WORKING PAPERS

N° 1049

November 2019

"Identification and inference in discrete choice models with imperfect information"

Cristina Gualdani and Shruti Sinha

Toulouse School of Economics

Identification and inference in discrete choice models with imperfect information*

Cristina Gualdani† Shruti Sinha‡

October 2019

Abstract

In this paper we study identification and inference of preference parameters in a single-agent, static, discrete choice model where the decision maker may face attentional limits precluding her to exhaustively process information about the payoffs of the available alternatives. By leveraging on the notion of one-player Bayesian Correlated Equilibrium in Bergemann and Morris (2016), we provide a tractable characterisation of the sharp identified set and discuss inference under minimal assumptions on the amount of information processed by the decision maker and under no assumptions on the rule with which the decision maker resolves ties. Simulations reveal that the obtained bounds on the preference parameters can be tight in several settings of empirical interest.

Keywords: Discrete choice model, Bayesian Persuasion, Bayesian Correlated Equilibrium, Incomplete Information, Partial Identification, Moment Inequalities.

*We acknowledge funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010.

†Email: cristina.gualdani@tse-fr.eu, Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France.

‡Email: shruti.sinha@tse-fr.eu, Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France.

1 Introduction

Attentional limits have long been recognized to play a critical role in decision problems by precluding agents' ability to exhaustively process information about the value of every possible alternative (e.g., Simon, 1955; 1959; Kaheman, 1973; Sims, 1998; 2003; 2006; Lacetera, Pope, and Sydnor, 2012; Des Los Santos, Hortaçsu, and Wildenbeest, 2012; Matějka and McKay, 2015; Caplin, Dean, and Leahy, 2018). In this paper we offer a robust and tractable method to explore the empirical content of attentional limits in decision problems. In particular, we study identification and inference of preferences in a single-agent, static, discrete choice model where the decision maker (hereafter, DM) may face attentional limits hampering her capacity to learn about the payoff generated by each of the available alternatives.

More formally, we consider a static setting where the DM chooses an alternative from a discrete feasible set. The payoff generated by the chosen alternative depends on the state of the world.[1] While in standard discrete choice models in the tradition of McFadden (1974) the DM is assumed to observe the state of the world before choosing an alternative, here the DM chooses an alternative possibly without being fully aware about the state of the world. Instead, the DM has a prior about the state of the world. Moreover, before choosing an alternative, the DM has the opportunity to investigate further the state of the world by processing additional information (hereafter, *information structure*). Such information structure takes the form of a noisy signal of the state of the world. Admissible information structures range from full revelation of the state of the world to no information whatsoever, depending on the DM's attentional limits. The DM uses the acquired information structure to update her prior about the state of the world and obtain a posterior. Lastly, the DM chooses an alternative maximising the expected payoff, where the expectation is computed via the posterior. If there is more than one maximising alternative, then the DM picks one of them according to some rule (hereafter, *selection rule*).

Our objective is to study identification and inference of the parameters governing the DM's preferences by using the empirical choice probabilities. In developing identification and inference arguments, we remain agnostic about agents' information structures and selection rules. This is because information structures and selection rules are unobserved by the researcher and potentially heterogenous across agents. In particular, heterogeneity of information structures comes from the fact that different agents could have different attentional limits and, thus, sustain different costs to gather information on the state of the world. Heterogeneity of selection rules comes from the fact that different agents could resolve in different ways indifferences between alternatives with the same expected payoff. Remaining agnostic about information structures and selection rules implies that empirical results and counterfactual outcomes obtained via our methodology are robust to restrictions imposed on agents' cognitive skills and

[1] The state of the world is defined by variables like attributes of the available alternatives, tastes of the DM, exogenous market shocks, etc.

on how agents respond to equally ranked alternatives. Moreover, remaining agnostic about information structures and selection rules permits our framework to nest various discrete choice models analysed in the literature. Examples include standard discrete choice models with complete information (e.g., McFadden, 1974; Berry, 1994) and discrete choice models with uncertainty (e.g., Barseghyan, et al., 2018).

Leaving information structures and selection rules unrestricted raises the possibility of partial identification of the preference parameters because the model is incomplete in the sense of Tamer (2003). Consequently, this poses the challenge of tractably characterising the region of preference parameter values that exhausts all the implications of the model and data (i.e., the sharp identified set). In fact, in order to determine whether a given value of the preference parameters belongs to the sharp identified set, the researcher needs to establish whether the empirical choice probabilities belong to the collection of choice probabilities predicted by the model under a range of information structures and selection rules. The difficulty here lies in the necessity of exploring all possible information structures and selection rules.

We approach the above problem by applying the notion of Bayesian Persuasion or one-player Bayesian Correlated Equilibrium provided in Kamenica and Gentzkow (2011) and Bergemann and Morris (2013; 2016). Specifically, we exploit Theorem 1 in Bergemann and Morris (2016) to claim that the collection of choice probabilities predicted by our model for a range of information structures and selection rules is equivalent to the collection of choice probabilities predicted by our model under the notion of one-player Bayesian Correlated Equilibrium. That is, it is equivalent to the collection of choice probabilities in a mediated decision problem where the mediator directly provides recommendations to the DM and these recommendations are incentive compatible. Further, such a collection is a convex set. Therefore, determining whether a given value of the preference parameters belongs to the sharp identified set amounts to finding whether the vector of empirical choice probabilities belongs to that convex set. By using insights from Beresteanu, Molchanov, and Molinari (2011), Magnolfi and Roncoroni (2017), and Syrgkanis, Tamer, and Ziani (2018), we argue that this corresponds to solving quadratically constrained linear programming problems or linear programming problems (depending on which objects the econometrician wants to recover). Thus, constructing the sharp identified set becomes a computationally tractable exercise. Lastly, after having reformulated the identifying restrictions as moment inequalities, we explain how inference on the sharp identified set can be conducted by using the generalised moment selection procedure in Andrews and Shi (2013).

One concern with the above method could be that the one-player Bayesian Correlated Equilibrium will have weak identification power because we make minimal assumptions. However, simulations reveal that the collection of conditional choice probabilities predicted by the one-player Bayesian Correlated Equilibrium is a strict and tight subset of the unit simplex under various data generating processes. Further, our model can be informative about the sign and magnitude of the preference parameters in several settings of economic interest.

Research questions similar to ours have been addressed in the empirical literature using two different approaches. The first approach consists in modelling the mechanism according to which the DM acquires an information structure. This approach may allow the analyst to obtain point identification of the preference parameters, however at the cost of possible misspecification of agents' information structures. For example, Mehta, Rajiv, and Srinivasan (2003), Honka and Chintagunta (2016), and Abaluck and Compiani (2019) consider search frameworks, where the DM follows a sequential protocol to learn about the payoffs generated by the available alternatives. Csaba (2018) adopts a rational inattention perspective, where the attentional costs sustained by the DM to process information structures are parametrically modelled, along the lines of Matějka and McKay (2015), Fosgerau, et al. (2017), and Caplin, Dean, and Leahy (2018).

The second approach relates to the econometric analysis of discrete choice models when the sets of alternatives actually considered by the DM (hereafter, *consideration sets*) could be strict subsets of the feasible set, heterogenous, and unobserved by the researcher. In fact, one key implication of attentional limits is that, since attention is a scarce resource, the DM may process an information structure inducing her to contemplate, in equilibrium, only a subset of the available alternatives, ignoring all the others (Caplin, Dean, and Leahy, 2018).[2,3] Several papers have considered the issue of identifying and estimating preferences in discrete choice models with heterogeneous and latent consideration sets. These papers can be grouped in three categories. The first group of papers rely on auxiliary data about the composition or probability distribution of consideration sets (e.g., Des Los Santos, Hortaçsu, and Wildenbeest, 2012; Conlon and Mortimer, 2013). The second group of papers rely on exclusion restrictions (e.g., Goeree, 2008; Gaynor, Propper, and Seiler, 2016). The third group of papers rely on assumptions on the consideration set formation process (e.g., Abaluck and Adams, 2018; Barseghyan, Molinari, and Teitelbaum, 2019; Barseghyan, et al., 2019; Cattaneo, et al., 2019; Crawford, Griffith, and Iaria, 2019). The papers in the third group are close in spirit to our framework. However, there are several differences. The papers in the third group are mostly concerned with settings where the DM is unaware of the existence of (or, voluntarily ignores) some alternatives in the feasible set. On the other hand, we consider the complementary problem of imperfect information at the level of payoffs. Further, the papers in the third group assume either that the DM is aware of the payoff generated by each alternative in her consideration set (e.g., Abaluck and Adams, 2018), or that the DM computes the expected payoff of the alternatives in her consideration set through distributions that are known, or estimable, by

[2]Recall that the DM's information structure takes the form of a noisy signal of the state of the world. Hence, an alternative belongs to the DM's consideration set if the subset of the signal's support inducing the DM to choose that alternative has positive measure (Caplin, Dean, and Leahy, 2018).

[3]Limited attention in choice is not the only mechanism that can induce endogenous considerations sets in discrete choice models. Consideration sets may arise also because of lack of awareness of some alternatives in the feasible set (e.g., Goeree, 2008), deliberately ignoring some alternatives in the feasible set (e.g., Wilson, 2008), incomplete product availability (e.g., Conlon and Mortimer, 2014), being offered the possibility of receiving program access from outside an experiment (e.g., Kamat, 2019), and absence of market clearing transfers in two-sided matching models (e.g., He, Sinha, and Sun, 2019).

the analyst (e.g., Barseghyan, et al., 2019). Instead, we proceed without imposing any such assumptions. In our framework the DM could still be imperfectly informed about the payoff generated by some or all alternatives in her consideration set, depending on the information structure that she processes, and about which the analyst remains ignorant.[4]

This paper also relates to two important works, Magnolfi and Roncoroni (2017) and Syrgkanis, Tamer, and Ziani (2018), that use the notion of Bayesian Correlated Equilibrium to tractably characterise the sharp identified set while remaining agnostic about information structures and selection rules. In particular, Magnolfi and Roncoroni (2017) use the notion of Bayesian Correlated Equilibrium in an entry game framework, where each firm may or may not know the payoff of the competing firms. Syrgkanis, Tamer, and Ziani (2018) use the notion of Bayesian Correlated Equilibrium in an auction framework, where each participant may or may not know the common value of the auctioned good (in common value auctions) or the value of the auctioned good for the other participants (in private value auctions). We contribute to this thread of literature by highlighting the empirical usefulness of the notion of Bayesian Correlated Equilibrium in a single-agent, static, discrete choice model with attentional limits.

The remaining of the paper is organised as follows. Section 2 describes the model. Section 3 discusses identification. Section 4 presents simulations. Section 5 illustrates inference. Section 6 concludes.

Notation Capital letters are used for random variables/vectors/matrices and small case letters for their realisations. Calligraphic capital letters are used for sets. Given a random vector Z, P_Z denotes its joint density when all the components of Z are continuous, mixed joint density when some components of Z are continuous and some discrete, and probability mass function when all the components of Z are continuous. However, for readability, sometimes in the paper we generically refer to P_Z as a *density*. Given a random vector Z, F_Z represents its cumulative distribution function. $\mathbb{R}^K_+$ denotes the K-dimensional positive real space. Given a set, $\mathcal{A}$, $\Delta(\mathcal{A})$ is the function space of all possible densities with support equal to or contained in $\mathcal{A}$. Given a set, $\mathcal{A}$, $|\mathcal{A}|$ denotes $\mathcal{A}$'s cardinality. Given two sets, $\mathcal{A}$ and $\mathcal{R} \subseteq \mathcal{A}$, $\mathcal{A} \setminus \mathcal{R}$ is the complement of $\mathcal{R}$ in $\mathcal{A}$. 0_L is the $L \times 1$ vector of zeros.

[4]There are other differences with the papers in the third group. For example, in Abaluck and Adams (2018), Barseghyan, Molinari, and Teitelbaum (2019), Cattaneo, et al. (2019), and Crawford, Griffith, and Iaria (2019) consideration sets are independent of preferences conditional on observables. In Barseghyan, et al. (2019) consideration sets have at least size $\kappa \geq 2$ with κ known by the analyst. In Abaluck and Adams (2018) and Barseghyan, Molinari, and Teitelbaum (2019) there are observables with large support. In Cattaneo, et al. (2019) the probability of a given consideration set decreases when the number of possible consideration sets decreases. In Dardanoni, et al., (2018) unobserved heterogeneity is assumed away. In Crawford, Griffith, and Iaria (2019) the analyst specifies how considerations sets evolve over time. In Lu (2018) there are restrictions on the smallest possible consideration set. We do not impose these assumptions here. However, recall that relaxing many of these assumptions simultaneously leads to partially identified preference parameters.

2 The model

In this section we describe a single-agent, static, discrete choice model, where the DM may face attentional limits hampering her capacity to learn about the payoff generated by each of the available alternatives. In such a framework, different restrictions on the amount of information processed by the DM typically lead to different optimal strategies. As what the DM knows is unobserved by the researcher and potentially heterogenous across agents, we use Theorem 1 in Bergemann and Morris (2016) to characterise the set of optimal strategies under minimal assumptions on the amount of information that is processed by the DM.

2.1 Baseline choice problem G

Suppose that the DM has to choose an alternative, y, from a finite set, $\mathcal{Y}$, possibly without having complete information about the state of the world. The state of the world is represented by a triplet, (x, e, v). In particular, x is a realisation of some covariates, X, with support $\mathcal{X}$. x is observed by the DM and the researcher. e is a realisation of some tastes of the DM, ϵ, with support conditional on x denoted by $\mathcal{E}$.[5] e is observed by the DM but not by the researcher. e is drawn by nature from the density of ϵ conditional on x, $P_{\epsilon|X}(\cdot|x)$. v is a realisation of some further (dis)value, V, that the DM can derive from the choice problem, with support conditional on (x, e) denoted by $\mathcal{V}$. v is not observed by the DM and the researcher. v is drawn by nature from the density of V conditional on (x, e), $P_{V|X,\epsilon}(\cdot|x, e)$, which constitutes the DM's prior about V.

