Identification in discrete choice models with imperfect information

Cristina Gualdani Shruti Sinha

December 2020

Abstract

We study identification of preferences in a class of single-agent, static, discrete choice models where decision makers may be imperfectly informed about the payoffs generated by the available alternatives. Decision makers rely on a common family of priors. They are allowed to update their priors by processing any possible signals, on which we remain agnostic. As we remain agnostic about signals, our methodology correctly recovers preferences under both perfect information and a range of models with information frictions. We leverage on the notion of one-player Bayes Correlated Equilibrium in Bergemann and Morris (2016) to provide a tractable characterisation of the identified set. We use our methodology and data on the 2017 UK general election to estimate a spatial model of voting under weak assumptions on the information that voters have about the returns to voting. Counterfactual exercises quantify the consequences of imperfect information in politics.

Keywords: Discrete choice model, Bayesian Persuasion, Bayes Correlated Equilibrium, Incomplete Information, Partial Identification, Moment Inequalities, Spatial Model of Voting.

JEL codes: C01, C25, D72, D80.

Acknowledgments : We would like to thank Andrew Caplin, Christian Hellwig, Vishal Kamat, Niccoló Lomys, and Thierry Magnac for their useful comments. We would also like to thank seminar participants at TSE, UCL, Sciences Po Paris, Tinbergen Institute Amsterdam, University of Cambridge, University of Essex, Queen Mary University of London, and Aarhus University. 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.

Authors' affiliations and email: Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France (cristina.gualdani@tse-fr.eu); Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France (shruti.sinha@tse-fr.eu).

1 Introduction

When facing decision problems, agents may encounter frictions which prevent them from learning the payoffs associated with the available alternatives. These frictions (hereafter, *information frictions*) stem from various sources, such as attentional and cognitive limits, financial constraints, spatial and temporal boundaries, cultural biases, and personal inclinations. Ideally, one would like to take into account information frictions in the empirical analysis of decision problems in order to evaluate how choices would respond to additional information. However, this is a challenging task because information frictions can be heterogenous across agents and are typically unobserved by the researcher.

A large part of the applied literature on decision problems ignores information frictions and imposes perfect information. A recent strand of the literature incorporates information frictions by *fully* specifying what are the beliefs of decision makers about their payoffs. This includes search models (for example, Mehta, Rajiv, and Srinivasan, 2003; Honka and Chintagunta, 2016; Ursu, 2018), models with rational inattention (for example, Caplin and Dean, 2015; Matějka and McKay, 2015; Fosgerau, Melo, de Palma, and Shum, 2017; Csaba, 2018; Caplin, Dean, and Leahy, 2019b; Brown and Jeon, 2020), and models with preferences for risk (for a review, see Barseghyan, Molinari, O'Donoghue, and Teitelbaum, 2018). In this paper, we impose weak restrictions on what are the agents' beliefs and develop a methodology to identify preferences from cross-sectional data whether agents are perfectly or partially informed.

More formally, we consider a static setting where the decision maker (hereafter, DM) has to choose an alternative from a finite set. The utility generated by each alternative is determined by the state of the world. The state of the world is defined by variables like attributes of the available alternatives, attributes and tastes of the DM, and exogenous market shocks. The DM chooses an alternative, possibly without being fully aware of the state of the world. However, the DM has a prior on it. Moreover, the DM has the opportunity to refine her prior by processing additional information. Such additional information takes the form of a signal randomly drawn from a certain distribution and is hereafter referred to as the *information structure* of the DM. This information structure can range from full revelation of the state of the world to no information whatsoever, depending on the information frictions encountered by the DM in the learning process. As the information frictions are unobserved by the researcher, our analysis proceeds by leaving the information structure processed by the DM completely unrestricted. The DM uses the acquired information structure to update her prior and obtain a posterior through the Bayes rule. Finally, the DM chooses an alternative maximising her expected utility, where the expectation is computed via the posterior.

Note that several discrete choice models studied in the literature - such as (Multinomial) Logit/Probit model, Nested Logit model, Mixed Logit model, discrete choice models with risk aversion, discrete choice models with rational inattention, and some discrete choice models with search - can be obtained within the above framework under additional assumptions on the information structure processed by the DM.

We assume that the researcher has data on choices made by many DMs facing the above decision problem and, possibly, on some covariates which are part of (or coincide with) the state of the world.[1] DMs rely on a common family of priors, but are allowed to process arbitrarily different and completely unrestricted information structures. Hence, they can base their decisions on any Bayes-consistent posteriors. Our objective is to study identification of the utility function and the common family of priors. Identifying such objects is useful, for instance, to evaluate how the choice probabilities change in response to changes in the availability of information about the state of the world and, in turn, assess the associated welfare benefits prior to conducting such intervention.

Studying identification of the preference parameters while remaining agnostic about information structures is challenging because the model is incomplete in the sense of Tamer (2003), thus raising the possibility of partially identified preference parameters. Tractably characterising the identified set is not an easy task. In fact, in order to determine whether a given parameter value belongs to the identified set, we need to establish whether the empirical choice probabilities belong to the collection of choice probabilities predicted by our model under a large range of possible information structures. The difficulty here lies in the necessity of exploring such a range of possible information structures because it is intractably large.

We approach the above problem by revisiting our framework through the lens of one-player Bayes Correlated Equilibrium (Bergemann and Morris, 2013; 2016). The concept of one-player Bayes Correlated Equilibrium is based on a theoretical setting where an omniscient mediator makes incentive-compatible recommendations to DMs as a function of the state of the world and compatible with their priors. If DMs follow such recommendations, then the resulting distribution of choices is a one-player Bayes Correlated Equilibrium. The concept of one-player Bayes Correlated Equilibrium is a powerful tool because it provides behavioural predictions that do not depend on (and, thus, are robust to) the specification of information structures. In particular, Theorem 1 in Bergemann and Morris (2016) shows that the collection of choice probabilities predicted by our model under a large range of possible information structures is equivalent to the collection of choice probabilities predicted by our model under the notion of one-player Bayes Correlated Equilibrium. Further, the latter collection is a convex set. Therefore, determining whether a given parameter value belongs to the identified set can be rewritten as a linear programming problem. In turn, constructing such identified set becomes a computationally tractable exercise.

Our framework is applicable to several settings, such as health (when choosing a pharmaceutical product, DMs might be uncertain about the equivalence between generics and branded prescription drugs), education (when choosing an educational career, DMs might be uncertain

[1] Our methodology allows for the state of the world to be fully observed, partly observed, or fully unobserved by the researcher. Further, we do not require the researcher to know more (or, less) than the DM about the state of the world.

about the returns to schooling), environment (when choosing a transport mode, DMs might be uncertain about the associated carbon footprints), and political economy (when voting in an election, DMs might be uncertain about the returns to voting for the various parties). In our empirical application we focus on the latter setting.

More precisely, we consider the spatial model of voting, which is an important framework in political economy to explain individual preferences for parties (Downs 1957; Black, 1958; Davis, Hinich, and Ordeshook, 1970; Enelow and Hinich 1984; Hinich and Munger, 1994). This model postulates that an agent has a most preferred policy and votes for the party whose position is closest to her ideal (i.e., she votes "ideologically"). In empirical analysis, it is typically implemented by estimating a classical parametric discrete choice model with perfect information (Alvarez and Nagler, 1995; 1998; 2000; Alvarez, Nagler, and Bowler, 2000). However, in reality, uncertainty pervades voting. That is, voters may be aware of their own and the parties' attitudes towards some popular issues, but they might be less prepared on how they themselves and the parties stand towards more technical or less debated topics, and on the traits of the candidates other than those publicly advertised. Further, their competence on these matters is likely to be arbitrarily different, depending, for example, on political sentiment, civic sense, intellectual preparation, attentional limits, media exposure, and the transparency of candidates. Our methodology allows us to incorporate such frictions in an empirical spatial voting framework under weak assumptions on the latent, heterogeneous, and potentially endogenous process followed by voters to gather and evaluate information.

In particular, we focus on a setting where the state of the world consists of distances between the voters and the parties' ideological positions on a few popular policy issues (for simplicity, the *systematic component*), and of voter-party-specific taste variables capturing evaluations of the candidates' qualities and of the parties' opinions on more complicated and less media-covered topics (for simplicity, the *idiosyncratic component*). We assume that each voter observes the realisation of the systematic component, but may be uncertain about the realisation of the idiosyncratic component. We estimate such a model using data from the British Election Study, 2017: Face-to-Face Post-Election Survey (Fieldhouse, et al., 2018) on the UK general election held on 8 June 2017. We compare our findings with the results one gets under the standard assumption that all voters are fully informed about the idiosyncratic component, and under the assumption that all voters are fully uninformed about the idiosyncratic component. Several conclusions on the utility parameters that are achieved under those two information environments are not unambiguously corroborated when we remain agnostic about the information structures of voters.

To better interpret our results, we perform various counterfactual exercises. For example, we investigate to what extent voter uncertainty affects vote shares. We do that by imagining an omniscient mediator who implements a policy that gives voters perfect information about the state of the world. We simulate the counterfactual vote shares and study how they change with respect to the factual scenario. This question has been debated at length in the literature. Political scientists have often answered it by arguing that a large population composed of possibly uninformed citizens act as if it was perfectly informed (for a review, see Bartels, 1996). Carpini and Keeter (1996), Bartels (1996), and Degan and Merlo (2011) provide quantitative evidence to disconfirm such claims; the first two by using auxiliary data on the level of information of the survey respondents as rated by the interviewers or assessed by test items, and the latter by parametrically specifying the probability that a voter is informed. We contribute to this thread of the literature by providing a way to construct counterfactual vote shares under perfect information, which neither requires the difficult task of measuring the level of knowledge of voters in the factual scenario, nor imposes parametric assumptions on the probability that a voter is informed. Among the various results, we quantify the value of information to voters in the sense of Blackwell and Girshick (1954) as captured through a reduction in abstention. We also find that the "losers" from the increase in voter awareness are the two historically dominating parties, i.e., the Conservative Party and the Labour Party. The observed changes in vote shares suggest that policy initiatives in the direction of perfect information (for example, transparency laws) can increase voter welfare by reducing ex-post regret.

Literature review Research questions similar to ours have been addressed so far in the empirical literature by fully specifying what are the beliefs of DMs about their payoffs. Some examples are search models, models with rational inattention, and models with preferences for risk (for references, see above). Here, instead, we impose weak restrictions on such beliefs by simply requiring them to be Bayes-consistent with a common family of priors. We further remark on an important difference between our framework and most of the search models considered in empirical work. In search models, individuals typically acquire perfect information about the option they choose. Instead, our framework allows for the possibility that individuals acquire partial information about any of the options in their choice set. This is reasonable when agents have to choose among "complex" products, such as a political party, an education career, or an insurance plan. Finally, a more precise connection between our framework and models with rational inattention is established in Example 3 of Section 2.

This paper also relates to the econometric literature on discrete choice models when the sets of alternatives actually considered by agents (hereafter, *consideration sets*) could be subsets of the entire set of alternatives, heterogenous, arbitrarily correlated with the payoff-relevant variables, and latent (for some recent contributions see, for example, Abaluck and Adams, 2018; Barseghyan, Coughlin, Molinari, and Teitelbaum, 2019; Barseghyan, Molinari, and Thirkettle, 2019; Cattaneo, Ma, Masatlioglu, and Suleymanov, 2019; Dardanoni, et al., 2020a;b). In fact, one implication of imperfect information at the level of payoffs is that agents may process information structures inducing them to contemplate, in equilibrium, only a subset of the available alternatives, ignoring all the others. Hence, in our model, consideration sets might arise endogenously (Caplin, Dean, and Leahy, 2019b).[2,3] Yet, there is an important difference between the literature on consideration sets and this paper. The consideration set literature focuses on recovering the consideration probabilities from the empirical choice probabilities, but parameterises the expected utilities. Instead, in this paper we allow the expected utilities to depend on any Bayes-consistent posteriors, but we do not recover the consideration probabilities. Thus, we can answer different types of questions.

This paper also relates to the literature concerned with evaluating the impact on choices of sending agents information about the state of the world (for example, Hastings and Tejeda-Ashton, 2008, studying retirement fund options in Mexico; Hastings and Weinstein, 2008 and Bettinger, et al., 2012, studying school choice; Kling, et al., 2012, studying Medicare Part D prescription drug plans in the United States). This literature typically exploits field experiments. Our methodology can offer complementary insights because it allows us to evaluate how the choice probabilities change in response to changes in the availability of information, prior to conducting such intervention.

More generally, this work relates to the literature concerned with relaxing assumptions about expectation formation and about the amount of information on which agents condition their expectations (see, for example, the seminal paper by Manski, 2004). In fact, by not restricting information structures, we allow agents to compute expected utilities with any Bayes-consistent posteriors. An interesting empirical paper that estimates a model of firm export participation under weak assumptions on firm expectation is Dickstein and Morales (2018), where non-sharp bounds are characterised using the revealed preference approach of Pakes, at al. (2015).

More recently, the notion of Bayes Correlated Equilibrium has been exploited to characterise the identified set in an entry game (Magnolfi and Roncoroni, 2017) and in an auction framework (Syrgkanis, Tamer, and Ziani, 2018). As in those two papers, we rely on Theorem 1 in Bergemann and Morris (2016), which establishes the robustness properties of Bayes Correlated Equilibrium and is applicable to any n-player games, including single-agent decision problems. Differently from those two papers, we investigate the identifying power of Bayes Correlated Equilibrium within a *stochastic multinomial choice* framework, where stochasticity is due to agents being uncertain about their *own* utility and deciding among alternatives based on noisy signals. Note that, when there is a single player, the framework of Magnolfi and Ron-

[2]Recall that the DM's information structure takes the form of a signal randomly drawn from a certain distribution. 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. More details are in Section 2.

[3]We remark that imperfect information at the level of payoffs is *not* the only mechanism that can induce endogenous consideration sets in decision problems. Consideration sets may arise also because of lack of awareness of some alternatives in the feasible set (for example, Goeree, 2008), deliberately ignoring some alternatives in the feasible set (for example, Wilson, 2008), incomplete product availability (for example, Conlon and Mortimer, 2014), being offered the possibility of receiving program access from outside an experiment (for example, Kamat, 2019), and absence of market clearing transfers in two-sided matching models (for example, He, Sinha, and Sun, 2019). These additional mechanisms leading to endogenous considerations sets are *not* nested in our framework.

coroni (2017) reduces to a binary choice model with perfect information, as the uncertainty that players face in their model depends only on the strategies adopted by the other players. Hence, our model is not nested by Magnolfi and Roncoroni (2017).

In our setting, the concept of Bayes Correlated Equilibrium is further related to the Bayesian Persuasion problem introduced in Kamenica and Gentzkow (2011). The Bayesian Persuasion problem consists of an information design problem, where the regulator picks an information structure to send to the DM. This is equivalent to selecting a one-player Bayes Correlated Equilibrium from the collection of admissible one-player Bayes Correlated Equilibria. For a discussion see Bergemann and Morris (2019).

This paper also contributes to the voting literature in political economy. There is a broad literature on spatial voting models (see references above and in Section 4). Further, there is a large body of work on uncertainty in voting (for example, Downs, 1957; Shepsle, 1972; Aldrich and McKelvey, 1977; Weisberg and Fiorina, 1980; Hinich and Pollard, 1981; Enelow and Hinich, 1981; Bartels, 1986; Baron, 1994; Alvarez and Nagler, 1995; Matsusaka 1995; Carpini and Keeter, 1996; Grossman and Helpman, 1996; Alvarez, 1998; Lupia and McCubbins, 1998; Feddersen and Pesendorfer, 1999; Degan and Merlo, 2011; Matějka and Tabellini, 2019). However, only a few empirical works have attempted to take into account voter sophistication while estimating a spatial voting framework (for example, Aldrich and McKelvey, 1977; Bartels, 1986; Palfrey and Poole, 1987; Franklin, 1991; Alvarez, 1998; Degan and Merlo, 2011; Tiemann, 2019). This has been done by exogenously and parametrically modelling how information frictions affect the perceptions of DMs about the returns to voting (for instance, via an additive, exogenous, and parametrically distributed evaluation error in the payoffs), or by parametrically specifying the probability of being informed versus uninformed when voting. Instead, our methodology permits us to incorporate uncertainty under weak assumptions on the latent, heterogeneous, and potentially endogenous process followed by voters to collect information.

The remainder of the paper is organised as follows. Section 2 describes the model. Section 3 discusses identification and some simulations. Section 4 presents the empirical application. Section 5 concludes. Proofs and further details are in the Appendices.

Notation Capital letters are used for random variables/vectors/matrices and lower case letters for their realisations. Calligraphic capital letters are used for sets. Given a set $\mathcal{Z} \subseteq \mathbb{R}^J$, $\Delta(\mathcal{Z})$ represents the collection of all possible densities/mixed joint densities/probability mass functions (depending on whether $\mathcal{Z}$ is finite or not) on $\mathcal{Z}$. An element of $\Delta(\mathcal{Z})$ is denoted by P_Z. When the nature of $\mathcal{Z}$ is unspecified, we generically refer to and treat P_Z as a density.

Consider two random variables, Z and X, with supports $\mathcal{Z}$ and $\mathcal{X}$, respectively. Given $x \in \mathcal{X}$, we denote the density of Z conditional on $X = x$ by $P_{Z|X}(\cdot|x) \in \Delta(\mathcal{Z})$. Further, we denote the family of densities of Z conditional on every realisation x of X by $\mathcal{P}_{Z|X}$, i.e., $\mathcal{P}_{Z|X} \equiv \{P_{Z|X}(\cdot|x)\}_{x\in\mathcal{X}}$.

The K-dimensional positive real space is denoted by $\mathbb{R}^K_+$. 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. 1_L is the $L \times 1$ vector of zeros. I_L is the $L \times L$ identity matrix.

$\mathbb{B}^{|\mathcal{Y}|}$ is the unit ball in $\mathbb{R}^{|\mathcal{Y}|}$, i.e., $\mathbb{B}^{|\mathcal{Y}|} \equiv \{b \in \mathbb{R}^{|\mathcal{Y}|} : b^T b \leq 1\}$. "$\times$" denotes the Cartesian product operator or is used to indicate vector dimensions. "$\cdot$" denotes the standard product operator.

2 The model

In this section we describe a class of single-agent, static, discrete choice models, where DM i may be partially aware of the utilities generated by the available alternatives.

Let DM i face the decision problem of choosing an alternative from a finite set, $\mathcal{Y}$, possibly under imperfect information about the state of the world. The state of the world consists of all the payoff-relevant variables. It can include, for example, attributes of the alternatives, attributes and tastes of DM i, and exogenous market shocks. It is represented by a vector, (x_i, e_i, v_i). We describe each component of this vector in Assumption 1 below.

Assumption 1. (*State of the world*)

1. x_i is a real vector (or, scalar) drawn at random from the density $P_X \in \Delta(\mathcal{X})$, where $\mathcal{X} \subseteq \mathbb{R}^{H_X}$ and H_X is the dimension of x_i. Hereafter, we denote by X_i the random vector (or, variable) with support $\mathcal{X}$ and density P_X. The realisation x_i of X_i is observed by DM i and the researcher.

2. e_i is a real vector (or, scalar) drawn at random from the conditional density $P_{\epsilon|X}(\cdot|x_i) \in \Delta(\mathcal{E})$, where $\mathcal{E} \subseteq \mathbb{R}^{H_\epsilon}$ and H_ϵ is the dimension of e_i. Hereafter, we denote by ϵ_i the random vector (or, variable) with support $\mathcal{E}$ and family of conditional densities $\mathcal{P}_{\epsilon|X} \equiv \{P_{\epsilon|X}(\cdot|x)\}_{x \in \mathcal{X}}$, with each $P_{\epsilon|X}(\cdot|x) \in \Delta(\mathcal{E})$. The realisation e_i of ϵ_i is observed by DM i but not by the researcher.

3. v_i is a real vector (or, scalar) drawn at random from the conditional density $P_{V|X,\epsilon}(\cdot|x_i, e_i) \in \Delta(\mathcal{V})$, where $\mathcal{V} \subseteq \mathbb{R}^{H_V}$ and H_V is the dimension of v_i. Hereafter, we denote by V_i the random vector (or, variable) with support $\mathcal{V}$ and family of conditional densities $\mathcal{P}_{V|X,\epsilon} \equiv \{P_{V|X,\epsilon}(\cdot|x, e)\}_{(x,e) \in \mathcal{X} \times \mathcal{E}}$, with each $P_{V|X,\epsilon}(\cdot|x, e) \in \Delta(\mathcal{V})$. The realisation v_i of V_i is not observed by DM i. v_i may or may not be observed by the researcher, depending on the specific setting at hand. Further, DM i has a correct *prior* on V_i conditional on $(X_i, \epsilon_i) = (x_i, e_i)$, which is $P_{V|X,\epsilon}(\cdot|x_i, e_i)$.

$\diamond$

Before making a choice, DM i has an opportunity to study better the state of the world and, in turn, resolve some uncertainty about the utilities generated by the alternatives. More precisely, DM i can refine her prior upon reception of a private signal which may or may not be informative about V_i, as specified by Assumption 2 below.

Assumption 2. (*Signal*) Before choosing an alternative from $\mathcal{Y}$, DM i receives a signal realisation, t_i, drawn at random from the conditional density $P^i_{T|X,\epsilon,V}(\cdot|x_i, e_i, v_i) \in \Delta(\mathcal{T}_i)$, where $\mathcal{T}_i \subseteq \mathbb{R}^{H_{T_i}}$ and H_{T_i} is the dimension of t_i. Hereafter, we denote by T_i the random vector (or, variable) with support $\mathcal{T}_i$ and family of conditional densities $\mathcal{P}^i_{T|X,\epsilon,V} \equiv \{P^i_{T|X,\epsilon,V}(\cdot|x, e, v)\}_{(x,e,v)\in\mathcal{X}\times\mathcal{E}\times\mathcal{V}}$, with each $P^i_{T|X,\epsilon,V}(\cdot|x, e, v) \in \Delta(\mathcal{T}_i)$. The realisation t_i of T_i is observed by DM i but not by the researcher. However, DM i does not know which conditional density t_i has been drawn from because she does not observe v_i. Instead, DM i is aware of the entire family of conditional densities, $\{P^i_{T|X,\epsilon,V}(\cdot|x_i, e_i, v)\}_{v\in\mathcal{V}}$. Hence, DM i uses t_i and $\{P^i_{T|X,\epsilon,V}(\cdot|x_i, e_i, v)\}_{v\in\mathcal{V}}$ to update $P_{V|X,\epsilon}(\cdot|x_i, e_i)$ via Bayes rule, and obtains the *posterior*, $P^i_{V|X,\epsilon,T}(\cdot|x_i, e_i, t_i) \in \Delta(\mathcal{V})$.[4] $\diamond$

Finally, DM i chooses alternative $y \in \mathcal{Y}$ maximising her expected utility computed under the posterior,

$$\int_{v\in\mathcal{V}} u(y, x_i, e_i, v) P^i_{V|X,\epsilon,T}(v|x_i, e_i, t_i) dv,$$

where $u : \mathcal{Y} \times \mathcal{X} \times \mathcal{E} \times \mathcal{V} \to \mathbb{R}$ is the utility function. If there is more than one maximising alternative (i.e., if there are ties), then DM i applies some tie-breaking rule. We provide a formal definition of the optimal strategy of DM i later in this section.

