for $\nu_i = 0$.*

[19] Probability distortions are featured in a number of other models, including prospect theory (Kahneman and Tversky 1979; Tversky and Kahneman 1992), rank-dependent expected utility theory (Quiggin 1982), Gul disappointment aversion theory (Gul 1991), and Kőszegi-Rabin reference-dependent utility theory (Kőszegi and Rabin 2006, 2007).

Assuming CARA has two key virtues. First, $u_i(\cdot)$ is fully characterized by a single parameter: the Arrow-Pratt coefficient of absolute risk aversion, $\nu_i \equiv -u_i''(w_i)/u_i'(w_i)$. Second, ν_i is a constant function of w_i, and hence we need not observe wealth to estimate $u_i(\cdot)$.

To keep the EU model and the DT model on "equal footing," we need $\Omega_i(\cdot)$ to be as parsimonious as $u_i(\cdot)$. This suggests a single-parameter specification. The literature contains many examples, and we run our analysis with the following one due to Prelec (1998):

Assumption 4.3 (Prelec's $\Omega(\cdot)$ function): $\Omega_i(\mu) = \exp(-(-\ln \mu)^{\omega_i})$, $\omega_i > 0$.

We also carry out our analysis using other utility functions for the EU type (one proposed by Cohen and Einav (2007) and one by Barseghyan et al. (2013)) and other probability distortion functions for the DT type (one put forward by Tversky and Kahneman (1992) and one by Barseghyan et al. (2016)). The results confirm the main takeaways reported here, and are available from the authors upon request.[20]

The EU and DT models are true alternative theories of decision making under risk.[21] Neither model is a special case of the other. DT preferences depart from EU preferences in two key ways. First, risk averse behavior is driven by distortions of probabilities for households with DT preferences, but by nonlinear evaluation of wealth for households with EU preferences. Second, narrow bracketing has behavioral implications for households with DT preferences, but not for households with EU preferences. In our framework, where the lotteries are independent across the brackets,[22] the choices of a household with EU preferences and CARA utility are independent of the scope of bracketing (e.g., Rabin and Weizsacker 2009). The well-known reason is the absence of wealth effects with CARA utility. In contrast, the choices of a household with DT preferences are not independent of the scope of bracketing, because of the rank-dependent nature of how probability distortions are applied.

Within context j, the resulting utility function is

$$U_i(\mathcal{L}(\mathtt{d}^{\ell j}, p^{\ell j}, \mu^j)) = \begin{cases} (1-\mu^j)u_i(w - p^{\ell j}) + \mu^j u_i(w - p^{\ell j} - \mathtt{d}^{\ell j}) & \text{if } t_i = 1 \text{ (EU)}, \\ (1-\Omega_i(\mu^j))(w - p^{\ell j}) + \Omega_i(\mu^j)(w - p^{\ell j} - \mathtt{d}^{\ell j}) & \text{if } t_i = 0 \text{ (DT)}. \end{cases} \tag{4.3}$$

While we obtain conditions for nonparametric point identification of $F(\cdot)$ and $G(\cdot)$, for tractability we estimate a fully parametric model via Maximum Likelihood.

Assumption 4.4 (Heterogeneity Restrictions):

[20] Vuong tests comparing the various models confirm the good fit of our preferred specification.

[21] Except when both degenerate into net present value calculations with $\nu_i = 0$ and $\omega_i = 1$.

[22] Independence results from the assumption that claims follow a Poisson distribution, which is imposed in estimating the probability of a claim (see Barseghyan et al. 2013, 2018b).

(I) Conditional on $t_i = 1$, ν_i follows a Beta distribution on $[0, 0.025]$ with parameter vector $(\gamma_{\nu 1}, \gamma_{\nu 2})$ and is independent of $[(\mu_i^j, \bar{p}_i^j), j = I, II]$.

(II) Conditional on $t_i = 0$, ω_i follows a Beta distribution on $[0, 1]$ with parameter vector $(\gamma_{\omega 1}, \gamma_{\omega 2})$ and is independent of $[(\mu_i^j, \bar{p}_i^j), j = I, II]$.

Assumption 4.4 specifies that the distributions of ν and ω are Beta distributions. The main attraction of the Beta distribution is its flexibility (e.g., Ghosal 2001). Its bounded support is a plus given our setting. A lower bound of zero rules out risk-loving preferences and seems appropriate for insurance markets that exist primarily because of risk aversion. Imposing an upper bound enables us to rule out absurd levels of risk aversion. The choice of 0.025 for CARA is conservative both as a theoretical matter and in light of prior empirical estimates in similar settings (e.g., Cohen and Einav 2007; Sydnor 2010; Barseghyan et al. 2011, 2013, 2016). Similarly, for the probability distortion function, the upper bound of 1 insures over-weighting of probabilities; the lower bound of 0 insures that it is a well-behaved function. None of these constraints is binding in our analysis.

We close the empirical model by restricting how $C_i \subseteq \mathcal{D} = \mathcal{D}^{\mathrm{I}} \times \mathcal{D}^{\mathrm{II}}$ is drawn:

Assumption 4.5 ((Broad) Alternative-Specific Consideration): *Household i draws a* consideration set $C_i \subseteq \mathcal{D}$ *s.t.*

$$\Pr(C_i = G) = \prod_{\mathcal{I} \in G} \phi_{\mathcal{I}} \prod_{\tilde{\mathcal{I}} \notin G} (1 - \phi_{\tilde{\mathcal{I}}}), \quad \forall G \subseteq \mathcal{D}$$

where $\phi_{\mathcal{I}} \equiv \Pr(\mathcal{I} \in C_i) = \Pr(\mathcal{I} \in C_i | t_i) \geqslant 0$, $\mathcal{I} \in \mathcal{D}$, *and* $\phi_{\mathcal{I}_{1,1}} = 1$.

Assumption 4.5 strengthens Assumption 2.5 by requiring consideration to be independent of type (in addition to being independent of households' preferences given type). This is *not* needed to establish identification, but we think it prudent to impose it in our application because, as further discussed below, $|\mathcal{D}| = 30$ and allowing for type-dependent consideration would add 60 rather than 30 consideration parameters to the model. Assumption 4.5 also adapts the Alternative-specific Random Consideration (ARC) model first proposed by Manski (1977) and later axiomatized by Manzini and Mariotti (2014), to hold over bundles of insurance deductibles across contexts. Each bundle $\mathcal{I} \in \mathcal{D}$ appears in the consideration set with probability $\phi_{\mathcal{I}}$ independently of other bundles. To avoid empty consideration sets, following Manski (1977), we assume that one bundle is always considered, and further impose that the always-considered bundle is the cheapest one.[23] Once the consideration set is drawn, the agent chooses the best alternative according to its preferences as in Eq. (2.3).

[23] Alternatively, we could assume that if the realized consideration sets is empty, agents choose one of the alternatives in $\mathcal{D}$ uniformly at random. Our estimation results are robust to this modeling assumption.

4.2 Data Description

We obtained the data from a large U.S. property and casualty insurance company. The company offers several lines of insurance, including auto. We focus on deductible choices in two lines of coverage: auto collision and auto comprehensive. Auto collision coverage pays for damage to the insured vehicle caused by a collision with another vehicle or object, without regard to fault. Auto comprehensive coverage pays for damage to the insured vehicle from all other causes (e.g., theft, fire, flood, windstorm, or vandalism), without regard to fault. The sets of offered deductibles are $\mathcal{D}^{\mathtt{I}} = \{\$100, \$200, \$250, \$500, \$1000\}$ in collision and $\mathcal{D}^{\mathtt{II}} = \{\$50, \$100, \$200, \$250, \$500, \$1000\}$ in comprehensive, for a total of 30 bundles of offered coverages in $\mathcal{D} = \mathcal{D}^{\mathtt{I}} \times \mathcal{D}^{\mathtt{II}}$.