Let $\mathcal{P}_{\epsilon|X} \equiv \{P_{\epsilon|X}(\cdot|x) \in \Delta(\mathcal{E}) : x \in \mathcal{X}\}$ be a family of densities of ϵ conditional on every realisation x of X. Here we would like to emphasise that our notation, $\mathcal{P}_{\epsilon|X}$, assumes that $\mathcal{P}_{\epsilon|X}$ contains one density for each $x \in \mathcal{X}$. The same notational convention is maintained for any family of conditional densities introduced below. Let $\mathcal{P}_{V|X,\epsilon} \equiv \{P_{V|X,\epsilon}(\cdot|x, e) \in \Delta(\mathcal{V}) : x \in \mathcal{X}, e \in \mathcal{E}\}$ be a family of densities of V conditional on every realisation (x, e) of (X, ϵ). In what follows, we refer to

$$G \equiv \Big(\mathcal{Y}, \mathcal{X}, \mathcal{E}, \mathcal{V}, u, \mathcal{P}_{V|X,\epsilon}, \mathcal{P}_{\epsilon|X}\Big),$$

as the *baseline choice problem*. As clarified later, G represents the minimal amount of information available to the DM before choosing.

2.2 Information structure S and augmented choice problem (G, S)

This section augments the baseline choice problem G by allowing the DM to refine her prior about V, $P_{V|X,\epsilon}(\cdot|x, e)$, upon reception of a private signal which may be informative about the realisation, v, of V drawn by nature. In particular, let T be a random variable representing the private signal received by the DM, with support conditional on (x, e, v) and density

[5] Note that $\mathcal{E}$ can vary across x. However, to keep the notation simple, we suppress this dependence. We adopt such notational convention for any conditional support considered in the paper.

conditional on (x, e, v) denoted by $\mathcal{T}$ and $P_{T|X,\epsilon,V}(\cdot|x, e, v)$, respectively. The DM observes t, uses $P_{T|X,\epsilon,V}(\cdot|x, e, v)$ to update her prior about V, and obtains the posterior, $P_{V|X,\epsilon,T}(\cdot|x, e, t)$. Then, the DM chooses an alternative, $y \in \mathcal{Y}$, maximising her expected payoff which is computed using the posterior. Finally, the DM receives the payoff $u(y, x, e, v)$, where $u : \mathcal{Y}\times\mathcal{X}\times\mathcal{E}\times\mathcal{V} \to \mathbb{R}$ is the payoff function.

Let $\mathcal{P}_{T|X,\epsilon,V} \equiv \{P_{T|X,\epsilon,V}(\cdot|x, e, v) \in \Delta(\mathcal{T}) : x \in \mathcal{X}, e \in \mathcal{E}, v \in \mathcal{V}\}$ be a family of densities of T conditional on every realisation (x, e, v) of (X, ϵ, V). In what follows, we refer to

$$S \equiv \left(\mathcal{T}, \mathcal{P}_{T|X,\epsilon,V}\right),$$

as the *information structure*. S represents the additional information that the DM processes to refine her prior about V.[6] Lastly, the pair (G, S) constitutes what will be hereafter called the *augmented choice problem*.

2.3 Optimal strategy of the augmented choice problem (G, S)

Let us define an optimal strategy of the DM in the augmented choice problem (G, S). A (mixed) strategy in the augmented choice problem (G, S) is a family of probability mass functions of Y conditional on every realisation (x, e, t) of (X, ϵ, T), i.e.,

$$\mathcal{P}_{Y|X,\epsilon,T} \equiv \{P_{Y|X,\epsilon,T}(\cdot|x, e, t) \in \Delta(\mathcal{Y}) : x \in \mathcal{X}, e \in \mathcal{E}, t \in \mathcal{T}\}.$$

$\mathcal{P}_{Y|X,\epsilon,T}$, is an optimal strategy of the augmented choice problem (G, S) if $P_{Y|X,\epsilon,T}(\cdot|x, e, t)$ maximises the DM' expected payoff, for each $x \in \mathcal{X}$, $e \in \mathcal{E}$, and $t \in \mathcal{T}$.

Definition 1. (*Optimal strategy of the augmented choice problem* (G, S)) $\mathcal{P}_{Y|X,\epsilon,T}$ is an optimal strategy of the augmented choice problem (G, S) if $\forall x \in \mathcal{X}$, $\forall e \in \mathcal{E}$, and $\forall t \in \mathcal{T}$,

$$\int_{\mathcal{V}} u(y, x, e, v)P_{T|X,\epsilon,V}(t|x, e, v)P_{V|X,\epsilon}(v|x, e)dv \geq \int_{\mathcal{V}} u(\tilde{y}, x, e, v)P_{T|X,\epsilon,V}(t|x, e, v)P_{V|X,\epsilon}(v|x, e)dv,$$

$\forall y \in \mathcal{Y}$ such that $P_{Y|X,\epsilon,T}(y|x, e, t) > 0$, and $\forall \tilde{y} \in \mathcal{Y} \setminus \{y\}$.

◇

Note that one can alternatively define an optimal strategy of the augmented choice problem (G, S) as follows.

Definition 2. (*Alternative definition*) Given $x \in \mathcal{X}$, $e \in \mathcal{E}$, and $t \in \mathcal{T}$, let $\mathcal{Y}^*_{x,e,t} \subseteq \mathcal{Y}$ be the

[6] We represent the latent variables (from the point of view of the analyst) constituting the state of the world by using two terms, ϵ and V, to obtain a flexible framework. In particular, the model nests settings where ϵ is a non-degenerate random variable and the DM *knows more* than the researcher about the state of the world. This is because, first, the DM observes e and, second, the DM might refine her knowledge about v by observing t. The model also nests settings where ϵ is a degenerate random variable and the DM *might know more* than the researcher about the state of the world by observing t.

set of alternatives maximising the expected payoff , i.e.,

$$\mathcal{Y}^*_{x,e,t} \equiv \text{argmax}_{y \in \mathcal{Y}} \int_{\mathcal{V}} u(y,x,e,v) P_{T|X,\epsilon,V}(t|x,e,v) P_{V|X,\epsilon}(v|x,e) dv.$$

Let $\mathcal{P}^*_{x,e,t}$ be the family of probability mass functions of Y conditional on (x,e,t) that are degenerate on $\mathcal{Y}^*_{x,e,t}$, i.e.,

$$\mathcal{P}^*_{x,e,t} \equiv \{P_{Y|X,\epsilon,T}(\cdot|x,e,t) \in \Delta(\mathcal{Y}) : P_{Y|X,\epsilon,T}(y|x,e,t) = 1, \, y \in \mathcal{Y}^*_{x,e,t}\}.$$

Let $\text{Conv}(\mathcal{P}^*_{x,e,t})$ be the convex hull of $\mathcal{P}^*_{x,e,t}$. Then, $\mathcal{P}_{Y|X,\epsilon,T}$ is an optimal strategy of the augmented choice problem (G,S) if $\forall x \in \mathcal{X}$, $\forall e \in \mathcal{E}$, and $\forall t \in \mathcal{T}$,

$$P_{Y|X,\epsilon,T}(\cdot|x,e,t) \in \text{Conv}(\mathcal{P}^*_{x,e,t}).$$

$\diamond$

Let $\mathcal{S}$ be the set of all admissible information structures. By using the continuity of the expected payoff in Y, it is possible to show that an optimal strategy of the augmented choice problem (G,S) exists for every $S \in \mathcal{S}$, even though it may not be unique.

Proposition 1. (*Existence of optimal strategy of the augmented choice problem* (G,S)) The augmented choice problem (G,S) admits an optimal strategy, $\mathcal{P}_{Y|X,\epsilon,T}$, for every $S \in \mathcal{S}$. $\diamond$

Note that the information structure has relevant implications on the behaviour of the DM. If the DM receives an information structure that is informative about the realisation of V drawn by nature, then her posterior and, hence, optimal strategy will reflect this information. Thus, the more informative the information structure, the more the DM will adjust her strategy according to the realisation of V. If the information structure is totally uninformative, then the DM will base her strategy on her posterior that will be equal to the prior, $P_{V|X,\epsilon}(\cdot|x,e)$.

2.4 Some examples of the augmented choice problem (G,S)

This section provides some examples of models and information structures that are nested in our framework. In such examples, we add the subscript i to our notation as a label for the DM with the purpose of emulating standard notation in the empirical literature on discrete choice models.

Multinomial Logit or Probit model Suppose that DM i must choose a transportation mode to get to work among $\mathcal{Y} \equiv \{0,1,...,L\}$, where "0" denotes the outside option of working from home. Let $X_i \equiv (X_{i,1},...,X_{i,L})$ be a $K \times L$ matrix, where $X_{i,y}$ is a vector of K characteristics of transportation mode $y \in \mathcal{Y} \setminus \{0\}$. These characteristics can be consumer specific (hence, the subscript i). For example, for each transportation mode $y \in \mathcal{Y} \setminus \{0\}$, X_{iy} could represent journey time, cost, and number of changes. Let $\epsilon_i \equiv (\epsilon_{i,0}, ..., \epsilon_{i,L})$ be an $(L+1) \times 1$ vector, where $\epsilon_{i,y}$ represents the DM's taste for transportation mode $y \in \mathcal{Y}$, such as comfort. Lastly, let $V_i \equiv (V_{i,0}, ..., V_{i,L})$ be an $(L+1) \times 1$ vector, where $V_{i,y}$ represents further (dis)value that product $y \in \mathcal{Y}$ may generate for the DM and about which the DM is uncertain, such as safety and pro-environmental features. In this framework, one can imagine that V_i is realised *before* the choice is made, but such realisation may be hidden to the DM. For each $y \in \mathcal{Y}$, the payoff function, u_i, is specified as

$$u_i(y, X_i, \epsilon_i, V_i) \equiv \begin{cases} \beta_y' X_{i,y} + \epsilon_{i,y} + V_{i,y} & \text{if } y \in \mathcal{Y} \setminus \{0\}, \\ \epsilon_{i,0} + V_{i,0} & \text{otherwise.} \end{cases}$$

Before choosing the DM can process additional information to refine her prior about V_i. For example, she can learn the technical features of each transportation mode, seek out reviews, and so on.

Suppose that the DM processes enough information to discover the realisation of V_i drawn by nature. That is, the DM's information structure features, for each $v \in \mathcal{V}$,

$$\mathcal{T} = \{v\}, P_{T_i|X_i,\epsilon_i,V_i}(v|x, e, v) = 1, \forall x \in \mathcal{X}, \forall e \in \mathcal{E},$$

(hereafter, *complete information structure*). Further, suppose that V_i is continuously distributed conditional on (X_i, ϵ_i). Then, an optimal strategy exists, is almost surely unique, and equal to

$$P_{Y_i|X_i,\epsilon_i,T_i}(\cdot|x, e, t) \in \Delta(\mathcal{Y}) \text{ s.t. } P_{Y_i|X_i,\epsilon_i,T_i}(y^*_{x,e,t}|x, e, t) = 1, \forall x \in \mathcal{X}, \forall e \in \mathcal{E}, \forall t \in \mathcal{T}, \tag{1}$$

where $y^*_{x,e,t} \equiv \text{argmax}_{y \in \mathcal{Y}}\, u_i(y, x, e, t)$. In particular, almost sure uniqueness follows from the fact that ties have *zero* measure because V_i is continuously distributed conditional on (X_i, ϵ_i). Furthermore, if $\{\epsilon_{i,y} + V_{i,y}\}_{\forall y \in \mathcal{Y}}$ are i.i.d. Gumbel with scale 1 and location 0 independent of X_i, this is the classical multinomial Logit model. If $\{\epsilon_{i,y} + V_{i,y}\}_{\forall y \in \mathcal{Y}}$ are i.i.d. standard normals independent of X_i, this is the classical multinomial Probit model. Hence, our framework nests discrete choice models in the tradition of McFadden (1974) where all agents are assumed to process the complete information structure.

Alternatively, suppose that the DM has no time to investigate further about V_i. That is, the DM's information structure features

$$\mathcal{T} = \{t\}, P_{T_i|X_i,\epsilon_i,V_i}(t|x, e, v) = 1, \forall x \in \mathcal{X}, \forall e \in \mathcal{E}, \forall v \in \mathcal{V},$$

for some $t \in \mathbb{R}$ (hereafter, *degenerate information structure*). Recall that under the degenerate information structure the DM's posterior is equal to the DM's prior about V_i. Then, Definition

1 reduces to: $\mathcal{P}_{Y_i|X_i,\epsilon_i}$ is an optimal strategy if for each $x \in \mathcal{X}$ and $e \in \mathcal{E}$,

$$\mathbb{E}[u_i(y, x, e, V_i)|X_i = x, \epsilon_i = e] \geq \mathbb{E}[u_i(\tilde{y}, x, e, V_i)|X_i = x, \epsilon_i = e].$$

for each $y \in \mathcal{Y}$ such that $P_{Y_i|X_i,\epsilon_i}(y|x, e) > 0$ and for each $\tilde{y} \in \mathcal{Y} \setminus \{y\}$, where expectations are computed using the DM's prior. Moreover, note that the set of optimal strategies could be singleton or there could be multiple optimal strategies. In fact, given $x \in \mathcal{X}$ and $e \in \mathcal{E}$, let $\mathcal{Y}^*_{x,e}$ be the set of alternatives maximising the expected payoff, i.e.,

$$\mathcal{Y}^*_{x,e} \equiv \text{argmax}_{y\in\mathcal{Y}}\ \mathbb{E}[u_i(y, x, e, V_i)|X_i = x, \epsilon_i = e].$$

Suppose that $\beta_y = 0_K$ for each $y \in \mathcal{Y} \setminus \{0\}$, V_i is independent of (X_i, ϵ_i), and $\{V_{i,y}\}_{y\in\mathcal{Y}}$ are i.i.d. standard normals. Then, for each $y \in \mathcal{Y}$, $x \in \mathcal{X}$, and $e \in \mathcal{E}$,

$$\mathbb{E}[u_i(y, x, e, V_i)|X_i = x, \epsilon_i = e] = e_y.$$

If ϵ_i is continuously distributed conditional on X_i, then $\mathcal{Y}^*_{x,e}$ is almost surely singleton for each $x \in \mathcal{X}$ and $e \in \mathcal{E}$. Consequently, the only optimal strategy is $\mathcal{P}_{Y_i|X_i,\epsilon_i}$ such that, for each $x \in \mathcal{X}$ and $e \in \mathcal{E}$,

$$P_{Y_i|X_i,\epsilon_i}(y^*_{x,e,}|x, e) = 1,$$

where $y^*_{x,e} \in \mathcal{Y}^*_{x,e}$. Instead, if ϵ_i has degenerate support $\{0_{L+1}\}$ conditional on X_i, then $\mathcal{Y}^*_{x,e} = \mathcal{Y}$ for each $x \in \mathcal{X}$ and $e \in \mathcal{E}$. In turn, this implies that the set of optimal strategies coincides with the entire collection of possible strategies.