Before proceeding, we add a few remarks on our framework.

Remark 1. (*Information frictions*) The informativeness of signal T_i about V_i (in the Blackwell sense) is inherently related to the frictions potentially encountered by DM i while investigating the state of the world. These frictions can stem from various sources, such as attentional and cognitive limits, financial constraints, spatial and temporal boundaries, cultural and personal biases, or values taken by the known components of the state of the world. When these frictions are severe, DM i may decide not to inform herself better about the state of the world up to the point where the payoffs are known with certainty.

For example, if DM i faces no information frictions, then she may process a signal revealing the exact realisation of V_i. A possible representation of that when $\mathcal{V}$ is finite is

$$\mathcal{T}_i \equiv \mathcal{V}, \;\; P^i_{T|X,\epsilon,V}(v|x_i, e_i, v) = 1 \;\forall v \in \mathcal{V}, \tag{1}$$

for a given realisation (x_i, e_i) of (X_i, ϵ_i). Instead, if DM i experiences considerable information frictions, then she may process a signal adding nothing to her prior on V_i. A possible representation of that is

$$\mathcal{T}_i \equiv \{0\}, \;\; P^i_{T|X,\epsilon,V}(0|x_i, e_i, v) = 1 \;\forall v \in \mathcal{V}, \tag{2}$$

for a given realisation (x_i, e_i) of (X_i, ϵ_i). Note that, under (2), the posterior of DM i is equal

[4] We remark here that T_i can be a random vector, with arbitrarily correlated components.

to her prior. A signal whose informativeness is between such two extremes is plausible as well. For instance, DM i may process a signal revealing whether v_i is in $[a,b] \subset \mathcal{V}$ or not. A possible representation of that is

$$\mathcal{T}_i \equiv \{0,1\}, \; P^i_{T|X,\epsilon,V}(1|x_i,e_i,v) = 1 \; \forall v \in [a,b],$$
$$P^i_{T|X,\epsilon,V}(0|x_i,e_i,v) = 1 \; \forall v \in \mathcal{V} \setminus [a,b],$$

for a given realisation (x_i, e_i) of (X_i, ϵ_i).

In a typical empirical application, the information frictions possibly encountered by DM i are not observed by the researcher. In turn, it is impossible to know which signal DM i processes and, specifically, how the conditional density of T_i varies across the realisations of (X_i, ϵ_i, V_i). Hence, our framework proceeds without assumptions on $\mathcal{T}_i$ and $\mathcal{P}^i_{T|X,\epsilon,V}$ in order to avoid misspecifications. $\diamond$

Remark 2. (*Heterogeneity*) Suppose we have data on a cross-section of DMs facing the above decision problem (as Assumption 3 in Section 3 formally imposes). Our framework accommodates two layers of heterogeneity. The first layer concerns the *realisations* of the random variables. In particular, the state of the world and the realisation of the signal can vary across DMs, as indicated by subscript "i" in (x_i, e_i, v_i, t_i). This layer is important but standard in empirical work. The second layer concerns the *densities* from which such realisations are randomly drawn. Our approach allows the prior and the family of conditional signal densities (together with their supports) to vary across DMs. Specifically, with regards to the prior, note that every agent i has a common functional form, given by $P_{V|X,\epsilon}(\cdot|x_i, e_i)$. However, because the realisation of (X_i, ϵ_i) can vary across every agent i, if agents i,j have $(x_i, e_i) \neq (x_j, e_j)$ then it could be that $P_{V|X,\epsilon}(\cdot|x_i, e_i) \neq P_{V|X,\epsilon}(\cdot|x_j, e_j)$. With regards to the family of conditional signal densities, we incorporate heterogeneity in a fully flexible way. In fact, even if agents i,j have $(x_i, e_i) = (x_j, e_j) \equiv (x, e)$, it could be that they use different families of conditional signal densities to compute their posteriors, i.e., $\{P^i_{T|X,\epsilon,V}(\cdot|x,e,v)\}_{v\in\mathcal{V}} \neq \{P^j_{T|X,\epsilon,V}(\cdot|x,e,v)\}_{v\in\mathcal{V}}$, as highlighted by superscripts "i,j".[5] We believe that allowing for arbitrary heterogeneity in conditional signal densities (and, thus, posteriors) is important to avoid misspecifications of information frictions and, in turn, design a robust econometric analysis. This is because different agents could encounter different information frictions and, consequently, process more or less informative signals as emphasised in Remark 1. $\diamond$

Remark 3. (*Common family of priors and Bayesian rationality*) As mentioned in Remark 2, DMs rely on a common family of priors, $\mathcal{P}_{V|X,\epsilon} \equiv \{P_{V|X,\epsilon}(\cdot|x,e)\}_{(x,e)\in\mathcal{X}\times\mathcal{E}}$. Further, DMs are Bayesian rational, i.e., their priors are correct and are updated through the Bayes rule. These are "default" assumptions in the theoretical literature on decision problems under imperfect

[5] Put another way, agents i,j featuring $(x_i, e_i) = (x_j, e_j) \equiv (x, e)$ (and, hence, having the same prior) could end up with different posteriors because $\{P^i_{T|X,\epsilon,V}(\cdot|x,e,v)\}_{v\in\mathcal{V}} = \{P^j_{T|X,\epsilon,V}(\cdot|x,e,v)\}_{v\in\mathcal{V}}$ but $t_i \neq t_j$, or because $t_i = t_j$ but $\{P^i_{T|X,\epsilon,V}(\cdot|x,e,v)\}_{v\in\mathcal{V}} \neq \{P^j_{T|X,\epsilon,V}(\cdot|x,e,v)\}_{v\in\mathcal{V}}$, or a combination of both.

information and, therefore, we deem this to be a good starting point for our econometric analysis (see Morris, 1995 for a general discussion). In particular, these assumptions are extensively used in the theoretical and applied literature on voting (for example, Feddersen and Pesendorfer, 1997; Knight and Schiff, 2010; McMurray, 2013; Matějka and Tabellini, 2019). They are also used in the applied literature on consumer behaviour (for example, Crawford and Shum, 2005; Moretti, 2011), school choice (for a review, see Agarwal and Somaini, 2020), insurance choice (for example, Brown and Jeon, 2020), occupational choice (for example, Gibbons and Waldman, 1999), and games (Magnolfi and Roncoroni, 2017; Syrgkanis, Tamer, and Ziani, 2018). They are typically justified by presuming that, prior to the decision process, there had been some learning opportunities faced by all DMs (conditional on sharing similar attributes). For instance, when voting in general elections, such learning opportunities may stem from observing past election outcomes and consequent behaviour of parties. When choosing an education career, aggregate admission chances and available statistics from previous years may help students understand the cutoff rules of the assignment mechanism. In a job matching problem, resumés, recommendation letters, and other information gathered by human resource departments can reveal preliminary information on an individual's ability. From a technical perspective, these assumptions allow us to reduce the degrees of freedom in the econometric analysis and are sources of the identifying power. $\diamond$

Remark 4. (*Distinction among* X_i, ϵ_i, V_i) We distinguish among X_i, ϵ_i, V_i in order to get a flexible framework nesting various settings. However, the researcher can omit any of the variables among X_i, ϵ_i, V_i by simply assuming degenerate distributions. Further, the researcher has the freedom to decide which components of the state of the world are observed by DM i before processing any signal (hereafter, "observed pre-signal"), i.e., which variables of the model should fall into X_i, ϵ_i, V_i. In some scenarios, the researcher may prefer to be very cautious and assume that none of the components of the state of the world are observed pre-signal by DM i. In other scenarios, the researcher may feel confident imposing that some components of the state of the world are observed pre-signal by DM i. This choice will have an impact on the identifying power of the model. In this respect, our methodology can also be used to perform a sensitivity analysis of the identifying power of the model to changes in the set of components of the state of the world observed pre-signal by DM i. $\diamond$

We provide a more compact representation of our framework. Following the terminology of Bergemann and Morris (2013; 2016), we define the *baseline choice problem* faced by DM i as

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

G contains what DM i knows before processing any signal, together with the specific realisation (x_i, e_i) of (X_i, ϵ_i) that DM i observes. We also define the *information structure* processed by DM i as

$$S_i \equiv \{\mathcal{T}_i, \mathcal{P}^i_{T|X,\epsilon,V}\}.$$

S_i represents the additional information gathered by DM i to learn about V_i, together with the specific realisation t_i of T_i that DM i observes. As discussed in Remark 1, we remain agnostic about S_i and, thus, allow S_i to freely depend on the frictions possibly faced by DM i while studying the payoffs. Hereafter, we refer to the information structure revealing the exact realisation of V_i for each $(x, e) \in \mathcal{X} \times \mathcal{E}$ as the *complete information structure* (for an example, see (1) in Remark 1), and to the information structure adding no information whatsoever on the realisation of V_i for each $(x, e) \in \mathcal{X} \times \mathcal{E}$ as the *degenerate information structure* (for an example, see (2) in Remark 1). Further, we denote by $\mathcal{S}$ the set of all admissible information structures, ranging from the complete to the degenerate information structure. Lastly, the pair $\{G, S_i\}$ constitutes the *augmented choice problem* faced by DM i. The augmented choice problem $\{G, S_i\}$ summarises our framework, together with the specific realisation (x_i, e_i, t_i) of (X_i, ϵ, T_i) that DM i observes.

We formally define the optimal strategy of DM i when she faces the augmented choice problem $\{G, S_i\}$. Let Y_i be a random variable representing the choice of DM i. A (mixed) strategy for DM i is a family of probability mass functions of Y_i conditional on $(X_i, \epsilon_i, T_i) = (x, e, t)$ across all possible $(x, e, t) \in \mathcal{X} \times \mathcal{E} \times \mathcal{T}_i$.[6] We denote it by $\mathcal{P}^i_{Y|X,\epsilon,T} \equiv \{P^i_{Y|X,\epsilon,T}(\cdot|x, e, t)\}_{(x,e,t)\in\mathcal{X}\times\mathcal{E}\times\mathcal{T}_i}$, with each $P^i_{Y|X,\epsilon,T}(\cdot|x, e, t) \in \Delta(\mathcal{Y})$.[7] This strategy is optimal if, for every $(x, e, t) \in \mathcal{X} \times \mathcal{E} \times \mathcal{T}_i$, DM i maximises her expected payoff by choosing any alternative $y \in \mathcal{Y}$ such that $P^i_{Y|X,\epsilon,T}(y|x, e, t) > 0$.

Definition 1. (*Optimal strategy of the augmented choice problem* $\{G, S_i\}$) The family of probability mass functions $\mathcal{P}^i_{Y|X,\epsilon,T}$ is an optimal strategy of the augmented choice problem $\{G, S_i\}$ if, $\forall (x, e, t) \in \mathcal{X} \times \mathcal{E} \times \mathcal{T}_i$,

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

$\forall \tilde{y} \in \mathcal{Y} \setminus \{y\}$, and $\forall y \in \mathcal{Y}$ such that $P^i_{Y|X,\epsilon,T}(y|x, e, t) > 0$, where $P^i_{V|X,\epsilon,T}(\cdot|x, e, t)$ is the posterior computed via Bayes rule as

$$P^i_{V|X,\epsilon,T}(v|x, e, t) = \frac{P^i_{T|X,\epsilon,V}(t|x, e, v) P_{V|X,\epsilon}(v|x, e)}{\int_{\tilde{v}\in\mathcal{V}} P^i_{T|X,\epsilon,V}(t|x, e, \tilde{v}) P_{V|X,\epsilon}(\tilde{v}|x, e) d\tilde{v}} \; \forall v \in \mathcal{V}.$$

$\diamond$

In Appendix A we provide an equivalent definition of the optimal strategy of DM i. In the same appendix we also give a definition of the consideration set of DM i which endogenously arises from her optimal strategy.

By using the continuity of the expected utility with respect to Y_i (in the discrete metric),

[6] A mixed strategy will arise in the presence of ties.

[7] The superscript "i" in $\mathcal{P}^i_{Y|X,\epsilon,T}$ highlights that agents i, j could choose different alternatives because $(x_i, e_i, t_i) \neq (x_j, e_j, t_j)$, or because $S_i \neq S_j$, or because they adopt different tie-breaking rules, or a combination of all such scenarios.

it is possible to show that an optimal strategy of the augmented choice problem $\{G, S_i\}$ exists for any $S_i \in \mathcal{S}$, even though it may not be unique. Further details are in Appendix C.

Various discrete choice models that have been analysed in the literature can be obtained within the above framework under additional assumptions on S_i. We provide some examples below.

Example 1. (*Nested Logit*) As a first example, we consider the Nested Logit model with one nest collecting all goods but the outside option. The payoff function, u, is

$$u(y, Z_i, \xi_i, \eta_i) \equiv \begin{cases} \beta' Z_{iy} + \xi_i + \lambda \eta_{iy} & \text{if } y \in \mathcal{Y} \setminus \{0\}, \\ \eta_{i0} & \text{if } y = 0, \end{cases} \tag{3}$$

where $0 \in \mathcal{Y}$ is the outside option, $\mathcal{Y}$ has cardinality L, Z_{iy} is an $M \times 1$ vector of covariates of good $y \in \mathcal{Y} \setminus \{0\}$, $Z_i \equiv (Z_{i1}, ..., Z_{iL-1})$ is an $M(L-1) \times 1$ vector of the inside goods' characteristics, ξ_i represents tastes of DM i. $\eta_i \equiv (\eta_{i0}, ..., \eta_{iL-1})$ captures other individual-alternative specific features of the available alternatives. The parameter $\lambda \in (0, 1)$ captures the correlation among the inside goods. The variables $\xi_i, \eta_{i0}, ..., \eta_{iL-1}$ are mutually independent, and independent of Z_i. The densities of ξ_i and η_{iy} are parameterised as in Cardell (1997), so that $\rho_{iy} \equiv \xi_i + \lambda \eta_{iy}$ has the standard Gumbel density and the CDF of $(\rho_{i1}, ..., \rho_{iL-1})$ evaluated at $(s_1, ..., s_{L-1})$ is $\exp(-(\sum_{y=1}^{L-1} \exp(-s_y/\lambda))^\lambda)$.[8] The researcher observes the choice made by DM i and the realisation of Z_i. Below we present two alternative scenarios that are allowed in our framework and that, under additional assumptions on the information structures processed by agents, collapse to the standard Nested Logit model.

First, suppose that DM i observes the realisation of (Z_i, ξ_i) but might be uncertain about the realisation of the other tastes, η_i. Hence, following our general notation, $X_i \equiv Z_i$, $\epsilon_i \equiv \xi_i$, and $V_i \equiv \eta_i$. DM i has a prior on V_i conditional on (X_i, ϵ_i), which is assumed to obey the Gumbel parameterisation above. Further, DM i processes an information structure, S_i, to update her prior. Note that this framework collapses to the Nested Logit model under the additional assumption that each agent in the population processes the complete information structure.

Second, for each $y \in \mathcal{Y}$, let Z_{iy}^1 and Z_{iy}^{-1} denote the first component and the residual $M-1$ components of the $M \times 1$ vector Z_{iy}, respectively. Further, let $Z_i^1 \equiv (Z_{i1}^1, ..., Z_{iL-1}^1)$ and $Z_i^{-1} \equiv (Z_{i1}^{-1}, ..., Z_{iL-1}^{-1})$. Suppose now that DM i observes the realisation of $(Z_i^{-1}, \xi_i, \eta_i)$ but might be uncertain about the realisation of Z_i^1. Hence, following our general notation, $X_i \equiv Z_i^{-1}$, $\epsilon_i \equiv (\xi_i, \eta_i)$, and $V_i \equiv Z_i^1$. DM i has a prior on V_i conditional on (X_i, ϵ_i) which is assumed equal to the empirical distribution of Z_i^1 conditional on Z_i^{-1}. Further, DM i processes an information structure, S_i, to refine her prior. As earlier, note that this framework collapses to the Nested Logit model under the additional assumption that each agent in the population processes the complete information structure.

[8] See also Galichon (2019) regarding the random utility representation of the Nested Logit model.

Similar considerations can be made for other discrete choice models, such as the (Multinomial) Logit/Probit model and the Mixed Logit model.

In the discussion above, we have interpreted the utility components unobserved by the researcher, (ξ_i, η_i), as tastes of DM i. That is, (ξ_i, η_i) capture, in some aggregate ways, additional latent determinants of the utility that DM i can get from the decision problem. Along the lines of McFadden (1981), an alternative interpretation of (ξ_i, η_i) is as "errors in judgment" made by DM i, which are due to the complexity of the choice problem and can lead to suboptimal outcomes. The latter interpretation recognises the importance of information frictions but, in contrast to our work, it treats their impact on the agent perception about the attainable utilities as exogenous and random. In this paper, when referring to classical parametric discrete choice models like the Nested Logit Model, we always interpret the utility components unobserved by the researcher as tastes of DM i.

◇

Example 2. (*Risk aversion*) As a second example, we consider a discrete choice model of insurance plans. Specifically, DM i faces an underlying risk of a loss (for example, a car accident) and can choose among L insurance plans. The loss event is denoted by C_i. $C_i = 1$ if the loss event occurs, and 0 otherwise. Each insurance plan $y \in \mathcal{Y}$ is characterised by a deductible, D_y, and a premium, P_{iy}. Further, DM i is endowed with some wealth (Wealth$_i$). The payoff function, u, belongs to the CARA family, i.e., for each $y \in \mathcal{Y}$,

$$u(y, P_i, D, \text{Wealth}_i, r_i, C_i) \equiv \begin{cases} \frac{1-\exp[-r_i \times (\text{Wealth}_i - P_{iy} - D_y)]}{r_i} & \text{if } C_i = 1, r_i \neq 0, \\ \frac{1-\exp[-r_i \times (\text{Wealth}_i - P_{iy})]}{r_i} & \text{if } C_i = 0, r_i \neq 0, \\ \text{Wealth}_i - P_{iy} - D_y & \text{if } C_i = 1, r_i = 0, \\ \text{Wealth}_i - P_{iy} & \text{if } C_i = 0, r_i = 0, \end{cases} \tag{4}$$

where $P_i \equiv (P_{i1}, ..., P_{iL})$, $D \equiv (D_1, ..., D_L)$, and r_i is the coefficient of absolute risk aversion. r_i is often assumed distributed according to some parametric distribution such as the Beta distribution. The researcher observes the choice made by DM i and the realisation of $(P_i, D, \text{Wealth}_i)$. In some cases, the researcher also observes the realisation of C_i from ex-post data on claims.

Before choosing an insurance plan, DM i is aware of the realisation of $(P_i, D, \text{Wealth}_i, r_i)$. However, DM i does not observe the realisation of C_i because it is realised after the insurance plan choice has been made. Hence, following our general notation, $X_i \equiv (P_i, D, \text{Wealth}_i)$, $\epsilon_i \equiv r_i$, and $V_i \equiv C_i$. DM i has a prior on V_i conditional on (X_i, ϵ_i),[9] which can be assumed to belong to some parametric family. For instance, one can use a simple Probit model or a more sophisticated Poisson-Gamma model (for an example of the latter see Barseghyan, Molinari,

[9] We can also condition the prior of DM i on a vector of individual-specific characteristics, Z_i, such as gender, age, insurance score, and rating territories, that are observed by the researcher. Z_i can be treated as fully observed by DM i and, hence, added to X_i.

O' Donoghue, and Teitelbaum, 2013; Barseghyan, Molinari, and Teitelbaum; 2016). Further, DM i processes an information structure, S_i, to update her prior. S_i incorporates any extra private information on the risky event at the disposal of DM i, other than her level of risk aversion, and can arbitrarily depend on DM i's risk aversion.

Under the additional restriction that each agent processes the degenerate information structure, note that this framework collapses to the standard risk aversion setting considered in the empirical literature, where individuals have no extra private information on the risky event.

$\diamond$

Example 3. (*Rational inattention*) As a third example, we consider the rational inattention framework by Caplin and Dean (2015) and Matějka and McKay (2015). In that setting, the decision problem has two stages. In the first stage, DM i optimally chooses an information structure to update her prior. Although DM i is free to choose any information structure, attention is a scarce resource and there is a cost of processing information. As a result, more informative signals are more costly. Such attentional costs are parameterised in various ways, for example, the Shannon entropy (Sims, 2003) and the posterior-separable function (Caplin, Dean, and Leahy, 2019a). Formally, in the first stage DM i observes the realisation (x_i, e_i) of (X_i, ϵ_i) and chooses an information structure $S_i \in \mathcal{S}$ such that

$$S_i \in \text{argmax}_{S \equiv \{\mathcal{T}, \mathcal{P}_{T|X,\epsilon,V}\} \in \mathcal{S}} \int_{(v,t)\in\mathcal{V}\times\mathcal{T}} \left[\max_{y\in\mathcal{Y}} \mathbb{E}_{S,t} u(y, x_i, e_i, V_i)\right] P_{T|X,\epsilon,V}(t|x_i, e_i, v) P_{V|X,\epsilon}(v|x_i, e_i) d(v,t) - C(S),$$

where $\mathbb{E}_{S,t} u(y, x_i, e_i, V_i)$ is the expected payoff from choosing $y \in \mathcal{Y}$ under the posterior induced by the information structure S and the signal realisation t, and $C(S)$ represents the parameterised attentional costs associated with the information structure S. Then, in the second stage, DM i observes a signal realisation, t_i, randomly drawn according to S_i. Lastly, DM i chooses alternative $y \in \mathcal{Y}$ maximising $\mathbb{E}_{S_i,t_i} u(y, x_i, e_i, V_i)$.

Note that this rational inattention framework can be obtained within our model under the additional assumption that DM i processes an information structure chosen as prescribed by the above first stage. Also, note that such an optimal information structure depends on the way in which attentional costs are parameterised.