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 (this is the same sample used by Barseghyan et al. (2021b)).[24] We observe households' deductible choices in auto collision and auto comprehensive, and the premiums they paid for these coverages. We also observe the household-coverage specific menus of deductible-premium combinations—i.e., the pricing menus—that were available to the households when they made their deductible choices.

Given that we assume that premiums are exogenous to preferences and exhibit substantial variation across households, it is important to understand the sources of such variation. First, we note that the company's rating plan is subject to state regulation and oversight. In particular, the regulations require that the company receive prior approval of its rating plan by the state insurance commissioner, and they prohibit the company and its agents from charging rates that depart from the plan. Second, we describe the procedure the company applies to rate a policy in each line of coverage. Under the plan, within each context j the company determines a household's base price $\bar{p}^j$ according to a coverage-specific rating function, which takes into account the household's coverage-relevant characteristics and any applicable discounts. Using the base price, the company then generates the household's pricing menu $\mathcal{M}^j \equiv \{(p^{\ell j}, \mathtt{d}^{\ell j}) : \mathtt{d}^{\ell j} \in \mathcal{D}^j\}$, which associates a premium $p^{\ell j}$ with each deductible $\mathtt{d}^{\ell j}$ in the coverage-specific set of deductible options $\mathcal{D}^j$, according to a (non-stochastic) coverage-specific multiplication rule, $p^{\ell j} = (g^{\ell j} \cdot \bar{p}^j) + \delta^j$, where $g^{\ell j}$ is increasing in ℓ and strictly greater than zero, $\delta^j > 0$, and $\mathtt{d}^{\ell j}$ is decreasing in ℓ. The multiplicative factors $\{g^{\ell j} : \mathtt{d}^{\ell j} \in \mathcal{D}^j\}$ are known as the deductible factors and δ^j is a small markup known as the expense fee. The deductible factors and the expense fee are coverage specific but household invariant. Naturally, the base prices $\bar{p}^{\mathtt{I}}$ and $\bar{p}^{\mathtt{II}}$ may exhibit substantial correlation due to

[24] As explained in Barseghyan et al. (2021b), 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.

Table 4.1: Collision and Comprehensive Deductible Choices, in %

	Comprehensive					
Collision	\$50	\$100	\$200	\$250	\$500	\$1,000
\$100	0.7	0.2	0	0	0	0
\$200	1.8	1.1	10	0	0.1	0
\$250	0.9	1.3	4.6	5.4	0	0
\$500	1.0	1.3	17.8	6.5	41	0
\$1,000	0	0.1	0.4	0.2	1.9	3.7

common factors in the rating function (this correlation equals 0.74 in our data), highlighting the importance of our weak requirement on variation in $\mathbf{x}$ stated in Assumption 3.4 – which in particular can hold when $\mathbf{x}^{\mathrm{I}}$ and $\mathbf{x}^{\mathrm{II}}$ are strongly correlated (see Figure 3.1).

Table 4.1 displays the deductible choices of the households in our sample. In each coverage, the modal deductible choice is \$500. Interestingly, virtually no household purchases a comprehensive deductible larger than their collision deductible. As we discuss in more detail below, this choice pattern cannot be rationalized by standard discrete choice models under the assumption of full consideration, but can easily be explained once one allows for limited consideration.

The top panel of Table 4.2 shows that base premiums vary dramatically in our sample. The ninety-ninth percentile of the \$500 deductible is more than ten times the corresponding first percentile in each line of coverage. While not reported in the table, here we summarize the pricing menus. The cost of decreasing the deductible from \$500 to \$250 is on average \$56 in collision and \$31 in comprehensive. The saving from increasing the deductible from \$500 to \$1,000 is on average \$42 in collision and \$23 in comprehensive.

The claim probabilities μ_i^j stem from Barseghyan et al. (2018b), who estimated them using coverage-by-coverage Poisson-Gamma Bayesian credibility models applied to a large auxiliary panel of more than one million observations. We treat estimated claim probabilities as if they were observed data. Predicted claim probabilities (summarized in the bottom panel of Table 4.2) exhibit substantial variation: the ninety-ninth percentile claim probability in collision (comprehensive) is 4.3 (12) times higher than the corresponding first percentile. Finally, the correlation between claim probabilities and premiums for the \$500 deductible is 0.38 for collision and 0.15 for comprehensive. Hence, there is independent variation in both (although our identification results only require independent variation in premiums).

Table 4.2: Descriptive statistics for premiums of \$500 deductible and claim probabilities

		Mean	**Std. Dev.**	**Qunatiles**						
				0.01	0.05	0.25	0.50	0.75	0.95	0.99
Premiums										
	Collision	187	104	53	74	117	162	227	383	565
	Comprehensive	117	86	29	41	69	99	141	242	427
Claim probs										
	Collision	0.081	0.026	0.036	0.045	0.062	0.077	0.096	0.128	0.156
	Comprehensive	0.023	0.012	0.005	0.008	0.014	0.021	0.030	0.045	0.062

4.3 Evidence in support of unobserved heterogeneity in C_i

As discussed in, e.g., Barseghyan et al. (2016, 2021b), standard models of risk preferences fail to rationalize some salient data patterns. First, in our data the pricing rule in collision coverage is such that (virtually) no household, regardless of their preference type and random coefficient, should choose the \$200 deductible under full consideration. The reason is that for agents with lower risk aversion (probability distortions) it is dominated by the \$250 deductible, and for agents with higher risk aversion (probability distortions) it is dominated by the \$100 deductible.[25] Limited consideration has no problems explaining such a pattern because it allows for the \$200 deductible to be considered without either \$100 or \$250.

Second, the joint probability mass function of choices across contexts (see Table 4.1) exhibits a striking pattern where virtually none of the 7,736 households purchase a deductible in comprehensive that exceeds the deductible they purchase in collision. Unless prices (and claim probabilities) exhibit strong negative correlation, a feature that does not occur in our data, standard models (e.g., a Mixed Logit with full consideration) under the assumption of context invariant preferences will struggle to replicate this pattern.

A final note pertains to modeling limited consideration as operating at the bundle level, rather than independently across contexts. A model where limited consideration operates independently across contexts may be successful in matching the marginal distribution of choices within each context, but not the joint (see the working paper Barseghyan et al. 2019, Section 7.3.3). The limited consideration model studied in this paper, by operating on the bundles, does have the capacity to match the joint distribution of choices. By doing so, it also resolves the preference stability debate discussed in, e.g., Barseghyan et al. (2011);

[25] An analogous fact can be established even if an i.i.d., type-specific, noise term were added to the utility function in Eq. (4.3) at the coverage level or, more broadly, for any model that abides a notion of generalized dominance formally defined in Barseghyan et al. (2021b).

Table 5.1: Estimated Probability of Considering each Deductibles Pair

	Comprehensive					
Collision	\$50	\$100	\$200	\$250	\$500	\$1,000
\$100	0.05	0.01	0.00	0.00	0.00	0.00
	[0.04 0.06]	[0.01 0.01]	[0.00 0.00]			
\$200	0.12	0.04	0.29	0.00	0.00	0.01
	[0.1 0.13]	[0.04 0.05]	[0.27 0.3]	[0.00 0.00]	[0.00 0.00]	[0.00 0.01]
\$250	0.04	0.03	0.08	0.09	0.00	0.00
	[0.03 0.05]	[0.02 0.03]	[0.07 0.08]	[0.09 0.10]	[0.00 0.00]	
\$500	0.13	0.06	0.46	0.18	0.83	0.00
	[0.11 0.15]	[0.05 0.07]	[0.44 0.47]	[0.18 0.20]	[0.81 0.84]	
\$1000	0.04	0.03	0.18	0.07	0.47	1.00
	[0.02 0.07]	[0.01 0.05]	[0.14 0.22]	[0.05 0.10]	[0.43 0.52]	

Notes: 95% confidence intervals obtained via subsampling in square brackets.