Nested Logit demand model Suppose that DM i must choose which product to buy among $\mathcal{Y} \equiv \{0, 1, ..., L\}$, where "0" denotes the outside option of purchasing none of the products. Let $X \equiv (X_1, ..., X_L)$ be a $K \times L$ matrix, where X_y is a vector of K characteristics of product $y \in \mathcal{Y} \setminus \{0\}$ that are not consumer specific. Let $\epsilon_i \equiv (\xi, \eta_i)$, where $\xi \equiv (\xi_1, ..., \xi_L)$ is an $L \times 1$ vector with ξ_y representing attributes of product $y \in \mathcal{Y} \setminus \{0\}$ unobserved by the researcher, and η_i represents the DM's taste. Lastly, let $V_i \equiv (V_{i,0}, ..., V_{i,L})$ be an $(L + 1) \times 1$ vector, where $V_{i,y}$ represents further (dis)value that product $y \in \mathcal{Y}$ may generate for the DM and about which the DM is uncertain. For each $y \in \mathcal{Y}$, the payoff function, u_i, is specified as

$$u_i(y, X, \epsilon_i, V_i) \equiv \begin{cases} \beta' X_y + \xi_y + \eta_i + \lambda V_{i,y} & \text{if } y \in \mathcal{Y} \setminus \{0\}, \\ V_{i,0} & \text{otherwise,} \end{cases}$$

where $\lambda \in (0, 1)$. Before choosing the DM can process additional information to refine her prior about V_i. Suppose that the DM processes enough information to discover the realisation of V_i drawn by nature, i.e., the DM is endowed with the complete information structure. Further, let the DM's tastes be i.i.d. and independent of (X, ξ), and the probability distribution of

$\eta_i + \lambda V_{i,y}$ be chosen to yield the familiar Nested Logit market share function, as illustrated in Cardell (1997). Then, this is the Nested Logit demand model of Berry (1994), where the "inside" goods are separated from the outside good.

Discrete choice models under risk Let $\mathcal{Y} \equiv \{1, ..., L\}$ be a menu of L insurance plans against auto collision among which DM i must choose. Each plan $y \in \mathcal{Y}$ is represented by a premium, $P_{i,y}$, and a deductible, D_y. The DM is endowed with a $K \times 1$ vector of demographic characteristics, Z_i, such as wealth (denoted by W_i below), gender, age, insurance score, etc. We collect $(P_{i,y}, D_y)$ for each plan $y \in \mathcal{Y}$ and Z_i in the vector X_i. The DM is also endowed with a coefficient of absolute risk aversion, ϵ_i. Finally, the DM is endowed with some other features, η_i, which determine the event of a claim together with Z_i and about which the DM is uncertain. In particular, let V_i be equal to 1 if the DM experiences a claim after the choice is made and 0 otherwise, with

$$V_i = \mathbb{1}\{Z_i'\beta + \eta_i \geq 0\}. \tag{2}$$

In contrast to the examples discussed above, here η_i (and, hence, V_i) is realised *after* an insurance plan has been chosen and this causes the DM's uncertainty about the realisation of V_i. The payoff function, u_i, belongs to the CARA family, i.e., for each $y \in \mathcal{Y}$,

$$u_i(y, X_i, \epsilon_i, V_i) \equiv \begin{cases} \frac{1-\exp[-\epsilon_i \times (W_i - P_{i,y} - D_y)]}{\epsilon_i} & \text{if } V_i = 1, \epsilon_i \neq 0, \\ \frac{1-\exp[-\epsilon_i \times (W_i - P_{i,y})]}{\epsilon_i} & \text{if } V_i = 0, \epsilon_i \neq 0, \\ W_i - P_{i,y} - D_y & \text{if } V_i = 1, \epsilon_i = 0, \\ W_i - P_{i,y} & \text{if } V_i = 0, \epsilon_i = 0. \end{cases}$$

Before choosing an insurance plan the DM has the opportunity of processing additional information to refine her prediction about the realisation of V_i drawn by nature and refine her prior. For example, she can check the technical features of her car, road and traffic conditions, etc. Let $S_i \equiv (\mathcal{T}_i, \mathcal{P}_{T_i|X_i,\epsilon_i,V_i})$ denote the DM's information structure and $\mathcal{P}_{V_i|X_i,\epsilon_i,T_i}$ the family of resulting posteriors. Then, given the realisation (x, e, t) of (X_i, ϵ_i, T_i), the DM chooses an insurance plan y such that

$$y \in \text{argmax}_{y \in \mathcal{Y}} \begin{cases} [1 - F_{\eta_i|X_i,\epsilon_i,T_i}(z'\beta|x, e, t)] \times \frac{1-\exp[-e \times (w - p_y - d_y)]}{e} \\ \qquad + F_{\eta_i|X_i,\epsilon_i,T_i}(z'\beta|x, e, t) \times \frac{1-\exp[-e \times (w - p_y)]}{e} & \text{if } e \neq 0, \\ [1 - F_{\eta_i|X_i,\epsilon_i,T_i}(z'\beta|x, e, t)] \times [w - p_y - d_y] \\ \qquad + F_{\eta_i|X_i,\epsilon_i,T_i}(z'\beta|x, e, t) \times [w - p_y] & \text{if } e = 0. \end{cases}$$

This example is a simplified version of the framework that has been used in the empirical literature on risk preferences. Moreover, there the analysis is performed under the assumptions that S_i is the degenerate information structure (i.e., the posterior and prior about V_i are the same) and the prior about V_i is parametrically specified (see Barseghyan, et al., 2018 for a review).

2.5 Robust predictions

Suppose the researcher knows that the DM faces the payoff environment G. Further, the researcher is only aware that the DM processes *some* information structure, $S \in \mathcal{S}$, to refine her prior about V before choosing an optimal strategy as outlined in Definition 1. The researcher is interested in characterising the collection of optimal strategies of the augmented choice problem (G, S) while leaving S unrestricted. This is because S is unobserved by the researcher and potentially heterogenous across agents. In particular, heterogeneity of information structures comes from the fact that different agents could have different attention constraints and, thus, sustain different costs to gather information on the state of the world. Therefore, specifying or imposing assumptions on S could lead the researcher to make misleading predictions about the DM's optimal behaviour.

In this section we construct such characterisation of densities by using the notion of Bayesian Persuasion or one-player Bayesian Correlated Equilibrium (hereafter, 1BCE) provided in Kamenica and Gentzkow (2011) and Bergemann and Morris (2013; 2016).[7] Specifically, we exploit Theorem 1 in Bergemann and Morris (2016) which shows that the collection of densities of (Y, V) conditional on (X, ϵ) that are predicted by the model if the DM were to process *some* information structure, $S \in \mathcal{S}$, is equal to the collection of 1BCE of the baseline choice problem G.[8] That is, it is equivalent to the collection of densities of (Y, V) conditional on (X, ϵ) in a mediated decision problem where the mediator directly provides recommendations to the DM and these recommendations are incentive compatible. We now define a 1BCE of the baseline choice problem G, as in Bergemann and Morris (2016).[9]

Definition 3. (*1BCE of the baseline choice problem G*) A family of densities of (Y, V) conditional on every realisation (x, e) of (X, ϵ),

$$\mathcal{P}_{Y,V|X,\epsilon} \equiv \{P_{Y,V|X,\epsilon}(\cdot|x, e) \in \Delta(\mathcal{Y} \times \mathcal{V}) : x \in \mathcal{X}, e \in \mathcal{E}\},$$

is a 1BCE of the baseline choice problem G if:

[7]The notions of Bayesian Persuasion and 1BCE coincide. Specifically, Kamenica and Gentzkow (2011) consider a framework where a sender (planner) chooses an information structure, $S \in \mathcal{S}$, to give to a receiver (DM) and then the receiver chooses an alternative. Instead of letting the sender choose an S, Bergemann and Morris (2019) finds that this is equivalent to letting the sender choose her favourite 1BCE.

[8]Recall that Theorem 1 in Bergemann and Morris (2016) is valid for a general n-player game, where $n \geq 1$. It is used here for a one-player game.

[9]The published version of Bergemann and Morris (2016) uses an equivalent, but slightly different, definition of Bayesian Correlated Equilibrium. We adopt the working paper definition because it is more convenient for our purpose of conducting identification and inference on the model's primitives, as in Magnolfi and Roncoroni (2017) and Syrgkanis, Tamer, and Ziani (2018).

1. It is *consistent* with the baseline choice problem G, i.e., when integrating $P_{Y,V|X,\epsilon}(\cdot|x,e)$ with respect to Y, one obtains the DM's prior, $P_{V|X,\epsilon}(\cdot|x,e)$, $\forall x \in \mathcal{X}$ and $\forall e \in \mathcal{E}$. That is,

$$\sum_{y \in \mathcal{Y}} P_{Y,V|X,\epsilon}(y, v|x, e) = P_{V|X,\epsilon}(v|x, e),\ \forall x \in \mathcal{X}, \forall e \in \mathcal{E}, \forall v \in \mathcal{V}.$$

2. It is *obedient*, i.e., the DM who is recommended alternative $y \in \mathcal{Y}$ by an omniscient mediator has no incentive to deviate. That is,

$$\int_{\mathcal{V}} u(y, x, e, v) P_{Y,V|X,\epsilon}(y, v|x, e) dv \geq \int_{\mathcal{V}} u(\tilde{y}, x, e, v) P_{Y,V|X,\epsilon}(y, v|x, e) dv,$$
$$\forall y \in \mathcal{Y}, \forall \tilde{y} \in \mathcal{Y} \setminus \{y\}, \forall x \in \mathcal{X}, \forall e \in \mathcal{E}.$$

◇

Note that, for each $x \in \mathcal{X}$ and $e \in \mathcal{E}$, the set of 1BCE of the baseline choice problem G is convex because it is characterised by linear equalities and inequalities, where the linearity is in $P_{Y,V|X,\epsilon}(\cdot|x,e)$. This property will be exploited in Section 3 to conduct identification on the primitives of interest.

We now state Theorem 1 in Bergemann and Morris (2016) for our model.

Theorem 1. (*Bergemann and Morris, 2016*) $\mathcal{P}_{Y,V|X,\epsilon}$ is a 1BCE of the baseline choice problem G if and only if there exists an information structure, $S \in \mathcal{S}$, and an optimal strategy, $\mathcal{P}_{Y|X,\epsilon,T}$, of the augmented choice problem (G, S), such that $\mathcal{P}_{Y,V|X,\epsilon}$ is induced by $\mathcal{P}_{Y|X,\epsilon,T}$. ◇

Theorem 1 allows the researcher to make robust predictions about the DM's optimal behaviour through the notion of 1BCE. This is because it captures all possible optimal behaviours if the DM had access to *some* information structure $S \in \mathcal{S}$ that is left unspecified.

Theorem 1 is also useful to see why a 1BCE of the baseline choice problem G exists. Indeed, take any information structure $S \in \mathcal{S}$. Let $\mathcal{P}_{Y|X,\epsilon,T}$ be an optimal strategy of the augmented choice problem (G, S), which exists by Proposition 1. Let $\mathcal{P}_{Y,V|X,\epsilon}$ be the family of densities of (Y, V) conditional on every realisation (x, e) of (X, ϵ) induced by $\mathcal{P}_{Y|X,\epsilon,T}$. Then, by Theorem 1, $\mathcal{P}_{Y,V|X,\epsilon}$ is a 1BCE of the baseline choice problem G. Therefore, the set of 1BCE of the baseline choice problem G is non-empty.

Furthermore, the set of 1BCE of the baseline choice problem G is typically non-singleton. Indeed, if the set of 1BCE was a singleton, then information would be essentially irrelevant, i.e., a certain alternative would be optimal regardless of any extra information that the DM might process.

3 Identification

Section 2 exploits Theorem 1 in Bergemann and Morris (2016) to provide robust predictions about the DM's optimal behaviour, i.e., to find a map from the model's primitives to the DM's optimal behaviour that does not depend on the specification of the information structure, $S \in \mathcal{S}$, that is processed by the DM. In this section we study the inverse of such a map to conduct robust identification, i.e., to see what can be learnt about the model's fundamentals given the DM's optimal behaviour without restrictions on S.

We start by discussing our assumptions about the data generating process (hereafter, DGP).

Assumption 1. (*DGP*) The DM faces the baseline choice problem

$$G^{\theta_0} \equiv (\mathcal{Y}, \mathcal{X}, \mathcal{E}, \mathcal{V}, u^{\theta_0}, \mathcal{P}^{\theta_0}_{V|X,\epsilon}, \mathcal{P}^{\theta_0}_{\epsilon|X}).$$

The sets $\mathcal{Y}$, $\mathcal{X}$, $\mathcal{E}$, and $\mathcal{V}$ are finite. The DM processes some information structure from the set, $\mathcal{S}$, of admissible information structures. Given such information structure, the DM chooses an alternative from $\mathcal{Y}$ according to the notion of optimal strategy given by Definition 1.

The payoff function, u^{θ_0}, has a parametric form indexed by the finite dimensional vector of parameters $\theta_{1,0} \in \Theta_1 \subseteq \mathbb{R}^{K_1}$. The probability mass functions collected in $\mathcal{P}^{\theta_0}_{V|X,\epsilon}$ and $\mathcal{P}^{\theta_0}_{\epsilon|X}$ belong to parametric families indexed by the finite dimensional vectors of parameters $\theta_{2,0} \in \Theta_2 \subseteq \mathbb{R}^{K_2}$ and $\theta_{3,0} \in \Theta_3 \subseteq \mathbb{R}^{K_3}$, respectively.