For an empirical counterpart of the rational inattention model, see Csaba (2018) and Brown and Jeon (2020). A few theoretical papers on rational inattention and, more generally, stochastic choice also grapple with the identification problem. However, those papers typically require either to have data on choices for every possible realisation of the state of the world (for example, Caplin and Martin, 2015), or to have data on choices for multiple menus (for example, Lu, 2016; Lin, 2019).[10] These data are rarely available outside of laboratory experiments, especially for "complex" products as targeted here. Instead, in our framework the state of the world can

[10] According to our notation, having data on choices for multiple menus would consist of observing the choices of agents for multiple sets $\mathcal{Y}$.

be fully observed, partly observed, or fully unobserved by the researcher. Moreover, we assume to have data on choices for a single menu. In other words, we rely on less rich data. On one hand, this reduces the sources of identifying power and makes the task of deriving identification arguments more challenging; on the other hand, it allows for a wider applicability of our methodology.[11]

Hébert and Woodford (2018) and Morris and Strack (2019) consider continuous-time models of sequential evidence accumulation and show that the resulting choice probabilities are identical to those of a static rational inattention model with posterior-separable attentional cost functions. That is, there is an equivalence between the information that is ultimately acquired in some search models and the information acquired in a static model of rational inattention, under a particular parameterisation of the attentional costs. Therefore, our setting also nests such search frameworks. $\diamond$

3 Identification

In this section we discuss identification of the primitives $u, \mathcal{P}_{\epsilon|X}, \mathcal{P}_{V|X,\epsilon}$ from observing the choices made by many DMs facing the decision problem described in Section 2. Before proceeding, we parameterise such primitives and index them by the vectors of parameters $\theta_u \in \Theta_u \subseteq \mathbb{R}^{K_u}$, $\theta_V \in \Theta_V \subseteq \mathbb{R}^{K_V}$, and $\theta_\epsilon \in \Theta_\epsilon \subseteq \mathbb{R}^{K_\epsilon}$, respectively. Hereafter, we represent them as

$$u(\cdot;\theta_u), \mathcal{P}^{\theta_V}_{V|X,\epsilon} \equiv \{P_{V|X,\epsilon}(\cdot|x,e;\theta_V)\}_{(x,e)\in\mathcal{X}\times\mathcal{E}}, \mathcal{P}^{\theta_\epsilon}_{\epsilon|X} \equiv \{P_{\epsilon|X}(\cdot|x;\theta_\epsilon)\}_{x\in\mathcal{X}}.$$

Further, we denote by θ the whole vector of parameters, i.e., $\theta \equiv (\theta_u, \theta_V, \theta_\epsilon) \in \Theta \equiv \Theta_u \times \Theta_V \times \Theta_\epsilon \subseteq \mathbb{R}^K$, where $K \equiv K_u + K_V + K_\epsilon$. Finally, observe that in the case where the sets $\mathcal{X}$, $\mathcal{E}$, and $\mathcal{V}$ are finite, the primitives of interest can be fully flexibly characterised by a finite number of parameters, one for each combination of (X_i, ϵ_i, V_i) values.

3.1 Data generating process

We formally outline our restrictions on the data generating process (hereafter, DGP). The relevant notation has been introduced in Section 2. Some objects have the superscript "0" in order to distinguish their true value from other possible values.

Assumption 3. (*DGP*) The sets $\mathcal{Y}$, $\mathcal{X}$, $\mathcal{E}$, and $\mathcal{V}$ are known by the researcher. $\mathcal{Y}$ and $\mathcal{X}$ are finite. The researcher has a random sample of observations $\{y_i, x_i\}_{i=1}^n$, where for each $i = 1, ..., n$:

[11] Note that when the researcher has data on choices for multiple menus, it is important to specify whether the agents' information structures are allowed to vary across menus. This is not relevant in our framework because we focus on a single menu.

1. DM i is endowed with the realisation (x_i, e_i, v_i) of (X_i, ϵ_i, V_i). The realisations x_i, e_i, and v_i are randomly drawn from P^0_X, $P_{\epsilon|X}(\cdot|x_i;\theta^0_\epsilon)$, and $P_{V|X,\epsilon}(\cdot|x_i, e_i;\theta^0_V)$, respectively. DM i observes (x_i, e_i). DM i does not observe v_i. However, DM i has a prior on V_i conditional on $(X_i, \epsilon_i) = (x_i, e_i)$, that is $P_{V|X,\epsilon}(\cdot|x_i, e_i;\theta^0_V)$. $G^0 \equiv \{\mathcal{Y}, \mathcal{X}, \mathcal{E}, \mathcal{V}, u(\cdot;\theta^0_u), \mathcal{P}^{\theta^0_\epsilon}_{\epsilon|X}, \mathcal{P}^{\theta^0_V}_{V|X,\epsilon}\}$ constitutes the baseline choice problem of DM i.

2. DM i processes an information structure, $S^0_i \equiv \{\mathcal{T}^0_i, \mathcal{P}^{i,0}_{T|X,\epsilon,V}\} \in \mathcal{S}$, to refine her prior. DM i observes a signal realisation, t_i, randomly drawn according to S^0_i and computes the posterior by applying the Bayes rule.

3. DM i chooses alternative y_i from $\mathcal{Y}$ according to the notion of an optimal strategy of the augmented choice problem $\{G^0, S^0_i\}$ provided in Definition 1.

$\diamond$

Assumption 3 summarises Assumptions 1 and 2 of Section 2 and adds further details. The set $\mathcal{X}$ is assumed finite as standard in empirical work with partial identification in order to easily transform identifying restrictions into unconditional moment inequalities. More details on this are in Section B. If $\mathcal{X}$ is not finite, then our identification analysis still goes through. However, for inference, one has to implement a method dealing with conditional moment inequalities. As anticipated in Section 2, Assumption 3 remains agnostic about the information structures processed by DMs and, thus, allows these information structures to freely depend on the underlying latent frictions possibly faced by DMs while learning about the state of the world. Further, these information structures can be arbitrarily different across DMs, ranging from the complete to the degenerate information structure. Being agnostic about the information structures processed by DMs means that DMs compute their expected utilities with any Bayes-consistent posteriors. Assumption 3 allows the priors of DMs to be heterogeneous, depending on the realisations of (X_i, ϵ_i). See also Remarks 2 and 3 on this. Assumption 3 does not impose any restriction on the tie-breaking rules adopted by DMs and these can vary across the population. Assumption 3 allows for correlation between ϵ_i and V_i. Further, it allows for correlation between X_i and (ϵ_i, V_i) and conditional heteroskedasticity. For example, one can impose that, conditional on $X_i = x$, (ϵ_i, V_i) are jointly distributed as a multivariate normal with mean vector $\mu^0(x)$ and variance-covariance matrix $\Sigma^0(x)$. In such a case, $(\theta^0_\epsilon, \theta^0_V) \equiv (\{\mu^0(x)\}_{x\in\mathcal{X}}, \{\Sigma^0(x)\}_{x\in\mathcal{X}})$.[12] Lastly, the probability mass function of (Y_i, X_i) which results from the decision problem is denoted by $P^0_{Y,X} \in \Delta(\mathcal{X} \times \mathcal{Y})$. Note that Assumption 3 does not require i.i.d. random sampling. All that is needed is for the law of large numbers to hold, so that $P^0_{Y,X}$ can be treated as known in the identification analysis.

Remarks 5 and 6 conclude our discussion of Assumption 3.

[12]If $\mathcal{X}$ is not finite, then it remains an open question whether one can incorporate correlation between X_i and (ϵ_i, V_i), for instance, by extending insights from the parametric control function literature to our setting (for example, Blundell and Smith, 1986; 1989).

Remark 5. (*When V_i is observed by the researcher*) In certain settings, some or all the components of the realisation, v_i, of V_i are observed by the researcher, together with (x_i, y_i) for $i = 1, ..., n$. For example, in models of insurance plans, the researcher often has data on the ex-post claim experience of the agents in the sample. In those cases, θ_V^0 could be identified directly from such additional data.[13] In our general discussion below, we focus on the scenario where v_i is unobserved to the researcher for $i = 1, ..., n$. This is the case considered in our empirical application to voting behaviour, as discussed in Section 4. It is also the case studied in several other papers with imperfect information, such as Crawford and Shum (2005) and Moretti (2011).

$\diamond$

Remark 6. (*Policy relevance of θ^0*) Our methodology does not attempt to recover the information structures of DMs and, rather, focuses on (partially) identifying θ^0. In fact, identifying θ^0 is sufficient to answer some standard questions in the empirical literature on decision problems. For instance, we can use the estimates of θ^0 to find how the choice probabilities change in response to changes in the realisation of X_i, while holding the information structures of DMs fixed.

Further, identifying θ^0 permits us to assess how choices respond to additional information. In particular, we can use the estimates of θ^0 to find how the choice probabilities change when the researcher gives some additional information to DMs that induce them to modify their information structures, while holding the state of the world fixed. Such a change in choice probabilities quantifies the extent to which uncertainty affects individual decisions and can help to measure the welfare benefits of the policy intervention, prior to conducting it. This is a question that has received lots of attention in the applied literature, as discussed in Section 1.

More details are provided when discussing the empirical application in Section 4. $\diamond$

3.2 A tractable characterisation of the identified set

Let us first introduce some useful notation. In what follows, given $x \in \mathcal{X}$, we denote by $P^0_{Y|X}(\cdot|x) \in \Delta(\mathcal{Y})$ the probability mass functions of Y_i conditional on $X_i = x$ induced by $P^0_{Y,X}$ and P^0_X. We use the same notation without superscript "0" to indicate a generic probability mass function of Y_i conditional on $X_i = x$. Lastly, given $\theta \in \Theta$, we denote by $G^\theta \equiv \{\mathcal{Y}, \mathcal{X}, \mathcal{E}, \mathcal{V}, u(\cdot; \theta_u), \mathcal{P}^{\theta_\epsilon}_{\epsilon|X}, \mathcal{P}^{\theta_V}_{V|X,\epsilon}\}$ the corresponding baseline choice problem.

We now discuss identification of θ^0 under Assumption 3. Due to the absence of restrictions on the information structures and tie-breaking rules of DMs, our model is incomplete in the sense of Tamer (2003). This raises the possibility of partial identification of θ^0 and, conse-

[13] For example, suppose that all the components of the realisation, v_i, of V_i are observed by the researcher for $i = 1, ..., n$. Then, under the additional assumption that V_i is independent of ϵ_i conditional on X_i, we can recover $\{P_{V|X}(\cdot|x)\}_{x\in\mathcal{X}}$ without parameterising it, simply from its empirical distribution.

quently, the challenge of tractably characterising the set of θs exhausting all the implications of the model and data, i.e., the identified set for θ^0.

Intuitively, the identified set for θ^0 is the set of θs for which the model predicts a probability mass function of Y_i conditional on $X_i = x_i$ that matches with $P^0_{Y|X}(\cdot|x)$, for each $x \in \mathcal{X}$. More formally, for every $\theta \in \Theta$ and $S \in \mathcal{S}$, let $\mathcal{R}^{\theta,S}$ be the collection of optimal strategies of the augmented choice problem $\{G^\theta, S\}$.[14] Further, for every $\theta \in \Theta$ and $x \in \mathcal{X}$, let $\bar{\mathcal{R}}^\theta_{Y|x}$ be the collection of probability mass functions of Y_i conditional on $X_i = x$ that are induced by the model's optimal strategies under θ, while remaining agnostic about information structures. That is,

$$\begin{aligned}
&\bar{\mathcal{R}}^\theta_{Y|x} \equiv \text{Conv}\Big\{ P_{Y|X}(\cdot|x) \in \Delta(\mathcal{Y}) : \\
&P_{Y|X}(y|x) = \int_{(t,v,e)\in \mathcal{T}\times\mathcal{V}\times\mathcal{E}} P_{Y|X,\epsilon,T}(y|x,e,t) P_{T|X,\epsilon,V}(t|x,e,v) P_{V|X,\epsilon}(v|x,e;\theta_V) P_{\epsilon|X}(e|x;\theta_\epsilon) d(t,v,e) \ \forall y \in \mathcal{Y}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \mathcal{P}_{Y|X,\epsilon,T} \in \mathcal{R}^{\theta,S}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad S \equiv \{\mathcal{T}, \mathcal{P}_{T|X,\epsilon,V}\} \in \mathcal{S} \Big\},
\end{aligned} \tag{5}$$

where we have used the fact that Y_i is independent of V_i conditional on (X_i, ϵ_i, T_i). Convexification (via the convex hull operator, $\text{Conv}\{\cdot\}$) allows us to include in $\bar{\mathcal{R}}^\theta_{Y|x}$ probability mass functions of Y_i conditional on $X_i = x$ that are *mixtures* across information structures. Importantly, this ensures that the information structures can be arbitrarily different across DMs. It follows that the identified set for θ^0 can be defined as

$$\Theta^* \equiv \{\theta \in \Theta : P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{R}}^\theta_{Y|x} \ \forall x \in \mathcal{X}\}. \tag{6}$$

Unfortunately, the definition of Θ^* in (6) seems hardly useful in practice. This is because constructing $\bar{\mathcal{R}}^\theta_{Y|x}$ is infeasible due to the necessity of exploring the large class $\mathcal{S}$ which contains infinite-dimensional objects. In what follows, we overcome such an issue by revisiting our framework through the lens of one-player Bayes Correlated Equilibrium (Bergemann and Morris, 2013; 2016). The concept of one-player Bayes Correlated Equilibrium is based on a theoretical setting where an omniscient mediator makes incentive-compatible recommendations to DMs as a function of the state of the world and compatible with their priors. If DMs follow such recommendations, then the resulting distribution of choices is a one-player Bayes Correlated Equilibrium. The concept of one-player Bayes Correlated Equilibrium is a powerful tool because it provides behavioural predictions that do not depend on the specification of information structures. In particular, Theorem 1 in Bergemann and Morris (2016) shows that the set of one-player Bayes correlated equilibria for a given baseline choice problem equals the set of optimal strategies that could arise when adding to that baseline choice problem any information structures. In turn, this allows us to characterise Θ^* in a more tractable way.

Our analysis proceeds in three steps. First, we give the definition of one-player Bayes

[14] Note that, if there are no ties, $\mathcal{R}^{\theta,S}$ contains only one optimal strategy.

Correlated Equilibrium (hereafter, 1BCE) of the baseline choice problem G^{θ}. Further, we highlight that the set of 1BCEs of the baseline choice problem G^{θ} is convex. Second, we introduce Theorem 1 in Bergemann and Morris (2016) which claims that the set of 1BCEs of the baseline choice problem G^{θ} is *equivalent* to the collection of optimal strategies of the augmented choice problem $\{G^{\theta}, S\}$ across every possible information structure $S \in \mathcal{S}$. Third, we combine the first and second steps to construct Θ^* (or, an outer set of Θ^*) via a collection of linear programming problems. Details on each step follow.

In order to give the definition of 1BCE of the baseline choice problem G^{θ}, let us consider a family of densities of (Y_i, V_i) conditional on $(X_i, \epsilon_i) = (x, e)$ across all possible $(x, e) \in \mathcal{X} \times \mathcal{E}$. We denote it by $\mathcal{P}_{Y,V|X,\epsilon} \equiv \{P_{Y,V|X,\epsilon}(\cdot|x, e)\}_{(x,e)\in\mathcal{X}\times\mathcal{E}}$, with each $P_{Y,V|X,\epsilon}(\cdot|x, e) \in \Delta(\mathcal{Y}\times\mathcal{V})$. Let the marginal of $\mathcal{P}_{Y,V|X,\epsilon}$ on $\mathcal{V}$ be equal to the prior of DM i. Further, imagine an omniscient mediator using $\mathcal{P}_{Y,V|X,\epsilon}$ to recommend DM i which alternative to choose in a way that is incentive compatible. Then, $\mathcal{P}_{Y,V|X,\epsilon}$ is a 1BCE of the baseline choice problem G^{θ}.

Definition 2. (*1BCE of the baseline choice problem G^{θ}*) Given $\theta \in \Theta$, the family of densities $\mathcal{P}_{Y,V|X,\epsilon}$ is a 1BCE of the baseline choice problem G^{θ} if:

1. It is *consistent* for the baseline choice problem G^{θ}, i.e., the marginal of $P_{Y,V|X,\epsilon}(\cdot|x, e)$ on $\mathcal{V}$ is equal to the prior, $P_{V|X,\epsilon}(\cdot|x, e; \theta_V)$, for every $x \in \mathcal{X}$ and $e \in \mathcal{E}$. That is,

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

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

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

$\diamond$

Note that, for each $(x, e) \in \mathcal{X} \times \mathcal{E}$, the collection of conditional densities $P_{Y,V|X,\epsilon}(\cdot|x, e)$ satisfying the consistency and obedience requirements of Definition 2 is convex. This is because the consistency and obedience requirements are linear in $P_{Y,V|X,\epsilon}(\cdot|x, e)$.

We now illustrate Theorem 1 in Bergemann and Morris (2016). Such a theorem highlights the robustness properties of 1BCE. Specifically, it shows that the set of 1BCEs of the baseline choice problem G^{θ} equals the set of optimal strategies of the augmented choice problem $\{G^{\theta}, S\}$ across every admissible information structure $S \in \mathcal{S}$. Therefore, it allows us to compactly characterise all possible optimal behaviours of an agent if she had access to any of the information structures in $\mathcal{S}$.

Theorem 1. (*Bergemann and Morris, 2016*) Given $\theta \in \Theta$, $\mathcal{P}_{Y,V|X,\epsilon}$ is a 1BCE of the baseline choice problem G^{θ} if and only if there exists an information structure, $S \equiv \{\mathcal{T}, \mathcal{P}_{T|X,\epsilon,V}\} \in \mathcal{S}$, and an optimal strategy, $\mathcal{P}_{Y|X,\epsilon,T}$, of the augmented choice problem $\{G^\theta, S\}$, such that $\mathcal{P}_{Y,V|X,\epsilon}$ is induced by $\mathcal{P}_{Y|X,\epsilon,T}$.[15] ◇

Note that Theorem 1 also implies that a 1BCE of the baseline choice problem G^θ exists. Indeed, fix any information structure $S \equiv \{\mathcal{T}, \mathcal{P}_{T|X,\epsilon,V}\} \in \mathcal{S}$. Let $\mathcal{P}_{Y|X,\epsilon,T}$ be an optimal strategy of the augmented choice problem $\{G^\theta, S\}$, which exists by Lemma C.1. Let $\mathcal{P}_{Y,V|X,\epsilon}$ be the family of densities of (Y_i, V_i) conditional on $(X_i, \epsilon_i) = (x, e)$ across all possible $(x, e) \in \mathcal{X} \times \mathcal{E}$ 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^θ. Therefore, the set of 1BCE of the baseline choice problem G^θ is non-empty. Furthermore, the set of 1BCE of the baseline choice problem G^θ is typically non-singleton. In fact, 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 agents might process.

We now exploit Theorem 1 to represent Θ^* in a more useful way. For each $\theta \in \Theta$, let $\mathcal{Q}^\theta$ be the convex set 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_i conditional on $X_i = x$ that are induced by the 1BCEs of the baseline choice problem G^θ. That is,

$$\bar{\mathcal{Q}}^\theta_{Y|x} \equiv \Big\{P_{Y|X}(\cdot|x) \in \Delta(\mathcal{Y}) : \; P_{Y|X}(y|x) = \int_{(e,v)\in\mathcal{E}\times\mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e) P_{\epsilon|X}(e|x;\theta_\epsilon) d(e,v) \; \forall y \in \mathcal{Y}, \\ \mathcal{P}_{Y,V|X,\epsilon}(\cdot|x,e) \in \mathcal{Q}^\theta\Big\}. \tag{7}$$

Note that $\bar{\mathcal{Q}}^\theta_{Y|x}$ is convex and, hence, agents in the population can obey arbitrarily different 1BCEs.

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

Proposition 1. (*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 3, $\Theta^* = \Theta^{**}$. ◇

We are now ready to outline a tractable procedure to construct Θ^* by leveraging on the convexity of $\bar{\mathcal{Q}}^\theta_{Y|x}$ for each $x \in \mathcal{X}$ and $\theta \in \Theta$. We distinguish two cases. The first case is when the sets $\mathcal{E}$ and $\mathcal{V}$ are finite. Recall that in such a case, $\mathcal{P}^{\theta_V}_{V|X,\epsilon}$, $\mathcal{P}^{\theta_\epsilon}_{\epsilon|X}$, and $\mathcal{P}_{Y,V|X,\epsilon}$ are families of conditional probability mass functions. Further, by Proposition 1 and Definition 2, note that $\theta \in \Theta^*$ if and only if the following linear programming problem has a solution with respect to $\mathcal{P}_{Y,V|X,\epsilon}$:

[15] Suppose $\mathcal{T}$ is finite. Then, by "induced" we mean

$$P_{Y,V|X,\epsilon}(y,v|x,e) = \sum_{t\in\mathcal{T}} P_{Y|X,\epsilon,T}(y|x,e,t) P_{T|X,\epsilon,V}(t|x,e,v) P_{V|X,\epsilon}(v|x,e;\theta_V),$$

$\forall y \in \mathcal{Y}$, $\forall v \in \mathcal{V}$, $\forall x \in \mathcal{X}$, and $\forall e \in \mathcal{E}$.

$$
\begin{aligned}
&\text{[1BCE-Consistency]:} && \sum_{y\in\mathcal{Y}} P_{Y,V|X,\epsilon}(y,v|x,e) = P_{V|X,\epsilon}(v|x,e;\theta_V)\ \forall v\in\mathcal{V},\forall e\in\mathcal{E},\forall x\in\mathcal{X},\\
&\text{[1BCE-Obedience]:} && -\sum_{v\in\mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e)[u(y,x,e,v;\theta_u)-u(y',x,e,v;\theta_u)]\le 0\ \forall y\in\mathcal{Y},\forall y'\in\mathcal{Y}\setminus\{y\},\forall e\in\mathcal{E},\forall x\in\mathcal{X},\\
&\text{[1BCE-Data match]:} && P^0_{Y|X}(y|x) = \sum_{(e,v)\in\mathcal{E}\times\mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e)P_{\epsilon|X}(e|x;\theta_\epsilon)\ \forall y\in\mathcal{Y},\forall x\in\mathcal{X}.
\end{aligned}
\tag{8}
$$

Therefore, one can construct Θ^* by checking whether the linear programming problem (8) has a solution with respect to $\mathcal{P}_{Y,V|X,\epsilon}$ for every $\theta\in\Theta$. In practice, this is done by appropriately selecting a finite subset of Θ (also called a grid) and solving the linear programming problem (8) for each θ in such a grid. In the simulations of Section 3.3, we know θ^0 and its dimension K is small. Thus, we can design a grid of values around each component of θ^0, take the Cartesian product of the K grids obtained, and consider this as our final grid. In the empirical application (where we do not know θ^0), we obtain a grid by using the simulated annealing algorithm as explained in Appendix E.

The second case is when the sets $\mathcal{E}$ and $\mathcal{V}$ are not finite. Recall that in such a case, $\mathcal{P}^{\theta_\epsilon}_{\epsilon|X}$, $\mathcal{P}^{\theta_V}_{V|X,\epsilon}$, and $\mathcal{P}_{Y,V|X,\epsilon}$ are families of conditional densities or conditional mixed joint densities. Hence, by Proposition 1 and Definition 2, $\theta\in\Theta^*$ if and only if the following system of equalities and inequalities has a solution with respect to $\mathcal{P}_{Y,V|X,\epsilon}$:

$$
\begin{aligned}
&\text{[1BCE-Consistency]:} && \sum_{y\in\mathcal{Y}} P_{Y,V|X,\epsilon}(y,v|x,e) = P_{V|X,\epsilon}(v|x,e;\theta_V)\ \forall v\in\mathcal{V},\forall e\in\mathcal{E},\forall x\in\mathcal{X},\\
&\text{[1BCE-Obedience]:} && -\int_{v\in\mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e)[u(y,x,e,v;\theta_u)-u(y',x,e,v;\theta_u)]dv\le 0\ \forall y\in\mathcal{Y},\forall y'\in\mathcal{Y}\setminus\{y\},\forall e\in\mathcal{E},\forall x\in\mathcal{X},\\
&\text{[1BCE-Data match]:} && P^0_{Y|X}(y|x) = \int_{(e,v)\in\mathcal{E}\times\mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e)P_{\epsilon|X}(e|x;\theta_\epsilon)d(e,v)\ \forall y\in\mathcal{Y},\forall x\in\mathcal{X}.
\end{aligned}
\tag{9}
$$

Note that (9) contains an uncountable number of constraints which cannot be feasibly implemented as a linear programming problem. To operationalise (9), we suggest to discretise $\mathcal{E}$ and $\mathcal{V}$ and approximate $\mathcal{P}^{\theta_\epsilon}_{\epsilon|X}$ and $\mathcal{P}^{\theta_V}_{V|X,\epsilon}$ by conditional probability mass functions. Specifically, we appropriately select some finite subsets, $\mathcal{E}^{\text{discr}}\subset\mathcal{E}$ and $\mathcal{V}^{\text{discr}}\subset\mathcal{V}$. Then, for each $x\in\mathcal{X}$, we construct the conditional probability mass function $P^{\text{discr}}_{\epsilon|X}(\cdot|x;\theta_\epsilon)$ as

$$
P^{\text{discr}}_{\epsilon|X}(e|x;\theta_\epsilon) \equiv \frac{P_{\epsilon|X}(e|x;\theta_\epsilon)}{\sum_{e\in\mathcal{E}^{\text{discr}}} P_{\epsilon|X}(e|x;\theta_\epsilon)}\ \forall e\in\mathcal{E}^{discr}. \text{[16]} \tag{10}
$$

We proceed similarly to construct $\{P^{\text{discr}}_{V|X,\epsilon}(\cdot|x,e;\theta_V)\}_{(x,e)\in\mathcal{X}\times\mathcal{E}^{\text{discr}}}$. In turn, we replace $\mathcal{E}$, $\mathcal{V}$,

[16] Note that here we approximate a continuous cumulative distribution function by a step function. There are other ways to do so, in addition to (10). For alternative discretisations see Magnolfi and Roncoroni (2017) and Syrgkanis, Tamer, and Ziani (2018). Many other methods can be found in Bracquemond and Gaudoin (2003), Lai (2013), and Chakraborty (2015), together with a discussion on how each method preserves important properties of the continuous case.