Einav et al. (2012); Barseghyan et al. (2016). This debate is centered around the fact that while households' risk aversion relative to their peers is correlated across lines of coverage, implying that households preferences have a stable component, 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. Under limited consideration, testing for preference stability amounts to asking whether one can find a consideration set and a random coefficient (preference parameter) which jointly rationalize an agent's choice, which is inherently weaker then asking whether one can find preferences that rationalize the agent's choice under full consideration (see, e.g., Barseghyan et al. 2021a).

5 Estimation Results

We begin our discussion of the estimates that we obtain through MLE by focusing on the type of limited consideration that we uncover, and its role in the results one obtains when estimating preferences. Table 5.1 reports the estimated consideration probabilities for each bundle (these are the $\phi_{\mathcal{I}}$ coefficients in Assumption 4.5), along with 95% confidence intervals obtained by subsampling.[26] The estimated model is very far from a full consideration one. Bundles where the collision deductible is strictly lower than the comprehensive one are almost never considered (the probability that the bundle (\$200, \$1000) is considered is 1/100, and all

[26]We use subsampling because the parameter vector is on the boundary of the parameter space.

others are zero).[27] The cheapest bundles, excluding the one where the collision deductible is lower than the comprehensive one, are considered most often (the consideration probabilities for ($\$500, \$500$) and ($\$1000, \$500$) are, respectively, 0.83 and 0.47).[28]

The presence of limited consideration alters inference about preference types and about the distribution of the random coefficient within each type in essentially every possible way. To illustrate these effects, we estimate preferences in a pure random coefficients model under three scenarios for the consideration set formation mechanism: limited consideration as in Assumption 4.5 (our proposed model); *triangular consideration*, where for $\mathcal{I} = (\ell^{\mathrm{I}}, \ell^{\mathrm{II}})$, $\phi_{\mathcal{I}} = 0$ when $\ell^{\mathrm{I}} < \ell^{\mathrm{II}}$ and $\phi_{\mathcal{I}} = 1$ when $\ell^{\mathrm{I}} \geqslant \ell^{\mathrm{II}}$; and full consideration, where $\phi_{\mathcal{I}} = 1$ for all $\mathcal{I} \in \mathcal{D}$. In all cases, we estimate a model where households choose their optimal bundle according to Eq. (2.3) with the utility function in Eq. (4.3).

Figure 5.1 depicts the resulting Prelec distortion function in Assumption 4.3 when ω_i equals the mean, median, 25th and 75th quantile of the distribution $G(\omega)$ estimated in the limited consideration model (left panel), in the triangular consideration model (center panel), and in the full consideration model (right panel), each with a mixture of types. As the figure illustrates, there is substantial variation in the function across these different values of ω, and all functions are substantially far from the 45^o line, indicating substantial over-weighting of small probabilities. Of notice is the fact that the over-weighting is larger in the limited consideration model than in the triangular or in the full consideration model.

Figure 5.2 depicts the cumulative distribution function $F(\cdot)$ in our limited consideration model (left panel), in the triangular consideration model (middle panel), and in the full consideration model (right panel). Each panel depicts $F(\cdot)$ for a model that assumes that all households are of the EU type (blue line), for our model with a mixture of EU and DT types (red line), and, for the mixture model, also the implied cumulative distribution function for the entire population, where the $(1-\alpha)$ share of DT households has $\nu = 0$. The important feature to notice is that in all panels of Figure 5.2, the risk aversion displayed is much higher for the EU households in the mixture model than in the single-type model, and the discrepancy grows from the limited to the triangular to the full consideration model.

In Table 5.2 we analyze the same interplay between consideration and preferences from a different angle. We report the estimated excess willingness to pay (WTP) of households in our sample to avoid a lottery where with probability 10% the household loses \$500 (hence, the total WTP equals \$50 plus the values reported in the table).

[27] Given the choice patterns in the data discussed in Section 4.3, this is not surprising, as MLE sets the consideration probability of never-chosen bundles to zero.

[28] Recall that we assume that ($\$1000, \$1000$) is considered with probability one.

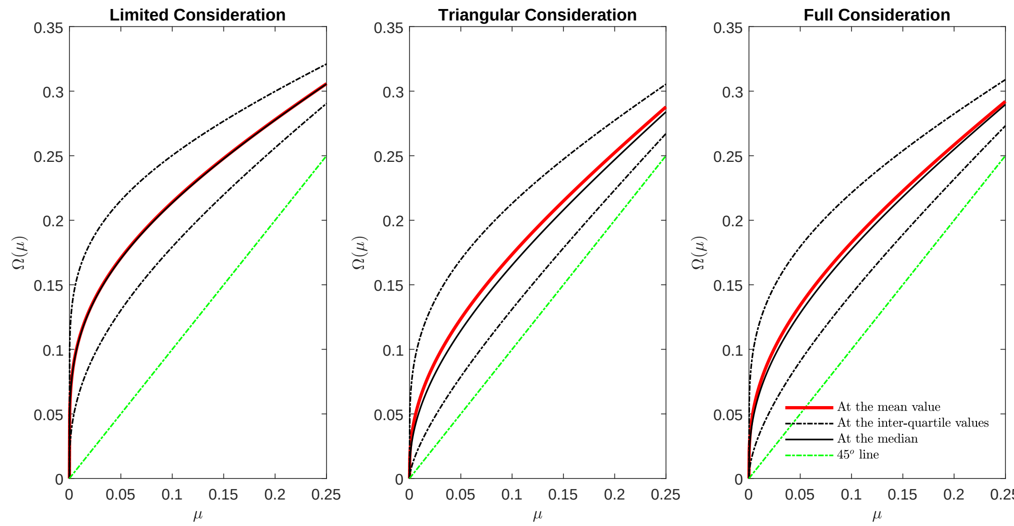

Figure 5.1: The function $\Omega(\mu)$ when ω equals its estimated mean, 25th, 50th, or 75th quantile, in a model with limited (left), triangular (middle), or full (right) consideration.

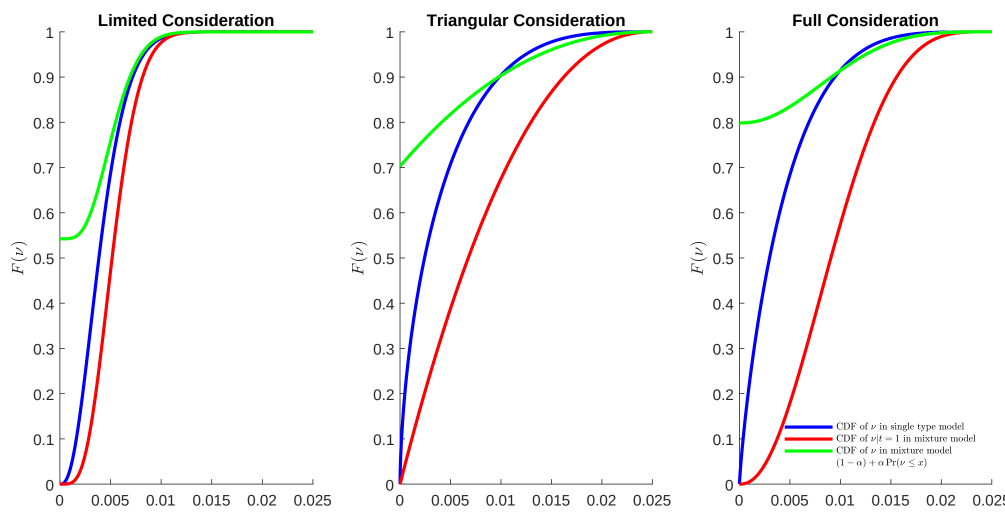

Figure 5.2: Estimated $F(\nu)$ in a model that assumes limited (left), triangular (middle), or full consideration (right).

