Robust Likelihood Ratio Tests for Incomplete Economic Models*

Hiroaki Kaido† Yi Zhang‡

December 2019

Abstract

This study develops a framework for testing hypotheses on structural parameters in incomplete models. Such models make set-valued predictions and hence do not generally yield a unique likelihood function. The model structure, however, allows us to construct tests based on the least favorable pairs of likelihoods using the theory of Huber and Strassen (1973). We develop tests robust to model incompleteness that possess certain optimality properties. We also show that sharp identifying restrictions play a role in constructing such tests in a computationally tractable manner. A framework for analyzing the local asymptotic power of the tests is developed by embedding the least favorable pairs into a model that allows local approximations under the limits of experiments argument. Examples of the hypotheses we consider include those on the presence of strategic interaction effects in discrete games of complete information. Monte Carlo experiments demonstrate the robust performance of the proposed tests.

Keywords: Incomplete models, Robust inference, Likelihood ratio tests, Limits of experiments

1 Introduction

Incomplete structures arise in a wide class of economic models when the researcher's theory does not fully describe how a particular outcome occurs given the primitives of the model. In this study, we consider a class of models in which, given structural parameter $\theta \in \Theta$ and latent variable $u \in U$, the model predicts the set $G(u|\theta)$ of values for discrete outcome s.[1] The researcher observes s, but his/her theory is silent about the mechanism that determines how s is selected from the predicted set. This class encompasses various models studied in the empirical literature. Examples include models of market entry (Bresnahan and Reiss, 1990, 1991; Berry, 1992; Ciliberto and Tamer, 2009) where the theory does not specify how a pure strategy Nash equilibrium is selected, models of self-selection (Heckman and Honoré, 1990; Mourifie, Henry, and Meango, 2018) where an individual's choice of the sector of activity interacts with unobserved skills, and models of English auction (Haile and Tamer, 2003; Aradillas-Lopez and Tamer, 2008) where the researcher wants to allow solutions that satisfy weak rationality restrictions.[2] In many of these settings, empirical questions can be investigated by testing hypotheses on the structural parameter. Given the incompleteness of the theory, it is desirable to conduct tests without adding assumptions on how selections operate. Despite the need for such robustness, the theoretical study of robust testing procedures and their properties has been limited.[3] This study develops a framework for hypotheses testing in incomplete models, shows how to construct robust and optimal tests, and provides asymptotic tools to evaluate their performance.

Each of the hypotheses we consider can be written as

$$H_0 : \varphi(\theta) \in K_0 \quad \text{v.s.} \quad H_1 : \varphi(\theta) \in K_1, \tag{1}$$

for some function $\varphi : \Theta \to \mathbb{R}^k$ and mutually exclusive sets $K_0, K_1 \subset \mathbb{R}^k$. Such hypotheses naturally arise in applications of incomplete models. For example, in an entry game, a key parameter is the strategic interaction effect, which measures the effect of an opponent firm's entry on a firm's profit. An important empirical question is whether the presence of such interaction effects can be supported by the observed data (de Paula and Tang, 2012). One way to address the question is to formally test the null hypothesis that the strategic interaction does not exist, namely $H_0 : \varphi(\theta) = 0$, against an alternative hypothesis that negative externalities exist, $H_1 : \varphi(\theta) < 0$, by choosing a suitable functional φ.

Such a hypothesis testing problem, however, faces several challenges. First, without further assumptions, the model permits multiple distributions of the observables even if each hypothesis fully specifies the value of θ. To see this, consider a simplified problem in which $H_0 : \theta = \theta_0 \ v.s. \ H_1 :$

---

[1]Introducing covariates does not fundamentally change the structure. We therefore treat this case in Section 5.2 as an extension.

[2]Other examples include models of voting (Kawai and Watanabe, 2013), choice of product variety (Eizenberg, 2014), and network formation models (Miyauchi, 2016).

[3]Studies of tests of the moment inequality model are related, but their model differs from the one we consider. See Section 1.1.

$\theta = \theta_1$. This problem may appear as testing a simple null hypothesis against a simple alternative. However, under each hypothesis, multiple distributions (of s) may be compatible with the theory because any distribution P of the outcome is consistent with θ as long as one can augment the model by finding a suitable selection mechanism that induces P. Therefore, even under the simplest setting, both null and alternative hypotheses can be composite (in terms of permitted distributions).[4] The problem becomes even more challenging when data are obtained from a sequence of experiments. If one stays agnostic about the selection, the unknown selection mechanism is allowed to be arbitrary across experiments. For example, across experiments, the true selection mechanism may vary with and be correlated through a specific variable; however, the researcher does not even know the identity of this variable. From the researcher's viewpoint, the resulting outcome sequence is then heterogeneous and dependent in an unknown way, which in turn makes it hard to characterize the large-sample distribution of test statistics and apply standard asymptotic tools to analyze the power of the tests.

We develop tests that overcome these challenges. For this, we exploit the fact that the sampling uncertainty and lack of understanding of the selection can be represented by a *belief function*, a capacity (or non-additive probability), which belongs to a broader class of two-monotone capacities. Capacities in this class are known to have properties useful for conducting robust statistical inference (Huber and Strassen, 1973, HS henceforth).[5] We start by demonstrating that for testing between simple hypotheses, robust tests can be constructed for any finite sample. The proposed test, which takes the form of a likelihood ratio (LR) test, controls the size in finite samples regardless of the unknown selection mechanism and maximizes a measure of power, which we call *lower power*. One may wonder how such an LR test can be constructed because incomplete models generally admit infinitely many likelihoods. A key observation is that HS's theory ensures that there exists the least favorable pair (LFP) of likelihoods: one compatible with the null that is the least favorable for size control and the other compatible with the alternative that is the least favorable for power maximization.[6] Distinguishing two such extreme distributions turns out to be the best way to test one parameter value against another while staying agnostic about the selection.

We then develop LR tests for repeated experiments. Our first main contribution is to show that despite the potential heterogeneity and dependence of the data, the LFP consists of product measures as long as the latent variables are independent across experiments. Heuristically, this means that under the least favorable distribution for size control (or power maximization), the observables can be viewed as independent across experiments, while the true data-generating process (DGP) may not satisfy such regularity. This leads to a number of desirable results. In particular, it allows us to construct robust LR tests that are optimal in the minimax sense, provide a simple critical value based on a large-sample Gaussian approximation, and develop an asymptotic framework for evaluating the power of the tests.

---

[4]Furthermore, the hypotheses in (1) allow the presence of additional nuisance parameters such as sub-components of θ. We address this issue separately as an extension of the base framework in Section 5.1.

[5]See also Huber and Strassen (1974) for corrigendum and Huber (1981) for the broader area of robust statistics.

[6]They also show that such a pair is unique up to its Radon-Nikodym derivative.