G^{θ_0} is known by the researcher up to $\theta_0 \equiv (\theta_{1,0}, \theta_{2,0}, \theta_{3,0}) \in \Theta \equiv \Theta_1 \times \Theta_2 \times \Theta_3 \subseteq \mathbb{R}^{K_1+K_2+K_3}$, where the subscript "0" denotes the primitives of the true model and the true underlying DGP. The information structure processed by the DM is unobserved by the researcher. Further, in case of multiple optimal strategies (i.e., ties with non-zero measure), the DM selects an optimal strategy according to a selection rule that is unobserved by the researcher.

The probability mass function of (Y, X) which results from the decision problem is denoted by $P^0_{Y,X} \in \Delta(\mathcal{X} \times \mathcal{Y})$. $P^0_{Y,X}$ is nonparametrically identified by the sampling process and, hence, treated as known in the identification analysis. $\diamond$

Assumption 1 is similar to what is discussed in Magnolfi and Roncoroni (2017) and Syrgkanis, Tamer, and Ziani (2018) for an entry game setting and an auction setting, respectively. It summarises the model of Section 2 and draws attention to the fact that the researcher is aware only of the minimal amount of information available to the DM, consisting of the baseline choice problem G^{θ_0}. The researcher remains agnostic about the information structure processed by the DM, i.e., about how the DM's prior on V may be updated. Also, the researcher is ignorant of the DM's selection rule, i.e., of how the DM resolves indifferences between alternatives with the same expected payoff. The supports of X, ϵ, and V are assumed finite in order to make the construction of the sharp identified set for θ_0 tractable. When this is not the case, one can discretise them, as is common in the empirical literature with partially identified parameters.

We choose to present a parametric setup to be consistent with most of the applied work on discrete choice models in the literature. In particular, parameterising $\mathcal{P}^{\theta_0}_{V|X,\epsilon}$ and $\mathcal{P}^{\theta_0}_{\epsilon|X}$ is necessary if the researcher wants to conduct welfare analysis via counterfactuals. This is because running such counterfactuals requires recovering $\mathcal{P}^{\theta_0}_{V|X,\epsilon}$ and $\mathcal{P}^{\theta_0}_{\epsilon|X}$. The researcher can proceed without parameterising $\mathcal{P}^{\theta_0}_{V|X,\epsilon}$ and $\mathcal{P}^{\theta_0}_{\epsilon|X}$ if interested only in the vector of payoff parameters, $\theta_{1,0}$. In the main text we focus on the first (hereafter, *parametric*) case. We analyse the second (hereafter, *semiparametric*) case in Appendix B. As discussed in Appendix B, the semiparametric case is computationally easier. Intuitively, this is because the semiparametric case requires recovering less objects. Specifically, while identification in the parametric case entails solving quadratically constrained linear programming problems, identification in the semiparametric case involves solving linear programming problems.

Note that Assumption 1 allows for endogeneity of X. In the parametric case, the dependence between X and (ϵ, V) should be completely specified by the researcher up to some (unknown) finite dimensional vector of parameters. In the semiparametric case the dependence between X and (ϵ, V) can be arbitrary.

Lastly, in order to nonparametrically identify $P^0_{Y,X}$, it is sufficient for the analyst to observe the covariates and the alternative chosen by the DM for a large number of i.i.d. replications of the decision problem. Note that the DM's information structure and selection rule can vary across such replications. In other words, we allow for heterogeneity of information structures and selection rules across different agents in the population. We further discuss this aspect in Equation (3) below.

Before continuing the analysis, let us introduce some useful notation. In what follows, we denote by $\mathcal{P}^0_{Y|X}$ the family of probability mass functions of Y conditional on every realisation x of X induced by $P^0_{Y,X}$. We denote by $\theta \equiv (\theta_1, \theta_2, \theta_3)$ a generic element of Θ. For each $x \in \mathcal{X}$ and $P_{Y|X}(\cdot|x) \in \Delta(\mathcal{Y})$, let us rearrange the one-to-one image set of the mapping $y \in \mathcal{Y} \mapsto P_{Y|X}(\cdot|x) \in \Delta(\mathcal{Y})$ into a $|\mathcal{Y}| \times 1$ dimensional vector. With some abuse of notation, let us still denote such a vector by $P_{Y|X}(\cdot|x)$. Lastly, let us label the elements of $\mathcal{Y}$ as $y^1, ..., y^{|\mathcal{Y}|-1}, y^{|\mathcal{Y}|}$.

Our objective is to investigate identification of θ_0 under Assumption 1. Given the absence of restrictions on information structures and selection rules, the model is incomplete in the sense of Tamer (2003). This raises the possibility of partial identification and, consequently, the challenge of tractably characterising the set of θs exhausting all the implications of the model and data, i.e., the sharp identified set for θ_0.

Intuitively, the sharp identified set for θ_0 is the set of θs for which the model predicts a probability mass function of (Y, X) that matches with $P^0_{Y,X}$. More formally, for each $\theta \in \Theta$ and $S \in \mathcal{S}$, let $\mathcal{R}^{\theta,S}$ be the collection of optimal strategies of the augmented choice problem (G^θ, S). Lastly, for each $\theta \in \Theta$ and $x \in \mathcal{X}$, let $\bar{\mathcal{R}}^{\theta}_{Y|x}$ be the collection of probability mass functions of Y conditional on the realisation x of X that are induced by the model's optimal strategies under

θ, while remaining agnostic about the DM's information structure and selection rule. That is,

$$\bar{\mathcal{R}}^{\theta}_{Y|x} \equiv \text{Conv}\Big\{P_{Y|X}(\cdot|x) \in \Delta(\mathcal{Y}) : \\ P_{Y|X}(y|x) = \int_{\mathcal{T}\times\mathcal{V}\times\mathcal{E}} P_{Y|X,\epsilon,T}(y|x,e,t)P_{T|X,\epsilon,V}(t|x,e,v)P^{\theta}_{V|X,\epsilon}(v|x,e)P^{\theta}_{\epsilon|X}(e|x)d(t,v,e)\ \forall y \in \mathcal{Y}, \\ \mathcal{P}_{Y|X,\epsilon,T} \in \mathcal{R}^{\theta,S}, S \in \mathcal{S}\Big\}, \tag{3}$$

where we have used the fact that Y is independent of V conditional on (X,ϵ,T), because the DM's information on V is captured in the signal T.

Convexification allows us to include in $\bar{\mathcal{R}}^{\theta}_{Y|x}$ probability mass functions of Y conditional on the realisation x of X that are mixtures across information structures and selection rules. Importantly, this implies that the DM's information structure and selection rule can vary across the many replications of the decision problem used to non parametrically identify $P^0_{Y,X}$. That is, information structures and selection rules can be heterogenous across different agents in the population.[10] It follows that the sharp identified set for θ_0 can be defined as

$$\Theta^* \equiv \{\theta \in \Theta : P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{R}}^{\theta}_{Y|x}\ \forall x \in \mathcal{X}\}. \tag{4}$$

Unfortunately, the definition of Θ^* in (4) seems hardly useful in practice. This is because computing $\bar{\mathcal{R}}^{\theta}_{Y|x}$ is intractable due to the necessity of exploring the large class $\mathcal{S}$. In what follows we explain how to overcome such an issue by using Theorem 1 of Section 2.5. In particular, for each $\theta \in \Theta$, let $\mathcal{Q}^{\theta}$ be the collection of 1BCEs of the baseline choice problem G_{θ}. Moreover, for each $\theta \in \Theta$ and $x \in \mathcal{X}$, let $\bar{\mathcal{Q}}^{\theta}_{Y|x}$ be the collection of probability mass functions of Y conditional on the realisation x of X that are induced by the 1BCEs of the baseline choice problem G_{θ}, while remaining agnostic about the DM's selection rule. That is,

$$\bar{\mathcal{Q}}^{\theta}_{Y|x} \equiv \Big\{P_{Y|X}(\cdot|x) \in \Delta(\mathcal{Y}) : \\ P_{Y|X}(y|x) = \sum_{(e,v)\in\mathcal{E}\times\mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e)P^{\theta}_{\epsilon|X}(e|x)\ \forall y \in \mathcal{Y}, \mathcal{P}_{Y,V|X,\epsilon} \in \mathcal{Q}^{\theta}\Big\}. \tag{5}$$

Theorem 1 of Section 2.5 implies that $\bar{\mathcal{R}}^{\theta}_{Y|x} = \bar{\mathcal{Q}}^{\theta}_{Y|x}$ $\forall x \in \mathcal{X}$ and $\forall \theta \in \Theta$. Therefore, one can rewrite Θ^* by using the notion of 1BCE, as formalised in Proposition 2.

Proposition 2. (*Characterisation of* Θ^* *through the notion of 1BCE*) Let

$$\Theta^{**} \equiv \{\theta \in \Theta : P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}\ \forall x \in \mathcal{X}\}.$$

Under Assumption 1, $\Theta^* = \Theta^{**}$. $\diamond$

[10]This means that in our framework different agents may choose different alternatives because of taste heterogeneity (different realisation of X and ϵ), different cognitive heterogeneity (different S), signal heterogeneity (different realisation of T, for a given S), different selection rule (different choice among equally ranked alternatives), etc.

As explained in Section 2.5, recall that the set of 1BCE of the baseline choice problem G^θ is convex because it is defined by linear equalities and inequalities. Therefore, for each $x \in \mathcal{X}$ and $\theta \in \Theta$, $\bar{\mathcal{Q}}^\theta_{Y|x}$ is also convex. Constructing Θ^* as characterised in Proposition 2 is computationally feasible by leveraging on the convexity of $\bar{\mathcal{Q}}^\theta_{Y|x}$ for each $x \in \mathcal{X}$ and $\theta \in \Theta$. This is formalised in Proposition 3.

Let $\mathbb{B}^{|\mathcal{Y}|}$ be the unit ball in $\mathbb{R}^{|\mathcal{Y}|}$, i.e., $\mathbb{B}^{|\mathcal{Y}|} \equiv \{b \in \mathbb{R}^{|\mathcal{Y}|} : b^T b \leq 1\}$. Let $\mathbb{S}^{|\mathcal{Y}|}$ be the unit sphere in $\mathbb{R}^{|\mathcal{Y}|}$, i.e., $\mathbb{S}^{|\mathcal{Y}|} \equiv \{b \in \mathbb{R}^{|\mathcal{Y}|} : b^T b = 1\}$.

Proposition 3. (*Construction of* Θ^**)* Under Assumption 1:

(i) For each $\theta \in \Theta$, $\theta \in \Theta^*$ if and only if

$$\max_{b \in \mathbb{B}^{|\mathcal{Y}|}} \min_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T [P^0_{Y|X}(\cdot|x) - P_{Y|X}(\cdot|x)] = 0, \tag{6}$$

$\forall x \in \mathcal{X}$.

(ii) For each $\theta \in \Theta$ and $x \in \mathcal{X}$, (6) can be rewritten as a quadratically constrained linear maximisation problem.

(iii) For each $\theta \in \Theta$, $\theta \in \Theta^*$ if and only if

$$-b^T \begin{pmatrix} P^0_{Y|X}(y^1|x) \\ \vdots \\ P^0_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix} + \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T \begin{pmatrix} P_{Y|X}(y^1|x) \\ \vdots \\ P_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix} \geq 0 \ \forall b \in \mathbb{S}^{|\mathcal{Y}|-1}, \tag{7}$$

$\forall x \in \mathcal{X}$.

$\diamond$

Proposition 3 (i) states that, in order to determine whether a candidate vector of parameters belongs to Θ^*, the researcher should solve (6) for each $x \in \mathcal{X}$. In fact, following Beresteanu, Molchanov, and Molinari (2011), one can exploit the convexity of $\bar{\mathcal{Q}}^\theta_{Y|x}$ to rewrite the condition $P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}$ as

$$\max_{b \in \mathbb{B}^{|\mathcal{Y}|}} [b^T P^0_{Y|X}(\cdot|x) - \sup_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T P_{Y|X}(\cdot|x)] = 0, \tag{8}$$

where the map

$$b \in \mathbb{R}^{|\mathcal{Y}|} \mapsto \sup_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T P_{Y|X}(\cdot|x) \in \mathbb{R},$$

is the support function of $\bar{\mathcal{Q}}^\theta_{Y|x}$. Some simple algebraic manipulations reveal that (8) is equal to (6). (6) is a max-min problem and can be computationally costly to solve. However, note that the inner constrained minimisation problem in (6) is linear in $P_{Y|X}(\cdot|x)$. Thus, it can be replaced by its dual, which consists of a linear constrained maximisation problem. Moreover, the outer constrained maximisation problem in (6) has a quadratic constraint, $b^T b \leq 1$. Therefore, as claimed by Proposition 3 (ii), (6) can be rewritten as a quadratically constrained linear maximisation problem which is a tractable exercise. Lastly, Proposition 3 (iii) expresses the identification problem as a collection of inequalities. Such alternative representation is exploited in Section 5 to conduct inference. Similar results to Proposition 3 are discussed in Section 3.3. and Appendix A of Magnolfi and Roncoroni (2017) for an entry game setting.

4 Simulations

In this section we investigate the informativeness of our model about the primitives of interest under various data generating processes.

As a first exercise, we construct the collection of choice probabilities predicted by 1BCEs under various DGPs. In particular, we consider the transport choice problem discussed in Section 2.4 with $\mathcal{Y} \equiv \{0, 1, 2\}$ and no covariates, under the following DGPs.

1. ϵ_i is independent of V_i. The probability mass function of ϵ_i is obtained as the density of a normal random vector with mean μ_ϵ and variance covariance matrix Σ_ϵ, discretised and truncated to have support $\mathcal{E} \equiv \{0, 1, ..., 5\}^3$, i.e.,

$$P_\epsilon(e) = \frac{\frac{1}{2\pi}\det(\Sigma_\epsilon)^{-\frac{1}{2}} \exp^{-\frac{1}{2}(e-\mu_\epsilon)'\Sigma_\epsilon^{-1}(e-\mu_\epsilon)}}{\sum_{e \in \mathcal{E}} \frac{1}{2\pi}\det(\Sigma_\epsilon)^{-\frac{1}{2}} \exp^{-\frac{1}{2}(e-\mu_\epsilon)'\Sigma_\epsilon^{-1}(e-\mu_\epsilon)}}, \quad (9)$$

for each $e \in \mathcal{E}$. The probability mass function of V_i is obtained in an analogous way. In particular, we consider:

(DGP1): $\mu_\epsilon = \mu_V \equiv (2.5, 2.5, 2.5)'$, $\Sigma_\epsilon = \Sigma_V \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. We denote by DGP1$_{\text{deg}}$ the case where V_i is distributed as in DGP1 but ϵ_i is degenerate with support $\mathcal{E} \equiv \{(0, 0, 0)'\}$.