Table 5.2: Excess willingness to pay to avoid lottery where with probability 10% agent loses $500

	Mean	Median	1^{st} Quar.	3^{rd} Quar.		Mean	Median	1^{st} Quar.	3^{rd} Quar.
Mixture:		**All Population**			**EU share**				
Limited Consideration	82.10	79.95	61.82	100.49	$\alpha = 0.46$				
Lower Triangular	73.27	72.38	42.33	101.64	$\alpha = 0.30$				
Full Consideration	73.49	72.42	54.31	91.72	$\alpha = 0.20$				
		EU type					**DT type**		
Limited Consideration	110.74	104.47	69.29	146.60		57.95	56.73	39.70	75.11
Lower Triangular	155.56	148.83	50.94	255.07		38.47	32.59	15.54	56.49
Full Consideration	194.83	205.93	127.16	267.12		42.88	38.72	21.52	60.59
Single Type:		**All Population EU**					**All Population DT**		
Limited Consideration	81.74	69.72	39.58	113.45		74.01	75.03	51.79	97.15
Lower Triangular	76.19	37.61	8.64	121.29		43.07	37.26	17.60	63.90
Full Consideration	80.87	49.17	15.01	127.35		47.16	42.99	23.38	67.57

Notes. Top panel: excess WTP in our model for the overall population and within each preference type. Bottom panel: excess WTP for a single-type model, where all agents are either EU or DT.

A first feature to notice is that the estimated share of EU types is much higher when the model allows for limited consideration than in models that assume triangular or full consideration (almost a half versus 30% and 20% respectively). The implied degree of aversion to risk changes for households of both preference types, but in opposite directions. The top left panel of Table 5.2 shows that if one disregards limited consideration, one infers that the risk aversion of EU types is much higher (more than 40% according to our metric) than under limited consideration, but the aversion to risk of DT types is about one third lower under full consideration (and similarly for triangular consideration). The cumulative effect of limited consideration in the overall population results in a near 12 percent higher willingness to pay to avoid the simple lottery relative to a model that imposes full consideration.[29]

We conclude by observing that both the full and the triangular consideration model cannot rationalize the choices of a substantial fraction of households in our data and in general deliver a poor fit. Even adding an Extreme Value Type I error term to the utility function in Eq. (4.3) and estimating a Mixed Logit model does not remedy this problem. Indeed, the Mixed Logits do not fit our data well, as shown in Figure 5.3, while our limited consideration model essentially replicates the observed shares.

[29]These results are sensitive to the choice of the simple lottery to benchmark willingness to pay. Changing the stakes will induce a non-linear response by the EU types but a linear one by the DT types. Changing the loss probability will induce a non-linear response by the DT types but a linear one by the EU types.

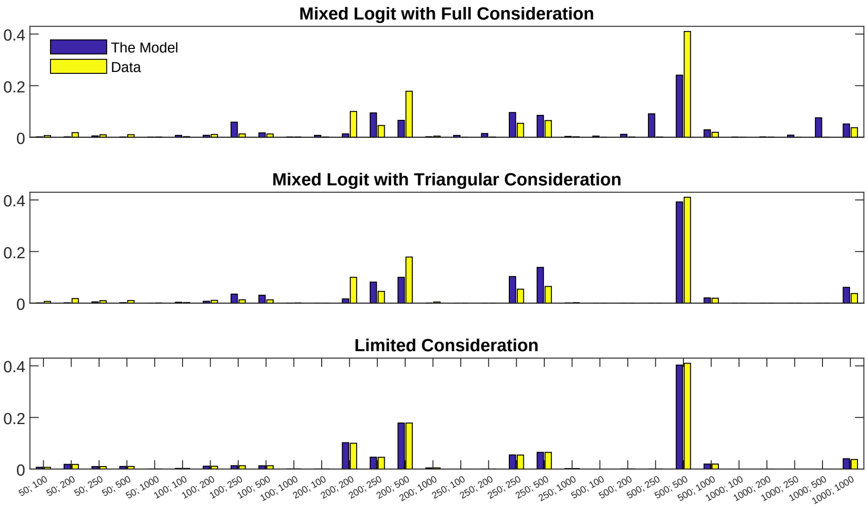

Figure 5.3: Top panel: predicted choice probabilities across deductible bundles for a mixed Logit model with full consideration; middle panel - a mixed Logit model with triangular consideration; bottom panel - our limited consideration model.

6 Implications for Welfare Analysis

In our setting, there are three channels for potential welfare losses. First, limited consideration may prevent agents from choosing their first best. Second, if the probability distortions are capturing a mismatch between subjective and objective beliefs about loss probabilities,[30] agents may not choose their objective first best, even if they consider it. Third, non-expected utility maximizing households (the DT type in our model) may be open to *nudging*, whereby modifications of market features that leave the behavior (and welfare) of EU households mostly unchanged may trigger large changes in behavior (and welfare) of DT households.

We therefore conduct two welfare exercises aimed at assessing the impact of each of these channels on the welfare of households purchasing auto deductible insurance. In the first exercise, we estimate the impact on welfare of all households having full consideration. To do so, we take the preferences estimated using our limited consideration model, predict each household's optimal choice from the entire menu $\mathcal{D}$, and compute each household's utility gain (in certainty equivalent terms). To carry out this exercise, we need to take a stand on how does the household value alternatives. For the EU type, we use their choice utility (also called decision utility), i.e., the CARA utility function (with ν distributed according to our estimate of the distribution F). For the DT types, we report results both for their choice utility, i.e., using the Prelec distortion function in Eq. (4.2) (with ω distributed according to our estimate of the distribution G); and for the case where the probability distortion function is completely removed, so that $\Omega(\mu) = \mu$ and the household values alternatives

[30] See, e.g., the model with imperfect information in Gualdani and Sinha (2023, Example 2).

based on their net present value (NPV). This also allows one to think about the effect of eliminating the mismatch between subjective and objective beliefs about loss probabilities, if this is what the probability distortion function captures.

In the second exercise, we propose a restructuring of the auto insurance market where collision and comprehensive coverage are offered as a single auto insurance product with

$$\begin{aligned}\mathcal{D}^{auto} &= \{100, 200, 250, 500, 1000\}\\ \mu^{auto} &\simeq \mu^{\mathtt{I}} + \mu^{\mathtt{II}}\\ p^{\ell\, auto} &= p^{\mathtt{I}\ell} + p^{\mathtt{II}\ell}\end{aligned}$$

where μ^{auto} is the probability of experiencing a claim in either collision or comprehensive (we disregard the probability that a claim occurs in both contexts within the policy period as this probability is extremely low in our data) and $p^{\ell\, auto}$ is the premium charged for an auto coverage that offers the same deductible in collision and comprehensive when firms operate under perfect competition or if they use a constant markup rule.

Again, we take the preferences estimated using our limited consideration model, predict each household's optimal choice, and compute the household's utility gain/loss (in certainty equivalent terms). However, to carry out the exercise not only do we need to take a stand on how does the household value alternatives, but, importantly, also on how does the household draw its consideration set after the intervention. For the former, we proceed as in our first welfare exercise, and report results where the EU types value alternatives based on their choice utility, and DT types based on both their choice utility and on the alternatives' NPV. For the latter, we report our results under several scenarios, detailed below. This exercise may help inform the debate on the need to "simplify insurance choice," and clarify the role of limited consideration in mediating nudging effects.