Our second contribution is on the practical side. While HS's theory ensures the existence of the LFP, in practice, one needs to find a way to compute it. We show that in the class of models we consider, the LFP can be computed by solving a finite-dimensional convex program in which the constraints of the program are the *sharp identifying restrictions* studied in the identification literature (Beresteanu, Molchanov, and Molinari, 2011; Galichon and Henry, 2011; Chesher and Rosen, 2017). These restrictions simplify the constraints by making them linear in the control variable, and they therefore play a crucial role in computing the LFP and implementing the robust optimal tests. While the restrictions are useful for characterizing sharp identified sets, little is known about whether they lead to statistically optimal inference. Our result shows they are indeed crucial for likelihood-based inference that has a certain optimality property. Our theoretical result on the LFP also has a practical implication. In particular, under mild conditions, the distributions forming the LFP are independently and identically distributed (i.i.d.) laws, and hence the researcher only needs to find the LFP in a "single" experiment rather than finding it from the entire sequence of experiments. This result also contributes to a significant reduction in the computational cost of our tests.

Our third main contribution is to provide a framework for analyzing the asymptotic power of the tests by embedding the product LFPs into a model that admits local approximations. Specifically, we show that under regularity conditions, a sequence of experiments characterized by the ratio of the LFPs, obtained from a null parameter value and a local alternative, converge to a limit in the sense of Le Cam (1972, 1986). We use this property to characterize the upper bound of the asymptotic lower power of the tests for one-sided hypotheses. Our approach uses the limits of experiments argument and can potentially be used in other statistical decision problems in incomplete models. The main advantage of this approach is that once the LFPs are embedded into a probabilistic model whose limit is tractable, most of the power analysis can be performed using standard tools.

Our framework, however, also incorporates some non-standard features. First, the underlying model in which we embed the LFPs may not satisfy the well-known differentiability in quadratic mean condition, which is sufficient for the local asymptotic normality (LAN) of the experiments over the entire local parameter space. Instead, the model is typically directionally differentiable (in the L^2 sense) and satisfies the LAN property separately on a collection of convex cones that partition the local parameter space. Second, perhaps more importantly, incomplete models may yield alternatives that are not robustly testable. Such an alternative admits a selection mechanism that makes the lower power of any level-α test weakly below the nominal level. We clarify the notion of robust testability and relate it to the observational equivalence concepts studied in the identification literature (Chesher and Rosen, 2017). To conduct a meaningful power analysis, we then introduce an extended notion of alternatives, which we call *shifted local alternatives*. The asymptotic power envelope is shown to be non-trivial against such alternatives.

We further extend our analysis to a model that permits the presence of nuisance components of the parameter vector. Setting up a statistical decision problem, we construct a robust LR test that minimizes a certain risk function. We call this test a Bayes–Dempster–Shafer (BDS) test as it

minimizes a risk that treats parameter uncertainty in a Bayesian way and incorporates ambiguity due to incompleteness through a belief function. Finally, we establish a minimax theorem for this setting, which suggests that a level-α test that maximizes a weighted average of lower power can be approximated using a sequence of BDS tests.

1.1 Relation to the Literature

Our study is most closely related to Epstein, Kaido, and Seo (2016) who developed a theoretical framework for modeling repeated experiments with incompleteness.[7] We adopt their framework and use the (product) belief function to characterize the set of joint distributions of outcomes across experiments. This allows us to study the robustness of tests even in settings where selections are heterogeneous and dependent in an unknown way. This study then takes a step further and develop ways to examine the optimality of tests in such settings.

Our study is also related to earlier work on incomplete models.[8] In particular, our framework for the single experiment builds on that of Jovanovic (1989), who pointed out that models with multiple equilibria lead to incomplete structures and face potential difficulty in identifying structural parameters. Tamer (2003) studied identifying restrictions in an incomplete simultaneous discrete response model with multiple equilibria. Since his seminal work, it has become common to use partially identifying inequality restrictions to bound parameters of interest. Galichon and Henry (2011), Beresteanu, Molchanov, and Molinari (2011), and Chesher and Rosen (2017) characterized sharp identifying restrictions for a wide range of incomplete models using the theory of random sets. We also use, as a central tool, the capacities associated with random sets. As discussed above, the sharp identifying restrictions play an important role in the construction of tests that achieve robustness and statistical optimality.

Commonly used identifying restrictions take the form of moment inequalities. As such, inference methods developed for moment inequality models (Chernozhukov, Hong, and Tamer (2007), Andrews and Soares (2010), Bugni (2010), Andrews and Barwick (2012)) have been commonly used. Some of them (e.g. Galichon and Henry, 2006, 2009, 2013) use test statistics based on capacities to construct confidence regions. These methods combine the implications of incomplete models on moments with an additional assumption on the sampling process (e.g., i.i.d. sampling). By contrast, our approach uses the model's implications on certain likelihoods and does not restrict the sampling process. Chen, Tamer, and Torgovitsky (2011) considered a sieve MLE-based inference, which can be applied to incomplete models. Their approach profiles out a non-parametric nuisance parameter (selection) from the likelihood function using a sieve. Our approach, which picks out the LFP, can also be interpreted as a way to average out the nuisance parameter, in which the weights are the least favorable priors, and averaging is carried out without explicitly introducing a

---

[7]Epstein and Seo (2015) provided axiomatic foundations for robust subjective inference and decision making in such a setting.

[8]The analysis of an incomplete system of equations dates back to the early work of Wald (1950). Here, we focus on reviewing more recent developments in models with multiple equilibria.

functional space.

The results on the optimality of the tests in related settings are somewhat limited. Within a moment inequality framework, Canay (2010) found that a test based on the empirical LR statistic is optimal with respect to the large deviations criterion. In a more specialized setting in which moment restrictions are convex in the parameter, Kaido and Santos (2014) showed that a test based on a semiparametrically efficient estimator of the identified set achieves the asymptotic power envelope against some local alternatives. In models characterized by conditional moment inequalities, Armstrong (2014, 2018) compared the relative power of the testing procedures based on Cramer–von Mises and weighted Kolmogorov–Smirnov statistics. These studies deal with testing problems in models characterized by moment inequalities, which differ from ours in terms of (i) the hypotheses they test and (ii) how they extend a single experiment to repeated experiments. For the former, these studies consider testing whether θ is in the identified set, while our focus is on testing hypotheses of the form in (1), which does not involve identified sets. For the latter, they assume that an i.i.d. sample is available, and hence the robustness issue against heterogeneity and the dependence of selection does not arise.[9]

Finally, our framework for inference is related to others that use limit theorems based on the thoery of random sets. As mentioned earlier, we use a Gaussian approximation to compute the critical value for the LR statistic, which is similar in spirit to the central limit theorem (CLT) in Epstein, Kaido, and Seo (2016), whereas a different tool is used to obtain this result because of the non-trivial difference between the LR statistic we use here and their Kolmogorov–Smirnov-type statistic. In a different class of models, in which observations are set-valued, Beresteanu and Molinari (2008) applied a central limit theorem for random sets to make their inference.