$\mathcal{P}^{\theta_\epsilon}_{\epsilon|X}$, and $\mathcal{P}^{\theta_V}_{V|X,\epsilon}$ in (9) with such discretised objects and obtain:

$$\begin{aligned}
&\text{[1BCE-Consistency]:} && \sum_{y\in\mathcal{Y}} P_{Y,V|X,\epsilon}(y,v|x,e) = P^{discr}_{V|X,\epsilon}(v|x,e;\theta_V)\ \forall v\in\mathcal{V}^{discr}, \forall e\in\mathcal{E}^{discr}, \forall x\in\mathcal{X},\\
&\text{[1BCE-Obedience]:} && -\sum_{v\in\mathcal{V}^{discr}} P_{Y,V|X,\epsilon}(y,v|x,e)[u(y,x,e,v;\theta_u)-u(y',x,e,v;\theta_u)]\leq 0\ \forall y\in\mathcal{Y}, \forall y'\in\mathcal{Y}\setminus\{y\}, \forall e\in\mathcal{E}^{discr}, \forall x\in\mathcal{X},\\
&\text{[1BCE-Data match]:} && P^0_{Y|X}(y|x) = \sum_{(e,v)\in\mathcal{E}^{discr}\times\mathcal{V}^{discr}} P_{Y,V|X,\epsilon}(y,v|x,e)P^{discr}_{\epsilon|X}(e|x;\theta_\epsilon)\ \forall y\in\mathcal{Y}, \forall x\in\mathcal{X}.
\end{aligned} \tag{11}$$

Finally, one can approximate Θ^* by checking whether the linear programming problem (11) has a solution with respect to $\mathcal{P}_{Y,V|X,\epsilon}$ for every $\theta\in\Theta$ (in practice, for each θ in a grid, as explained above). In our simulations and empirical application we have tried a few different discretisations of $\mathcal{E}$ and $\mathcal{V}$ and obtained negligible differences among the resulting outer sets, provided that $\mathcal{E}^{\text{discr}}$ and $\mathcal{V}^{\text{discr}}$ contain the extreme values of $\mathcal{E}$ and $\mathcal{V}$, respectively.

In what follows, with some abuse of notation, we do not report the superscript "discr" when discussing the practical implementation of (9) for the case where the sets $\mathcal{E}$ and $\mathcal{V}$ are not finite, it being understood that all the necessary discretisations have been carried out.

Lastly, in Appendix D we describe a case where one can avoid doing a grid search over Θ in order to construct the projection of Θ^* along each of its dimensions.

3.3 Simulations

We implement the developed procedure in some simulations with the purpose to investigate how the shape of the identified set is sensitive to variations in the underlying DGP.

[1] Set of choice probabilities predicted by 1BCEs for a given $\boldsymbol{\theta}$

The objective of this paragraph is to get a preliminary understanding of the identifying power of 1BCE. In particular, for a given value of θ, we construct the collection of conditional choice probabilities that are rationalised by 1BCE and compare it with the unit simplex.

We consider the Nested Logit model introduced in Example 1 of Section 2, when $X_i\equiv Z_i$, $\epsilon_i\equiv\xi_i$, $V_i\equiv\eta_i$, and $L=3$. Hereafter, we refer to this DGP as DGP1. Given $\theta\equiv(\beta,\lambda)=(0,0.5)$, let $\bar{\mathcal{R}}^{\theta,\text{com}}_Y$ be the collection of choice probabilities induced by the model's optimal strategies when the researcher assumes that all DMs process the complete information structure.[17] Let $\bar{\mathcal{R}}^{\theta,\text{deg}}_Y$ be the collection of choice probabilities that are induced by the model's optimal strategies when the researcher assumes that all DMs process the degenerate information structure. Finally, recall that $\bar{\mathcal{Q}}^{\theta}_Y$ is the collection of choice probabilities that are induced by 1BCEs, as defined in (7). Figure 1 represents $\bar{\mathcal{Q}}^{\theta}_Y$ (black region), $\bar{\mathcal{R}}^{\theta,\text{com}}_Y$ (red point), and $\bar{\mathcal{R}}^{\theta,\text{deg}}_Y$ (blue point). By Theorem 1, $\bar{\mathcal{R}}^{\theta,\text{com}}_Y$ and $\bar{\mathcal{R}}^{\theta,\text{deg}}_Y$ are subsets of $\bar{\mathcal{Q}}^{\theta}_Y$. Further, note that $\bar{\mathcal{Q}}^{\theta}_Y$ is a strict subset of the unit simplex, which suggests that the notion of 1BCE may have some empirical content in this setting.

[17] Since $\beta=0$, in this exercise we compute unconditional choice probabilities.

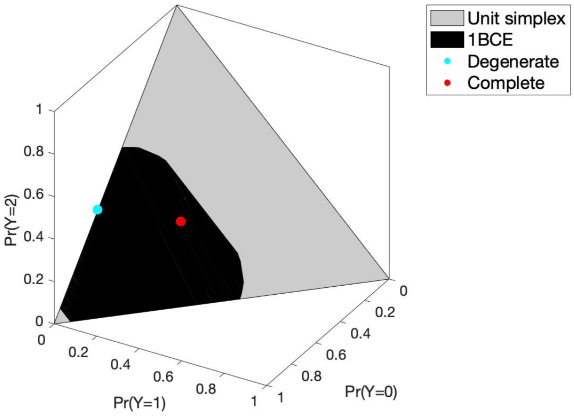

Figure 1: The figure represents $\bar{\mathcal{Q}}_Y^{\theta}$ (black region), $\bar{\mathcal{R}}_Y^{\theta,\text{com}}$ (red point), and $\bar{\mathcal{R}}_Y^{\theta,\text{deg}}$ (blue point) under DGP1.

[2] Information structures

The objective of this paragraph is to investigate how the shape of the identified set depends on the information structures processed by DMs in the underlying DGP.

We consider again the Nested Logit model introduced in Example 1 of Section 2, when $X_i \equiv Z_i$, $\epsilon_i \equiv \xi_i$, and $V_i \equiv \eta_i$. We set $L = 4$, $M = 1$, $\beta = 1.6$, and $\lambda = 0.5$. We randomly draw covariates from a probability mass function, which is constructed by taking a trivariate normal and then discretising it to have support $\{-1, 0, 1\}^3$. The empirical conditional choice probabilities are derived under three alternative information environments: (i) $\frac{1}{10}$ of the population processes the complete information structure and $\frac{9}{10}$ of the population processes the degenerate information structure (hereafter, DGP2); (ii) $\frac{1}{2}$ of the population processes the complete information structure and $\frac{1}{2}$ of the population processes the degenerate information structure (hereafter, DGP3); (iii) each agent processes the complete information structure (hereafter, DGP4). Lastly, in the presence of ties, DMs select one of the maximisers uniformly at random.

Figure 2 represents the identified set (black region), the true value of the parameters (red point), and the value of the parameters that is identified when all DMs are assumed to process the complete information structure as standard in the Nested Logit framework (blue point); under DGP2 (first picture from the left), DGP3 (second picture from the left), and DGP4 (last picture from the left). For each of these DGPs, Table 1 reports the true value of the parameters (second column), the value of the parameters that is identified when all DMs are assumed to process the complete information structure (third column), and the projection of the identified set along every dimension (fourth column).

Assuming that all DMs are fully informed, as standard in the Nested Logit framework, leads to recovering one parameter value which is contained in the identified set. When this assumption is misspecified, the recovered parameter value can be different from the truth, as it is the case under DGP2 and DGP3, where the blue and red points are quite close with regards to λ but relatively far apart with regards to β. Such a result should warn analysts to be cautious about imposing restrictions on the information environment because these can have consequences on the empirical conclusions, if incorrect. Instead, under DGP4, the red and blue points coincide as expected, because the assumption on the information structures of DMs is not misspecified.

Across the three DGPs considered, the model has the least identifying power under DGP2, i.e., when the majority of DMs process the degenerate information structure. As soon as a significant proportion of DMs process the complete information structure, the identifying power of the model improves significantly. This is because, under DGP3 and DGP4, many (all) DMs take their decisions based on the true payoffs. Instead, under DGP2, all DMs take their decisions based on the expected payoffs computed by relying on the same prior. Hence, under DGP3 and DGP4, there is more variation in the agent responses for each value of the covariates. This leads to tighter bounds for β. It is because data featuring less variation in the agent responses can be reproduced by more combinations of parameters, information structures, and tie-breaking rule.[18] The projection for λ is narrower under DGP3 and DGP4 than under DGP2, but it seems less sensitive to the underlying information environment.

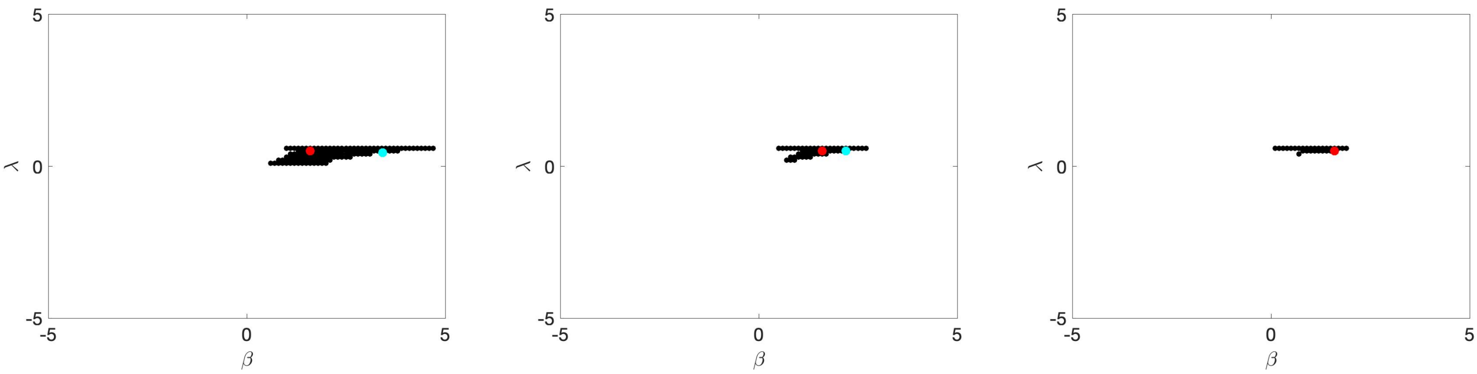

Figure 2: From the left: the first figure is based on DGP2; the second figure is based on DGP3; the last figure is based on DGP4. In each figure: the black region is the identified set; the red point is the true value of the parameters; the blue point is the value of the parameters that is identified when all DMs are assumed to process the complete information structure.

	True	Complete	Identified set
DGP2		$\beta = 3.4184, \lambda = 0.4503$	$\beta \in [0.6, 4.7], \lambda \in [0.1, 0.6]$
DGP3	$\beta = 1.6, \lambda = 0.5$	$\beta = 2.1979, \lambda = 0.5052$	$\beta \in [0.5, 2.7], \lambda \in [0.2, 0.6]$
DGP4		$\beta = 1.6, \lambda = 0.5$	$\beta \in [0.1, 1.9], \lambda \in [0.4, 0.6]$

Table 1: Results under DGP2-DGP4.

[18] To see this better, ignore ϵ_i in the payoff specification and generate the data under the assumption that all agents process the degenerate information structure. Then, the choices of DMs are driven only by the covariate realisations. Assume, for simplicity, that there are no ties. Hence, for every given realisation of X_i, we will see all agents choosing the same alternative. For example, suppose that all agents pick the alternative with the highest realisation of X_{iy}. Then, the data can be replicated under any $\beta > 0$ combined with the degenerate information structure.

[3] Variation in X_i

The objective of this paragraph is to investigate how the shape of the identified set depends on the variation of X_i in the underlying DGP.

We now consider a Multinomial Probit framework with payoff function specified as

$$u(y, X_i, \epsilon_i, V_i) \equiv \begin{cases} \beta' X_{iy} + \epsilon_i + V_{iy} & \text{if } y \in \mathcal{Y} \setminus \{0\}, \\ V_{i0} & \text{if } y = 0, \end{cases} \tag{12}$$

where the notation follows Example 1 of Section 2 and (ϵ_i, V_i) is distributed as an $(L+1)$-variate standard normal, independent of X_i. We set $L = 4$, $M = 1$, and $\beta = -2.5$. X_i is randomly drawn from a probability mass function, which is constructed by taking a trivariate normal and then discretising it to have a given support, $\mathcal{X}$. In particular, we implement four alternative scenarios, where we progressively increase the cardinality of $\mathcal{X}$: (i) $\mathcal{X} \equiv \{0,1\}^3$ (hereafter, DGP5); (ii) $\mathcal{X} \equiv \{-1,0,1\}^3$ (hereafter, DGP6); (iii) $\mathcal{X} \equiv \{-2,-1,0,1,2\}^3$ (hereafter, DGP7); (iv) $\mathcal{X} \equiv \{-3,-2,-1,0,1,2,3\}^3$ (hereafter, DGP8). In all the three DGPs, each agent processes the complete information structure. The first four rows of Table 2 report the identified set of β under DGP5-DGP8. We find that the identified set shrinks as the cardinality of $\mathcal{X}$ increases. This is because variation of exogenous variables that are "observed pre-signal" by the researcher and DMs helps identification of β, as it is the case in standard parametric analysis.

[4] Without ϵ_i

The objective of this paragraph is to investigate how the shape of the identified set depends on the presence of individual-specific tastes that are private information of DMs.

We consider (12) without ϵ_i entering the payoffs from the inside alternatives. Again, we set $L = 4$, $M = 1$, and $\beta = -2.5$. X_i is randomly drawn from a probability mass function, which is constructed by taking a trivariate normal and then discretising it to have support $\{-1,0,1\}^3$. Each agent processes the complete information structure. Hereafter, we refer to this as DGP9. Note that DGP9 is equal to DGP6, except for that the absence of ϵ_i. The fifth row of Table 2 reports the identified set of β under DGP9. We find that the identified set is larger than the identified set obtained under DGP6. This is because individual specific tastes induce more variation in the agent responses for each value of the covariates, which in turn helps identification of β, as discussed for DGP2-DGP4.

[5] Correlation between the components of V_i

The objective of this paragraph is to investigate how the shape of the identified set depends on the presence of correlation between the components of the vector V_i as part of the common prior.

We consider (12), where V_i is distributed as an L-variate normal with mean 0_L and variance-covariance matrix $I_L + \rho(1_L - I_L)$. We set $L = 4$, $M = 1$, and $\beta = 2.5$. X_i is randomly drawn from a probability mass function, which is constructed by taking a trivariate normal and then discretising it to have support $\{0, 1\}^3$. The empirical choice probabilities are derived under three alternative scenarios, where we progressively increase ρ: (i) $\rho = 0$ (hereafter, DGP10); (ii) $\rho = 0.2$ (hereafter, DGP11); (iii) $\rho = 0.6$ (hereafter, DGP12). In all the three DGPs, each agent processes the complete information structure. The last three rows of Table 2 report the identified set of β under DGP10-DGP12, while we pretend ρ to be point identified. We find that the identified set shrinks as ρ increases. This is because, when ρ is high, the realisations of $\{V_{iy}\}_{y\in\mathcal{Y}}$ are "partly" predictable, which in turn helps identification of β.

		True	Identified set
Cardinality of $\mathcal{X}$	DGP5	$\beta = -2.5$	$\beta \in [-3.13, -0.76]$
	DGP6		$\beta \in [-3.13, -0.83]$
	DGP7		$\beta \in [-2.84, -0.85]$
	DGP8		$\beta \in [-2.71, -0.92]$
Without ϵ_i	DGP9	$\beta = -2.5$	$\beta \in [-3.28, 0]$
Correlation in V_i	DGP10	$\beta = 2.5$	$\beta \in [1.01, 2.86]$
	DGP11		$\beta \in [1.08, 2.81]$
	DGP12		$\beta \in [1.62, 2.53]$

Table 2: Results under DGP5-DGP11.

4 Empirical application

In this section we use our methodology to study the determinants of voting behaviour during the UK general election held on 8 June 2017 and perform some counterfactual exercises.

4.1 Setting and model specification

The spatial model of voting is a dominant framework in political economy to explain individual preferences for parties and, in turn, how such preferences shape the policies implemented by democratic societies (see Section 1 for references). This model posits that an agent has a most preferred policy (also called "bliss point") and casts her vote in favour of the party whose position is closest to her ideal (i.e., she votes "ideologically"). In empirical analysis, it is typically implemented by estimating a classical parametric discrete choice model with perfect information (see Section 1 for references). That is, it is assumed that each DM i processes the complete information structure and votes for party $y \in \mathcal{Y}$ maximising her utility,

$$u(y, X_i, \omega_i) \equiv \beta' Z_{iy} + \gamma_y' W_i + \omega_{iy}, \tag{13}$$

where $Z_{iy} \equiv |Z_i - Z_y|$ is an $M \times 1$ vector observed by the researcher, representing the distance between DM i's opinion (Z_i) and party y's opinion (Z_y) on M issues, as measured in some common M-dimensional ideological metric space. W_i is a $J \times 1$ vector of individual-specific covariates observed by the researcher. $X_i \equiv (Z_{iy}\ \forall y \in \mathcal{Y}, W_i)$ collects the ideological distances and the individual-specific covariates. ω_{iy} is a scalar capturing the tastes of DM i for each party/candidate that are unknown to the researcher, whose distribution belongs to a parametric family, thereby outlining a (Multinomial) Logit model, Probit model, Nested Logit model, etc. If voters vote ideologically, then β is expected to be negative so that DM i's utility declines with increasing distance between Z_i and Z_y.[19] Finally, note that according to the spatial model of voting, citizens receive utility from truthfully expressing their preferences, plausibly due to a sense of civic responsibility. For the same reason, citizens care about evaluating the returns to voting as accurately as possible.

The above model is scientifically appealing because of its elegance and simplicity but has limitations. Importantly, many papers highlight that uncertainty affects voting (see references in Section 1). That is, voters may be unsure about their own and the parties' ideological positions and, more generally, about the qualities of the candidates. This is because of the inevitable difficulty of making precise political judgments and understanding associated returns, or because the parties deliberately obfuscate information in order to attract voters with different preferences and expand electoral support. One of the most prominent early works states that:

> *The democratic citizen is expected to be well informed about political affairs. He is supposed to know what the issues are, what their history is, what the relevant facts are, what alternatives are proposed, what the party stands for, what the likely consequences are. By such standards, the voter falls short.* (Berelson, Lazarfeld, and McPhee, 1954, p.308).

More plausibly, in the wake of election campaigns, voters are conscious of their own and the parties' attitudes towards some popular issues, but might be uncertain about how they themselves and the parties stand towards more technical or less debated topics, and about the traits of the candidates other than those publicly advertised. Further, they may attempt to fill such gaps in information with various degrees of success and in different ways, depending on a priori inclination for certain parties, political sentiments, interest in specific issues, civic sense, intellectual preparation, attentional limits, participation in seminars, candidates' transparency, opinion makers, and media exposure. In turn, some individuals might become much more informed, others less, giving rise to heterogeneity in the public understanding of politics. In fact, it has been observed that:

> [...] *in a world of imperfect information, a world in which there are costs associated with gathering and evaluating new information, the voter, faced with a serious decision such as deciding which candidate would make a better president, is forced to utilize a shortcut method to arrive at his choice.* (Enelow and Hinich, 1981, p.489).

[19] See Degan and Merlo (2009) and Henry and Mourifié (2013) for a characterisation of conditions under which the hypothesis that individuals vote ideologically is falsifiable. See Merlo and De Paula (2017) for nonparametric identification and estimation of preferences from aggregate data when individuals vote ideologically.

Also,

> *It is not reasonable to suppose that the voter who is exceedingly well informed about politics and the one who is largely ignorant of it would enumerate potentially relevant considerations with the same exhaustiveness; or frame alternative considerations with the same precision; or foresee consequences of alternative choices with the same distinctness; or coordinate calculations, both about alternative means and alternative ends, with the same exactness.* (Sniderman, Brody, and Tetlock, 1991, p.165-166).

Despite the acknowledgement of the central role played by the sophistication of voters in determining voting patterns, only a few empirical works have attempted to take it into account while estimating a spatial voting framework. This has been done, for instance, through an additive, exogenous, and parametrically distributed error in the payoffs representing the evaluation mistakes made by voters, or a parametric specification of the variance of the perceived party position across voters, or a parametric specification of the probability of being informed versus uninformed when voting (for references, see Section 1). By contrast, our methodology permits us to incorporate uncertainty under weak assumptions on the latent, heterogeneous, and potentially endogenous process followed by voters to gather and evaluate information.

In particular, we focus on the following setting. We assume that, when assessing the returns to voting for each party $y \in \mathcal{Y}$, DM i is:

- Aware of the distances between her position and party y's position on highly debated topics (for instance, EU integration). These distances are captured in the vector $Z_{iy} \equiv |Z_i - Z_y|$. The realisation of Z_{iy} is also observed by the researcher, as typically available in survey data.

- Potentially uncertain about her tastes towards party y's opinion on more complicated and less media-covered issues (for instance, public expenditure management and reactions to pandemics) and towards the candidates' qualities that have been less advertised (for instance, disclosure of assets, liabilities, and any conflict of interests). These tastes are captured, in some aggregate way, by the scalar V_{iy}. DM i has a prior on V_{iy}. Further, DM i has access to a learning technology that allows her to become more informed about V_{iy}. The realisation of V_{iy} is not observed by the researcher, as typically unavailable in survey data.