Before presenting the results of these two exercises, we explain why EU and DT households may respond differently to an intervention that combines collision and comprehensive into a single coverage. A defining feature of the DT model is that it is non-linear in probabilities. Hence, offering insurance as a bundle or as a single product may have a first order impact on DT households' choices and welfare. To see why, suppose the probability distortion function is strictly sub-additive (as is the case in our estimated model). Then, under the maintained assumption of narrow bracketing (Assumption 2.3), the agent's willingness to pay to avoid a \$500 loss which occurs with a 10 percent chance, is strictly lower than twice their willingness to pay to avoid the same loss with 5 percent chance. Put differently, a single insurance product against two (mutually exclusive) identical losses, instead of a bundle of two products, reduces the degree of over-weighting of loss probabilities. At the same time, combining insurance products into one line of insurance limits choice, and may eliminate the

Table 6.1: Welfare implications of limited consideration and of combining collision and comprehensive into a single coverage (95% confidence intervals in parenthesis)

	Choice Utility	**CARA NPV**	**As a % of Average price for (1000,1000) ($238)**	
All at Full Consideration	30.1	18.2	12.7	7.6
	[28.3 31.9]	[16.0 20.3]	[11.9 13.4]	[6.7 8.5]
Bundled Auto Insurance:				
Worst Case Consideration	-3.2	-18.7	-1.3	-7.9
	[-8.0 1.5]	[-22.2 -15.2]	[-3.4 0.6]	[-9.3 -6.4]
Middle Case Consideration	48.1	26.5	20.2	11.1
	[45.9 50.3]	[25.4 27.7]	[19.3 21.1]	[10.7 11.7]
All at Full Consideration	52.3	29.6	22	12.5
	[50.0 54.7]	[28.2 31.1]	[21 23]	[11.8 13.1]

first best alternative. Ceteris paribus, for a fully rational agent making choices according to the EU model, this can only be welfare reducing. Interestingly, there are examples of insurance products that are indeed sold both as a single coverage and as a bundle, such as single limit liability coverage versus bodily injury and property damage in auto insurance.

In summary, our first welfare exercise addresses the question: what is the (average) welfare cost associated with limited consideration? Our second welfare exercise addresses the question: what are the welfare implications of combining collision and comprehensive into a single product, and how does the presence of limited consideration alter these implications?

The top panel of Table 6.1 reports our estimates of the welfare losses due to limited consideration. Using the choice utility for each preference type, the welfare losses are about $30, or 12.7% of the average price of the cheapest bundle. The effect is smaller ($18 or 7.6%) if for DT types we use the alternatives' NPV as their value (i.e., we shut down the probability distortion). This is expected, since all utilities and utility differences decrease.

The bottom panel of Table 6.1 reports estimated welfare changes associated with combining collision and comprehensive insurance into a single product. We carry out the exercise for three different ways in which consideration sets may be drawn after the market intervention. In the worst case scenario, in the sense that consideration is lowest, the probability that deductible $\mathtt{d}$ is considered equals the estimated consideration probability for bundle $(\mathtt{d}, \mathtt{d}), \mathtt{d} \in \mathcal{D}^{auto}$. In this case, the impact of the intervention is negative, although the magnitude of the effect depends substantially on how the welfare of DT types is evaluated. This is because under choice utility, following the intervention, DT types overweight the overall loss probability to a lesser degree than they did with separate coverages, and this effect attenuates substantially the welfare reduction from not being able to choose from a larger menu.

On the other hand, when welfare of DT types is evaluated according to NPV, although the overweighting of loss probabilities affects choice, it does not enter the welfare calculations.

Under full consideration, the best case scenario, the welfare gains for both evaluation approaches are positive and large. Relative to the worst case scenario, this is, of course, expected. What is more interesting is that the welfare gains are higher than those obtained in the counterfactual of full consideration that maintains the status-quo separation between collision and comprehensive insurance. This is because under full consideration, the EU types are worse off when the collision and comprehensive are combined into a single product (for them, the choice set is being reduced without any associated benefit); however, the DT types, despite facing a smaller choice set, benefit from such a reduction because in making choices they overweight losses by a smaller degree. The latter effect dominates, more so when welfare is computed based on choice utility rather than on NPV.

For completeness we also report welfare changes for a case that we label "middle consideration," in which each deductible in the combined single coverage is considered with a probability equal to the sum of the probability that it is considered either as collision or comprehensive deductible (or with probability one if the sum exceeds one). The results are reported in the middle row of the bottom panel of Table 6.1. Even with this intermediate consideration level, the welfare gains are substantial.

Based on these welfare exercises, we argue that the interplay between features of the decision making process at the utility evaluation level and of the consideration mechanism cannot be ignored when analyzing possible market interventions. In the second welfare exercise carried out above, reducing the feasible set may lead to unambiguous welfare gains, provided consideration increases. However, if consideration does not increase, the same intervention can lead to welfare losses that exceed the gains stemming from nudging the non-expected utility maximizers in the population.

7 Discussion

This paper provides semi-nonparametric point identification results for a model of discrete choice under risk that allows for unobserved heterogeneity in preference types, unobserved heterogeneity within each type, and unobserved heterogeneity in consideration sets, and assumes a data environment that is common when data come from a single insurance company. The feasible alternatives are characterized by a covariate (the deductible) that is fixed across agents. The other covariates that are relevant to the decision problem –premia and claim probabilities– vary stochastically only across agents but not across alternatives (premia of different alternatives are a fixed affine function of each other, and claim probabilities do not vary across alternatives). This limited amount of variation in covariates, relative to the richness of unobserved heterogeneity allowed for, creates challenges for identification analysis. The paper shows how these challenges can be overcome, provided that the researcher observes agents making choices in at least two contexts, and that premia display independent variation across the contexts.

The choice environment that we study in this paper is similar to that studied in Barseghyan et al. (2021b). They offer a comprehensive analysis of the implications of the Spence-Mirlees single crossing property for semi-nonparametric identification of a model of discrete choice under risk that features a single preference type and unobserved heterogeneity in consideration sets. They also illustrate the tradeoff between the common exclusion restrictions and the restrictions on consideration set formation required for semi-nonparametric point identification. Their work is the closest to ours. However, in our model consideration sets are formed at the bundle level (i.e., across contexts), and hence the single crossing property that both Barseghyan et al. (2021b) and we assume to hold within a context, may not necessarily hold across tuples of alternatives. This is because bundles may not be monotonically ranked (with respect to preference parameters) against each other. Hence, the results in Barseghyan et al. (2021b) do not apply and in this paper we develop a new approach to obtain point identification of the distribution of preferences, of the shares of preferences types, and of features of the distribution of consideration sets given type.[31] In Barseghyan and Molinari (2023) we show that in a richer data environment where the researcher observes a characteristic for each alternative that displays independent variation both across agents and across alternatives, and that affects utility but not consideration, semi-nonparametric point identification holds for a flexible pure random coefficients model with unrestricted dependence between the random coefficients and consideration set formation mechanism.

The challenges posed to identification of discrete choice models by unobserved heterogeneity in consideration sets have long been recognized (e.g., Manski 1977).[32] It is not uncommon for the problem to be ignored, as a textbook assumption is that agents pick an alternative to maximize their utility over the entire feasible set. When heterogeneity in consideration sets is allowed for, point identification of the model often relies on the availability of auxiliary information about the composition or distribution of agents' consideration sets. Examples include Draganska and Klapper (2011); De los Santos et al. (2012); Conlon and Mortimer

[31] As we allow for multiple preference types, our analysis extends that of Barseghyan et al. (2021b) even in the simplified framework where consideration is independent across contexts.