Throughout, for any metric space A, we let Σ_A denote its Borel σ-algebra. We then denote the set of Borel probability measures on A by $\Delta(A)$ and equip it with the topology of weak convergence. Let $N(\mu, V)$ denote the law of a normal random vector with mean $\mu \in \mathbb{R}^k$ and variance-covariance matrix $V \in \mathbb{R}^{k \times k}$. For any integrable random vector X, we let $E_P[X]$ denote its expectation with respect to probability measure P.

The remainder of the paper is organized as follows. Section 2 introduces the model and provides illustrative examples. In Section 3, after reviewing the theory of Huber and Strassen (1973) (Section 3.1), we present our main result on minimax tests in repeated experiments (Section 3.3). We also discuss the robust testability of the hypotheses and computational aspects (Sections 3.1–3.2). Section 3.4 provides results on the local asymptotic power of the tests. Section 4 then provides simulation evidence. Section 5 contains extensions of the baseline framework and Section 6 concludes. Appendices B and C collect the proofs of the theoretical results.

---

[9]See Epstein, Kaido, and Seo (2016) for this distinction as well as Molchanov and Molinari (2018) (Section 5.3).

2 Setup

Let S be a finite set of observable outcomes and let $u \in U$ denote a variable unobservable to the researcher, where U is assumed to be a Polish space. Let Θ denote the parameter space. We let $m = \{m_\theta, \theta \in \Theta\}$ denote a family of Borel probability measures on U. For each $\theta \in \Theta$, let $G(\cdot|\theta) : U \twoheadrightarrow S$ be a weakly measurable correspondence. This map shows how latent variable u is mapped to a set of permissible outcomes. Observable outcome s is then a measurable selection of random set $G(u|\theta)$. As such, the model does not impose any restrictions on how s is selected. One may also introduce observable covariates to this model. As the core analysis remains unaffected, we defer the analysis of this case to Section 5.2.

The incomplete structure above is summarized by tuple $(S, U, m, \Theta; G)$. Such structures arise in various economic models. To fix the ideas, we present several examples based on simplifications of well-known models. The first example is a binary response game, which is commonly used to analyze environments such as firms' entry into markets and households' joint labor supply decisions (Bresnahan and Reiss, 1990, 1991; Berry, 1992; Ciliberto and Tamer, 2009).

Example 1 (Binary response game)**.** Consider a two-player binary response game with the following payoff:

	out	in
out	$0, 0$	$0, u^{(2)}$
in	$u^{(1)}, 0$	$u^{(1)} + \theta^{(1)}, u^{(2)} + \theta^{(2)}$

The effect of the other player's action (e.g., entry) on player k's payoff is represented by $\theta^{(k)}$. Throughout, we call $\theta = (\theta^{(1)}, \theta^{(2)})' \in \Theta \subset \mathbb{R}^2$ the players' *strategic interaction effects*. Let $U = \mathbb{R}^2$. The latent payoff shifter $u = (u^{(1)}, u^{(2)})'$ follows a continuous distribution m_θ. Consider pure strategy Nash equilibria in this game when $\theta^{(1)} \leq 0$ and $\theta^{(2)} \leq 0$.[10] There are four possible equilibrium outcomes: $S = \{(0,0), (1,1), (1,0), (0,1)\}$. How u and θ are mapped to the equilibrium outcomes is summarized by the following correspondence:

$$
G(u|\theta) = \begin{cases}
\{(0,0)\} & u^{(1)} < 0, \ u^{(2)} < 0 \\
\{(1,1)\} & u^{(1)} \geq -\theta^{(1)}, \ u^{(2)} \geq -\theta^{(2)} \\
\{(1,0)\} & u \in U_1, \\
\{(0,1)\} & u \in U_2, \\
\{(1,0),(0,1)\} & 0 \leq u^{(1)} < -\theta^{(1)}, \ 0 \leq u^{(2)} < -\theta^{(2)},
\end{cases}
\tag{2}
$$

where $U_1 = \{u : u^{(1)} \geq -\theta^{(1)}, u^{(2)} < -\theta^{(2)}) \cup \{u : 0 \leq u^{(1)} < -\theta^{(1)}, u^{(2)} < 0\}$ and $U_2 = \{u : 0 \leq u^{(1)} < -\theta^{(1)}, u^{(2)} \geq -\theta^{(2)}\} \cup \{u : u^{(1)} < 0, u^{(2)} \geq 0\}$. The model predicts multiple equilibria when each player's latent payoff shifter is between the two thresholds (0 and $-\theta^{(k)}$, $k = 1, 2$).

---

[10]For simplicity, we focus on games with strategic substitutes throughout. Games with strategic complements, in which $\theta^{(1)} > 0, \theta^{(2)} > 0$, can be analyzed similarly.

The second example is the (binary) Roy model studied in Mourifie, Henry, and Meango (2018).

Example 2 (Roy model). Consider an individual who chooses a sector of activity $D \in \{0, 1\}$ and whether to work $Y \in \{0, 1\}$ in the sector. The binary outcome is given by $Y = Y_1 D + Y_0(1 - D)$, where selection indicator D is determined by binary potential outcomes (Y_0, Y_1) through the following structure:

$$
D = \begin{cases} 1 & Y_1 > Y_0 \\ 0 \text{ or } 1 & Y_1 = Y_0 \\ 0 & Y_1 < Y_0. \end{cases} \tag{3}
$$

Binary potential outcome Y_d represents whether one has good economic prospects in sector $d \in \{0, 1\}$.[11] The sector choice is not uniquely determined if $Y_1 = Y_0$. This model can be mapped to the present framework by letting $s = (y, d) \in S = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ be observable outcomes and $u = (Y_0, Y_1) \in U \equiv \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ be latent variables. Since u is discrete, we take the probability mass function of u as a parameter vector. For this, let $\theta = (\theta^{(0,0)}, \theta^{(0,1)}, \theta^{(1,0)})' \in \Theta$, where $\theta^{(0,0)} = m_\theta((Y_0, Y_1) = (0, 0))$, for instance, and $\Theta = \{\theta \in [0, 1]^3 : \theta^{(0,0)} + \theta^{(0,1)} + \theta^{(1,0)} \leq 1\}$. The Roy selection in (3) then yields the following correspondence:

$$
G(u) = \begin{cases} \{(0, 0), (0, 1)\} & u = (0, 0) \\ \{(1, 1)\} & u = (0, 1) \\ \{(1, 0)\} & u = (1, 0) \\ \{(1, 0), (1, 1)\} & u = (1, 1). \end{cases} \tag{4}
$$

The model implies a unique outcome only if the potential outcomes are ordered (e.g., an individual works in sector 1 when $Y_0 = 0$ and $Y_1 = 1$). Otherwise, it predicts multiple outcome values.