In turn, the payoff from voting party $y \in \mathcal{Y}$ is specified as

$$u(y, X_i, \omega_i) \equiv \beta' Z_{iy} + \gamma_y' W_i + \epsilon_i + \sigma V_{iy}, \tag{14}$$

where (W_i, ϵ_i) control for other individual characteristics in the information set of DM i (resp. observed and unobserved by the researcher), $X_i \equiv (Z_{iy}\ \forall y \in \mathcal{Y}, W_i)$, $V_i \equiv (V_{iy}\ \forall y \in \mathcal{Y})$, $\omega_i \equiv (\epsilon_i, V_i)$, and $\sigma > 0$ to allow for arbitrary variance. Our objective is conducting inference on $(\beta, \gamma_y\ \forall y \in \mathcal{Y}, \sigma)$.

We close the model specification with a discussion on how we parameterise the prior of voters about V_i. We follow the literature on voting under uncertainty which favours modelling priors according to normal distributions (see, for example, Knight and Schiff, 2010; Matějka and Tabellini, 2019; Yuksel, 2020). In particular, we assume that V_i is distributed as an L-variate standard normal, independent of (X_i, ϵ_i), where L is the cardinality of $\mathcal{Y}$.[20] There is no way to test for the normality of the prior. Nevertheless, recall that this is still much weaker than estimating a classic Multinomial Probit model with perfect information.Finally, we assume that ϵ_i has a standard normal distribution independent of X_i.

Before concluding this section, note that we could also allow the prior of voters on V_i to be heterogenous across X_i, for instance by assuming that V_i conditional on $X_i = x$ is distributed as an L-variate normal with mean $\mu(x)$ and variance-covariance matrix $\Sigma(x)$ for each $x \in \mathcal{X}$. Similarly, we could allow the prior of voters on V_i to be heterogenous across ϵ_i, for instance by assuming that (V_i, ϵ_i) are jointly normal and correlated. We have opted for a common prior (regardless of the realisations of X_i and ϵ_i) in order to have a lighter specification and speed up computation.

4.2 Data

We estimate our model by using data on the UK general election held on 8 June 2017. Specifically, we use data from the British Election Study, 2017: Face-to-Face Post-Election Survey (Fieldhouse, et al., 2018). The survey took place immediately after the election. It asks questions concerning key contemporary problems about political representation, accountability, and engagement, and aims to explain changes in party support. The interviewees constitutes an address-based random probability sample of eligible voters living in 468 wards in 234 Parliamentary Constituencies across England, Scotland, and Wales.

We believe that such data fit the framework described in Section 4.1 for four main reasons. First, the UK parties were clearly deployed with regards to the Brexit and focused their election campaign around a few topics such as public health and austerity, thus inducing potential uncertainty among voters with respect to many other factors (Hutton, 2017; Snowdon and Demianyk, 2017). Second, the UK political scene is dominated by historical parties. Hence, past election outcomes and consequent behaviour of parties can justify the common prior assumption on V_i. We also refer the reader to Remark 3 for a discussion on the use of the common prior assumption and Bayesian rationality in the voting literature. Third, the survey reports the positions of the respondents on topics that were debated at length before the election, which we discuss more precisely below. The survey also asks respondents to state the parties's positions with respect to those topics and the answers provided are substantially aligned. This suggests that those topics can be used to construct the vector Z_{iy} for each party

[20] Recall that V_{iy} can also have negative realisations because it captures not only ideological distances but also quality of candidates.

$y \in \mathcal{Y}$, whose realisation is observed by the researcher and voters. The survey does not contain data on other factors that might have induced uncertainty among voters. Hence, it is natural to treat the realisation of V_i as unobserved by the researcher. As in Knight and Schiff (2010), here we consider a setting where the researcher does not know more than voters about the returns to voting. Fourth, the survey asks respondents to declare if they voted tactically. Only 2.16% of the respondents answer affirmatively. We delete them from our final sample, in order for the assumption that voters vote ideologically to be appropriate.

We now describe the data in more details. To limit the impact of Scottish and Welsh independentist fronts on our results, we focus on the respondents who reside in England. We consider the answers of respondents on which party they have voted for among the Conservative Party, Labour Party, Liberal Democrats, United Kingdom Independence Party (UKIP), Green Party, and none.[21]

We collect in Z_{iy} the distances between DM i's position and party y's position on four dimensions: EU integration, taxation and social care, income inequality, and left-right political orientation. These are the most publicised topics on which the election was contested. More precisely, we select the answers of the respondents to the following questions (summarised with respect to the original version, for brevity):

1. [EU integration]: On a scale from 0 to 10, do you think that Britain should do all it can to unite fully with the European Union (0), or do all it can to protect its independence from the European Union (10)? Provide also the positions of the parties on the same scale.

2. [Taxation and social care]: On a scale from 0 to 10, do you think that government should cut taxes a lot and spend much less on health and social services (0), or that government should raise taxes a lot and spend much more on health and social services (10)? Provide also the positions of the parties on the same scale.

3. [Income inequality]: On a scale from 0 to 10, do you think that government should make much greater efforts to make people's incomes more equal (0), or that government should be much less concerned about how equal people's incomes are (10)? Provide also the positions of the parties on the same scale.

4. [Left-right political orientation] Where would you place yourself on a scale from 0 to 10 where 0 denotes left political attitudes and 10 denotes right political attitudes? Provide also the positions of the parties on the same scale.

[21] The original questionnaire includes among the possible answers also the Scottish National Party (i.e, the Scottish nationalist social-democratic party in Scotland), Plaid Cymru (i.e., the Welsh nationalist social-democratic party in Wales), other unspecified minor parties, and "Refused to declare". None of the respondents who reside in England have voted for the Scottish National Party. Only one of the respondents who reside in England has voted for Plaid Cymru. 4 respondents who reside in England have voted for other unspecified minor parties. Lastly, 3.52% of the respondents who reside in England have refused to declare who they have voted for. We have dropped all these observations.

Following the literature (for example, Alvarez and Nagler, 1995; 1998; 2000; Alvarez, Nagler, and Bowler, 2000), we set party y's position on dimensions 1-4 equal to the median placement of the party on each dimension across the sample, although as noticed above there is substantial alignment among the respondents' answers.

We collect in W_i some demographic characteristics of respondents. In particular, we focus on gender, socio-economic class, and total income before tax. Recall that γ_y captures the impact of W_i on the vote shares. We allow this impact to be heterogenous across the parties. To be parsimonious on the number of parameters to estimate, we further parameterise γ_y by requiring that $\gamma_y \equiv \gamma Z_y^{\text{LR}}$ for every party y, where Z_y^{LR} is the position of party y with regards to left-right orientation. In other words, we assume that the aforementioned heterogeneity is driven by the position of each party in the left-right political spectrum. Lastly, we consider abstention as base category and normalise its payoff to zero as, for example, in Knight and Schiff (2010).

After omitting observations with missing data, our sample is made up of $1,217$ individuals. Of these, 36.48% have voted for the Labour Party, 36.65% for the Conservative Party, 6.41% for the Liberal Democrats, 1.73% for UKIP, 1.56% for the Green Party, and 17.17% did not vote.[22] Table 3 presents some descriptive statistics. The second column refers to the positions of the respondents on dimensions 1-4 and reports the mean (rounded to the nearest integer), median, and standard deviation across the sample. The remaining columns reports Z_y for each party y. As expected, the Conservative Party and UKIP are more right-wing, less concerned with income inequality, more Eurosceptic, and stronger supporters of low taxes and a minimal welfare state, than the Labour Party and the Green Party. The Liberal Democrats are more centrist.

	Self (Mean, Median, St.Dev.)	Conservative	Labour	Lib. Dem.	UKIP	Green
EU	5 5 3.355	7	4	3	10	3
Social care	7 7 2.051	5	7	6	4	6
Inequality	4 4 2.743	6	3	4	5	3
Left-right	5 5 2.059	8	2	5	9	3

Table 3: Descriptive statistics on the ideological positions.

The sample is gender balanced, with 48.97% of males and 51.03% of females. We assign

[22] Before omitting observations with missing data, the percentages are: 34.74% for the Labour Party, 33.78% for the Conservative Party, 5.39% for the Liberal Democrats, 1.91% for UKIP, 1.63% for the Green Party, and 22.56% did not vote. The survey seems slightly skewed towards supporters of the Labour Party. The actual vote shares in England were 41.9% for the Labour Party, 45.6% for the Conservative Party, 7.8% for the Liberal Democrats, 2.1% for UKIP, 1.9% for the Green Party. The election results led to a hung parliament and the Conservative Party formed a minority government supported by an agreement with the Northern Ireland's Democratic Unionist Party. The vote shares of each party (including minorities) can be obtained, for example, here `https://www.bbc.co.uk/news/election/2017/results/england`. See also `https://en.wikipedia.org/wiki/Opinion_polling_for_the_2017_United_Kingdom_general_election#2017` for opinion polls organised by various organisations to gauge voting intentions.

label 1 to females and 0 to males. In the original data, the socio-economic class is divided into seven categories, following the Standard Occupation Classification 2010: professional occupations; managerial and technical occupations; skilled occupations - non-manual; skilled occupations - manual; partly skilled occupations; unskilled occupations; armed forces. To lessen the computational burden, we reorganise these categories into three groups. The first group is assigned label 0 and collects professional occupations, managerial and technical occupations, skilled occupations - non-manual, and armed forces (68.04% of the sample). The second group is assigned label 1 and collects skilled occupations - manual and partly skilled occupations (29.42% of the sample). The third group is assigned label 2 and collects unskilled occupations (2.54% of the sample). Similarly, in the original data, the total income before tax is bracketed into 14 categories. We reorganise these categories into four groups, which we construct by approximately following the UK income tax rates. The first group is for income between £0 and £15,599 (21.78% of the sample). The second group is for income between £15,600 and £49,999 (51.68% of the sample). The third group is for income between £50,000 and £99,999 (21.45% of the sample). The fourth group is for income above £100,000 (5.09% of the sample). To each of the four groups, we assign as value the logarithm of the median income across the respondents belonging to that group (9.4727, 10.4282, 11.1199, and 12.6115, respectively). We summarise these numbers in Table 4.

Gender	Socio-economic class	Income
Males (0): 48.97%	First group (0): 68.04%	First group (9.4727): 21.78%
Females (1): 51.03%	Second group (1): 29.42%	Second group (10.4282): 51.68%
	Third group (2): 2.54%	Third group (11.1199): 21.45%
		Fourth group (12.6115): 5.09%

Table 4: Descriptive statistics on the demographic characteristics.

4.3 Implementation of the methodology and results

Table 5 presents the results. In particular, the second column reports the maximum likelihood estimates (hereafter, $\hat{\theta}_{\text{com}}$) and standard errors under the traditional assumption that all voters process the complete information structure (that is, each voter observes the realisation of V_i). The third column reports the maximum likelihood estimates (hereafter, $\hat{\theta}_{\text{deg}}$) and standard errors under the assumption that all voters process the degenerate information structure (that is, each voter votes according to the normal prior);[23] the last column is based on our methodology (that is, each voter can vote by using any posterior compatible with the normal prior) and reports the projections of the asymptotically half-median-unbiased estimated identified set (hereafter, $C_{n,0.50}$), along each dimension. See Appendix B on how to construct an asymptotically half-median-unbiased estimated set for Θ^*. Further inference results are discussed in

[23] We discuss in Appendix E how $\hat{\theta}_{\text{com}}$ and $\hat{\theta}_{\text{deg}}$ are computed.

Appendix E.

		Complete $\hat{\theta}_{\text{com}}$	Degenerate $\hat{\theta}_{\text{deg}}$	Our methodology
β_1	(EU)	-1.3269 (0.3204)	-0.0137 (0.0018)	$[-1.5017, 0]$
β_2	(Social care)	-0.1056 (0.2913)	-0.0224 (0.0042)	$[-3.0802, 0]$
β_3	(Inequality)	-0.6550 (0.2662)	-0.0127 (0.0030)	$[-2.6685, 0]$
β_4	(Left-right)	-2.6916 (0.5413)	-0.0225 (0.0018)	$[-3.5286, 0]$
γ_1	(Gender)	-0.1042 (0.1489)	-0.0013 (0.0018)	$[-3.5512, 2.3588]$
γ_2	(Socio-economic class)	-0.7551 (0.1857)	-0.0065 (0.0018)	$[-2.2181, 2.4728]$
γ_3	(Gross income)	0.0788 (0.0160)	0.0011 (0.0001)	$[-0.0396, 0.5197]$
$\log(\sigma^2)$		5.6787 (0.4450)	Not identified	$[4.4684, 6.0000]$

Table 5: Inference results.

Under the assumption that voters are fully informed (second column), all the β coefficients, except β_2, are significantly different from zero at 5%. This suggests that DMs vote ideologically on the EU, inequality, and left-right dimensions. That is, the smaller is the distance between DM i and party y's ideological positions on those dimensions, the more likely DM i votes for party y, ceteris paribus. Further, β_4 has the highest magnitude among the β coefficients. That is, voters particularly dis-value casting their votes in favour of a party which is ideologically distant on the left-right axis. More precisely, a one unit increase in the ideological distance on the left-right axis produces a payoff decrease that is roughly 2, 25, and 4 times bigger than the payoff decrease produced by a one unit increase in the ideological distance on the EU, social care, and inequality dimensions, respectively, ceteris paribus.

Under the assumption that voters are fully uninformed (third column), all the β coefficients are significantly different from zero at 5%, but they are much closer to zero and similar in magnitude than in the complete information case. β_4 has still the highest magnitude among the β coefficients, but now a one unit increase in the ideological distance on the left-right axis produces a payoff decrease that is only slightly bigger than the payoff decrease produced by a one unit increase in the ideological distance on any one of the other three dimensions, ceteris paribus.

When we remain agnostic about the level of voter sophistication (fourth and fifth columns), all the projections for the β coefficients include zero. Therefore, differently from above, we cannot reject the possibility that the ideological distances on the EU, social care, inequality, and left-right dimensions are statistically insignificant. In fact, voters may actually ignore theirs and the parties' opinions on the most debated policy issues and, rather, take their voting decisions based on some other latent factors. With regards to the magnitude of the β

coefficients, β_4 can have the highest magnitude among the β coefficients, in agreement with the results of the second and third columns. The fact that this finding on β_4 is robust to the restrictions on the information environment is in line with several post-election descriptive studies run by political experts, which emphasise that the traditional left-right values, rather than specific policy issues, have been the main driver of the British electoral behaviour in 2017. For example:

> [...] *the 2017 election resulted in the resurgence of two-party politics based on contestation along the classic economic left–right dimension* [...] (Hobolt, 2018, p.1-2).

> *This should not be surprising given that post-war party competition in Britain, and in most of Western Europe, has been organized around the economic left–right dimension. Moreover, given the nature of the election campaign where the two parties took very distinct positions on these economic issues – after two decades of ideological convergence – it is understandable economic left–right attitudes were also salient to voters.* (Hobolt, 2018, p.7).

Analysts highlight that the Brexit issue has played a critical role as well and, in fact, they often refer to the 2017 election as the "Brexit election" (for example, Mellon, et al., 2018). However, this fact does not come out clearly from our projection for β_1, whose lower bound is the smallest in absolute value among the β coefficients. The reason could be that the 2017 pre-election period saw a substantial increase in the relationship between EU referendum choice and Labour versus Conservative vote choice, with a sort of alignment of the remain-leave axis with the traditional left-right axis. The parties with the clearest positions against the Brexit (the Liberal Democrats) and in favour of the Brexit (UKIP) lost many supporters. These switched en masse to the Labour Party, offering a "soft Brexit" and the Conservative Party, offering a "hard Brexit", respectively (Mellon, et al., 2018; Heat and Goodwin, 2017). Such a tendency might have obfuscated the role of β_1 in determining the preferences of voters.

With regards to the γ coefficients,[24] under the assumption that voters are fully informed (second column), voting for a right-wing party when being a woman generates a lower payoff than voting for a left-wing party, ceteris paribus. Further, the lower is the socio-economic class, the lower is the payoff from voting a right-wing party than the payoff from voting a left-wing party, ceteris paribus. Finally, the higher is the gross income, the higher is the payoff from voting a right-wing party than the payoff from voting a left-wing party, ceteris paribus. Similar findings are obtained under the assumption that voters are fully uninformed (third column), even though the coefficients are much closer to zero than in the complete information case. Instead, when we remain agnostic about the level of voter sophistication (fourth and fifth columns), these facts are not validated, as the projections for γ_1, γ_2, and γ_3 of our estimated

[24] Recall that, to interpret the γ coefficients, one has to interact them with the ideological position of each party on the left-right ideological axis.

set lie on the positive and negative real line. Only the negative sign of the gross income is partially confirmed, as the projection for γ_3 of our estimate set (fourth column) lies mostly on the positive real line.

We also highlight that the projections for the β and γ coefficients in the fourth column contain the estimates in the second and third columns. This is in line with our identification results, according to which the sets of parameter values recovered under specific information structures belong to Θ^* when they are non-empty.

4.4 Policy experiments

To better interpret the magnitude of our results we perform two counterfactual experiments. Various political experts sustain that, while at the beginning of the 2017 election campaign the Conservative Party had a sizeable lead in the opinion polls over the Labour Party, as the campaign progressed the Labour Party recovered ground because it strengthened its left ideological position on social spending and nationalization of key public services (for example, Heath and Goodwin, 2017; Mellon, at al., 2018). To evaluate this, we reset the Labour Party's placement on dimension 2 (social care) to be two points less (i.e., 5 instead of 7) and study how the vote share of the Labour Party changes (hereafter, *counterfactual 1*).

A few steps should be implemented to operationalise this exercise. First, we introduce some useful notation. For each $x \in \mathcal{X}$, we denote by $\hat{x}$ the corresponding transformed realisation. Hereafter, computations done at x will be referred to as the factual scenario and computations done at $\hat{x}$ will be referred to as the fictional scenario. We collect all such $|\mathcal{X}|$ pairs, $(x, \hat{x})$, in $\bar{\mathcal{X}}$. Hereafter, a generic element of $\bar{\mathcal{X}}$ is interchangeably indicated by $\bar{x}$ or $(x, \hat{x})$. We denote by y_{Labour} the element of $\mathcal{Y}$ referred to the Labour Party.

Second, we establish a way to summarise the impact of the counterfactual intervention across the parameter values in $C_{n,0.50}$ and across the conditional choice probabilities induced by 1BCEs. We proceed as follows. For each $\bar{x} \in \bar{\mathcal{X}}$ and $\theta \in C_{n,0.50}$, we focus on the maximum attainable vote shares of the Labour Party across those predicted by 1BCEs, as a best-case scenario for the Labour Party. We denote such maximum attainable vote shares in the fictional and factual scenarios by $\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})$ and $\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|x)$, respectively. We repeat the same, now focusing on the minimum attainable vote shares of the Labour Party across those predicted by 1BCEs, as a worst-case scenario for the Labour Party. We denote such minimum attainable vote shares in the fictional and factual scenarios by $\underline{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})$ and $\underline{P}^{\theta}_{Y|X}(y_{\text{Labour}}|x)$, respectively. Then, we take the difference between $\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})$ and $\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|x)$, and between $\underline{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})$ and $\underline{P}^{\theta}_{Y|X}(y_{\text{Labour}}|x)$. We integrate out the covariates and report the maxima and minima of the differences obtained across the parameter values in $C_{n,0.50}$. More precisely, let

$$\bar{\Delta}^{\theta}_{\text{Labour}} \equiv \sum_{\bar{x} \in \bar{\mathcal{X}}} [\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|x) - \bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})] P^0_X(\bar{x}), \tag{15}$$

and

$$\underline{\Delta}^{\theta}_{\text{Labour}} \equiv \sum_{\bar{x} \in \bar{\mathcal{X}}} [\underline{P}^{\theta}_{Y|X}(y_{\text{Labour}}|x) - \underline{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})] P^{0}_{X}(\bar{x}). \tag{16}$$

Note that $\bar{\Delta}^{\theta}_{\text{Labour}}$ and $\underline{\Delta}^{\theta}_{\text{Labour}}$ are the changes in the "best-case scenario" vote share and the "worst-case scenario" vote share of the Labour Party, respectively. We report

$$\bar{\mathcal{I}}_{\text{Labour}} \equiv \left[\min_{\theta \in C_{n,0.50}} \bar{\Delta}^{\theta}_{\text{Labour}}, \max_{\theta \in C_{n,0.50}} \bar{\Delta}^{\theta}_{\text{Labour}} \right],$$

which is the interval where the change in the "best-case scenario" vote share of the Labour Party can lie, and

$$\underline{\mathcal{I}}_{\text{Labour}} \equiv \left[\min_{\theta \in C_{n,0.50}} \underline{\Delta}^{\theta}_{\text{Labour}}, \max_{\theta \in C_{n,0.50}} \underline{\Delta}^{\theta}_{\text{Labour}} \right].$$

which is the interval where the change in the "worst-case scenario" vote share of the Labour Party can lie.

Third, we explain how $\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})$, $\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|x)$, $\underline{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})$, and $\underline{P}^{\theta}_{Y|X}(y_{\text{Labour}}|x)$ in (15) and (16) are computed. Let us start from $\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})$. In order to calculate $\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})$, we should decide whether, in the fictional scenario, the information structures of DMs stay fixed at their factual level or are allowed to vary. We proceed by assuming that the information structures of DMs stay fixed. In fact, it is plausible that modifying the ideological position of the Labour Party does not affect how voters learn about payoffs. We thus construct $\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})$ as outlined in Theorem 1 by Bergemann, Brooks, and Morris (2019), which establishes how to obtain the fictional choice probabilities while holding DMs' information structures unchanged. We briefly summarise the procedure. Bergemann, Brooks, and Morris (2019) suggest to consider the set of 1BCEs of the "double choice problem" where DM i chooses alternative y of the factual choice problem and alternative $\hat{y}$ of the fictional choice problem in a way that is consistent with the prior, obedient, and compatible with the empirical conditional choice probabilities. They show that when marginalising the 1BCEs of the double choice problem on the action space, $\mathcal{Y}$, one gets the fictional choice probabilities under constant information structures. More formally, we define the double choice problem, $\bar{G}^{\theta} \equiv \{\bar{\mathcal{Y}}, \bar{\mathcal{X}}, \mathcal{E}, \mathcal{V}, \bar{u}(\cdot;\theta), P_V, P_\epsilon)$, where $\bar{\mathcal{Y}} \equiv \mathcal{Y}^2$ and $\bar{u}(\bar{y}, \bar{x}, e, v; \theta) \equiv u(y, x, e, v; \theta) + u(\hat{y}, \hat{x}, e, v; \theta)$ for each $\bar{y} \equiv (y, \hat{y}) \in \mathcal{Y}^2$, $\bar{x} \in \bar{\mathcal{X}}$, $e \in \mathcal{E}$, and $v \in \mathcal{V}$.[25] By Theorem 1 in Bergemann, Brooks, and Morris (2019),

[25] Here we follow the notation of Sections 2 and 3. Also, recall that Section 4.1 imposes that ϵ_i is independent of X_i and V_i is independent of (X_i, ϵ_i). Therefore, the families of conditional densities $\mathcal{P}_{\epsilon|X}$ and $\mathcal{P}_{V|X,\epsilon}$ can be replaced by the unconditional densities of ϵ_i and V_i. The latter are denoted by P_ϵ and P_V, respectively, and are known by assumption. Further, all the discretisations discussed at the end of Section 3 are assumed to be implemented.