(DGP2): $\mu_\epsilon \equiv (4.074, 4.529, 0.635)'$, $\mu_V \equiv (4.567, 3.162, 0.488)'$, $\Sigma_\epsilon = \Sigma_V \equiv \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. We denote by DGP2$_{\text{deg}}$ the case where V_i is distributed as in DGP2 but ϵ_i is degenerate with support $\mathcal{E} \equiv \{(0, 0, 0)'\}$.

(DGP3): $\mu_\epsilon \equiv (3.221, 1.904, 3.315)'$, $\mu_V \equiv (2.591, 3.834, 4.669)'$, $\Sigma_\epsilon \equiv \begin{pmatrix} 3.164 & 0.977 & 0.378 \\ 0.977 & 3.235 & 0.361 \\ 0.377 & 0.361 & 3.544 \end{pmatrix}$, $\Sigma_V \equiv \begin{pmatrix} 3.090 & 0.215 & 0.865 \\ 0.215 & 3.239 & 0.659 \\ 0.865 & 0.658 & 3.031 \end{pmatrix}$. We denote by DGP3$_{\text{deg}}$ the case where V_i is distributed as in DGP3 but ϵ_i is degenerate with support $\mathcal{E} \equiv \{(0, 0, 0)'\}$. .

2. (DGP4): $\{\epsilon_{i,0}, \epsilon_{i,1}, \epsilon_{i,1}, V_{i,0}, V_{i,1}, V_{i,1}\}$ are mutually independent and distributed as Poissons with support $\{0, 1, ..., 5\}$ and rates equal to $\{0.4, 0.1, 0.9, 0.2, 0.8, 0.5\}$, respectively. We denote by DGP4$_{\text{deg}}$ the case where V_i is distributed as in DGP4 but ϵ_i is degenerate with support $\mathcal{E} \equiv \{(0, 0, 0)'\}$.

3. (DGP5): $\{\epsilon_{i,0}, \epsilon_{i,1}, \epsilon_{i,1}, V_{i,0}, V_{i,1}, V_{i,1}\}$ are mutually independent and distributed as Binomials with number of trials equal to 5 and probability of success in each trial equal to $\{0.587, 0.123, 0.611, 0.001, 0.801, 0.521\}$, respectively. We denote by DGP5$_{\text{deg}}$ the case where V_i is distributed as in DGP5 but ϵ_i is degenerate with support $\mathcal{E} \equiv \{(0, 0, 0)'\}$.

Let $\bar{\mathcal{R}}_Y^{\theta_0,\text{comp}}$ be the collection of choice probabilities induced by the model's optimal strategies when the researcher assumes that agents are endowed with the complete information structure. Let $\bar{\mathcal{R}}_Y^{\theta_0,\text{deg}}$ be the collection of choice probabilities that are induced by the model's optimal strategies when the researcher assumes that agents are endowed with the degenerate information structure. Finally, recall that $\bar{\mathcal{Q}}_Y^{\theta_0}$ is the collection of choice probabilities that are induced by 1BCEs, as defined in Equation (5). Figure 1 represents the sets $\bar{\mathcal{Q}}_Y^{\theta_0}$ (black region), $\bar{\mathcal{R}}_Y^{\theta_0,\text{comp}}$ (red region), and $\bar{\mathcal{R}}_Y^{\theta_0,\text{deg}}$ (blue region) under the DGPs described above. The pictures in the left column are for DGP1, DGP2, DGP3, DGP4, and DGP5, respectively. The pictures in the right column are for DGP1$_{\text{deg}}$, DGP2$_{\text{deg}}$, DGP3$_{\text{deg}}$, DGP4$_{\text{deg}}$, and DGP5$_{\text{deg}}$ respectively. By Theorem 1, the collections of choice probabilities induced by the model's optimal strategies under the complete and the degenerate information structures are contained in the collection of choice probabilities induced by 1BCEs. That is, $\bar{\mathcal{R}}_Y^{\theta_0,\text{comp}}$ and $\bar{\mathcal{R}}_Y^{\theta_0,\text{comp}}$ are subsets of $\bar{\mathcal{Q}}_Y^{\theta_0}$. Further, in all the DGPs considered, except DGP1$_{\text{deg}}$, $\bar{\mathcal{Q}}_Y^{\theta_0}$ is a strict subset of the unit simplex. Moreover, the collection of choice probabilities induced by 1BCEs is a relatively small subset of all possible choice probabilities.

As a second exercise, we construct the sharp identified set for θ_0 in various models. In particular, we consider the transport choice problem discussed in Section 2.4 with $\mathcal{Y} \equiv \{0, 1, 2, 3\}$, $X_{i,y}$ scalar for each transport option $y \in \mathcal{Y} \setminus \{0\}$. To generate the data we obtain the probability mass function of X_i as the density of a normal random vector with mean and variance covariance matrix

$$\mu_X \equiv (0.629, 0.812, -0.746)', \ \Sigma_X \equiv \begin{pmatrix} 3.913 & 0.455 & 0.531 \\ 0.455 & 3.547 & 0.558 \\ 0.531 & 0.558 & 3.971 \end{pmatrix},$$

respectively, discretised and truncated to have support $\mathcal{X} \equiv \{-1, 0, 1\}$ as done in (9). Further, we impose

$$\beta_0 \equiv (-2.5, -1.8 - 0.9)'.$$

Also, we assume that (ϵ_i, V_i) is independent of X_i and that ϵ_i is independent of V_i. The probability mass function of ϵ_i is obtained as the density of a normal random vector with mean and variance covariance matrix

$$\mu_\epsilon \equiv 0_{3\times 1},\ \Sigma_\epsilon \equiv \begin{pmatrix} 1/2 & \rho_0 & \rho_0 \\ \rho_0 & \sigma_0^2 & \rho_0 \\ \rho_0 & \rho_0 & \sigma_0^2 \end{pmatrix},$$

respectively, discretised and truncated to have support $\mathcal{E} \equiv \{0, 1, ..., 5\}^3$ as done in (9). In particular, we impose $\rho_0 \equiv 0.25$ and $\sigma_0^2 \equiv 1.5$. The variance of $\epsilon_{i,1}$ is normalised to 1/2. The probability mass function of V_i is obtained in an analogous way. Finally, the empirical choice probabilities are derived under the assumption that half of the population processes the complete information structure (i.e., half of the population observes the realisation of V_i) and half of the population processes the degenerate information structure (i.e., half of the population does not observe the realisation of V_i and decides by using a posterior equal to the prior). Hereafter, we refer to this DGP as DGP6. The black regions in Figure 2 represent the projections of the sharp identified set for $\theta_0 \equiv (\beta_{0,1}, \beta_{0,2}, \beta_{0,3}, \rho_0, \sigma_0^2)$ along the axis of $\beta_{0,1}, \beta_{0,2}, \beta_{0,3}$. The red dots in Figure 2 represent $\beta_{0,1}, \beta_{0,2}, \beta_{0,3}$. The black regions are tight and informative about the signs and magnitudes of $\beta_{0,1}, \beta_{0,2}, \beta_{0,3}$. Lastly, the projection of the sharp identified for θ_0 along the axis of ρ_0 and σ_0^2 are $[-0.15, 0.45]$ and $[0.25, \infty)$.

We also consider the insurance choice problem discussed in Section 2.4 with $\mathcal{Y} \equiv \{1, 2, 3, 4\}$ and deductibles $D \equiv (100, 200, 500, 1000)$. For each insurance plan $y \in \mathcal{Y}$, we generate the data by assuming that the premium, $P_{i,y}$, is equal to $P_i^{\text{base}} \times \lambda_y$, where $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) \equiv (5/6, 7/10, 3/10, 1/10)$ and P_i^{base} is uniformly distributed on $\{100, 200, 300\}$. Given that the utility function belongs to the CARA family, payoffs can be computed without observing W_i. We assume that Z_i entering (2) is scalar and uniformly distributed on $\{2.0000, 2.4000, 2.8000\}$. We also assume that η_i entering (2) is equal to $\xi_i + \tau_i$, where (ξ_i, τ_i) is distributed as a bivariate normal with mean and variance covariance matrix

$$\mu_{\eta,\tau} \equiv (0, 5)',\ \Sigma_{\eta,\tau} \equiv \begin{pmatrix} 2 & 0 \\ 0 & 0.5 \end{pmatrix},$$

respectively. We impose $\beta_0 \equiv 0.5$ in (2). The probability mass function of the coefficient of risk aversion, ϵ_i, is obtained as the density of a Beta distribution with parameters $\mu_{0,1} \equiv 1$ and $\mu_{0,2} \equiv 2$, discretised and truncated in order to have support $\mathcal{E} \equiv \{0.005, 0.010, 0.015, ..., 0.100\}$. That is,

$$P_\epsilon(e) = \frac{\frac{e^{\mu_{0,1}-1}(1-e)^{\mu_{0,2}-1}}{B(\mu_{0,1},\mu_{0,2})}}{\sum_{e\in\mathcal{E}} \frac{e^{\mu_{0,1}-1}(1-e)^{\mu_{0,2}-1}}{B(\mu_{0,1},\mu_{0,2})}},$$

for each $e \in \mathcal{E}$, where $B(\mu_{0,1}, \mu_{0,2}) \equiv \frac{\Gamma(\mu_{0,1})\Gamma(\mu_{0,2})}{\Gamma(\mu_{0,1}+\mu_{0,2})}$ and Γ is the Gamma function. Finally, the empirical choice probabilities are derived under the assumption that one third of the population processes the complete information structure (i.e., one third of the population observes η_i), one third of the population process the degenerate information structure (i.e., one third of the population does not observe η_i and decides by using a posterior equal to the prior), and one third of the population observes ξ_i but not τ_i. Hereafter, we refer to this DGP as DGP7. The black regions in Figure 3 represent the projections of the sharp identified set for $\theta_0 \equiv (\beta_0, \mu_{0,1}, \mu_{0,2})$ along each axis. The red dots in Figure 3 represent $\beta_0, \mu_{0,1}, \mu_{0,2}$. The black regions are informative about the signs of all the parameters. Further, the projections for β_0 and $\mu_{0,1}$ are bounded and tight, while the projection for $\mu_{0,2}$ is unbounded above.

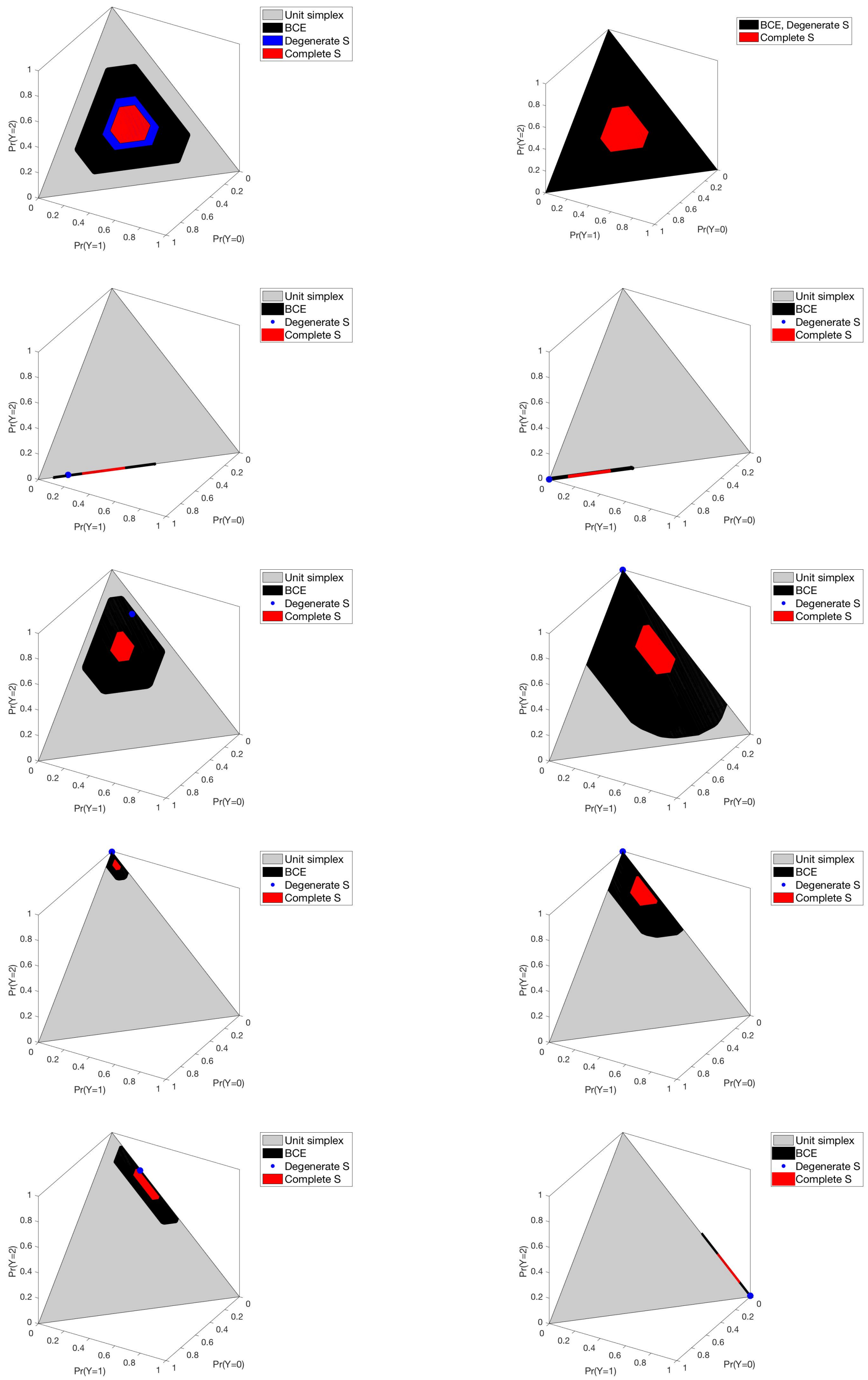

Figure 1: The pictures in the left column represent the sets $\bar{\mathcal{Q}}_Y^{\theta_0}$ (black region), $\bar{\mathcal{R}}_Y^{\theta_0,\text{comp}}$ (red region), and $\bar{\mathcal{R}}_Y^{\theta_0,\text{deg}}$ (blue region) under DGP1, DGP2, DGP3, DGP4, and DGP5, respectively. The pictures in the right column are for DGP1_{deg}, DGP2_{deg}, DGP3_{deg}, DGP4_{deg}, and DGP5_{deg} respectively.