The third example is an incomplete model of an English auction (Haile and Tamer, 2003).

Example 3 (English auction). For each auction, there are $k = 1, \cdots, \bar{N}$ potential bidders whose valuations $u^{(k)}, k = 1, \cdots, \bar{N}$ are drawn independently from common distribution F_θ with support $[\underline{u}, \bar{u}] \subset \mathbb{R}$, which is indexed by parameter $\theta \in \Theta$. There is reserve price r and minimum bid increment $\bar{\Delta} > 0$. Each bidder's set of actions is $\{r, r + \bar{\Delta}, r + 2\bar{\Delta}, \cdots, r + K\bar{\Delta}\}$, where $K \in \mathbb{N}$ is such that $r + K\bar{\Delta} > \bar{u}$. Bidders with valuations above the reserve price bid in the auction. Let $N \leq \bar{N}$ be the number of such bidders. Haile and Tamer (2003) assumed the following weak restrictions on observed bids $s = (s^{(1)}, \cdots, s^{(N)})$: (i) bidders do not bid more than their valuations, implying $s^{(k)} \leq u^{(k)}, k = 1, \cdots, N$, and (ii) bidders do not allow an opponent to win at a price

---

[11]We focus on the case in which the potential outcomes are binary. Mourifie, Henry, and Meango (2018) extended their analysis to more general settings in which Y_d is discrete or continuous (or both). As discussed in Section 2 of their paper, one could also think of the binary Roy model as a consequence of a two-step decision process in which D is determined first by potential wage Y_d^* in sector d, and whether to work in section d is determined by whether Y_d^* crosses a threshold.

they can beat, which implies $u^{(N-1,N)} \leq s^{(N,N)} + \bar{\Delta}$, where $x^{(k,N)}$ denotes the k-th (ascending) order statistic within a sample $(x^{(1)}, \cdots, x^{(N)})$.

Let $S = (\emptyset \cup \{r, r + \bar{\Delta}, r + 2\bar{\Delta}, \cdots, r + K\bar{\Delta}\})^{\bar{N}}$ be the set of bids. Let $U = [\underline{u}, \overline{u}]^{\bar{N}}$ be the set of valuations and $m_\theta = F_\theta^N$ be the (N-fold) product measure on U, which represents the joint distribution of private valuations. The prediction of the model is then given by

$$G(u) = \left\{ s \in S : s^{(k)} \leq u^{(k)}, \ u^{(N-1,N)} \leq s^{(N,N)} + \bar{\Delta}, \ k = 1, \cdots, N \right\}. \tag{5}$$

2.1 Set of Permitted Distributions and Robustness

To develop tests for incomplete models, we start by defining the family of probability distributions compatible with the model structure. For each $\theta \in \Theta$, define

$$\mathcal{P}_\theta \equiv \left\{ P \in \Delta(S) : P = \int_U P_u \, dm_\theta(u), \ \text{for some } P_u \in \Delta(G(u|\theta)) \right\}, \tag{6}$$

where P_u is a conditional law of s (supported on $G(u|\theta)$), which represents the unknown *selection mechanism*. This set collects probability distributions P, for which one can find a suitable selection mechanism and make it consistent with a given parameter value θ and the model structure. Economic theory rarely provides any guidance on selection. The researcher therefore views any distribution in $\mathcal{P}_\theta$ as consistent with θ.

Within this model, consider testing parameter value θ_0 against another value θ_1 on the basis of observed outcome $s \in S$. This is equivalent to testing the null hypothesis, $P \in \mathcal{P}_{\theta_0}$, against the alternative hypothesis, $P \in \mathcal{P}_{\theta_1}$. Note that $\mathcal{P}_{\theta_0}$ and $\mathcal{P}_{\theta_1}$ may contain multiple (typically infinitely many) elements because the selection is left unspecified. Therefore, even for testing a single value of θ against another value, the hypotheses are composite in terms of the permitted distributions.[12] Given this challenge, we pursue a robust approach to inference. That is, we construct tests that (i) control the size uniformly across distributions permitted under the null and (ii) maximize certain measures of power under the alternative.

3 Robust Tests for Incomplete Models

We provide the main theoretical results below. For this, we start with preliminaries including an introduction of the key technical tools and the important extension of the Neyman–Pearson lemma presented by Huber and Strassen (1973). We then discuss the computational aspects of our LR tests, novel results on minimax tests, and LFPs in repeated experiments as well as a local asymptotic power analysis, which builds on our main theorem (Theorem 3.1).

---

[12]The composite nature of the hypotheses arises because the unknown selection is a nuisance parameter. It is possible to allow some components of structural parameter θ to be additional nuisance parameters. We analyze this extension in Section 5.1.

3.1 Preliminaries

Belief Functions

For any $\mathcal{P} \subseteq \Delta(S)$, define the upper and lower probabilities of $\mathcal{P}$ pointwise by $\nu^*(A) \equiv \sup_{P \in \mathcal{P}} P(A)$ and $\nu(A) \equiv \inf_{P \in \mathcal{P}} P(A), A \subset S$, respectively. These functions are conjugate to each other in the sense that $\nu^*(A) = 1 - \nu(A^c)$ for any $A \subset S$. Under mild restrictions on $\mathcal{P}$, they define set functions called *capacities*.[13]

For each $\theta \in \Theta$ and $A \subset S$, define ν_θ and ν_θ^* as the lower and upper probabilities of $\mathcal{P}_\theta$ defined in (6):

$$\nu_\theta(A) \equiv \inf_{P \in \mathcal{P}_\theta} P(A), \qquad \text{and} \qquad \nu_\theta^*(A) \equiv \sup_{P \in \mathcal{P}_\theta} P(A). \tag{7}$$

The key factor for our analysis is that the lower probability ν_θ of $\mathcal{P}_\theta$ is a *belief function* (or infinitely monotone capacity).[14] From Choquet's theorem (e.g., Choquet, 1954; Philippe, Debs, and Jaffray, 1999; Molchanov, 2006), it is related to the probability distribution of random set $G(u|\theta)$ as follows:

$$\nu_\theta(A) = m_\theta(G(u|\theta) \subset A), \quad \text{for any } A \subset S. \tag{8}$$

This representation allows us to obtain ν_θ without explicitly solving the minimization (or maximization) in (7) by computing the right-hand side of (8) directly. Another key property of the belief function is that $P \in \mathcal{P}_\theta$ is equivalent to the following statement:

$$\nu_\theta(A) \leq P(A), \ A \subset S. \tag{9}$$

Galichon and Henry (2011) used the restrictions above to characterize the smallest possible (or "sharp") identification region of the parameters.[15] Following the literature, we call these the *sharp identifying restrictions* (see also Beresteanu, Molchanov, and Molinari, 2011; Chesher and Rosen, 2017).[16]

Theory of Huber and Strassen (1973)

Our starting point is an analog of the Neyman–Pearson framework, which builds upon HS. For $\theta_0, \theta_1 \in \Theta$ such that $\mathcal{P}_{\theta_0}$ and $\mathcal{P}_{\theta_1}$ are disjoint, consider testing a simple null hypothesis, $H_0 : \theta = \theta_0$,

---

[13]Appendix A provides the details. Some authors distinguish a capacity from its conjugate (co-capacity). For simplicity, we call both of these "capacities" throughout.