$\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})$ can be computed as

$$\begin{aligned}
\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x}) = & \max_{P_{\bar{Y},V|\bar{X},\epsilon}(\cdot|\bar{x},e)\in\mathbb{R}^{|\bar{\mathcal{Y}}|\cdot|\mathcal{V}|},\forall e\in\mathcal{E}} \sum_{y\in\mathcal{Y},v\in\mathcal{V},e\in\mathcal{E}} P_{\bar{Y},V|\bar{X},\epsilon}(y,y_{\text{Labour}},v|\bar{x},e)P_{\epsilon}(e),\\
& \text{s.t.}\\
& \text{[1BCE-Consistency]:} \quad \sum_{\bar{y}\in\bar{\mathcal{Y}}} P_{\bar{Y},V|\bar{X},\epsilon}(\bar{y},v|\bar{x},e) = P_V(v)\ \forall v\in\mathcal{V},\forall e\in\mathcal{E},\\
& \text{[1BCE-Obedience]:} \quad -\sum_{v\in\mathcal{V}} P_{\bar{Y},V|\bar{X},\epsilon}(\bar{y},v|\bar{x},e)[\bar{u}(\bar{y},\bar{x},e,v;\theta)-\bar{u}(\bar{y}',\bar{x},e,v;\theta)]\leq 0,\\
& \qquad \forall \bar{y}\in\bar{\mathcal{Y}},\forall\bar{y}'\in\bar{\mathcal{Y}}\setminus\{\bar{y}\},\forall e\in\mathcal{E},\\
& \text{[1BCE-Data match]:} \quad P^0_{Y|X}(y|x) = \sum_{\hat{y}\in\mathcal{Y},v\in\mathcal{V},e\in\mathcal{E}} P_{\bar{Y},V|\bar{X},\epsilon}(y,\hat{y},v|\bar{x},e)P_{\epsilon}(e)\ \forall y\in\mathcal{Y}.
\end{aligned} \tag{17}$$

Moreover, we compute $\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|x)$ by following the definition of 1BCE of the baseline choice problem G^{θ} (Definition 2).[26] That is,

$$\begin{aligned}
\bar{P}^{\theta}_{Y|X}(y_{\text{Labour}}|x) = & \max_{P_{Y,V|X,\epsilon}(\cdot|x,e)\in\mathbb{R}^{|\mathcal{Y}|\cdot|\mathcal{V}|},\forall e\in\mathcal{E}} \sum_{v\in\mathcal{V},e\in\mathcal{E}} P_{Y,V|X,\epsilon}(y_{\text{Labour}},v|x,e)P_{\epsilon}(e),\\
& \text{s.t.}\\
& \text{[1BCE-Consistency]:} \quad \sum_{y\in\mathcal{Y}} P_{Y,V|X,\epsilon}(y,v|x,e) = P_V(v)\ \forall v\in\mathcal{V},\forall e\in\mathcal{E},\\
& \text{[1BCE-Obedience]:} \quad -\sum_{v\in\mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e)[u(y,x,e,v;\theta)-u(y',x,e,v;\theta)]\leq 0,\\
& \qquad \forall y\in\mathcal{Y},\forall y'\in\mathcal{Y}\setminus\{y\},\forall e\in\mathcal{E}.
\end{aligned} \tag{18}$$

Lastly, we compute $\underline{P}^{\theta}_{Y|X}(y_{\text{Labour}}|\hat{x})$ and $\underline{P}^{\theta}_{Y|X}(y_{\text{Labour}}|x)$ by solving (17) and (18), with min in place of max.

We also calculate the difference between the fictional and factual choice probabilities under the assumption that all voters process the complete information structure and under the assumption that all voters process the degenerate information structure. Specifically, we report

$$\mathcal{I}^{\text{com}}_{\text{Labour}} \equiv \sum_{\bar{x}\in\bar{\mathcal{X}}}[P^{\hat{\theta}_{\text{com}}}_{Y|X}(y_{\text{Labour}}|x) - P^{\hat{\theta}_{\text{com}}}_{Y|X}(y_{\text{Labour}}|\hat{x})]P^0_X(\bar{x}),$$

and

$$\mathcal{I}^{\text{deg}}_{\text{Labour}} \equiv \sum_{\bar{x}\in\bar{\mathcal{X}}}[P^{\hat{\theta}_{\text{deg}}}_{Y|X}(y_{\text{Labour}}|x) - P^{\hat{\theta}_{\text{deg}}}_{Y|X}(y_{\text{Labour}}|\hat{x})]P^0_X(\bar{x}),$$

where all the probabilities entering $\mathcal{I}^{\text{com}}_{\text{Labour}}$ and $\mathcal{I}^{\text{deg}}_{\text{Labour}}$ are calculated following the standard Multivariate Probit formulas.

Table 6 presents the results for counterfactual 1. Under the assumption that voters are fully informed (first row), we find that holding a position equal to 7 on the social care dimension leads to an almost unnoticeable increase in the vote share of the Labour Party with respect to

[26] Note that here we do not have to impose the data match condition as in (17) because it has been already incorporated in the construction of $C_{n,0.50}$, as explained in Appendix B.1.

$\mathcal{I}^{\text{com}}_{\text{Labour}}$	0.0008
$\mathcal{I}^{\text{deg}}_{\text{Labour}}$	0.0339
$\bar{\mathcal{I}}_{\text{Labour}}$	$[-0.3179, 0.0457]$
$\underline{\mathcal{I}}_{\text{Labour}}$	$[-0.1123, 0]$

Table 6: Results of counterfactual 1.

holding a position equal to 5. Under the assumption that voters are fully uninformed (second row), the increase is more pronounced. This supports the claim that, by strengthening its left ideological position on the social care dimension, the Labour Party gained some votes during the election campaign. However, such a result is not confirmed when we remain agnostic about the level of voter sophistication. In fact, we find that holding a position equal to 7 on the social care dimension might lead to a significant decrease in the vote share of the Labour Party with respect to holding a position equal to 5. Our results highlight that imposing strong assumptions on the information environment can drive the types of conclusions we reach.

We now implement a second counterfactual experiment (hereafter, *counterfactual 2*). The uncertainty about the payoffs resulting from voting can occur due to deliberate strategies of the candidates who "*becloud*" their characteristics and opinions "*in a fog of ambiguity*" (Downs, 1957, p.136), in order to expand the electoral support by attracting groups of voters with different political preferences (Campbell, 1983; Dahlberg, 2009; Tomz and van Houweling, 2009; Somer-Topcu, 2015). It remains unclear, however, to what extent such uncertainty affects the vote shares and, in turn, influences the election results in democratic societies. A better understanding of it is important for designing transparency laws that can improve citizens' welfare. We investigate this question by imagining an omniscient mediator who implements a policy that gives voters complete information. This can be achieved, for instance, by organising a massive campaign in schools that develops democratic knowledge and political literacy skills,[27] forcing candidates to publicly disclose their assets, liabilities, and criminal records, enforcing a strict regulation regarding campaign spending and airtime, etc. We simulate the counterfactual vote shares under complete information and study how they change with respect to the factual scenario.

Before presenting our results, we emphasise that this question has been largely debated in the literature. As explained by Bartels (1996), political scientists have often answered it by arguing that a large population composed of possibly uninformed citizens act as if it was fully

[27] See, for example, Niemi and Junn (1998), Hooghe and Wilkenfeld (2007), and Pontes, Henn, and Griffiths (2019) on the impact of civic education on political engagement.

informed, either because each voter uses cues and information shortcuts helping her to figure out what she needs to know about the political world; or because individual deviations from fully informed voting cancel out in a large election, producing the same aggregate election outcome as if voters were fully informed. Carpini and Keeter (1996) and Bartels (1996) are the first studies to use quantitative evidence to disconfirm such claims. They simulate counterfactual vote shares under complete information using data on the level of information of the survey respondents as rated by the interviewers or assessed by test items. Several analysis along similar lines have then followed, for example, Althaus (1998), Gilens (2001), and Sekhon (2004). Degan and Merlo (2011) propose an alternative approach, which is closer to ours. As mentioned above, they consider a spatial model of voting with latent uncertainty. Differently from us, they estimate such a model by parametrically specifying the probability that a voter is informed. They use their estimates to obtain counterfactual vote shares under complete information and find that making citizens more informed about electoral candidates decreases abstention. We contribute to this thread of the literature by providing a way to construct counterfactual vote shares under complete information, which neither requires the difficult task of measuring voters' level of information in the factual scenario, nor imposes parametric assumptions on the probability that a voter is informed.

More precisely, for each $\theta \in C_{n,0.50}$, we simulate data from P_ϵ, P_V, and P^0_X. We let individuals vote under complete information. We then compute the fictional vote shares and denote them by $P^{\theta,\text{com}}_{Y|X}(y|x)$ for each $y \in \mathcal{Y}$ and $x \in \mathcal{X}$. We obtain the best-case and worst-case factual vote shares, $\bar{P}^\theta_{Y|X}(y|x)$ and $\underline{P}^\theta_{Y|X}(y|x)$, as outlined in (18) for each $y \in \mathcal{Y}$ and $x \in \mathcal{X}$. We calculate

$$\bar{\Delta}^\theta_y \equiv \sum_{x\in\mathcal{X}} [P^{\theta,\text{com}}_{Y|X}(y|x) - \bar{P}^\theta_{Y|X}(y|x)]P^0_X(x),$$

$$\underline{\Delta}^\theta_y \equiv \sum_{x\in\mathcal{X}} [P^{\theta,\text{com}}_{Y|X}(y|x) - \underline{P}^\theta_{Y|X}(y|x)]P^0_X(x),$$

and

$$\Delta^\theta_y \equiv \sum_{x\in\mathcal{X}} [P^{\theta,\text{com}}_{Y|X}(y|x) - P^0_{Y|X}(y|x)]P^0_X(x),$$

for each $y \in \mathcal{Y}$. Here, $\bar{\Delta}^\theta_y$ and $\underline{\Delta}^\theta_y$ are the gain/loss in vote share under complete information as compared to the best-case and the worst-case factual scenarios, respectively. Further, Δ^θ_y is the gain/loss in vote share under complete information as compared to the empirical factual scenario. Finally, we report the intervals

$$\bar{\mathcal{I}}_y \equiv \left[\min_{\theta\in C_{n,0.50}} \bar{\Delta}^\theta_y, \max_{\theta\in C_{n,0.50}} \bar{\Delta}^\theta_y\right],$$

$$\underline{\mathcal{I}}_y \equiv \left[\min_{\theta\in C_{n,0.50}} \underline{\Delta}^\theta_y, \max_{\theta\in C_{n,0.50}} \underline{\Delta}^\theta_y\right],$$

and

$$\mathcal{I}_y \equiv \left[\min_{\theta \in C_{n,0.50}} \Delta_y^\theta, \max_{\theta \in C_{n,0.50}} \Delta_y^\theta\right],$$

for each $y \in \mathcal{Y}$ in Tables 7 and 8.

	$\bar{\mathcal{I}}_y$ (Best-case factual scenario)	$\underline{\mathcal{I}}_y$ (Worst-case factual scenario)
Abstention	$[-0.9598, -0.0334]$	$[-0.2392, 0.0317]$
Conservatives	$[-0.7968, 0.0050]$	$[0.0950, 0.3006]$
Labour	$[-0.7961, 0.0593]$	$[0.0531, 0.2057]$
Lib. Dem	$[-0.7970, 0.0142]$	$[0.0674, 0.2006]$
UKIP	$[-0.7962, 0.0293]$	$[0.1138, 0.2450]$
Green	$[-0.7968, 0.0396]$	$[0.0838, 0.2032]$

Table 7: Results of counterfactual 2.

	$\mathcal{I}_y$(Empirical factual scenario)
Abstention	$[-0.1669, -0.0392]$
Conservatives	$[-0.1935, -0.0373]$
Labour	$[-0.3118, -0.1591]$
Lib. Dem	$[0.1144, 0.1721]$
UKIP	$[0.1381, 0.3121]$
Green	$[0.0681, 0.1875]$

Table 8: Results of counterfactual 2.

We outline a few guidelines on how to read Table 7. First, each row has to be considered separately because the best-case and worst-case factual vote shares for each $y \in \mathcal{Y}$, $\bar{P}^\theta_{Y|X}(y|x)$ and $\underline{P}^\theta_{Y|X}(y|x)$, are achieved under 1BCEs that can differ across parties. For example, Table 7 reveals that, when electors are fully informed, the Conservative Party may lose votes with respect to the best-case factual scenario (second column). However, we should not necessarily expect such a negative effect to be counterbalanced by a positive effect in other rows of the second column. That is, we should not necessarily expect some other parties to gain votes in the second column. Second, note that, for each $\theta \in C_{n,0.50}$ and $y \in \mathcal{Y}$, it holds that $\bar{\Delta}^\theta_y \leq \underline{\Delta}^\theta_y$. In turn, the lower and upper bounds of $\bar{\mathcal{I}}_y$ are smaller than the lower and upper bounds of $\underline{\mathcal{I}}_y$, respectively, which is the case in Table 7.[28]

We now comment on the results in Table 7. An important finding is that, under complete information, abstention drops with respect to the best-case factual scenario, as the interval in the second column lies entirely on the negative part of $[-1, 1]$. Likewise, abstention is likely to drop with respect to the worst-case factual scenario, as the interval in the third column lies

[28]As a minor remark, we also emphasise that $\bar{\Delta}^\theta$ is *not* necessarily non-positive by construction at *each* $\theta \in C_{n,0.50}$. In fact, if there is a $\theta \in C_{n,0.50}$ such that the complete information structure is supported by a 1BCE matching with the data, then $\bar{\Delta}^\theta$ is non-positive at that specific θ, but not necessarily at other parameter values in our estimated set. For similar reasons, $\bar{\Delta}^\theta$ is *not* necessarily non-negative at *each* $\theta \in C_{n,0.50}$.

mostly on the negative part of $[-1, 1]$. Hence, informed citizens are less prone to abstain. This confirms the empirical results in Degan and Merlo (2011), which also reveal that increasing the awareness of voters decreases abstention. Moving to the parties' vote shares, we see that, under complete information, the parties are likely to lose votes with respect to the best-case factual scenario, as the intervals in the second column lie mostly on the negative part of $[-1, 1]$. This can be due to the fact that, when there is no uncertainty on any payoff-relevant information, the parties are no longer able to obfuscate their weaknesses and hence lose support. The opposite mechanism can explain the positive signs in the third column.

We now move to Table 8. As before, we outline a few guidelines on how to read Table 8. Differently from Table 7, each row has to be read together with the others because $\mathcal{I}_y$ is computed using the same empirical factual vote shares for each party $y \in \mathcal{Y}$. For example, Table 8 reveals that, when electors are fully informed, the Conservative Party loses votes with respect to the empirical factual case (second column). Here, such a negative effect should be counterbalanced by a positive effect in other rows of the second column. That is, we should expect some other parties to gain votes in the second column. Further, recall that $C_{n,0.50}$ has been constructed by selecting all the values of θ such that the collection of conditional choice probabilities predicted by the model under 1BCE contains the empirical conditional choice probabilities. Hence, for each $\theta \in C_{n,0.50}$ and $y \in \mathcal{Y}$, we expect $\bar{\Delta}^\theta_y \leq \Delta^\theta{}_y \leq \underline{\Delta}^\theta_y$. In turn, we expect the lower bound of $\mathcal{I}_y$ to be between the lower bound of $\bar{\mathcal{I}}_y$ and the lower bound of $\underline{\mathcal{I}}_y$, and the upper bound of $\mathcal{I}_y$ to be between the upper bound of $\bar{\mathcal{I}}_y$ and the upper bound of $\underline{\mathcal{I}}_y$. This is not always the case when comparing Tables 7 and 8, but it does not indicate a mistake in the procedure. In fact, we remind the reader that implementing our empirical strategy involves several approximations and finite-sample issues which may induce small violations to the above relations.

We now comment on the results in Table 8. Again, we find that, under complete information, abstention drops with respect to the empirical factual scenario, as the interval in the second column lies entirely on the negative real line. We also find that the "losers" from the policy intervention are the two biggest parties, i.e., the Conservative Party and the Labour Party. Conversely, the "winners" from the policy intervention are the other minor parties, i.e., the Liberal Democrats, UKIP, and the Green Party. This suggests that there exists some payoff-relevant information unobserved by voters, and the historically dominating parties in the British political scene benefit the most from such uncertainty.[29]

We conclude this section by emphasising that our second counterfactual exercise robustly quantifies the consequences of incomplete information in politics. Even if perfect information might not be an achievable scenario, it suggests that policy initiatives in that direction can

[29]One might wonder whether the observed drop in abstention might be a mechanical effect of the normalisation to zero of the payoff from not voting. This is not the case because the observed realisation of V_{iy} could add to or subtract from the component of the payoff "observed pre-signal", $\beta' Z_{iy} + \gamma'_y W_i + \epsilon_i$. Further, note that it could be that many/all voters in the population already observe the realisation of V_i, in which case we should expect no significant change in the abstention share, regardless of the normalisation adopted.

increase citizens' welfare by reducing ex-post regret.

5 Conclusions

In this paper we study identification of preferences in a class of single-agent, static, discrete choice models where DMs may be imperfectly informed about the payoffs generated by the available alternatives. We conclude our analysis by summarising two key findings. First, as previous studies have highlighted for discrete games, the notion of Bayes Correlated Equilibrium can also deliver informative bounds for the preference parameters in stochastic choice frameworks. These bounds are "robust", in the sense that they work under any possible Bayes-consistent beliefs (including perfect information). Second, since we allow for imperfect information, our methodology allows us to answer questions that cannot be addressed using conventional methods. In particular, we can quantify how choices respond to additional information and, in turn, assess the associated welfare benefits prior to conducting such intervention.

In some settings, the assumptions that DMs rely on a common family of priors and are Bayes rational can be deemed as being restrictive. However, note that these assumptions are still much weaker than the classic assumption of perfect information and hence may be preferable to conduct counterfactual experiments. Further, these assumptions are "default" restrictions in the theoretical literature on decision problems under imperfect information and, therefore, they are a good starting point for our econometric analysis. Relaxing them can be an interesting direction for future work, even though it poses the difficult problem of determining where rationality should end and how it should be replaced.

References

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

Agarwal, N., and P. Somaini (2020): "Revealed Preference Analysis of School Choice Models,", *Annual Review of Economics*, 12, 471–501.

Aldrich, J.H., and R.D. McKelvey (1977): "A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections," *American Political Science Review*, 71(1), 111–130.

Althaus, S.L. (1998): "Information Effects in Collective Preferences," *American Political Science Review*, 92(3), 545–558.

Alvarez, R.M. (1998): *Information and Elections*, University of Michigan Press, Ann Arbor, MI.

Alvarez, R.M., and J. Nagler (1995): "Economics, Issues and the Perot Candidacy: Voter Choice in the 1992 Presidential Election," *American Journal of Political Science*, 39(3), 714–744.

Alvarez, R.M., and J. Nagler (1998): "When Politics and Models Collide: Estimating Models of Multiparty Elections," *American Journal of Political Science*, 42(1), 55–96.

Alvarez, R.M., and J. Nagler (2000): "A New Approach for Modelling Strategic Voting in Multiparty Elections," *British Journal of Political Science*, 30(1), 57–75.

Alvarez, R.M., J. Nagler, and S. Bowler (2000): "Issues, Economics, and the Dynamics of Multiparty Elections: the British 1987 General Election," *American Political Science Review*, 94(1), 131–149.

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

Baron, D.P. (1994): "Electoral Competition with Informed and Uniformed Voters," *American Political Science Review*, 88(1), 33–47.

Barseghyan, A., F. Molinari, T. O'Donoghue, and J.C. Teitelbaum (2013): "The Nature of Risk Preferences: Evidence from Insurance Choices," *American Economic Review*, 103(6), 2499–2529.

Barseghyan, A., F. Molinari, and J.C. Teitelbaum (2016): "Inference under Stability of Risk Preferences," *Quantitative Economics*, 7(2), 367–409.

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

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

Bartels, L.M. (1986): "Issue Voting Under Uncertainty: An Empirical Test," *American Journal of Political Science*, 30(4), 709–728.

Bartels, L.M. (1996): "Uninformed Votes: Information Effects in Presidential Elections," *American Journal of Political Science*, 40(1), 194–230.

Berelson, B.R., P.F. Lazarsfeld, and W.N. McPhee (1954): *Voting: A Study of Opinion Formation in a Presidential Campaign*, Chicago: University of Chicago Press.

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

Bergemann, D., B. Brooks, and S. Morris (2019): "Counterfactuals with Latent Information," Cowles Foundation Discussion Paper 2162.

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.

Bergemann, D. and S. Morris (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57(1), 44–95.

Bettinger, E.P., B.T. Long, P. Oreopoulos, L. Sanbonmatsu (2012): "The Role of Application Assinstance and Information in College Decisions: Results from the H&R Block Fafsa Experiment," *Quarterly Journal of Economics*, 127(3), 1205–1242.

Black, A. (1998): *Information and Elections*, Michigan University Press, Ann Arbor, MI.

Blackwell, D., and M.A. Girshick (1954): *Theory of Games and Statistical Decisions*, John Wiley & Sons.

Blundell, R.W., and R.J. Smith (1986): "An Exogeneity Test for a Simultaneous Tobit Model," *Econometrica*, 54(3), 679–685.

Blundell, R.W., and R.J. Smith (1989): "Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models," *Review of Economic Studies*, 56(1) 37–58.

Bracquemond, C., and O. Gaudoin (2003): "A Survey on Discrete Life Time Distributions," *International Journal of Reliability, Quality and Safety Engineering*, 10(1) 69–98.

Brown, Z.Y., and J. Jeon (2020): "Endogenous Information and Simplifying Insurance Choice," Working Paper.

Campbell, J.E. (1983): "The Electoral Consequences of Issue Ambiguity: An Examination of the Presidential Candidates' Issue Positions from 1968 to 1980," *Political Behavior*, 5(3), 277-291.

Caplin, A., and M. Dean (2015): "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 105(7), 2183–2203.

Caplin, A., M. Dean, and J. Leahy (2019a): "Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy," Working Paper.

Caplin, A., M. Dean, and J. Leahy (2019b): "Rational Inattention, Optimal Consideration Sets and Stochastic Choice," *Review of Economic Studies*, 86(3), 1061–1094.

Caplin, A., and D. Martin (2015): "A Testable Theory of Imperfect Perception," *The Economic Journal*, 125(582), 184–202.

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.

Carpini Delli, M.X., and S. Keeter (1996): *What Americans Know about Politics and Why It Matters*, Yale University Press.

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

Chakraborty, S. (2015): "Generating Discrete Analogues of Continuous Probability Distributions - A Survey of Methods and Constructions," *Journal of Statistical Distributions and Applications*, 2(6).

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791–1828.

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., and M. Shum (2005): "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, 73(4), 1137–1173.

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

Dahlberg, S. (2009): “Political Parties and Perceptual Agreement: The Influence of Party Related Factors on Voters’ Perceptions in Proportional Electoral Systems,” *Electoral Studies*, 28(2), 270–278.