[32] Many important papers in the theory literature—including papers on revealed preference analysis under limited attention, limited consideration, rational inattention, and other forms of bounded rationality that manifest in unobserved heterogeneity in consideration sets—also grapple with the identification problem (e.g., Masatlioglu et al. 2012; Manzini and Mariotti 2014; Caplin and Dean 2015; Lleras et al. 2017; Cattaneo et al. 2020). However, these papers generally assume rich datasets—e.g., observed choices from every possible subset of the feasible set—that often are not available in applied work, especially outside of the laboratory.

(2013); Honka et al. (2017); Honka and Chintagunta (2017). Alternatively, researchers bring to bear two-way exclusion restrictions, whereby certain variables impact consideration but not preferences and vice versa. Examples include Goeree (2008); van Nierop et al. (2010); Gaynor et al. (2016); Hortaçsu et al. (2017); Heiss et al. (2016). A third approach relies primarily on restrictions to the consideration set formation process. Recent examples include Abaluck and Adams (2020); Crawford et al. (2021); Dardanoni et al. (2020); Lu (2022).

When such assumptions may not be credible and one does not have access to auxiliary data or valid exclusion restrictions, Barseghyan et al. (2021a) provide a method to obtain informative sharp identification regions for the parameters of discrete choice models, even when preferences and consideration sets may depend on each other, under the assumption that agents' consideration sets include at least two alternatives. Cattaneo et al. (2020, 2021) provide revealed preference theory, testable implications, and partial identification results for preference orderings and attention frequency, in very general models of limited consideration, albeit with no heterogeneity in preferences, under the assumption that one observes agents repeated choices (in a single context) while facing varying choice sets.

We next discuss the relation of our work with the extant literature on estimation of risk preferences. Unobserved heterogeneity within a single preference type has received much attention in the recent literature that estimates risk preferences using agent-level market choice data (for a review, see Barseghyan et al. 2018a). Multiple preference types are a focus of the literature that estimates risk preferences using experimental data (e.g., Bruhin et al. (2010); Conte et al. (2011); Harrison et al. (2010)), although preferences are homogeneous within each type, at most conditioning on some observed demographic characteristics. Yet, a large literature in experimental economics reports substantial heterogeneity in risk preferences within type; see, e.g., Choi et al. (2007) and references therein. And it documents that while some people exhibit behavior consistent with standard expected utility (EU) theory, others exhibit behavior that systematically deviates from it (e.g., Starmer 2000). In response to the latter finding, non-expected utility (NEU) theories of decision making under risk and uncertainty have emerged. These models feature non-linearity in how agents evaluate wealth and/or non-linearity in how agents evaluate risk.[33] As such, most NEU models embed the EU model as a special case, and hence it is expected that NEU models can better explain observed behavior, see, e.g., Barseghyan et al. (2013).

In contrast, our mixture model allows both for a mixture of non-nested preference types, with unobserved heterogeneity within type, and for unobserved heterogeneity in consideration sets. Our empirical findings highlight the importance of using a model that allows for such rich heterogeneity. As noted in Barseghyan et al. (2021a), the behavioral literature

[33] Additionally, NEU models in prospect theory, e.g., Kahneman and Tversky (1979) and Tversky and Kahneman (1992), feature reference-dependent utility as in loss aversion.

aimed at explaining choices under risk, has focused on developing and estimating models that depart from expected utility theory in their specification of *how* agents evaluate risky alternatives. Our findings here provide evidence in addition to what we reported in Barseghyan et al. (2021a,b) on the importance of developing models that differ in their specification of *which* alternatives agents evaluate, especially when the goal of the analysis is to evaluate counterfactual policies.

A Appendix: Proof of Theorem 3.1 and Corollary 3.1

Proof of Theorem 3.1. Fix $\nu \in [\nu^*, \nu^{**}]$ and the corresponding $\mathbf{X}^1(\nu)$. By Assumption 3.4, $\mathbf{X}^1(\nu)$ is non-empty and there is an ϵ-ball around it of positive density. By Definition 3.1, for any $(\mathbf{x}^{\mathrm{I}}, \mathbf{x}^{\mathrm{II}}) \in \mathbf{X}^1(\nu)$, $\mathcal{V}^{1,1}_{2,1}(\mathbf{x}^{\mathrm{I}}) = \mathcal{V}^{1,1}_{1,2}(\mathbf{x}^{\mathrm{II}}) = \nu$. Along with Assumption 3.3, this implies that there are vectors $\mathbf{x}' = (\mathbf{x}^{\mathrm{I}\prime}, \mathbf{x}^{\mathrm{II}\prime}) \in \mathsf{B}_\epsilon(\mathbf{X}^1(\nu))$ and $\mathbf{x}'' = (\mathbf{x}^{\mathrm{I}\prime\prime}, \mathbf{x}^{\mathrm{II}\prime\prime}) \in \mathsf{B}_\epsilon(\mathbf{X}^1(\nu))$ such that $\nu = \mathcal{V}^{1,1}_{2,1}(\mathbf{x}^{\mathrm{I}\prime}) < \mathcal{V}^{1,1}_{1,2}(\mathbf{x}^{\mathrm{II}\prime})$ and $\mathcal{V}^{1,1}_{2,1}(\mathbf{x}^{\mathrm{I}\prime\prime}) > \mathcal{V}^{1,1}_{1,2}(\mathbf{x}^{\mathrm{II}\prime\prime})$. We claim that

$$\lim_{\mathbf{x}', \mathbf{x}'' \to \mathbf{x}} \left(\frac{\partial \Pr(\mathcal{I}^*_i = \mathcal{I}_{1,1} | \mathbf{x}')}{\partial \mathbf{x}^{\mathrm{I}}} - \frac{\partial \Pr(\mathcal{I}^*_i = \mathcal{I}_{1,1} | \mathbf{x}'')}{\partial \mathbf{x}^{\mathrm{I}}} \right) = \alpha f(\nu) \cdot h_1(\mathbf{x}, \mathcal{O}_1), \tag{A.1}$$

where $h_1(\mathbf{x}, \mathcal{O}_1)$ is a function of $\mathbf{x}$ and of the consideration probabilities given by:

$$\begin{aligned} h_1(\mathbf{x}, \mathcal{O}_1) &= \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,1}\}; \varnothing) \frac{\partial \mathcal{V}^{1,1}_{2,1}(\mathbf{x})}{\partial \mathbf{x}^{\mathrm{I}}} + \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \mathcal{I}_{2,1}) \frac{\partial \mathcal{V}^{1,1}_{2,2}(\mathbf{x})}{\partial \mathbf{x}^{\mathrm{I}}} \\ &\quad - \left(\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \mathcal{I}_{1,2}) \frac{\partial \mathcal{V}^{1,1}_{2,2}(\mathbf{x})}{\partial \mathbf{x}^{\mathrm{I}}} + \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,1}\}; \{\mathcal{I}_{2,2}, \mathcal{I}_{1,2}\}) \frac{\partial \mathcal{V}^{1,1}_{2,1}(\mathbf{x})}{\partial \mathbf{x}^{\mathrm{I}}} \right) \end{aligned} \tag{A.2}$$

Under Assumption 3.5-(I), Eq. (A.2) simplifies to[34]

$$h_1(\mathbf{x}, \mathcal{O}_1) = \Big(\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,1}\}; \varnothing) - \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,1}\}; \{\mathcal{I}_{2,2}, \mathcal{I}_{1,2}\}) \Big) \frac{\partial \mathcal{V}^{1,1}_{2,1}(\mathbf{x})}{\partial \mathbf{x}^{\mathrm{I}}} \neq 0 \tag{A.3}$$