[14]The infinite monotonicity of ν_θ follows from Philippe, Debs, and Jaffray (1999) (Theorem 3). The foundations of belief functions are given by Dempster (1967) and Shafer (1982). See Gul and Pesendorfer (2014) and Epstein and Seo (2015) for the axiomatic foundations of the use of belief functions in incomplete models.

[15]Galichon and Henry (2011) used the conjugate of ν_θ, which yields equivalent identifying restrictions.

[16]While the restrictions play a role in constructing robust tests, the sharp identified set does not play a role as the latter is an object of interest when the sampling processes reveals the unique data generating process in the limit, which is not guaranteed in our setting. See Epstein, Kaido, and Seo (2016) for a discussion.

against a simple alternative hypothesis, $H_1 : \theta = \theta_1$. In *complete* models, in which G is singleton-valued, a well-defined reduced form induces a unique likelihood function (Tamer, 2003). In such settings, an optimal test is an LR test, as is well known from the Neyman–Pearson lemma. In incomplete models, however, the model generally admits a (non-singleton) set $\mathcal{P}_\theta$ of likelihoods, which prevents us from directly applying the Neyman–Pearson lemma.

In this setting, it is useful to consider *minimax tests* (see Lehmann and Romano, 2006, Ch. 8 for the general principles). Let $\phi : S \mapsto [0, 1]$ denote a (possibly randomized) test. For each P on (S, Σ_S), the rejection probability of ϕ is

$$E_P[\phi(s)] = \int \phi(s) dP. \tag{10}$$

Let $\pi_{\theta_1}(\phi) \equiv \inf_{P_1 \in \mathcal{P}_{\theta_1}} E_P[\phi(s)]$ be the *lower power* of ϕ under θ_1. This is the power value certain to be obtained regardless of the unknown selection mechanism. We then call test ϕ a *level-α minimax test* if it satisfies the following conditions:

$$\sup_{P \in \mathcal{P}_{\theta_0}} E_P[\phi(s)] \leq \alpha , \tag{11}$$

and

$$\pi_{\theta_1}(\phi) \geq \pi_{\theta_1}(\tilde{\phi}), \ \forall \tilde{\phi} \text{ satisfying (11).} \tag{12}$$

The condition in (11) imposes a uniform size control requirement. In (12), tests are ranked in terms of their lower power. This reflects the researcher's preference for tests that exhibit robust power performance across selections.

A belief function (and its conjugate) is a special case of two-monotone (and two-alternating) capacities whose properties have proven powerful for conducting robust inference (Huber, 1981).[17] For a class of models whose lower probabilities are two-monotone, HS showed that the rejection region of a minimax test takes the form $\{s : \Lambda(s) > t\}$ for a measurable function, $\Lambda : S \to \mathbb{R}$, which they called the Radon–Nikodym derivative of $\nu_{\theta_1}^*$ with respect to $\nu_{\theta_0}^*$. Further, they showed that there exists an LFP of distributions $(Q_0, Q_1) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_1}$ such that for all $t \in \mathbb{R}_+$,

$$Q_0(\Lambda > t) = \nu_{\theta_0}^*(\Lambda > t), \tag{14}$$

and

$$Q_1(\Lambda > t) = \nu_{\theta_1}(\Lambda > t), \tag{15}$$

---

[17]Capacity ν is said to be *monotone of order k* or, for short, *k-monotone* if for any $A_i \subset S, i = 1 \cdots, k$,

$$\nu\left(\cup_{i=1}^k A_i\right) \geq \sum_{I \subseteq \{1, \cdots, k\}, I \neq \emptyset} (-1)^{|I|+1} \nu\left(\cap_{i \in I} A_i\right). \tag{13}$$

Conjugate $\nu^*(A) = 1 - \nu(A^c)$ is then called a *k-alternating* capacity.

where Λ can be taken to be a version of the Radon–Nikodym derivative:

$$\frac{dQ_1}{dQ_0} = \left\{ \frac{q_1}{q_0} : q_j \in \frac{dQ_j}{d\upsilon}, \; q_j \geq 0, \; j = 0, 1, \; q_0 + q_1 > 0 \right\}, \tag{16}$$

where υ is a measure that dominates $Q_j, j = 0, 1$. Below, we take υ to be the counting measure.

Heuristically, this means that Q_0 is the probability distribution consistent with the null parameter value, under which the size of the test is maximal. Similarly, Q_1 is the distribution consistent with the alternative parameter value, which is the least favorable for power maximization. The following extension of the classic Neyman–Pearson lemma (tailored to our setting) then follows from HS.

Lemma 3.1. *Let $\mathcal{P}_{\theta_0}$ and $\mathcal{P}_{\theta_1}$ be defined as in (6) with $\theta = \theta_0$ and $\theta = \theta_1$, respectively. Then, there is a level-α minimax test $\phi: S \to [0, 1]$ such that*

$$\phi(s) = \begin{cases} 1 & if \quad \Lambda(s) > C \\ \gamma & if \quad \Lambda(s) = C \\ 0 & if \quad \Lambda(s) < C, \end{cases} \tag{17}$$

where $\Lambda \in dQ_1/dQ_0$ is a version of the Radon–Nikodym derivative of the LFP $(Q_0, Q_1) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_1}$, and (C, γ) solves $E_{Q_0}[\phi(s)] = \alpha$.

Lemma 3.1 characterizes a level-α minimax test as an LR test in which the ratio is formed by the LFP. Recall that Q_1 is the least favorable for maximizing the test's power, while Q_0 is the least favorable for controlling the size. Heuristically, a large value of their ratio can then be taken as evidence against the null hypothesis. Lemma 3.1 states that it is indeed optimal in the minimax sense to reject H_0 when this ratio is sufficiently high.[18]

Lemma 3.1 is an existence and characterization result useful for obtaining the more general results below. To implement LR tests in practice, one needs to compute the LFPs (typically in a single experiment). We discuss the computational aspects in Section 3.2.