Dardanoni, V., P. Manzini, M. Mariotti, and C.J. Tyson (2020a): “Inferring Cognitive Heterogeneity From Aggregate Choices,” *Econometrica*, 88(3), 1269–1296.

Dardanoni, V., P. Manzini, M. Mariotti, H. Petri, and and C.J. Tyson (2020b): “Mixture Choice Data: Revealing Preferences and Cognition,” Working Paper.

Davis, O.A., M.J. Hinich, and P.C. Ordeshook (1970): “An Expository Development of a Mathematical Model of the Electoral Process,” *American Political Science Review*, 64(2), 426–448.

Degan, A., and A. Merlo (2009): “Do Voters Vote Ideologically?,” *Journal of Economic Theory*, 144(5), 1868–1894.

Degan, A., and A. Merlo (2011): “A Structural Model of Turnout and Voting in Multiple Elections,” *Journal of the European Economic Association*, 9(2), 209–245.

Dickstein, M.J., and E. Morales (2018): “What do Exporters Know?,” *Quarterly Journal of Economics*, 133(4), 1753–1801.

Downs, A. (1957): *An Economic Theory of Democracy*, Harper and Row, New York.

Enelow, J., and M.J. Hinich (1981): “A New Approach to Voter Uncertainty in the Downsian Spatial Model,” *American Journal of Political Science*, 25(3), 483–493.

Enelow, J., and M.J. Hinich (1984): *The Spatial Theory of Voting: An Introduction*, Cambridge University Press, New York, NY.

Feddersen, T., and W. Pesendorfer (1997): “Voting Behavior and Information Aggregation in Elections With Private Information,” *Econometrica*, 65(5), 1029–1058.

Feddersen J., and W. Pesendorfer (1999): “Abstention in Elections with Asymmetric Information and Diverse Preferences,” *American Political Science Review*, 93(2), 381–398.

Fieldhouse, E., Green, J. Evans G., Schmitt H., van der Eijk C., Mellon J., and Prosser C. (2019): *British Election Study, 2017: Face-to-Face Post-Election Survey*, available at https://www.britishelectionstudy.com/.

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

Franklin, C.H. (1991): "Eschewing Obfuscation? Campaigns and the Perception of U.S. Senate Incumbents," *American Political Science Review*, 85(4), 1193–1214.

Galichon, A. (2019): "On the Representation of the Nested Logit Model," arXiv:1907.08766.

Gibbons, R., and M. Waldman (1999): "A Theory of Wage and Promotion Dynamics Inside Firms," *Quarterly Journal of Economics*, 114(4), 1321–1358.

Gilens, M. (2001): "Political Ignorance and Collective Policy Preferences," *American Political Science Review*, 95(2), 379–396.

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

Grossman, G.M., and E. Helpman (1996): "Electoral Competition and Special Interest Politics," *Review of Economic Studies*, 63(2), 265–286.

Hastings, J.S., and L. Tejeda-Ashton (2008): "Financial Literacy, Information, and Demand Elasticity: Survey and Experimental Evidence from Mexico," NBER Working Paper 14538.

Hastings, J.S., and J.M. Weinstein (2008): "Information, School Choice, and Academic Achievement: Evidence from Two Experiments," *The Quarterly Journal of Economics*, 123(4), 1373–1414.

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

Heath, O., and M. Goodwin (2017): "The 2017 General Election, Brexit and the Return to Two- Party Politics: An Aggregate-level Analysis of the Result," *Political Quarterly*, 88(3), 345–358.

Hébert, B., and M. Woodford (2018): "Information Costs and Sequential Information Sampling," NBER Working Paper 25316.

Henry, M., and I. Mourifié (2013): "Euclidean Revealed Preferences: Testing the Spatial Voting Model," *Journal of Applied Econometrics*, 28(4), 650–666.

Hinish, M.J., and M.C. Munger (1994): *Ideology and the Theory of Political Choice*, University of Michigan Press, Ann Arbor, MI.

Hinich, M.J., and W. Pollard (1981): "A New Approach to the Spatial Theory of Electoral Competition," *American Journal of Political Science*, 25(2), 323–341.

Hobolt, S.B. (2018): "Brexit and the 2017 UK General Election," *Journal of Common Market Studies*, 56(S1), 39–50.

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

Hooghe, M., and B. Wilkenfeld (2007): "The Stability of Political Attitudes and Behaviors Across Adolescence and Early Adulthood: A Comparison of Survey Data on Adolescents and Young Adults in Eight Countries," *Journal of Youth and Adolescence*, 37(2), 155–167.

Hutton, R. (2017): "General Election 2017: What We've Learned So Far," Bloomberg, accessed *here*.

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.

Kling, Jeffrey R., S. Mullainathan, E. Shafir, L.C. Vermeulen, M.V. Wrobel (2012): "Comparison Friction: Experimental Evidence from Medicare Drug Plans," *Quarterly Journal of Economics*, 127(1), 199–235.

Knight, B., and N. Schiff (2010): "Momentum and Social Learning in Presidential Primaries," *Journal of Political Economy*, 118(6), 1110–1150.

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.

Lai, C.D. (2013): "Issues Concerning Constructions of Discrete Lifetime Models," *Quality Technology & Quantitative Management*, 10(2), 251–262.

Lin, Y.-H. (2019): "Stochastic Choice and Rational Inattention," Working Paper.

Lu, J. (2016): "Random Choice and Private Information," *Econometrica*, 84(6), 1983–2027.

Lupia, A., and M.D. McCubbins (1998): *The Democratic Dilemma. Can Citizens Learn What They Need to Know*, Cambridge University Press.

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

Manski, C.F. (2004): "Measuring Expectations," *Econometrica*, 72(5), 1329–1376.

Stochastic choice and consideration sets P Manzini, M Mariotti Econometrica 82 (3), 1153-1176

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.

Matějka, F., and G. Tabellini (2019): "Electoral Competition with Rationally Inattentive Voters," Working Paper.

Matsusaka, J.G. (1995): "Explaining Voter Turnout Patterns: An Information Theory," *Public Choice*, 84, 91–117.

McFadden, D.L. (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D.L. McFadden, MIT Press, Cambridge, 198–272.

McMurray, J.C. (2013): "Aggregating Information by Voting: The Wisdom of the Experts versus the Wisdom of the Masses," *Review of Economic Studies*, 80(1), 277–312.

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.

Mellon, J., G. Evans, E. Fieldhouse, J. Greeb, and C. Prosser (2018): "Brexit or Corbyn? Campaign and Inter-Election Vote Switching in the 2017 UK General Election," *Parliamentary Affairs*, 71(4), 719–737.

Merlo, A., and A. De Paula (2017): "Identification and Estimation of Preference Distributions When Voters Are Ideological," *Review of Economics Studies*, 84(3), 1238–1263.

Moretti, E. (2011): "Social Learning and Peer Effects in Consumption: Evidence from Movie Sales," *Review of Economic Studies*, 78(1), 356–393.

S. Morris (1995): "The Common Prior Assumption in Economic Theory," *Economics and Philosophy*, 11, 227–253.

Morris, S., and P. Strack (2019): "The Wald Problem and the Relation of Sequential Sampling and Ex-Ante Information Costs," SSRN Working Paper.

Niemi, R.G., and J. Junn (1998): *Civic Education: What Makes Students Learn*, New Haven, CT: Yale University Press.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2015): "Moment Inequalities and Their Application," *Econometrica*, 83(1), 315–334.

Palfrey, T.R., and K.T. Poole (1987): "The Relationship between Information, Ideology, and Voting Behavior," *American Journal of Political Science*, 31(3), 511–530.

Pontes, A.I., M. Henn, and M.D. Griffiths (2019): "Youth Political (Dis)engagement and the Need for Citizenship Education: Encouraging Young People's Civic and Political Participation through the Curriculum," *Education, Citizenship and Social Justice*, 14(1), 3–21.

Sekhon, J.S. (2004): "The Varying Role of Voter Information across Democratic Societies," Working Paper.

Shepsle, A.K. (1972): "The Strategy of Ambiguity: Uncertainty and Electoral Competition," *American Political Science Review*, 66(2), 555–568.

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

Sniderman, P.M., Brody R.A., and Tetlock, P. (1991): *Reasoning and Choice: Explorations in Political Psychology*, Cambridge, UK: Cambridge University Press.

Snowdon, K. (2017): "General Election 2017: 10 Forgotten Issues During The Campaign," HuffPost UK, accessed *here*.

Somer-Topcu (2015): "Everything to Everyone: The Electoral Consequences of the Broad-Appeal Strategy in Europe," *American Journal of Political Science*, 59(4), 841–854.

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.

Tiemann, G. (2019): "The Shape of Utility Functions and Voter Attitudes Towards Risk," *Electoral Studies*, 61, 102051.

Tomz, M., and R. P. van van Houweling (2009): "The Electoral Implications of Candidate Ambiguity," *American Political Science Review*, 103(1), 83–98.

Ursu, R. M. (2018): "The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions," *Marketing Science*, 37(4), 530–552.

Weisberg, H.F., and M.P. Fiorina (1980): "Candidate Preference under Uncertainty: An Expanded View of Rational Voting," in John C. Pierce and John L. Sullivan, eds., *The Electorate Reconsidered*, Beverly Hills: Sage.

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.

Yuksel, S. (2020): "Specialized Learning and Political Polarization," Working Paper.

A Some remarks on Definition 1

We add some remarks on Definition 1. First, note that we can equivalently define an optimal strategy of the augmented choice problem $\{G, S_i\}$ as follows. Given $(x, e, t) \in \mathcal{X} \times \mathcal{E} \times \mathcal{T}_i$, let $\mathcal{Y}^i_{x,e,t} \subseteq \mathcal{Y}$ be the set of alternatives maximising DM i's expected payoff, i.e.,

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

Let $\mathcal{P}^i_{x,e,t}$ be the family of probability mass functions of Y_i conditional on $(X_i, \epsilon_i, t_i) = (x, e, t)$ that are degenerate on each of element of $\mathcal{Y}^i_{x,e,t}$. Let $\text{Conv}(\mathcal{P}^i_{x,e,t})$ be the convex hull of $\mathcal{P}^i_{x,e,t}$. Then, $\mathcal{P}^i_{Y|X,\epsilon,T}$ is an optimal strategy of the augmented choice problem $\{G, S_i\}$ if $P^i_{Y|X,\epsilon,T}(\cdot|x, e, t) \in \text{Conv}(\mathcal{P}^i_{x,e,t})$ $\forall (x, e, t) \in \mathcal{X} \times \mathcal{E} \times \mathcal{T}_i$.

Second, note that Definition 1 allows to formally defines DM i's consideration set. In fact, following Caplin, Dean, and Leahy (2019b), DM i's consideration set, $\mathcal{C}_i$, arises endogenously from her optimal strategy, $\mathcal{P}^i_{Y|X,\epsilon,T}$. In particular, $\mathcal{C}_i$ collects every alternative such that the subset of the signal's support inducing DM i to choose that alternative has positive measure. For example, when $\mathcal{T}_i$ and $\mathcal{V}$ are finite,

$$\mathcal{C}_i \equiv \{y \in \mathcal{Y} : \sum_{t \in \mathcal{T}_i} P^i_{Y|X,\epsilon,T}(y|x_i, e_i, t) \sum_{v \in \mathcal{V}} P^i_{T|X,\epsilon,V}(t|x_i, e_i, v) P_{V|X,\epsilon}(v|x_i, e_i) > 0\},$$

where (x_i, e_i) are the realisations of (X_i, ϵ_i) assigned by nature to DM i. Crucially, note that considerations sets can be heterogenous across agents and arbitrarily dependent on (X_i, ϵ_i) as we leave the conditional signal densities fully unrestricted.

B Inference

Identification of the true parameter vector, θ^0, relies on the assumption that the true probability mass function 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, $\{Y_i, X_i\}_{i=1}^n$, and take into account the sampling variation. Given $\alpha \in (0, 1)$, this section illustrates how to construct a uniformly asymptotically valid $(1 - \alpha)$% confidence region, $C_{n,1-\alpha}$, for any $\theta \in \Theta^*$. In particular, we suggest to reformulate our problem using conditional moment inequalities and 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).[30]

$C_{n,1-\alpha}$ is obtained by running a test with null hypothesis $\text{H}_0 : \theta^0 = \theta$, for every $\theta \in \Theta$, and then collecting all the values of θ which are not rejected. For a given θ, the test rejects H_0 if

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

$\mathrm{TS}_n(\theta) > \hat{c}_{n,1-\alpha}(\theta)$, where $\mathrm{TS}_n(\theta)$ is a test statistic and $\hat{c}_{n,1-\alpha}(\theta)$ is a corresponding critical value. Thus,

$$C_{n,1-\alpha} \equiv \{\theta \in \Theta\colon\ \mathrm{TS}_n(\theta) \leq \hat{c}_{n,1-\alpha}(\theta)\}. \tag{B.1}$$

The remainder of the section explains how to compute $\mathrm{TS}_n(\theta)$ and $\hat{c}_{n,1-\alpha}(\theta)$ for any given $\theta \in \Theta$.

In order to define the test statistic, $\mathrm{TS}_n(\theta)$, let us first rewrite the linear programming (8) as a collection of conditional moment inequalities. To do so, we label the elements of $\mathcal{Y}$ as $y^1, ..., y^{|\mathcal{Y}|-1}, y^{|\mathcal{Y}|}$. Also, recall from the notation paragraph in Section 1 that $\mathbb{B}^{|\mathcal{Y}|-1}$ is the unit ball in $\mathbb{R}^{|\mathcal{Y}|-1}$.

Proposition B.1. (*Conditional moment inequalities)* Under Assumption 3, for each $\theta \in \Theta$, $\theta \in \Theta^*$ if and only if

$$\mathbb{E}[m(Y_i, X_i; b, \theta)|X_i = x] \leq 0\ \forall b \in \mathbb{B}^{|\mathcal{Y}|-1}, \forall x \in \mathcal{X},$$

where

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

◇

Proposition B.1 comes from the fact that, following BMM, one can express the condition $P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}$ as

$$b^T P^0_{Y|X}(\cdot|x) - \sup_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T P_{Y|X}(\cdot|x) \leq 0\ \forall b \in \mathbb{R}^{|\mathcal{Y}|}, \tag{B.2}$$

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 $\bar{\mathcal{Q}}^\theta_{Y|x}$. By exploiting the positive homogeneity of the support function and some algebraic manipulations, (B.2) is equal to the collection of conditional moment inequalities stated in Proposition B.1.

Second, we rewrite the conditional moment inequalities in Proposition B.1 as unconditional moment inequalities. Here, we use Lemma 2 in AS which shows that conditional moment inequalities can be transformed into 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_i. Thus,

$$\theta \in \Theta^* \Leftrightarrow \mathbb{E}[m(Y_i, X_i; b, \theta, h)] \leq 0\ \forall b \in \mathbb{B}^{|\mathcal{Y}|-1}, \forall h \in \mathcal{H} \text{ a.s.}, \tag{B.3}$$

where

$$m(Y_i, X_i; b, \theta, h) \equiv m(Y_i, X_i; b, \theta) \times h(X_i).$$

Third, observe that $\mathbb{E}[m(Y_i, X_i; b, \theta, h)]$ evaluated at $b \equiv 0_{|\mathcal{Y}|-1}$ is 0. Therefore, (B.3) is equivalent to

$$\theta \in \Theta^* \Leftrightarrow \max_{b \in \mathbb{B}^{|\mathcal{Y}|-1}} \mathbb{E}[m(Y_i, X_i; b, \theta, h)] = 0 \ \forall h \in \mathcal{H} \text{ a.s.}$$

In light of the three steps above, following Appendix B.1 of BMM, we can use as test statistic

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

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

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

Intuitively, $\text{TS}_n(\theta)$ is built by imposing a penalty for each h such that the maximum of $\mathbb{E}[m(Y_i, X_i; b, \theta, h)]$ across $b \in \mathbb{B}^{|\mathcal{Y}|-1}$ is different from zero. Moreover, given that the support of X_i is finite, the analyst can replace Γ with the uniform probability measure on $\mathcal{X}$ as suggested by Example 5 in Appendix B of AS. That is,

$$\text{TS}_n(\theta) \equiv \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \left[\sqrt{n} \max_{b \in \mathbb{B}^{|\mathcal{Y}|-1}} \bar{m}_n(b, \theta, x)\right]^2, \tag{B.4}$$

where

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

Lastly, we compute the critical value, $\hat{c}_{n,1-\alpha}(\theta)$, by following AS's bootstrap method consisting of the following steps. First, we draw W_n bootstrap samples using nonparametric i.i.d. bootstrap. Second, for each $w = 1, ..., W_n$, we compute the recentered test statistic

$$\text{TS}_n^w(\theta) \equiv \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \left[\sqrt{n} \max_{b \in \mathbb{B}^{|\mathcal{Y}|-1}} (\bar{m}_n^w(b, \theta, x) - \bar{m}_n(b, \theta, x))\right]^2, \tag{B.5}$$

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. Third, $\hat{c}_{n,1-\alpha}(\theta)$ is set equal to the $(1-\alpha)$ quantile of $\{\text{TS}_n^w(\theta)\}_{w=1}^{W_n}$. Once $\text{TS}_n(\theta)$ and $\hat{c}_{n,1-\alpha}(\theta)$ are computed for each $\theta \in \Theta$ (or, in practice, for each θ belonging to a grid), the confidence region, $C_{n,1-\alpha}$, defined in (B.1) can be constructed.

In Appendix B.1 we provide more details on the computation of (B.4) and (B.5). In particular, we show that computing (B.4) and (B.5) amounts to solving some quadratically constrained linear programming problems.

We conclude by highlighting that, in addition to reporting a confidence region, it is often useful to report an estimated set, so as to reveal how much of the volume of the confidence region is due to randomness and how much is due to a large identified set. In this respect, AS show that $C_{n,0.50}$ is an asymptotically half-median-unbiased estimated set.

B.1 Some computational simplifications

We first discuss a way to simplify the computation of the test statistic, $\text{TS}_n(\theta)$, as defined in (B.4). Observe that

$$\bar{m}_n(b,\theta,x) = P_X^0(x)b^T\Big(\tilde{P}^0_{Y|X}(\cdot|x) - \max_{P_{Y|X}(\cdot|x)\in\bar{\mathcal{Q}}^\theta_{Y|x}} \tilde{P}_{Y|X}(\cdot|x)\Big), \tag{B.6}$$

where $\tilde{P}^0_{Y|X}(\cdot|x) \equiv \begin{pmatrix} P^0_{Y|X}(y^1|x) \\ \vdots \\ P^0_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix}$ and $\tilde{P}_{Y|X}(\cdot|x) \equiv \begin{pmatrix} P_{Y|X}(y^1|x) \\ \vdots \\ P_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix}$, for each $x \in \mathcal{X}$ and $b \in \mathbb{B}^{|\mathcal{Y}|-1}$.

Therefore, (B.4) is equal to

$$\text{TS}_n(\theta) \equiv \frac{1}{|\mathcal{X}|}\sum_{x\in\mathcal{X}}\Big[\sqrt{n}\max_{b\in\mathbb{B}^{|\mathcal{Y}|-1}} b^T\Big(P_X^0(x)\tilde{P}^0_{Y|X}(\cdot|x) - P_X^0(x)\max_{P_{Y|X}(\cdot|x)\in\bar{\mathcal{Q}}^\theta_{Y|x}}\tilde{P}_{Y|X}(\cdot|x)\Big)\Big]^2. \tag{B.7}$$

To compute (B.7), the researcher should calculate, for each $x \in \mathcal{X}$,

$$\max_{b\in\mathbb{B}^{|\mathcal{Y}|-1}} b^T\Big(P_X^0(x)\tilde{P}^0_{Y|X}(\cdot|x) - P_X^0(x)\max_{P_{Y|X}(\cdot|x)\in\bar{\mathcal{Q}}^\theta_{Y|x}}\tilde{P}_{Y|X}(\cdot|x)\Big),$$

which is equivalent to

$$\max_{b\in\mathbb{B}^{|\mathcal{Y}|-1}}\ \min_{P_{Y|X}(\cdot|x)\in\bar{\mathcal{Q}}^\theta_{Y|x}} b^T\Big(P_X^0(x)\tilde{P}^0_{Y|X}(\cdot|x) - P_X^0(x)\tilde{P}_{Y|X}(\cdot|x)\Big). \tag{B.8}$$

(B.8) is a max-min problem which can be simplified as follows. Note that the inner constrained minimisation problem in (B.8) 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 (B.8) has a quadratic constraint, $b^Tb \leq 1$. Therefore, (B.8) can be rewritten as a quadratically constrained linear maximisation problem which is a tractable exercise. This is described in detail below.

By Definition 2, (B.8) is equivalent to

$$
\begin{aligned}
\max_{b\in\mathbb{R}^{|\mathcal{Y}|-1}} \quad & \min_{\substack{P_{Y|X}(\cdot|x)\,\in\,\mathbb{R}_+^{|\mathcal{Y}|}\\ P_{Y,V|X,\epsilon}(\cdot|x,e)\,\in\,\mathbb{R}_+^{|\mathcal{Y}|\cdot|\mathcal{V}|},\forall e\in\mathcal{E}}} \quad b^T[P^0_X(x)\tilde{P}^0_{Y|X}(\cdot|x) - P^0_X(x)\tilde{P}_{Y|X}(\cdot|x)],\\
\text{s.t.} \quad & [b\in\mathbb{B}^{|\mathcal{Y}|-1}]: \quad b^Tb\leq 1,\\
& [\text{1BCE-Consistency}]: \quad \sum_{y\in\mathcal{Y}} P_{Y,V|X,\epsilon}(y,v|x,e) = P_{V|X,\epsilon}(v|x,e;\theta_V)\ \forall v\in\mathcal{V},\forall e\in\mathcal{E},\\
& [\text{1BCE-Obedience}]: \quad -\sum_{v\in\mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e)[u(y,x,e,v;\theta_u)-u(y',x,e,v;\theta_u)]\leq 0\ \forall y\in\mathcal{Y},\forall y'\in\mathcal{Y}\setminus\{y\},\forall e\in\mathcal{E},\\
& [\text{1BCE-Choice prediction}]: \quad P_{Y|X}(y|x) = \sum_{(e,v)\in\mathcal{E}\times\mathcal{V}} P_{Y,V|X,\epsilon}(y,v|x,e)P_{\epsilon|X}(e|x;\theta_\epsilon)\ \forall y\in\mathcal{Y}.
\end{aligned}
\tag{B.9}
$$

We simplify (B.9) by introducing new variables. Let $W_1 \equiv P^0_X(x)(P^0_{Y|X}(\cdot|x) - P_{Y|X}(\cdot|x))$. Note that W_1 is a $|\mathcal{Y}|\times 1$ vector. Further, let W_2 be the $(|\mathcal{Y}|\cdot|\mathcal{V}|\cdot|\mathcal{E}|)\times 1$ vector collecting $P_{Y,V|X,\epsilon}(\cdot|x,e)$ across every $e\in\mathcal{E}$. Lastly, let W be the $(|\mathcal{Y}|+|\mathcal{Y}|\cdot|\mathcal{V}|\cdot|\mathcal{E}|)\times 1$ vector collecting W_1 and W_2. (B.9) can be rewritten as

$$
\begin{aligned}
\max_{b\in\mathbb{R}^{|\mathcal{Y}|-1}} \ & \min_{\substack{W_1\,\in\,\mathbb{R}^{|\mathcal{Y}|}\\ W_2\,\in\,\mathbb{R}_+^{|\mathcal{Y}|\cdot|\mathcal{V}|\cdot|\mathcal{E}|}}} \begin{bmatrix} b^T & 0^T_{1+|\mathcal{Y}|\cdot|\mathcal{V}|\cdot|\mathcal{E}|}\end{bmatrix} W,\\
\text{s.t.}\ & b^Tb\leq 1,\\
& A_{\text{eq}}\, W = B_{\text{eq}},\\
& A_{\text{ineq}}\, W \leq 0_{d_{\text{ineq}}},
\end{aligned}
\tag{B.10}
$$

where A_{eq} is the matrix of coefficients multiplying W in the equality constraints of (B.9) with d_{eq} rows, B_{eq} is the vector of constants appearing in the equality constraints of (B.9), and A_{ineq} is the matrix of coefficients multiplying W in the inequality constraints of (B.9) with d_{ineq} rows.