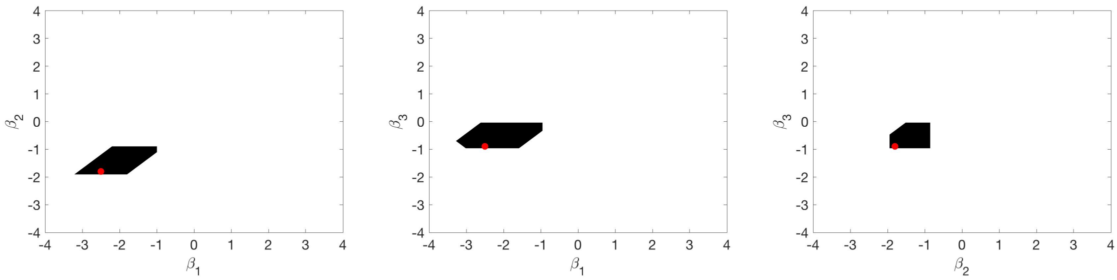

Figure 2: The figure is based on DGP6. The black regions represent the projections of the sharp identified set for $\theta_0 \equiv (\beta_{0,1}, \beta_{0,2}, \beta_{0,3}, \rho_0, \sigma_0^2)$ along the axis of $\beta_1, \beta_2, \beta_3$. The red dot represents $\beta_{0,1}, \beta_{0,2}, \beta_{0,3}$.

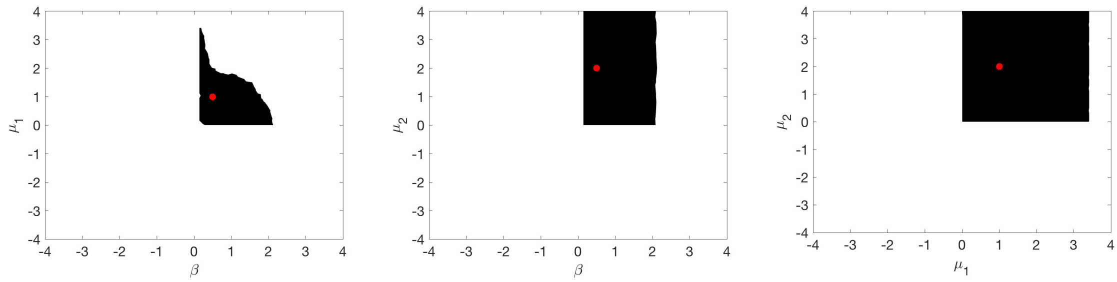

Figure 3: The figure is based on DGP7. The black regions represent the projections of the sharp identified set for $\theta_0(\beta_0, \mu_{0,1}, \mu_{0,2})$ along each axis. The red dot represents $\beta_0, \mu_{0,1}, \mu_{0,2}$.

5 Inference

Identification of the true parameter vector, θ_0, relies on the assumption that the true density of the observables, $P^0_{Y,X}$, is known by the researcher. However, when doing an empirical analysis, the researcher should replace $P^0_{Y,X}$ with its sample analogue resulting from having i.i.d. observations, $\{Y_i, X_i\}_{i=1}^n$. Given $\alpha \in (0,1)$, this section illustrates how to construct a uniformly asymptotically valid $(1-\alpha)$ confidence region, $C_{n,1-\alpha}$, for each $\theta \in \Theta^*$. In particular, we suggest to apply the generalised moment selection procedure by Andrews and Shi (2013) (hereafter, AS), as detailed in Appendix B.1 of Beresteanu, Molchanov, and Molinari (2011) (hereafter, BMM).[11] $C_{n,1-\alpha}$ is obtained by inverting a test with null hypothesis $\text{H}_0 : \theta_0 = \theta$ for every $\theta \in \Theta$. Such a test rejects H_0 if $\text{TS}_n > \hat{c}_{n,1-\alpha}(\theta)$, where TS_n is a test statistic and $\hat{c}_{n,1-\alpha}(\theta)$ is a corresponding critical value. Thus, $C_{n,1-\alpha} \equiv \{\theta \in \Theta\colon \text{TS}_n(\theta) \leq \hat{c}_{n,1-\alpha}(\theta)\}$. The remainder of the section explains how to compute $\text{TS}_n(\theta)$ and $\hat{c}_{n,1-\alpha}(\theta)$ for any $\theta \in \Theta$.

First, as claimed by Proposition 3 (iii), our model generates conditional moment inequalities. In fact, Proposition 3 (iii) can be rewritten as

$$\theta \in \Theta^* \Leftrightarrow \mathbb{E}[m(Y, X; b, \theta)|X = x] \geq 0 \;\forall b \in \mathbb{S}^{|\mathcal{Y}|-1}, \forall x \in \mathcal{X},$$

where

$$m(Y, x; b, \theta) \equiv -b^T \begin{pmatrix} \mathbb{1}\{Y = y^1\} \\ \dots \\ \mathbb{1}\{Y = y^{|\mathcal{Y}|-1}\} \end{pmatrix} + \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T \begin{pmatrix} P_{Y|X}(y^1|x) \\ \dots \\ P_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix}.$$

Second, Lemma 2 in AS shows that conditional moment inequalities can be transformed into equivalent unconditional moment inequalities by choosing appropriate instruments, $h \in \mathcal{H}$, where $\mathcal{H}$ is a collection of instruments and h is a function of X. In particular,

$$\theta \in \Theta^* \Leftrightarrow \mathbb{E}[m(Y, X; b, \theta, h)] \geq 0 \;\forall b \in \mathbb{S}^{|\mathcal{Y}|-1}, \forall h \in \mathcal{H} \text{ a.s.}, \tag{10}$$

where

$$m(Y, X; b, \theta, h) \equiv m(Y, X; b, \theta) \times h(X).$$

Further, observe that (10) is equivalent to

$$\theta \in \Theta^* \Leftrightarrow \min\left\{0, \min_{b \in \mathbb{S}^{|\mathcal{Y}|-1}} \mathbb{E}[m(Y, X; b, \theta, h)]\right\} = 0 \;\forall h \in \mathcal{H} \text{ a.s.}$$

[11] Note that the characterisation of Θ^* in Proposition 2 is equivalent to the characterisation in Theorem 2.1 of BMM. This is because the Aumann expectation of the random closed set of 1BCE alternative predictions is equal to $\bar{\mathcal{Q}}^\theta_{Y|x}$, for each $\theta \in \Theta$ and $x \in \mathcal{X}$.

In light of these remarks, BMM proposes as test statistic

$$\text{TS}_n(\theta) \equiv \int_{\mathcal{H}} \min\left\{0, [\min_{b \in \mathbb{S}^{|\mathcal{Y}|-1}} \sqrt{n}\bar{m}_n(b,\theta,h)]^2\right\} d\Gamma(h),$$

where Γ is a probability measure on $\mathcal{H}$ as explained in Section 3.4 of AS, and

$$\bar{m}_n(b,\theta,h) \equiv \frac{1}{n}\sum_{i=1}^{n} m(Y_i, X_i; b, \theta, h).$$

Theorem B.2 in BMM shows that, under some regularity conditions, $\text{TS}_n(\theta)$ satisfies Assumptions S1-S4 and M2 of AS. This implies that AS's procedure is applicable. Moreover, given that the set $\mathcal{X}$ is finite, the analyst can use the uniform probability measure as suggested by Example 5 in Appendix B of AS. That is,

$$\text{TS}_n(\theta) \equiv \frac{1}{|\mathcal{X}|}\sum_{x \in \mathcal{X}} \min\left\{0, [\min_{b \in \mathbb{S}^{|\mathcal{Y}|-1}} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(Y_i, X_i; b, \theta)\mathbb{1}\{X_i = x\}]^2\right\}. \tag{11}$$

In practice, to compute (11), the researcher should calculate, for each $x \in \mathcal{X}$,

$$\min_{b \in \mathbb{S}^{|\mathcal{Y}|-1}} \frac{1}{\sqrt{n}} \sum_{\substack{i \text{ s.t.} \\ x_i = x}} [-b^T \mathbb{1}_i + \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} b^T \tilde{P}_{Y|X}(\cdot|x)], \tag{12}$$

where $\mathbb{1}_i \equiv \begin{pmatrix} \mathbb{1}\{y_i = y^1\} \\ \dots \\ \mathbb{1}\{y_i = y^{|\mathcal{Y}|-1}\} \end{pmatrix}$ and $\tilde{P}_{Y|X}(\cdot|x) \equiv \begin{pmatrix} P_{Y|X}(y^1|x) \\ \dots \\ P_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix}$. By rearranging terms, Expression (12) becomes

$$\min_{b \in \mathbb{S}^{|\mathcal{Y}|-1}} \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} b^T[-\frac{1}{\sqrt{n}} \sum_{\substack{i \text{ s.t.} \\ x_i = x}} \mathbb{1}_i + \frac{n_{x_i=x}}{\sqrt{n}} \tilde{P}_{Y|X}(\cdot|x)], \tag{13}$$

where $n_{x_i=x}$ is the number of observations featuring $x_i = x$. (13) can be rewritten as a quadratically constrained linear minimisation problem by following similar steps to the proof of Proposition 3 (ii). Once (13) is computed for each $x \in \mathcal{X}$, the analysts easily obtains $\text{TS}_n(\theta)$.

To compute the critical value, we follow AS's bootstrap method consisting of the following steps. Specifically, for each $x \in \mathcal{X}$, let

$$\bar{m}_n(b,\theta,x) \equiv \frac{1}{n}\sum_{i=1}^{n} m(Y_i, X_i; b, \theta)\mathbb{1}\{X_i = x\}.$$

We draw W_n bootstrap samples using nonparametric i.i.d. bootstrap. For each $w = 1, ..., W_n$, we compute

$$TS_{n,w}(\theta) \equiv \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \min \Big\{ 0, [\min_{b \in \mathbb{S}^{|\mathcal{Y}|-1}} (\sqrt{n}(\bar{m}^*_{n,w}(b,\theta,x) - \bar{m}_n(b,\theta,x)) + \varphi_n(b,\theta,x))]^2 \Big\},$$

where $\bar{m}^*_{n,w}(b,\theta,x)$ is calculated just as $\bar{m}_n(b,\theta,x)$, but with the bootstrap sample in place of the original sample, $\varphi_n(b,\theta,h) \equiv \mathbb{1}\{\frac{1}{\kappa_n}\sqrt{n}\bar{m}_n(b,\theta,h) > 1\} \times B_n$, and $\{\kappa_n\}_{n\in\mathbb{N}}$, $\{B_n\}_{n\in\mathbb{N}}$ are sequences of constants satisfying Assumption G.1 in AS. In particular, we use $\kappa_n \equiv (0.3\log(n))^{1/2}$ and $B_n \equiv \left(\frac{0.4\log(n)}{\log(\log(n))}\right)^{1/2}$ as suggested in Section 9 of AS. Lastly, $\hat{c}_{n,1-\alpha}(\theta)$ is the $(1-\alpha)$ sample quantile of $\{TS_{n,w}(\theta)\}_{w=1}^{W_n}$.

6 Conclusions

In this paper we consider a single-agent, static, discrete choice model in which agents can face attentional limits. This implies that the DM may have imperfect information about the payoffs of the available alternatives that will affect her choice. Instead of explicitly modelling the information constraints, which can be susceptible to misspecification, we study identification and inference in settings where the researcher remains agnostic about the mechanism determining the amount of information processed by the DM. Moreover, we put no restriction on how the DM resolves ties. We exploit Theorem 1 in Bergemann and Morris (2016) to provide a tractable characterisation of the sharp identified set for the preference parameters and study inference in our imperfect information set up. Simulations reveal that the obtained bounds on the preference parameters can be tight in several settings of economic interest.

We are currently working on an empirical illustration to real data. We also leave to future research the possibility of using our methodology to test for the true information assumption underlying the DGP.

References

Abaluck, J., and A. Adams (2018): "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses," Working Paper.

Abaluck, J., and G. Compiani (2019): "A Method to Estimate Discrete Choice Models that is Robust to Consumer Search," Working Paper.

Andrews, D.W.K., and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81(2), 609–666.

Barseghyan, L., M. Coughlin, F. Molinari, and J.C. Teitelbaum (2019): "Heterogeneous Choice Sets and Preferences," arXiv:1907.02337.

Barseghyan, A., F. Molinari, T. O'Donoghue, and J.C. Teitelbaum (2018): "Estimating Risk Preferences in the Field," *Journal of Economic Literature*, 56(2), 501–564.

Barseghyan, L., F. Molinari, and J.C. Teitelbaum (2019): "Discrete Choice Under Risk with Limited Consideration," arXiv:1902.06629.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification in Models with Convex Moment Predictions," *Econometrica*, 79(6), 1785–1821.

Bergemann, D. and S. Morris (2013): "Robust Predictions in Games With Incomplete Information," *Econometrica*, 81(4), 1251–1308.

Bergemann, D. and S. Morris (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11(2), 487–522.

Berry, S. (1994), "Estimating Discrete-Choice Models of Product Differentiation," *The RAND Journal of Economics*, 25(2), 242–262.

Bergemann, D. and S. Morris (2019): "Information Design:

Caplin, A., M. Dean, and J. Leahy (2018): "Rational Inattention, Optimal Consideration Sets and Stochastic Choice," forthcoming at *Review of Economic Studies.*

Cardell, N.S. (1997): "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity," *Econometric Theory*, 13(2), 185–213.

Cattaneo, M., X. Ma, Y. Masatlioglu, and E. Suleymanov (2019): "A Random Attention Model," arXiv:1712.03448.

Conlon, C.T., and J.H. Mortimer (2013): "Demand Estimation Under Incomplete Product Availability," *American Economic Journal: Microeconomics*, 5(4), 1–30.