Under Assumption 3.5-(II), Eq. (A.2) simplifies to

$$h_1(\mathbf{x}, \mathcal{O}_1) = \Big(\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \mathcal{I}_{2,1}) - \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \mathcal{I}_{1,2}) \Big) \frac{\partial \mathcal{V}^{1,1}_{2,2}(\mathbf{x})}{\partial \mathbf{x}^{\mathrm{I}}} \neq 0 \tag{A.4}$$

[34] These derivations are based on repeated use of facts such as

$$\begin{aligned} \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \varnothing) &= \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}, \mathcal{I}_{2,1}\}; \varnothing) + \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \{\mathtt{I}_{2,1}\}) \\ &= \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}, \mathcal{I}_{1,2}\}; \varnothing) + \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \{\mathtt{I}_{1,2}\}) \end{aligned}$$

To derive the expression for $h_1(\mathbf{x}, \mathcal{O}_1)$ in Eq. (A.2), we return to Eq. (3.6), which states

$$\frac{\partial \Pr(\mathcal{I}_i^* = \mathcal{I}_{1,1}|\mathbf{x})}{\partial \mathbf{x}^{\mathrm{I}}} = \alpha \sum_{(k,r)\neq(1,1)} \mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{k,r}\}; \mathbb{B}(\mathcal{I}_{1,1}, \mathbf{x}; \mathcal{V}_{k,r}^{1,1}))f(\mathcal{V}_{k,r}^{1,1})\frac{\partial \mathcal{V}_{k,r}^{1,1}}{\partial \mathbf{x}^{\mathrm{I}}}$$
$$+ (1-\alpha) \sum_{(k,r)\neq(1,1)} \mathcal{O}_0(\{\mathcal{I}_{1,1}, \mathcal{I}_{k,r}\}; \mathbb{B}(\mathcal{I}_{1,1}, \mathbf{x}; \mathcal{W}_{k,r}^{1,1}))g(\mathcal{W}_{k,r}^{1,1})\frac{\partial \mathcal{W}_{k,r}^{1,1}}{\partial \mathbf{x}^{\mathrm{I}}}$$

Under Assumptions 3.3-(II) and 3.4, when $\mathbf{x}'$ and $\mathbf{x}''$ are sufficiently close to $\mathbf{x}$, the relative order of the cutoffs for type $t_i = 0$ preferences, $\mathcal{W}_{k,r}^{1,1}$, does not change. For type $t_i = 1$ preferences, it changes only for the cutoffs involving bundles $\{\mathcal{I}_{2,1}, \mathcal{I}_{1,2}, \mathcal{I}_{2,2}\}$. Hence,

$$\frac{\partial \Pr(\mathcal{I}_i^* = \mathcal{I}_{1,1}|\mathbf{x}')}{\partial \mathbf{x}^{\mathrm{I}}} = \alpha\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,1}\}; \varnothing)f(\mathcal{V}_{2,1}^{1,1})\frac{\partial \mathcal{V}_{2,1}^{1,1}(\mathbf{x}')}{\partial \mathbf{x}^{\mathrm{I}}}$$
$$+ \alpha\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \mathcal{I}_{2,1})f(\mathcal{V}_{2,2}^{1,1})\frac{\partial \mathcal{V}_{2,2}^{1,1}(\mathbf{x}')}{\partial \mathbf{x}^{\mathrm{I}}} + R(\mathbf{x}') \tag{A.5}$$

$$\frac{\partial \Pr(\mathcal{I}_i^* = \mathcal{I}_{1,1}|\mathbf{x}'')}{\partial \mathbf{x}^{\mathrm{I}}} = \alpha\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,2}\}; \mathcal{I}_{1,2})f(\mathcal{V}_{2,2}^{1,1})\frac{\partial \mathcal{V}_{2,2}^{1,1}(\mathbf{x}'')}{\partial \mathbf{x}^{\mathrm{I}}}$$
$$+ \alpha\mathcal{O}_1(\{\mathcal{I}_{1,1}, \mathcal{I}_{2,1}\}; \{\mathcal{I}_{1,2}, \mathcal{I}_{2,2}\})f(\mathcal{V}_{2,1}^{1,1})\frac{\partial \mathcal{V}_{2,1}^{1,1}(\mathbf{x}'')}{\partial \mathbf{x}^{\mathrm{I}}} + R(\mathbf{x}'') \tag{A.6}$$

where $R(\cdot)$ is a collection of terms that are continuous functions of their argument around $\mathbf{x}$. Consequently, in the limit where both $\mathbf{x}', \mathbf{x}''$ tend to $\mathbf{x}$, $R(\mathbf{x}')$ and $R(\mathbf{x}'')$ are identical to each other, and Eq. (A.2) follows by subtracting Eq. (A.6) from Eq. (A.5).

Next, observe that $h_1(\mathbf{x}, \mathcal{O}_1)$ equals a non-zero constant multiplied with $\frac{\partial \mathcal{V}_{2,1}^{1,1}(\mathbf{x})}{\partial \mathbf{x}^{\mathrm{I}}}$ (or $\frac{\partial \mathcal{V}_{2,2}^{1,1}(\mathbf{x})}{\partial \mathbf{x}^{\mathrm{I}}}$). The latter term is a known function of the data and is different from zero. Consequently, the density function of the random coefficient for type $t_i = 1$ agents evaluated at ν, $f(\nu)$, is identified up to a non-zero constant (α multiplied with a non-zero linear combination of consideration probabilities that does not depend on ν). If $[\nu^*, \nu^{**}] = [0, \bar{\nu}]$, then using that $f(\nu)$ integrates to one over its support identifies $\alpha \cdot h_1(\mathbf{x}, \mathcal{O}_1)$, and consequently the entire function $f(\cdot)$. The same argument applies to establish identification of $g(\cdot)$. $\square$

Proof of Corollary 3.1. Once $\alpha f(\nu) \cdot h_1(\mathbf{x}, \mathcal{O}_1)$ and $(1-\alpha)g(\omega) \cdot h_0(\mathbf{x}, \mathcal{O}_0)$ are identified, so are $f(\nu)$ and $g(\omega)$ provided there is large support. Under Assumption 3.5, $h_1(\mathbf{x}, \mathcal{O}_1)$ and $h_0(\mathbf{x}, \mathcal{O}_0)$ can be decomposed into a product of two terms, one known and another entirely dependent on consideration, see Eqs. (A.3)-(A.4). Moreover, these terms will be identical, as long as $\mathcal{O}_1(\cdot;\cdot) = \mathcal{O}_0(\cdot;\cdot)$ for all relevant combinations of $\{\mathcal{I}_{1,1}, \mathcal{I}_{1,2}, \mathcal{I}_{2,1}, \mathcal{I}_{2,2}\}$ in part (i) (respectively, part (ii)) of the assumptions stated in Corollary 3.1. Hence, the ratio of $\alpha f(\nu) \cdot h_1(\mathbf{x}, \mathcal{O}_1)$ and $(1-\alpha)g(\omega) \cdot h_0(\mathbf{x}, \mathcal{O}_0)$ identifies α. $\square$

References

Abaluck, J. and A. Adams (2020): "What do consumers consider before they choose? Identification from asymmetric demand responses," *Quarterly Journal of Economics*, forthcoming.

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

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

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

Barseghyan, L., M. Coughlin, F. Molinari, and J. C. Teitelbaum (2021a): "Heterogeneous Choice Sets and Preferences," *Econometrica*, 89, 2015–2048.

Barseghyan, L. and F. Molinari (2023): "Discrete Choice Models with Heterogeneous Preferences and Consideration," Tech. rep., mimeo.

Barseghyan, L., F. Molinari, T. O'Donoghue, and J. C. Teitelbaum (2013): "The Nature of Risk Preferences: Evidence from Insurance Choices," *American Economic Review*, 103, 2499–2529.