Further, the inner constrained minimisation problem in (B.10) is linear. Hence, by strong duality, it can be replaced with its dual. This allows us to solve one unique maximisation problem. Precisely, the solution of (B.10) is equivalent to the solution of

$$
\begin{aligned}
\max_{\substack{b\,\in\,\mathbb{R}^{|\mathcal{Y}|-1}\\ \tau_{\text{eq}}\,\in\,\mathbb{R}^{d_{\text{eq}}}\\ \tau_{\text{ineq}}\,\in\,\mathbb{R}_+^{d_{\text{ineq}}}}} \ & \begin{bmatrix} -B^T_{\text{eq}} & 0^T_{d_{\text{ineq}}}\end{bmatrix}\tau,\\
\text{s.t.}\ & b^Tb\leq 1,\\
& [A^T]_{1:|\mathcal{Y}|}\tau = \begin{pmatrix} -b\\ 0\end{pmatrix},\\
& -[A^T]_{|\mathcal{Y}|+1:|\mathcal{Y}|+|\mathcal{Y}|\cdot|\mathcal{V}|\cdot|\mathcal{E}|}\tau \leq 0_{|\mathcal{Y}|\cdot|\mathcal{V}|\cdot|\mathcal{E}|},
\end{aligned}
\tag{B.11}
$$

where τ is the $(d_{\text{eq}}+d_{\text{ineq}})\times 1$ vector collecting τ_{eq} and τ_{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.

Note that (B.11) is a quadratically constrained linear maximisation problem. In particular, the first constraint in (B.11) is quadratic. The objective function and the remaining constraints in (B.11) are linear. Close derivations are discussed in Magnolfi and Roncoroni (2017) for an entry game setting.

We now discuss a way to simplify the computation of bootstrap test statistic, $\text{TS}_n^w(\theta)$, as defined in (B.5). Similarly to (B.6), by rearranging terms it holds that

$$\bar{m}_n^w(b,\theta,x) = P_X^{0,w}(x) b^T \Big(\tilde{P}_{Y|X}^{0,w}(\cdot|x) - \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}_{Y|x}^{\theta}} \tilde{P}_{Y|X}(\cdot|x) \Big),$$

where the superscript "w" distinguishes the probabilities within the bootstrap sample from the original ones. Therefore,

$$\begin{aligned}
&\bar{m}_n^w(b,\theta,x) - \bar{m}_n(b,\theta,x) \\
&= P_X^{0,w}(x) b^T \Big(\tilde{P}_{Y|X}^{0,w}(\cdot|x) - \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}_{Y|x}^{\theta}} \tilde{P}_{Y|X}(\cdot|x) \Big) - P_X^{0}(x) b^T \Big(\tilde{P}_{Y|X}^{0}(\cdot|x) - \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}_{Y|x}^{\theta}} \tilde{P}_{Y|X}(\cdot|x) \Big), \\
&= b^T \Big[P_X^{0,w}(x) \tilde{P}_{Y|X}^{0,w}(\cdot|x) - P_X^{0}(x) \tilde{P}_{Y|X}^{0}(\cdot|x) - (P_X^{0,w}(x) - P_X^{0}(x)) \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}_{Y|x}^{\theta}} \tilde{P}_{Y|X}(\cdot|x) \Big].
\end{aligned}$$

To simplify the notation, let us rename

$$A_x^w \equiv P_X^{0,w}(x) \tilde{P}_{Y|X}^{0,w}(\cdot|x) - P_X^{0}(x) \tilde{P}_{Y|X}^{0}(\cdot|x),$$

and

$$C_x^w \equiv P_X^{0,w}(x) - P_X^{0}(x).$$

Therefore, (B.5) is equal to

$$\text{TS}_n^w(\theta) \equiv \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \Big[\sqrt{n} \max_{b \in \mathbb{B}^{|\mathcal{Y}|-1}} b^T \Big(A_x^w - C_x^w \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}_{Y|x}^{\theta}} \tilde{P}_{Y|X}(\cdot|x) \Big) \Big]^2. \tag{B.12}$$

To compute (B.12), the researcher should calculate, for each $x \in \mathcal{X}$,

$$\max_{b \in \mathbb{B}^{|\mathcal{Y}|-1}} b^T \Big(A_x^w - C_x^w \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}_{Y|x}^{\theta}} \tilde{P}_{Y|X}(\cdot|x) \Big),$$

which is equivalent to

$$\max_{b \in \mathbb{B}^{|\mathcal{Y}|-1}} \min_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}_{Y|x}^{\theta}} b^T \Big(A_x^w - C_x^w \tilde{P}_{Y|X}(\cdot|x) \Big). \tag{B.13}$$

(B.13) can be rewritten as a quadratically constrained linear maximisation problem as done for (B.8). Once (B.13) is computed for each $x \in \mathcal{X}$, the analyst easily obtains $\text{TS}_n^w(\theta)$.

C Proofs

Lemma C.1. (*Existence of an optimal strategy of the augmented choice problem* $\{G, S_i\}$) The augmented choice problem $\{G, S_i\}$ admits an optimal strategy, $\mathcal{P}^i_{Y|X,\epsilon,T}$, $\forall S_i \in \mathcal{S}$. $\diamond$

Proof of Lemma C.1 We proceed by construction. Take any $S_i \equiv \{\mathcal{T}_i, \mathcal{P}^i_{T|X,\epsilon,V}\} \in \mathcal{S}$. 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 $(x, e, v) \in \mathcal{X} \times \mathcal{E} \times \mathcal{V}$. Hence, the map $y \mapsto \int_{v\in\mathcal{V}} u(y, x, e, v) P^i_{V|X,\epsilon,T}(v|x, e, t)dv$ is also continuous for each $x \in \mathcal{X}$, $e \in \mathcal{E}$, and $t \in \mathcal{T}_i$. Therefore, Weierstrass theorem ensures the existence of the minimum and maximum of such a map. Given $(x, e, t) \in \mathcal{X} \times \mathcal{E} \times \mathcal{T}_i$, let $y^i_{x,e,t} \in \mathcal{Y}$ be one of the maximisers. Then, an optimal strategy is $\mathcal{P}^i_{Y|X,\epsilon,T}$ such that for each $(x, e, t) \in \mathcal{X} \times \mathcal{E} \times \mathcal{T}_i$,

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

Proof of Proposition 1 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 $\mathcal{P}_{Y,V|X,\epsilon} \in \mathcal{Q}^\theta$ inducing $P_{Y|X}(\cdot|x)$. By Theorem 1, it follows that there exists $S \equiv \{\mathcal{T}, \mathcal{P}_{T|X,\epsilon,V}\} \in \mathcal{S}$ 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_i conditional $X_i = x$ that are induced by the model's optimal strategies under θ, while remaining agnostic about information structures. That is,

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

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 $S \equiv \{\mathcal{T}, \mathcal{P}_{T|X,\epsilon,V}\} \in \mathcal{S}$ 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 $\bar{\mathcal{Q}}^\theta_{Y|x}$.

We 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 B.1 Fix any $\theta \in \Theta$ and $x \in \mathcal{X}$. Observe that

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

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

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

By (C.2), (C.1) is equivalent to

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

Moreover, given that $\bar{\mathcal{Q}}^\theta_{Y|x}$ is closed and bounded, (C.3) is equivalent to

$$P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x} \;\Leftrightarrow\; b^T P^0_{Y|X}(\cdot|x) - \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T P_{Y|X}(\cdot|x) \leq 0 \;\forall b \in \mathbb{B}^{|\mathcal{Y}|}. \tag{C.4}$$

Lastly, given that $\bar{\mathcal{Q}}^\theta_{Y|x}$ is a subset of the $(|\mathcal{Y}|-1)$-dimensional simplex, (C.4) is equivalent to

$$P^0_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x} \;\Leftrightarrow\; b^T \begin{pmatrix} P^0_{Y|X}(y^1|x) \\ \vdots \\ P^0_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix} - \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T \begin{pmatrix} P_{Y|X}(y^1|x) \\ \vdots \\ P_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix} \leq 0 \;\forall b \in \mathbb{B}^{|\mathcal{Y}|-1}. \tag{C.5}$$

Therefore, by combining Proposition 1 with (C.5), we get that

$$\theta \in \Theta^* \Leftrightarrow b^T \begin{pmatrix} P^0_{Y|X}(y^1|x) \\ \vdots \\ P^0_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix} - \max_{P_{Y|X}(\cdot|x) \in \bar{\mathcal{Q}}^\theta_{Y|x}} b^T \begin{pmatrix} P_{Y|X}(y^1|x) \\ \vdots \\ P_{Y|X}(y^{|\mathcal{Y}|-1}|x) \end{pmatrix} \leq 0 \;\forall b \in \mathbb{B}^{|\mathcal{Y}|-1}, \tag{C.6}$$

which is equivalent to

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

as claimed in Proposition B.1, where

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

D A case where grid search can be avoided

Suppose that the sets $\mathcal{X}$, $\mathcal{E}$, and $\mathcal{V}$ are finite, so that one can focus on identifying the finite-dimensional vector of parameters, θ^0, which collects the image values of the true primitives $u^0, \mathcal{P}^0_{V|X,\epsilon}, \mathcal{P}^0_{\epsilon|X}$, without the need to impose additional parameterisations. Further, suppose that the image values of the function u^0 are known by the researcher. Lastly, instead of separately identifying the conditional marginal distributions, $\mathcal{P}^0_{V|X,\epsilon}$ and $\mathcal{P}^0_{\epsilon|X}$, note that it is sufficient for our purposes to back out the joint distribution, $\mathcal{P}^0_{V,\epsilon|X} \equiv \{P^0_{V,\epsilon|X}(\cdot|x)\}_{\forall x \in \mathcal{X}}$, where $P^0_{V,\epsilon|X}(v, e|x) \equiv P^0_{V|X,\epsilon}(v|x, e)P^0_{\epsilon|X}(e|x)$ for every $(x, e, v) \in \mathcal{X} \times \mathcal{E} \times \mathcal{V}$.

Let θ^0 collect the image values of $\mathcal{P}^0_{V,\epsilon|X}$, with dimension K. Then, one can construct the identified set for θ^0 without performing a grid search.

To see why, note that, given a generic θ, finding if (8) admits a solution with respect to $\mathcal{P}_{Y,V|X,\epsilon}$ is equivalent to finding if the following linear programming problem admits a solution with respect to $\mathcal{P}_{Y,V,\epsilon|X} \equiv \{P_{Y,V,\epsilon|X}(\cdot|x)\}_{\forall x\in\mathcal{X}}$, where each $P_{Y,V,\epsilon|X}(\cdot|x) \in \Delta(\mathcal{Y} \times \mathcal{V} \times \mathcal{E})$ is a probability mass function of (Y_i, V_i, ϵ_i) conditional on $X_i = x_i$:

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

Now, recall that $u(y, x, e, v) - u(y', x, e, v)$ entering the obedience constraint is known by the researcher for every $y \in \mathcal{Y}$, $y' \in \mathcal{Y} \setminus \{y\}$, $x \in \mathcal{X}$, $e \in \mathcal{E}$, and $v \in \mathcal{V}$. Therefore, (D.1) is linear with respect to θ. Hence, one can find the feasible region of (D.1) with respect to $(\theta, \mathcal{P}_{Y,V,\epsilon|X})$ by solving a unique linear programming problem and then take its projection for θ.

E Empirical application: implementation details

We describe some steps implemented to obtain the results of Section 4.3.

Computation of $\hat{\theta}_{\text{com}}$ and $\hat{\theta}_{\text{deg}}$ Note that, under the assumption that all agents process the complete information structure, our framework resembles a classical Multinomial Probit model, where θ^0 is point identified and can be estimated by running a maximum likelihood procedure. Hence, we obtain $\hat{\theta}_{\text{com}}$ as

$$\begin{aligned}\hat{\theta}_{\text{com}} = \text{argmin}_{\theta} - \frac{1}{n}\sum_{i=1}^{n}\Big[\sum_{y\in\mathcal{Y}\setminus\{\emptyset\}} \\ \mathbb{1}\{y_i = y\} \times \log Pr(\beta' Z_{iy} + \gamma' Z_y^{\text{LR}} W_i + \epsilon_i + \sigma V_{iy} \geq \beta' Z_{ik} + \gamma' Z_k^{\text{LR}} W_i + \epsilon_i + \sigma V_{ik}\ \forall k \in \mathcal{Y}\setminus\{\emptyset, y\}, \\ \beta' Z_{iy} + \gamma' Z_y^{\text{LR}} W_i + \epsilon_i + \sigma V_{iy} \geq 0 | X_i = x_i) \\ + \mathbb{1}\{y_i = \emptyset\} \times \log Pr(0 \geq \beta' Z_{iy} + \gamma' Z_y^{\text{LR}} W_i + \epsilon_i + \sigma V_{iy}\ \forall y \in \mathcal{Y}\setminus\{\emptyset\} | X_i = x_i)\Big],\end{aligned} \tag{E.1}$$

where the minimisation is done using the Matlab solver FMINUNC, n is the sample size, $\emptyset$ represents the baseline category (abstention), and the integrals inside the log function are computed based on the fact that (ϵ_i, V_i) are jointly distributed as an L-variate standard normal, independent of X_i. Similarly, we obtain $\hat{\theta}_{\text{deg}}$ as

$$\begin{aligned}\hat{\theta}_{\text{deg}} \in \text{argmin}_{\theta} - \frac{1}{n}\sum_{i=1}^{n}\Big[\sum_{y\in\mathcal{Y}\setminus\{\emptyset\}} \\ \mathbb{1}\{y_i = y\} \times \log Pr(\beta' Z_{iy} + \gamma' Z_y^{\text{LR}} W_i + \epsilon_i \geq \beta' Z_{ik} + \gamma' Z_k^{\text{LR}} W_i + \epsilon_i\ \forall k \in \mathcal{Y}\setminus\{\emptyset, y\}, \\ \beta' Z_{iy} + \gamma' Z_y^{\text{LR}} W_i + \epsilon_i \geq 0 | X_i = x_i) \\ + \mathbb{1}\{y_i = \emptyset\} \times \log Pr(0 \geq \beta' Z_{iy} + \gamma' Z_y^{\text{LR}} W_i + \epsilon_i\ \forall y \in \mathcal{Y}\setminus\{\emptyset\} | X_i = x_i)\Big],\end{aligned} \tag{E.2}$$

where $\beta' Z_{iy} + \gamma' Z_y^{\text{LR}} W_i + \epsilon_i$ is the expected payoff from choosing $y \in \mathcal{Y}\setminus\{\emptyset\}$ under the prior and the integrals inside the log function are computed using that ϵ_i is distributed as a standard normal, independent of X_i. Note that in the latter case σ is not identified because, according to the priors of voters, V_i has expected value 0 and hence has no impact on the expected payoffs. Further, we use the symbol "$\in$" in (E.2) because the argmin may not be unique. In our implementation, we have run the solver FMINUNC from several starting values and obtained a unique argmin.

Construction of the grid Section B explains how to construct a confidence region by inverting a test with null hypothesis $\text{H}_0 : \theta^0 = \theta$, for every $\theta \in \Theta$. In practice, we do that by designing a grid of values for θ and inverting the test for each θ in the grid.

Recall that the test statistic, $\text{TS}_n(\theta)$, is expected to be lower for values of θ in the identified set because the moment inequalities should be approximatively satisfied. Hence, we design our grid by exploring the parameter space around the global infimum of $\text{TS}_n(\theta)$, for example, as in Ciliberto and Tamer (2009). Specifically:

1. We consider the maximum likelihood estimates $\hat{\theta}_{\text{com}}$ and $\hat{\theta}_{\text{deg}}$, obtained as discussed above. Recall that $\hat{\theta}_{\text{com}}$ and $\hat{\theta}_{\text{deg}}$ are $1 \times K$ vectors and $K = 8$ in our empirical application.

2. We construct an Halton set of 10^6 points around $\hat{\theta}_{\text{com}}$. We draw 100 points at random

from such a set. We construct an Halton set of 10^6 points around $\hat{\theta}_{\text{deg}}$. We draw 100 points at random from such a set. We stack these points, together with $\hat{\theta}_{\text{com}}$ and $\hat{\theta}_{\text{deg}}$, into an $201 \times K$ matrix, A.

3. We minimise $\text{TS}_n(\theta)$ with respect to θ by running the simulated annealing algorithm from each row of A as starting point with various initial temperatures, for a maximum of 10^4 iterations. We save every parameter value encountered (say, R values) in the course of the algorithm. We stack all the saved parameter values in an $R \times K$ matrix, G. G constitutes our final grid.

Discretisation of the supports of ϵ_i and V_i In order to compute the test statistic and critical value for a given θ as outlined in Section B, one has to discretise the supports of ϵ_i and V_i. We discretise the support of ϵ_i by collecting in $\mathcal{E}^{\text{discr}}$ q_ϵ equally spaced quantiles of the univariate standard normal CDF between 0.001 and 0.999. We discretise the support of V_i by first collecting in $\mathcal{V}_y^{\text{discr}}$ q_V equally spaced quantiles of the univariate standard normal CDF between 0.001 and 0.999, and then taking the Cartesian product $\mathcal{V}^{\text{discr}} \equiv \times_{y=1}^{L-1} \mathcal{V}_y^{\text{discr}}$.[31]

Parallelisation In order to compute the test statistics, recall that the quadratically constrained linear maximisation problem (B.11) has to be solved for each $x \in \mathcal{X}$ and for each θ in our grid. We parallelise the computation across x by using *parfor* in Matlab. We parallelise the computation across θ by running parallel array jobs in an HPC cluster. In order to compute the critical values, recall that the quadratically constrained linear maximisation problem (B.13) has to be solved for each $x \in \mathcal{X}$, for each bootstrap sample, and for each θ in our grid. We parallelise the computation across x by using *parfor*. We parallelise the computation across bootstrap samples and θ by running parallel array jobs in an HPC cluster.

[31]Recall that we normalise the payoff of the baseline category (abstention) to zero.

Sampling variability In Figure E.1, we report the projections of $\hat{\theta}_{\text{com}}$ (red), $\hat{\theta}_{\text{deg}}$ (blue), $C_{n,0.50}$ (black), and $C_{n,0.95}$ (gray) along some relevant pairs of dimensions. The projections of $C_{n,0.95}$ are quite wide, perhaps due to the limited sample size.

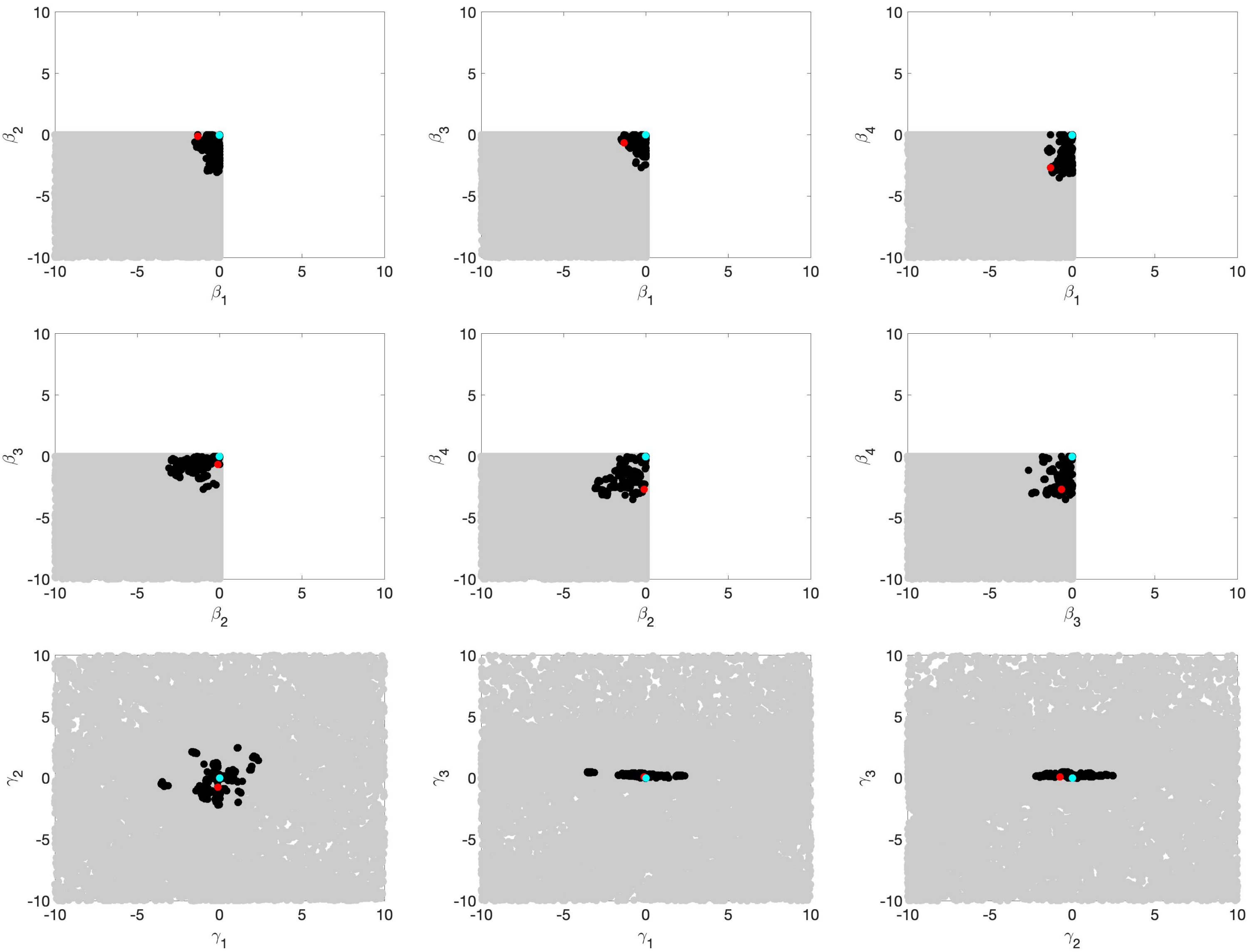

Figure E.1: The pictures report the projections of $\hat{\theta}_{\text{com}}$ (red), $\hat{\theta}_{\text{deg}}$ (blue), $C_{n,0.50}$ (black), and $C_{n,0.95}$ (gray) along some relevant pairs of dimensions.