Testability of Hypotheses

Before proceeding further, we comment on the testability of the hypotheses. The theory of HS requires that $\mathcal{P}_{\theta_0}$ and $\mathcal{P}_{\theta_1}$ are disjoint. Otherwise, any test is vacuous from the minimax viewpoint because probability distribution $P \in \mathcal{P}_{\theta_0} \cap \mathcal{P}_{\theta_1}$ is consistent with both hypotheses. If this is the case, we say θ_1 is not *robustly testable* relative to θ_0 because the lower power of any level-α test cannot exceed α. This issue does not arise in complete models as long as the likelihood function satisfies $f(s; \theta_0) \neq f(s; \theta_1), a.s.$ for any $\theta_0 \neq \theta_1$. One should therefore expect non-trivial lower power only if

---

[18]In addition, the binary experiment $(S, \Sigma_S, P \in \{Q_0, Q_1\})$ in which one tests Q_0 against Q_1 is the hardest (or least informative) in terms of Bayes risk among all binary experiments such that $(S, \Sigma_S, P \in \{P_0, P_1\})$ with $P_j \in \mathcal{P}_{\theta_j}, j = 0, 1$ (Bednarski, 1982).

an alternative hypothesis induces set $\mathcal{P}_{\theta_1}$ that does not intersect with $\mathcal{P}_{\theta_0}$. For this, there needs to be an event $\bar{A} \subset S$ such that $\nu^*_{\theta_0}(\bar{A}) < \nu_{\theta_1}(\bar{A})$ (or $\nu^*_{\theta_1}(\bar{A}) < \nu_{\theta_0}(\bar{A})$).[19]

The lack of robust testability is also related to the notion of *observational equivalence* (see Chesher and Rosen, 2017, and references therein). Let s follow distribution P and suppose P is known. Consider parameter values $\theta, \theta' \in \Theta$ such that $\theta \neq \theta'$, $P \in \mathcal{P}_\theta$, and $P \in \mathcal{P}_{\theta'}$. In other words, the true distribution can be justified by structure θ augmented with some selection or by another structure θ' (again augmented with some selection). When this holds, θ and θ' are said to be observationally equivalent with respect to P. In incomplete models, P is not in general identifiable, as the sampling process does not necessarily reveal it even asymptotically (Maccheroni and Marinacci, 2005; Epstein, Kaido, and Seo, 2016). Following Chesher and Rosen (2017), we say that θ and θ' are *potentially observationally equivalent* if two structures are observationally equivalent for some P. Clearly, any pair of potentially observationally equivalent parameter values are not robustly testable, as $\mathcal{P}_\theta$ and $\mathcal{P}_{\theta'}$ share a distribution in common. This feature of the model raises a challenge for analyzing the local power of the tests because some local alternatives may not be robustly testable. Evaluating the power of the tests under such alternatives does not lead to a meaningful comparison. We therefore introduce a suitably modified notion of local alternatives if such an issue arises (see Section 3.4).

3.2 Computing LFPs

A key step toward implementing our tests is the computation of the LFPs, in which the sharp identifying restrictions play a role. Let $H : [0,1] \to \mathbb{R}$ be a twice-continuously differentiable convex function. Our proposal is to find the LFP through the following characterization:

$$(Q_0, Q_1) = \underset{(P_0, P_1) \in \Delta(S)^2}{\arg\min} \int H\Big(\frac{dP_0}{d(P_0 + P_1)}\Big) d(P_0 + P_1) \tag{18}$$

$$s.t. \ \nu_{\theta_0}(A) \leq P_0(A), \ A \subset S$$

$$\nu_{\theta_1}(A) \leq P_1(A), \ A \subset S,$$

where the constraints on (P_0, P_1) are the sharp identifying restrictions.[20] The number of restrictions can be reduced further by restricting the class of events to the *core determining class* (see Galichon and Henry, 2011; Luo and Wang, 2017a). This is a convex program with a convex objective function and linear constraints.[21]

---

[19] In Example 1, $\bar{A} = \{(1,1)\}$ (or $\bar{A} = \{(1,0), (0,1)\}$) constitutes such an event for testing $H_0 : \theta = 0$ against $H_1 : \theta = \theta_1$ with $\theta_1 < 0$ when u is continuously distributed over $\mathbb{R}^2$.

[20] An alternative approach would be to use the sharp identifying restrictions of Beresteanu, Molchanov, and Molinari (2011), which also yield a finite number of linear restrictions. While we do not pursue that here, the insights presented in this paper may be useful for constructing optimal tests in models with endogeneity. Such models are studied by Chesher and Rosen (2017), who obtained sharp identifying restrictions using generalized instrumental variables.

[21] The convexity of the objective function follows from the convexity of the perspective $g(x,t) = tH(x/t)$ on its domain (Boyd and Vandenberghe, 2004, Sec. 3.2.6).

Our proposal builds on Theorem 6.1 in HS, which characterizes the LFP as a solution to a more general and abstract optimization problem in which (P_0, P_1) is constrained to $\mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_1}$. However, because of the presence of an unknown selection in the definition of $\mathcal{P}_\theta$ (see (6)), directly imposing such constraints does not lead to a tractable program. Restating the constraints using the sharp identifying restrictions, we may reduce the problem to a convex one with linear constraints, which can then be solved using efficient algorithms (e.g., Boyd and Vandenberghe, 2004). Computing ν_{θ_0} and ν_{θ_1} in the constraints is often straightforward (see Example 1 and the supplementary material of Epstein, Kaido, and Seo (2016)).

Emphasizing the role of the sharp identifying restrictions is worthwhile. Instead of using them to characterize the set of identifiable parameter values, we use them to obtain the LFP. To the best of our knowledge, this way of using the sharp identifying restrictions is new. Further, imposing only a subset of them in (18) does not generally yield an LFP. In this sense, these restrictions are crucial for robust and optimal inference.

Remark 3.1. Since S is finite, the program in (18) can be simplified further. Let p_0 denote the probability mass function of $P_0 \in \mathcal{P}_{\theta_0}$ and p_1 be defined similarly. For simplicity, suppose $p_0(s) > 0$ for all $s \in S$ and let $H(x) = -\ln x$. Then, one may solve

$$(q_0, q_1) = \underset{(p_0, p_1) \in \Delta(S)^2}{\arg\min} \sum_{s \in S} \ln\left(\frac{p_0(s) + p_1(s)}{p_0(s)}\right)(p_0(s) + p_1(s)) \tag{19}$$

$$s.t. \quad \nu_{\theta_0}(A) \leq \sum_{s \in A} p_0(s), \ A \subset S$$

$$\nu_{\theta_1}(A) \leq \sum_{s \in A} p_1(s), \ A \subset S.$$

In this finite-dimensional convex program, one minimizes Kullback–Leibler divergence $D_{KL}(p_0 + p_1 \| p_0)$ subject to linear constraints on (p_0, p_1). One may then use efficient numerical solvers (e.g., `CVX`) to obtain the LFP.

We illustrate the computation of an LFP and minimax test using Example 1.