——— (2018a): "Estimating Risk Preferences in the Field," *Journal of Economic Literature*, 56, 501–564.

Barseghyan, L., F. Molinari, and J. C. Teitelbaum (2016): "Inference Under Stability of Risk Preferences," *Quantitative Economics*, 7, 367–409.

Barseghyan, L., F. Molinari, and M. Thirkettle (2019): "Discrete Choice Under Risk With Limited Consideration," *CEMMAP Working Paper*, CWP08/19, available at http://dx.medra.org/10.1920/wp.cem.2019.0819.

——— (2021b): "Discrete Choice under Risk with Limited Consideration," *American Economic Review*, 111, 1972?2006.

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

Barseghyan, L., J. C. Teitelbaum, and L. Xu (2018b): "Different Contexts, Different Risk Preferences?" SSRN Working Paper No. 3200657.

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

Bruhin, A., H. Fehr-Duda, and T. Epper (2010): “Risk and Rationality: Uncovering Heterogeneity in Probability Distortion,” *Econometrica*, 78, 1375–1412.

Caplin, A. and M. Dean (2015): “Revealed Preference, Rational Inattention, and Costly Information Acquisition,” *American Economic Review*, 105, 2183–2203.

Cattaneo, M. D., P. Cheung, X. Ma, and Y. Masatlioglu (2021): “Attention Overload,” Tech. rep., available at https://arxiv.org/abs/2110.10650.

Cattaneo, M. D., X. Ma, Y. Masatlioglu, and E. Suleymanov (2020): “A Random Attention Model,” *Journal of Political Economy*, 128, 2796–2836.

Chen, Q. and Z. Fang (2019): “Improved inference on the rank of a matrix,” *Quantitative Economics*, 10, 1787–1824.

Chiappori, P.-A., B. Salanié, F. Salanié, and A. Gandhi (2019): “From aggregate betting data to individual risk preferences,” *Econometrica*, 87, 1–36.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007): “Consistency and Heterogeneity of Individual Behavior under Uncertainty,” *American Economic Review*, 97, 1921–1938.

Cohen, A. and L. Einav (2007): “Estimating Risk Preferences from Deductible Choice,” *American Economic Review*, 97, 745–788.

Conlon, C. T. and J. H. Mortimer (2013): “Demand Estimation under Incomplete Product Availability,” *American Economic Journal: Microeconomics*, 5, 1–30.

Conte, A., J. D. Hey, and P. G. Moffatt (2011): “Mixture Models of Choice under Risk,” *Journal of Econometrics*, 162, 79–88.

Crawford, G. S., R. Griffith, and A. Iaria (2021): “A Survey of Preference Estimation with Unobserved Choice Set Heterogeneity,” *Journal of Econometrics*, 222, 4–43.

Dardanoni, V., P. Manzini, M. Mariotti, and C. J. Tyson (2020): “Inferring Cognitive Heterogeneity from Aggregate Choices,” *Econometrica*, 88, 1269–1296.

De los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2012): “Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior,” *American Economic Review*, 102, 2955–2980.

Draganska, M. and D. Klapper (2011): “Choice Set Heterogeneity and the Role of Advertising: An Analysis with Micro and Macro Data,” *Journal of Marketing Research*, 48, 653–669.

Einav, L., A. Finkelstein, I. Pascu, and M. R. Cullen (2012): “How General are Risk Preferences? Choice under Uncertainty in Different Domains,” *American Economic Review*, 102, 2606–2638.

GAYNOR, M., C. PROPPER, AND S. SEILER (2016): "Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service," *American Economic Review*, 106, 3521–57.

GHOSAL, S. (2001): "Convergence Rates for Density Estimation with Bernstein Polynomials," *Annals of Statistics*, 29, 1264–1280.

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

GUALDANI, C. AND S. SINHA (2023): "Identification in Discrete Choice Models with Imperfect Information," Tech. rep., available at https://arxiv.org/abs/1911.04529.

GUL, F. (1991): "A Theory of Disappointment Aversion," *Econometrica*, 59, 667–686.

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

HARRISON, G. W., S. J. HUMPHREY, AND A. VERSCHOOR (2010): "Choice under Uncertainty: Evidence from Ethiopia, India and Uganda," *The Economic Journal*, 120, 80–104.

HEISS, F., D. MCFADDEN, J. WINTER, A. WUPPERMANN, AND B. ZHOU (2016): "Inattention and Switching Costs as Sources of Inertia in Medicare Part D," Working Paper 22765, National Bureau of Economic Research.

HONKA, E. AND P. CHINTAGUNTA (2017): "Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry," *Marketing Science*, 36, 21–42.

HONKA, E., A. HORTAÇSU, AND M. A. VITORINO (2017): "Advertising, Consumer Awareness, and Choice: Evidence from the U.S. Banking Industry," *RAND Journal of Economics*, 48, 611–646.

HORTAÇSU, A., S. A. MADANIZADEH, AND S. L. PULLER (2017): "Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market," *American Economic Journal: Economic Policy*, 9, 192–226.

KAHNEMAN, D. (2003): "Maps of Bounded Rationality: Psychology for Behavioral Economics," *American Economic Review*, 93, 1449–1475.

KAHNEMAN, D. AND A. TVERSKY (1979): "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47, 263–291.

KŐSZEGI, B. AND M. RABIN (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121, 1133–1166.

——— (2007): "Reference-Dependent Risk Attitudes," *American Economic Review*, 97, 1047–1073.

LLERAS, J. S., Y. MASATLIOGLU, D. NAKAJIMA, AND E. Y. OZBAY (2017): "When More is Less: Limited Consideration," *Journal of Economic Theory*, 170, 70–85.

LU, Z. (2022): "Estimating multinomial choice models with unobserved choice sets," *Journal of Econometrics*, 226, 368–398.

MANSKI, C. F. (1977): "The Structure of Random Utility Models," *Theory and Decision*, 8, 229–254.

MANZINI, P. AND M. MARIOTTI (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82, 1153–1176.

MAS-COLELL, A., M. D. WHINSTON, AND J. R. GREEN (1995): *Microeconomic Theory*, New York: Oxford University Press.

MASATLIOGLU, Y., D. NAKAJIMA, AND E. Y. OZBAY (2012): "Revealed Attention," *American Economic Review*, 102, 2183–2205.

MCFADDEN, D. L. (1974): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka, New York: Academic Press, 105–142.

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

PRELEC, D. (1998): "The Probability Weighting Function," *Econometrica*, 66, 497–527.

QUIGGIN, J. (1982): "A Theory of Anticipated Utility," *Journal of Economic Behavior and Organization*, 3, 323–343.

RABIN, M. AND G. WEIZSACKER (2009): "Narrow Bracketing and Dominated Choices," *American Economic Review*, 99, 1508–43.

READ, D., G. LOEWENSTEIN, AND M. RABIN (1999): "Choice Bracketing," *Journal of Risk and Uncertainty*, 19, 171–197.

SPENCE, A. M. (1974): "Market signaling: Informational transfer in hiring and related screening processes," *Harvard University Press*, 143.

STARMER, C. (2000): "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," *Journal of Economic Literature*, 38, 332–382.

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

TVERSKY, A. AND D. KAHNEMAN (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

VAN NIEROP, E., B. BRONNENBERG, R. PAAP, M. WEDEL, AND P. H. FRANSES (2010): "Retrieving Unobserved Consideration Sets from Household Panel Data," *Journal of Marketing Research*, 47, 63–74.

WILCOX, N. T. (2008): *Stochastic models for binary discrete choice under risk: a critical primer and econometric comparison*, Emerald Group Publishing Limited, 197–292.

YAARI, M. E. (1987): "The Dual Theory of Choice under Risk," *Econometrica*, 55, 95–115.