Crawford, G.S., R. Griffith, and A. Iaria (2019): "Preference Estimation with Unobserved Choice Set Heterogeneity using Sufficient Sets," Working Paper.

Csaba, D. (2018): "Attentional Complements," Working Paper.

Dardanoni, V., P. Manzini, M. Mariotti, and C.J. Tyson (2018): "Inferring Cognitive Heterogeneity from Aggregate Choices," Working Paper Series 1018, Department of Economics, University of Sussex Business School.

Fosgerau, M., E. Melo, A. de Palma, and M. Shum (2017): "Discrete Choice and Rational Inattention: a General Equivalence Result," arXiv:1709.09117.

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(11), 3521–3557.

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

He, Y., S. Sinha, and X. Sun (2019): "Identification and Estimation in Many-to-One Two-Sided Matching without Transfers," Working Paper.

Honka, E. and P. Chintagunta (2016): "Simultaneous or Sequential? Search Strategies in the US Auto Insurance Industry," *Marketing Science*, 36(1), 21–42.

Kahneman, D. (1979): *Attention and Effort*, Prentice-Hall Englewood Cliffs, NJ.

Kamat, V. (2019): "Identification with Latent Choice Sets," arXiv:1711.02048.

Kamenica, E., and M. Gentzkow (2011): "Basyesian Persuasion," *American Economic Review*, 101(6), 2590–2615.

Lacetera, N., D.G. Pope, and J.R. Sydnor (2012): "Heuristic Thinking and Limited Attention in the Car Market," *American Economic Review*, 102(5), 2206–2236.

Lu, Z. (2018): "Estimating Multinomial Choice Models with Unobserved Choice Sets," Working Paper.

Magnolfi, L. and C. Roncoroni (2017): "Estimation of Discrete Games with Weak Assumptions on Information," Working Paper.

Matějka, F., and A. McKay (2015): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105(1), 272–298.

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.

Mehta, N., S. Rajiv, and K. Srinivasan (2003): "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation," *Marketing Science*, 22(1), 58–84.

Des 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(6), 2955–2980.

Simon, H.A. (1955): "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69(1), 99–118.

Simon, H.A. (1959): "Theories of Decision-Making in Economics and Behavioral Science," *American Economic Review*, 49(3), 253–83.

Sims, C.A. (1998): "Stickiness," *Carnegie-Rochester Conference Series on Public Policy*, 49(1), 317–356.

Sims, C.A. (2003): "Implications of Rational Inattention," *Journal of Monetary Economics* , 50(3), 665–690.

Sims, C.A. (2006): "Rational Inattention: Beyond the Linear-Quadratic Case," *American Economic Review*, 96(2), 158–163.

Syrgkanis, V., E. Tamer, and J. Ziani (2018): "Inference on Auctions with Weak Assumptions on Information," arXiv:1710.03830.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70(1), 147–165.

Wilson, C.J. (2008): "Consideration Sets and Political Choices: A Heterogeneous Model of Vote Choice and Sub-National Party Strength," *Political Behavior*, 30(2), 161–183.

A Proofs

Proof of Proposition 1 We proceed by construction. Take any $S \equiv (\mathcal{T}, \mathcal{P}_{T|X,\epsilon,V}) \in \mathcal{S}$. First, note that the set $\mathcal{Y}$ is finite and, hence, compact. Second, the map $y \in \mathcal{Y} \mapsto u(y, x, e, v) \in \mathbb{R}$ is continuous using the discrete metric for each $x \in \mathcal{X}$, $e \in \mathcal{E}$, and $v \in \mathcal{V}$. Given the linearity of integrals, the map $y \mapsto \int_{\mathcal{V}} u(y, x, e, v) P_{T|x,e,v}(t) P_{V|x,e}(v) dv$ is also continuous for each $x \in \mathcal{X}$, $e \in \mathcal{E}$, and $t \in \mathcal{T}$. Therefore, Weierstrass theorem ensures the existence of the minimum and maximum of such a map. Given $x \in \mathcal{X}$, $e \in \mathcal{E}$, and $t \in \mathcal{T}$, let $y^*_{x,e,t} \in \mathcal{Y}$ be one of the maximisers. Then, an optimal strategy is $\mathcal{P}_{Y|X,\epsilon,T}$ such that for each $x \in \mathcal{X}$, $e \in \mathcal{E}$, and $t \in \mathcal{T}$,

$$P_{Y|X,\epsilon,T}(y^*_{x,e,t}|x, e, t) = 1 \text{ and } P_{Y|X,\epsilon,T}(\tilde{y}|x, e, t) = 0 \ \forall \tilde{y} \in \mathcal{Y} \setminus \{y^*_{x,e,t}\}.$$

Proof of Proposition 2 Take any $\theta \in \Theta$ and $x \in \mathcal{X}$. We show that if $P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}$, then $P_{Y|X}(\cdot|x) \in \bar{\mathcal{R}}^{\theta}_{Y|x}$. If $P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}$, then, by definition of $\bar{\mathcal{Q}}^{\theta}_{Y|x}$, there exists $\mathcal{P}_{Y,V|X,\epsilon} \in \bar{\mathcal{Q}}^{\theta}$ inducing $P_{Y|X}(\cdot|x)$. By Theorem 1, it follows that there exists $S \in \mathcal{S}$ and $\mathcal{P}_{Y|X,\epsilon,T} \in \mathcal{R}^{\theta,S}$ such that $\mathcal{P}_{Y|X,\epsilon,T}$ induces $\mathcal{P}_{Y,V|X,\epsilon}$. Thus, $\mathcal{P}_{Y|X,\epsilon,T}$ induces $P_{Y|X}(\cdot|x)$ by the transitive property. Therefore, by definition of $\bar{\mathcal{R}}^{\theta}_{Y|x}$, $P_{Y|X}(\cdot|x) \in \bar{\mathcal{R}}^{\theta}_{Y|x}$.

Conversely, we show that $P_{Y|X}(\cdot|x) \in \bar{\mathcal{R}}^{\theta}_{Y|x}$, then $P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}$. First, let $\tilde{\mathcal{R}}^{\theta}_{Y|x} \subseteq \bar{\mathcal{R}}^{\theta}_{Y|x}$ be the non-convexified collection of probability mass functions of Y conditional on the realisation x of X that are induced by the model's optimal strategies under θ. That is,

$$\begin{aligned}\tilde{\mathcal{R}}^{\theta}_{Y|x} \equiv \Big\{ & P_{Y|X}(\cdot|x) \in \Delta(\mathcal{Y}) : \\ & P_{Y|X}(y|x) = \int_{\mathcal{T}\times\mathcal{V}\times\mathcal{E}} P_{Y|X,\epsilon,T}(y|x, e, t) P_{T|X,\epsilon,V}(t|x, e, v) P^{\theta}_{V|X,\epsilon}(v|x, e) P^{\theta}_{\epsilon|X}(e|x) d(t, v, e) \ \forall y \in \mathcal{Y}, \\ & \qquad \mathcal{P}_{Y|X,\epsilon,T} \in \mathcal{R}^{\theta,S}, S \in \mathcal{S} \Big\}.\end{aligned}$$

Take $P_{Y|X}(\cdot|x) \in \tilde{\mathcal{R}}^{\theta}_{Y|x}$. Then, by definition of $\tilde{\mathcal{R}}^{\theta}_{Y|x}$, there exists $S \in \mathcal{S}$ and $\mathcal{P}_{Y|X,\epsilon,T} \in \mathcal{R}^{\theta,S}$ such that $\mathcal{P}_{Y|X,\epsilon,T}$ induces $P_{Y|X}(\cdot|x)$. By Theorem 1, it follows that there exists $\mathcal{P}_{Y,V|X,\epsilon} \in \mathcal{Q}^{\theta}$ inducing $\mathcal{P}_{Y|X,\epsilon,T}$. Thus, $\mathcal{P}_{Y,V|X,\epsilon}$ induces $P_{Y|X}(\cdot|x)$ by the transitive property. Hence, by definition of $\bar{\mathcal{Q}}^{\theta}_{Y|x}$, $P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}$. Now, take any K elements from $\tilde{\mathcal{R}}^{\theta}_{Y|x}$, for any K. Denote such elements by $P^1_{Y|X}(\cdot|x) \in \tilde{\mathcal{R}}^{\theta}_{Y|x}, ..., P^K_{Y|X}(\cdot|x) \in \tilde{\mathcal{R}}^{\theta}_{Y|x}$. Given the arguments above, it holds that $P^1_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|X}(\cdot|x), ..., P^K_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}$. Moreover, any convex combination of $P^1_{Y|X}(\cdot|x), ..., P^K_{Y|X}(\cdot|x)$ belongs to $\bar{\mathcal{Q}}^{\theta}_{Y|x}$ because $\bar{\mathcal{Q}}^{\theta}_{Y|x}$ is convex. Therefore, every $P_{Y|X}(\cdot|x) \in \bar{\mathcal{R}}^{\theta}_{Y|x}$ is also contained in $\bar{\mathcal{Q}}^{\theta}_{Y|x}$.

One can conclude that $\bar{\mathcal{R}}^{\theta}_{Y|x} = \bar{\mathcal{Q}}^{\theta}_{Y|x}$ $\forall \theta \in \Theta$ and $\forall x \in \mathcal{X}$. This implies $\Theta^* = \Theta^{**}$.

Proof of Proposition 3 Step 1 shows Proposition 3 (i). Step 2 shows Proposition 3 (ii). Step 3 shows Proposition 3 (iii). Similar derivations are discussed in Section 3.3. and Appendix A of Magnolfi and Roncoroni (2017) for an entry game setting.

Step 1 Fix any $\theta \in \Theta$ and $x \in \mathcal{X}$. By convexity of $\bar{\mathcal{Q}}^{\theta}_{Y|x}$,

$$P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x} \Leftrightarrow b^T P^0_{Y|X}(\cdot|x) - \sup_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} b^T P_{Y|X}(\cdot|x) \leq 0 \ \forall b \in \mathbb{R}^{|\mathcal{Y}|}. \tag{A.1}$$

By the positive homogeneity of the support function, $\forall b \in \mathbb{R}^{|\mathcal{Y}|}$,

$$b^T P^0_{Y|X}(\cdot|x) - \sup_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} b^T P_{Y|X}(\cdot|x) \leq 0 \Leftrightarrow \frac{b^T}{||b||} P^0_{Y|X}(\cdot|x) - \sup_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} \frac{b^T}{||b||} P_{Y|X}(\cdot|x) \leq 0. \tag{A.2}$$

Hence, by (A.2), (A.1) is equivalent to

$$P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x} \Leftrightarrow b^T P^0_{Y|X}(\cdot|x) - \sup_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} b^T P_{Y|X}(\cdot|x) \leq 0 \ \forall b \in \mathbb{B}^{|\mathcal{Y}|}. \tag{A.3}$$

Note that

$$b^T P^0_{Y|X}(\cdot|x) - \sup_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} b^T P_{Y|X}(\cdot|x) = 0 \text{ when } b = 0_{|\mathcal{Y}|}. \tag{A.4}$$

Hence, by (A.4), (A.3) is equivalent to

$$P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x} \Leftrightarrow \max_{b \in \mathbb{B}^{|\mathcal{Y}|}} [b^T P^0_{Y|X}(\cdot|x) - \sup_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} b^T P_{Y|X}(\cdot|x)] = 0. \tag{A.5}$$

Lastly, given that $\bar{\mathcal{Q}}^{\theta}_{Y|x}$ is closed and bounded, (A.5) is equivalent to

$$P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x} \Leftrightarrow \max_{b \in \mathbb{B}^{|\mathcal{Y}|}} [b^T P^0_{Y|X}(\cdot|x) - \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} b^T P_{Y|X}(\cdot|x)] = 0. \tag{A.6}$$

Also, $\forall b \in \mathbb{R}^{|\mathcal{Y}|}$,

$$- \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} b^T P_{Y|X}(\cdot|x) = \min_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} -b^T P_{Y|X}(\cdot|x). \tag{A.7}$$

Hence, by (A.7), (A.6) is equivalent to

$$P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x} \Leftrightarrow \max_{b \in \mathcal{B}^{|\mathcal{Y}|}} \min_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} b^T [P^0_{Y|X}(\cdot|x) - P_{Y|X}(\cdot|x)] = 0 \tag{A.8}$$

Therefore, by combining Proposition 2 with (A.8), we get that

$$\theta \in \Theta^* \Leftrightarrow \max_{b \in \mathcal{B}^{|\mathcal{Y}|}} \min_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} b^T [P^0_{Y|X}(\cdot|x) - P_{Y|X}(\cdot|x)] = 0 \ \forall x \in \mathcal{X}, \tag{A.9}$$

which concludes our proof of Proposition 3 (i).

Step 2 We first introduce some useful notation. For each $\theta \in \Theta$, $x \in \mathcal{X}$, $e \in \mathcal{E}$, and $P_{Y,V|X,\epsilon}(\cdot|x,e) \in \Delta(\mathcal{Y} \times \mathcal{V})$, we rearrange the one-to-one image set of the mapping $(y,v) \in \mathcal{Y} \times \mathcal{V} \mapsto P_{Y,V|X,\epsilon}(\cdot|x,e) \in \Delta(\mathcal{Y} \times \mathcal{V})$ into a $(|\mathcal{Y}| \cdot |\mathcal{V}|) \times 1$ dimensional vector. With some abuse of notation, we still denote such a vector by $P_{Y,V|X,\epsilon}(\cdot|x,e)$. Moreover, for simplicity of exposition and without loss of generality, we assume that ϵ and V are jointly independent of X and that V is independent of (X,ϵ). The proof can be replicated without such independence, at the cost of increasing notational complexity.