Example 1 (Binary response game (continued)). Let $0 < \alpha < 1/2$. Consider testing $H_0 : \theta = 0$ against $H_1 : \theta = \theta_1$, where $\theta_1^{(k)} < 0, k = 1, 2$. Suppose that u follows the standard bivariate normal distribution $N(0, I_2)$.

It is straightforward to calculate $\nu_\theta(A)$. As discussed earlier, a key feature of the belief function is that it is related to a probability distribution of a random set in (8). This allows us to compute $\nu_\theta(A)$ analytically. For example, let $A = \{(1, 0), (1, 1)\}$. From (2) and (8), the lower probability of

A is then

$$
\begin{aligned}
\nu_\theta\big(\{(1,0),(1,1)\}\big) &= m_\theta\big(G(u|\theta) \subseteq \{(1,0),(1,1)\}\big) \\
&= m_\theta\big(G(u|\theta) = \{(1,0)\}\big) + m_\theta\big(G(u|\theta) = \{(1,1)\}\big) + m_\theta\big(G(u|\theta) = \{(1,0),(1,1)\}\big) \\
&= \frac{1}{4} + \frac{\Phi(\theta^{(1)})}{2}, \quad (20)
\end{aligned}
$$

where Φ is the CDF of a standard normal random variable (see Table D.1 in Appendix D for $\nu_\theta(A)$ for other events). In more complex models, simulation-based methods can be used (Galichon and Henry, 2011; Ciliberto and Tamer, 2009; Epstein, Kaido, and Seo, 2016).

Suppose that $\Phi(\theta_1^{(k)})(1 - \Phi(\theta_1^{(-k)})) \leq \frac{1}{4}$ for $k = 1, 2.$[22] Solving (19), we obtain the following probability mass functions of the LFP (Q_0, Q_1):

$$
(q_0(0,0), q_0(1,1), q_0(1,0), q_0(0,1)) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \tag{21}
$$

$$
(q_1(0,0), q_1(1,1), q_1(1,0), q_1(0,1)) = \left(\frac{1}{4}, \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}), \frac{3 - 4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})}{8}, \frac{3 - 4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})}{8}\right). \tag{22}
$$

The LR statistic Λ is then given by

$$
\Lambda(s) = \begin{cases}
1 & s = (0,0) \\
4\Phi(\theta^{(1)})\Phi(\theta^{(2)}) & s = (1,1) \\
\frac{3 - 4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})}{2} & s = (1,0) \\
\frac{3 - 4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})}{2} & s = (0,1).
\end{cases} \tag{23}
$$

An LR test based on Λ is level-α when $C = \frac{3 - 4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})}{2}$ and $\gamma = 2\alpha$. Hence, the test can be simplified as

$$
\phi(s) = \begin{cases}
2\alpha & s = (1,0) \text{ or } (0,1) \\
0 & \text{otherwise.}
\end{cases} \tag{24}
$$

This test rejects the null hypothesis with probability $\gamma = 2\alpha$ when either $s = (1,0)$ or $(0,1)$ is observed. Otherwise, the null hypothesis is retained.

The intuition behind this test is as follows. When H_0 is true ($\theta^{(j)} = 0$ for both players), the model is indeed complete. The four possible outcomes occur with equal probabilities because of $u \sim N(0, I_2)$ (Figure 1, left). When H_1 is true, there exists a region of incompleteness: the set of values of u for which multiple equilibria $\{(1,0),(0,1)\}$ are predicted. While the model is silent about the exact allocation of the probabilities across equilibria, it predicts a higher probability of

---

[22]The form of the minimax test depends on the relative magnitude of $\theta_1^{(1)}$ and $\theta_1^{(2)}$. This assumption is made for analyzing one of the subcases. See Section 3.4 for the full description of the minimax test in Example 1.

$s \in \{(1,0),(0,1)\}$ under H_1 than H_0 (Figure 1, right). The robust LR test then interprets $s = (1,0)$ or $s = (0,1)$ as evidence of the presence of strategic interaction and rejects the null hypothesis with a positive probability. This mechanism does not rely on any knowledge of the selection.

Remark 3.2. Consider a special case of the example above in which the alternative hypothesis is symmetric: $\theta_1^{(1)} = \theta_1^{(2)} = \theta$. The minimax test in (24) does not depend on the value of θ under the alternative. Hence, it can be interpreted as a "uniformly most powerful" test in terms of the lower power for testing $H_0 : \theta = 0$ against $H_1 : \theta < 0$.[23]

Figure 1: Level sets of G under H_0 (left) and H_1 (right)
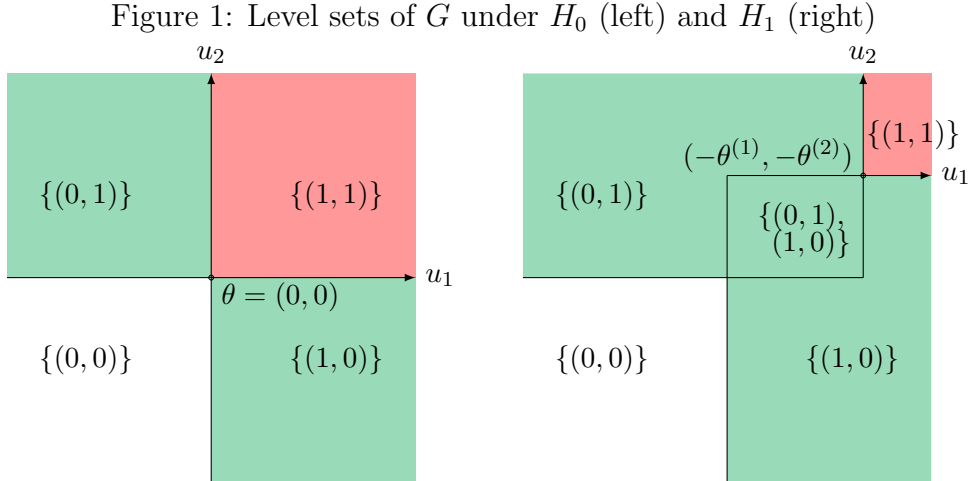
Note: The area in red represents the values of u under which $\{(1,1)\}$ is predicted. The area in green represents the values of u under which $\{(0,1)\}$, $\{(1,0)\}$, or their union is predicted.

3.3 Minimax Tests in Repeated Experiments

The sequence of outcomes $s^n = (s_1, s_2, \cdots, s_n)$ is commonly generated from repeated experiments. In this section, we present a set of theoretical results that characterize a minimax test in such a setting and provide an asymptotic Gaussian approximation to its (upper) rejection probability.