Fix any $\theta \in \Theta$ and $x \in \mathcal{X}$. Consider the max-min problem on the right-hand-side of (A.8),

$$\max_{b \in \mathcal{B}^{|\mathcal{Y}|}} \min_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^{\theta}_{Y|x}} b^T [P^0_{Y|X}(\cdot|x) - P_{Y|X}(\cdot|x)] \tag{A.10}$$

By using Definition 3 to write explicitly the feasible set $\bar{\mathcal{Q}}^{\theta}_{Y|x}$, (A.10) is

$$\begin{aligned}
\max_{b \in \mathbb{R}^{|\mathcal{Y}|}} \quad & \min_{\substack{P_{Y|X}(\cdot|x) \in \mathbb{R}^{|\mathcal{Y}|}_+ \\ P_{Y,V|X,\epsilon}(\cdot|x,e) \in \mathbb{R}^{|\mathcal{Y}|\cdot|\mathcal{V}|}_+, \forall e \in \mathcal{E}}} && b^T [P^0_{Y|X}(\cdot|x) - P_{Y|X}(\cdot|x)], \\
\text{s.t.} \quad & [b \in \mathbb{B}^{|\mathcal{Y}|}]: && b^T b \le 1, \\
& \text{[1BCE-Consistency]:} && \sum_{y \in \mathcal{Y}} P_{Y,V|X,\epsilon}(y,v|x,e) = P^{\theta}_V(v) \ \forall v \in \mathcal{V}, \forall e \in \mathcal{E}, \\
& \text{[1BCE-Obedience]:} && -\sum_{v \in \mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e)[u^{\theta}(y,x,e,v) - u^{\theta}(y',x,e,v)] \le 0 \\
& && \forall y \in \mathcal{Y}, \forall y' \in \mathcal{Y} \setminus \{y\}, \forall e \in \mathcal{E}, \\
& \text{[1BCE-model predictions]:} && P_{Y|X}(y|x) = \sum_{(e,v) \in \mathcal{E} \times \mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e) P^{\theta}_{\epsilon}(e) \ \forall y \in \mathcal{Y}, \\
& \text{[Probability requirements]:} && \sum_{(y,v) \in \mathcal{Y} \times \mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e) = 1 \ \forall e \in \mathcal{E}.
\end{aligned} \tag{A.11}$$

We simplify (A.11) by introducing new variables. Let Z_1 be the $|\mathcal{Y}| \times 1$ vector $P^0_{Y|X}(\cdot|x) - P_{Y|X}(\cdot|x)$. Let Z_2 be the $(|\mathcal{Y}| \cdot |\mathcal{V}| \cdot |\mathcal{E}|) \times 1$ vector collecting $P_{Y,V|X,\epsilon}(\cdot|x,e)$ for each $e \in \mathcal{E}$. Lastly, let Z be the $(|\mathcal{Y}| + |\mathcal{Y}| \cdot |\mathcal{V}| \cdot |\mathcal{E}|) \times 1$ vector collecting Z_1 and Z_2. Given that $\bar{\mathcal{Q}}^{\theta}_{Y|x}$ is a subset of the $(|\mathcal{Y}| - 1)$-dimensional simplex, (A.11) can be rewritten as

$$\begin{aligned}
\max_{b \in \mathbb{R}^{|\mathcal{Y}|-1}} \ & \min_{\substack{Z_1 \in \mathbb{R}^{|\mathcal{Y}|} \\ Z_2 \in \mathbb{R}^{|\mathcal{Y}|\cdot|\mathcal{V}|\cdot|\mathcal{E}|}_+}} \begin{bmatrix} b^T & 0 & 0^T_{|\mathcal{Y}|\cdot|\mathcal{V}|\cdot|\mathcal{E}|} \end{bmatrix} Z, \\
\text{s.t. } & b^T b \le 1, \\
& A_{\text{eq}}\, Z = B_{\text{eq}}, \\
& A_{\text{ineq}}\, Z \le 0_{d_{\text{ineq}}},
\end{aligned} \tag{A.12}$$

where A_{eq} is the matrix of coefficients multiplying Z in the equality constraints of (A.11) with d_{eq} rows, B_{eq} is the vector of constants appearing in the equality constraints of (A.11), and A_{ineq} is the matrix of coefficients multiplying Z in the inequality constraints of (A.11) with

d_{ineq} rows.

Further, the inner constrained minimisation problem in (A.12) is linear. Hence, by strong duality, can be replaced with its dual. This allows us to solve one unique maximisation problem. Precisely, the solution of (A.12) is equivalent to the solution of

$$\begin{aligned} \max_{\substack{b \in \mathbb{R}^{|\mathcal{Y}|-1} \\ \lambda_{\text{eq}} \in \mathbb{R}^{d_{\text{eq}}} \\ \lambda_{\text{ineq}} \in \mathbb{R}_+^{d_{\text{ineq}}}}} & \begin{bmatrix} -B_{\text{eq}}^T & 0_{d_{\text{ineq}}}^T \end{bmatrix} \lambda, \\ \text{s.t. } & b^T b \leq 1, \\ & [A^T]_{1:|\mathcal{Y}|} \lambda = \begin{pmatrix} -b \\ 0 \end{pmatrix}, \\ & -[A^T]_{|\mathcal{Y}|+1:|\mathcal{Y}|+|\mathcal{Y}|\cdot|\mathcal{V}|\cdot|\mathcal{E}|} \lambda \leq 0_{|\mathcal{Y}|\cdot|\mathcal{V}|\cdot|\mathcal{E}|}, \end{aligned} \tag{A.13}$$

where λ is the $(d_{\text{eq}} + d_{\text{ineq}}) \times 1$ vector collecting λ_{eq} and λ_{ineq}, A is the $(d_{\text{eq}} + d_{\text{ineq}}) \times (|\mathcal{Y}| + |\mathcal{Y}| \cdot |\mathcal{V}| \cdot |\mathcal{E}|)$ matrix obtained by stacking one on top of the other the matrices A_{eq} and A_{ineq}, and $[A]_{i:j}$ denotes the sub-matrix of A containing the rows $i, i+1, ..., j$ of A.

Therefore, the solution of (A.10) is equivalent to the solution of (A.13). Moreover, (A.13) is a quadratically constrained linear maximisation problem. In particular, the first constraint in (A.13) is quadratic. The objective function and the remaining constraints in (A.13) are linear.

Step 3 Fix any $\theta \in \Theta$ and $x \in \mathcal{X}$. (A.1) is equivalent to

$$P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x} \Leftrightarrow -b^T P^0_{Y|X}(\cdot|x) + \sup_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T P_{Y|X}(\cdot|x) \geq 0 \ \forall b \in \mathbb{R}^{|\mathcal{Y}|}. \tag{A.14}$$

By (A.2), (A.14) is equivalent to

$$P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x} \Leftrightarrow -b^T P^0_{Y|X}(\cdot|x) + \sup_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T P_{Y|X}(\cdot|x) \geq 0 \ \forall b \in \mathbb{S}^{|\mathcal{Y}|}. \tag{A.15}$$

Moreover, given that $\bar{\mathcal{Q}}^\theta_{Y|x}$ is closed and bounded, (A.15) is equivalent to

$$P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x} \Leftrightarrow -b^T P^0_{Y|X}(\cdot|x) + \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T P_{Y|X}(\cdot|x) \geq 0 \ \forall b \in \mathbb{S}^{|\mathcal{Y}|}. \tag{A.16}$$

Lastly, given that $\bar{\mathcal{Q}}^\theta_{Y|x}$ is a subset of the $(|\mathcal{Y}|-1)$-dimensional simplex, (A.16) is equivalent to

$$-b^T \begin{pmatrix} P^0_{Y|X}(y^1|x) \\ \vdots \\ P^0_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix} + \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T \begin{pmatrix} P_{Y|X}(y^1|x) \\ \vdots \\ P_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix} \geq 0 \ \forall b \in \mathbb{S}^{|\mathcal{Y}|-1}. \tag{A.17}$$

Therefore, by combining Proposition 2 with (A.17), we get that

$$\theta \in \Theta^* \Leftrightarrow -b^T \begin{pmatrix} P^0_{Y|X}(y^1|x) \\ \vdots \\ P^0_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix} + \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T \begin{pmatrix} P_{Y|X}(y^1|x) \\ \vdots \\ P_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix} \geq 0 \ \forall b \in \mathbb{S}^{|\mathcal{Y}|-1}. \quad \text{(A.18)}$$

which concludes our proof of Proposition 3 (iii).

B Semiparametric case

Suppose that the researcher is interested in identifying only the vector of payoff parameters, $\theta_{1,0}$. In such a case, the researcher can proceed without parameterising the probability mass functions contained in $\mathcal{P}_{V|X,\epsilon}$ and $\mathcal{P}_{\epsilon|X}$. Further, determining whether a candidate vector of parameters belongs to the sharp identified sets amounts to solving some linear programming problems. We formalise these arguments below.

Let $\theta_0 \equiv \theta_{1,0}$. Following Proposition 2, the sharp identified set for θ_0 can be characterised as

$$\Theta^* = \{\theta \in \Theta : \ \exists \ \mathcal{P}_{\epsilon|X} \text{ and } \mathcal{P}_{V|X,\epsilon} \text{ s.t. } P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}_x^{\theta, \mathcal{P}_{\epsilon|X}, \mathcal{P}_{V|X,\epsilon}} \ \forall x \in \mathcal{X}\}. \quad \text{(B.1)}$$

Proposition B.1 explains how one can construct Θ^*.

Proposition B.1. (*Construction of* Θ^**)* For each $\theta \in \Theta$ and $x \in \mathcal{X}$, consider the following linear program with unknowns $P_{Y|X}(\cdot|x) \in \Delta(\mathcal{Y})$, $P_{Y,V|X,\epsilon}(\cdot|x,e) \in \Delta(\mathcal{Y} \times \mathcal{V}) \ \forall e \in \mathcal{E}$, $P_{V|x,e}(\cdot|x,e) \in \Delta(\mathcal{V}) \ \forall e \in \mathcal{E}$, and $P_{\epsilon|X}(\cdot|x) \in \Delta(\mathcal{E})$:

$$\begin{aligned}
&\text{[Data=model predictions]} && P^0_{Y|X}(y|x) = P_{Y|X}(y|x) \ \forall y \in \mathcal{Y}, \\
&\text{[1BCE-Consistency]:} && \sum_{y \in \mathcal{Y}} P_{Y,V|X,\epsilon}(y,v|x,e) = P_{V|X,\epsilon}(v|x,e) \ \forall v \in \mathcal{V}, \forall e \in \mathcal{E}, \\
&\text{[1BCE-Obedience]:} && -\sum_{v \in \mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e)[u^\theta(y,x,e,v) - u^\theta(y',x,e,v)] \leq 0 \\
& && \forall y \in \mathcal{Y}, \forall y' \in \mathcal{Y} \setminus \{y\}, \forall e \in \mathcal{E}, \\
&\text{[1BCE-model predictions]:} && P_{Y|X}(y|x) = \sum_{(e,v) \in \mathcal{E} \times \mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e) P_{\epsilon|X}(e|x) \ \forall y \in \mathcal{Y}, \\
&\text{[Probability requirements]:} && \sum_{(y,v) \in \mathcal{Y} \times \mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e) = 1 \ \forall e \in \mathcal{E}.
\end{aligned} \quad \text{(B.2)}$$

Under Assumption 1, for each $\theta \in \Theta$,

$$\theta \in \Theta^* \ \Leftrightarrow \ \text{(B.2) is feasible } \forall x \in \mathcal{X}.$$

$\diamond$

Proposition B.1 states that, in order to determine whether a candidate vector of parameters belongs to Θ^*, the researcher should determine whether the linear program (B.2) is feasible.

A similar result is discussed in Section 2 of Syrgkanis, Tamer, and Ziani (2018) for an auction setting. We now provide the proof of Proposition B.1.

Proof of Proposition B.1 By (B.1), a candidate vector of parameters, $\theta \in \Theta$, belongs to Θ^* if and only if, for each $x \in \mathcal{X}$, there exists $P_{Y|X}(\cdot|x) \in \Delta(\mathcal{Y})$, $P_{Y,V|X,\epsilon}(\cdot|x,e) \in \Delta(\mathcal{Y} \times \mathcal{V})$ $\forall e \in \mathcal{E}$, $P_{V|x,e}(\cdot|x,e) \in \Delta(\mathcal{V})$ $\forall e \in \mathcal{E}$, and $P_{\epsilon|X}(\cdot|x) \in \Delta(\mathcal{E})$ such that $P_{Y|X}(\cdot|x) = P^0_{Y|X}(\cdot|x)$ and $P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}_x^{\theta,\mathcal{P}_{\epsilon|X},\mathcal{P}_{V|X,\epsilon}}$. Further, by Definition 3, the condition $P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}_x^{\theta,\mathcal{P}_{\epsilon|X},\mathcal{P}_{V|X,\epsilon}}$ consists of these constraints:

$$\begin{aligned}
&\text{[1BCE-Consistency]:} && \sum_{y\in\mathcal{Y}} P_{Y,V|X,\epsilon}(y,v|x,e) = P_{V|X,\epsilon}(v|x,e)\ \forall v \in \mathcal{V}, \forall e \in \mathcal{E},\\
&\text{[1BCE-Obedience]:} && -\sum_{v\in\mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e)[u^\theta(y,x,e,v) - u^\theta(y',x,e,v)] \le 0\\
& && \forall y \in \mathcal{Y}, \forall y' \in \mathcal{Y} \setminus \{y\}, \forall e \in \mathcal{E}, \qquad\qquad (B.3)\\
&\text{[1BCE-model predictions]:} && P_{Y|X}(y|x) = \sum_{(e,v)\in\mathcal{E}\times\mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e) P_{\epsilon|X}(e|x)\ \forall y \in \mathcal{Y},\\
&\text{[Probability requirements]:} && \sum_{(y,v)\in\mathcal{Y}\times\mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e) = 1\ \forall e \in \mathcal{E}.
\end{aligned}$$

Therefore, $\theta \in \Theta^*$ if and only, for each $x \in \mathcal{X}$, there exists $P_{Y|X}(\cdot|x) \in \Delta(\mathcal{Y})$, $P_{Y,V|X,\epsilon}(\cdot|x,e) \in \Delta(\mathcal{Y} \times \mathcal{V})$ $\forall e \in \mathcal{E}$, $P_{V|x,e}(\cdot|x,e) \in \Delta(\mathcal{V})$ $\forall e \in \mathcal{E}$, and $P_{\epsilon|X}(\cdot|x) \in \Delta(\mathcal{E})$ such that $P_{Y|X}(\cdot|x) = P^0_{Y|X}(\cdot|x)$ and (B.3) is satisfied. This amounts to checking whether (B.2) is feasible for each $x \in \mathcal{X}$.