For any set B, let B^n denote the n-fold Cartesian product of B. For each $n \in \mathbb{N}$, let S^n and U^n be the sets of outcome sequences $s^n = (s_1, s_2, \cdots, s_n)$ and latent variable sequences $u^n = (u_1, \cdots, u_n)$, respectively. Below, we use the abbreviation m^n to denote the family $\{m_\theta^n\}_{\theta \in \Theta}$ of u^n's joint laws permitted by the model. We then make the following assumption on each member of this family.

Assumption 3.1. *For each $\theta \in \Theta$, $m_\theta^n \in \Delta(U^n)$ is a product measure.*

This assumption requires that u_i's are distributed independently across experiments. A leading case is that $(u_1, \ldots, u_n)$ are i.i.d. This can also accommodate heteroskedasticity and other types of heterogeneity across cross-sectional units or clusters of them (e.g., group-specific effects).

---

[23]If $\theta_1^{(1)} = \theta_1^{(2)}$ is not imposed, the form of the minimax test depends on the relative magnitude of the interaction effects. See Table 1.

Without further assumptions, s^n takes values in the Cartesian product of the sets of permissible outcome values:

$$G^n(u^n|\theta) = \prod_{i=1}^{n} G(u_i|\theta), \tag{25}$$

where $G(\cdot|\theta)$ is given as in (2).[24] This set collects outcome sequences that are compatible with the model and θ. We represent the repeated experiments by the tuple $(S^n, U^n, \Theta, G^n; m^n)$. Although u^n is assumed to be independent, the outcome sequence s^n can be dependent because the model does not restrict the selection mechanism. Similarly, even if one makes the stronger assumption that the u_i's are i.i.d., the distribution of outcome s_i may be heterogeneous because of the potential heterogeneity of selection across experiments. This feature arises because the joint selection mechanism (across all experiments) is left unspecified and the joint distribution of the outcome sequence depends on this incidental parameter.

For each $\theta \in \Theta$ and $n \in \mathbb{N}$, the set of distributions compatible with the model is

$$\mathcal{P}_\theta^n = \left\{ P \in \Delta(S^n) : P = \int P_u dm_\theta^n, \;\; \text{for some } P_u \in \Delta(G^n(u^n|\theta)) \right\}. \tag{26}$$

This set collects all the distributions of s^n consistent with θ. P_u is unrestricted in the sense that the selection may be heterogeneous and dependent across experiments. Hence, $\mathcal{P}_\theta^n$ contains a broad range of distributions that can exhibit arbitrary dependence and heterogeneity. In particular, $\mathcal{P}_\theta^n$ allows measures under which the distributions of sample moments are not well approximated by classical limit theorems—even in large samples (see Epstein, Kaido, and Seo (2016)).

Finding the LFP in such a rich set of distributions may be challenging. However, under Assumption 3.1 and with the correspondence in (25), the model has a tractable "product" structure, which significantly simplifies the characterization of the LFPs.

Let $\nu_\theta^{n,*}$ and ν_θ^n denote the upper and lower probabilities of $\mathcal{P}_\theta^n$. For each $i \in \{1, \ldots, n\}$ and $\theta \in \Theta$, let

$$\mathcal{P}_{\theta,i} \equiv \left\{ P \in \Delta(S) : P = \int_U P_u dm_{\theta,i}(u), \text{ for some } P_u \in \Delta(G(u|\theta)) \right\}, \tag{27}$$

where $m_{\theta,i}$ is the i-th marginal distribution of m_θ^n. The following theorem shows that the minimax test in repeated experiments is an LR test and that the LFPs are product measures.

Theorem 3.1. *Suppose Assumption 3.1 holds. Then, (i) an LFP $(Q_0^n, Q_1^n) \in \mathcal{P}_{\theta_0}^n \times \mathcal{P}_{\theta_1}^n$ exists such*

---

[24]It is possible to allow the functional form of $G(u_i|\theta)$ to vary across i as well. For notational simplicity, we do not explicitly consider this extension here. We introduce the heterogeneity of G due to covariates in Section 5.2.

that for all $t \in \mathbb{R}_+$,

$$\nu_{\theta_0}^{*,n}(\Lambda_n > t) = Q_0^n(\Lambda_n > t) \qquad (28)$$

$$\nu_{\theta_1}^{n}(\Lambda_n > t) = Q_1^n(\Lambda_n > t), \qquad (29)$$

where Λ_n is a version of the Radon–Nikodym derivative dQ_1^n/dQ_0^n. The LFP consists of the product measures:

$$Q_0^n = \bigotimes_{i=1}^{n} Q_{0,i}, \quad and \quad Q_1^n = \bigotimes_{i=1}^{n} Q_{1,i}, \qquad (30)$$

where, for each $i \in \mathbb{N}$, $(Q_{0,i}, Q_{1,i}) \in \mathcal{P}_{\theta_0,i} \times \mathcal{P}_{\theta_1,i}$ is the LFP in the i-th experiment:

(ii) A minimax test ϕ_n: $S^n \to [0,1]$ can be constructed as

$$\phi_n(s^n) = \begin{cases} 1 & if \quad \Lambda_n(s^n) > C_n \\ \gamma_n & if \quad \Lambda_n(s^n) = C_n \\ 0 & if \quad \Lambda_n(s^n) < C_n, \end{cases} \quad with \quad \Lambda_n(s^n) = \prod_{i=1}^{n} \Lambda_i, \qquad (31)$$

where $\Lambda_i \in dQ_{1,i}/dQ_{0_i}$ for all i, and C_n and γ_n are chosen so that $E_{Q_0^n}[\phi_n(s^n)] = \alpha$.

The LFP consists of the product measures.[25] Heuristically, this means that either for controlling size or maximizing power, the least favorable distribution in $\mathcal{P}_{\theta_0}^n$ (or $\mathcal{P}_{\theta_1}^n$) is a law that multiplies up the least favorable distributions in the individual experiments. When the u_i's are i.i.d., this characterization has a particularly useful implication for the implementation. To construct a minimax test, it suffices to find the LFP $(Q_0, Q_1) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_1}$ in a *single* experiment (S, U, Θ, G, m). One may then obtain the LR statistic by taking the product of their ratios across experiments. We state this result as a corollary.

Corollary 3.1. *Suppose $(u_1, \ldots, u_n)$ are identically and independently distributed. Then, a minimax test ϕ_n: $S^n \to [0,1]$ can be constructed as*

$$\phi_n(s^n) = \begin{cases} 1 & if \quad \Lambda_n(s^n) > C_n \\ \gamma_n & if \quad \Lambda_n(s^n) = C_n \\ 0 & if \quad \Lambda_n(s^n) < C_n, \end{cases} \quad with \quad \Lambda_n(s^n) = \prod_{i=1}^{n} \Lambda_i, \qquad (32)$$

where $\Lambda_i \in dQ_1/dQ_0$, $(Q_0, Q_1) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_1}$ is the LFP in (S, U, Θ, G, m), and C_n and γ_n are chosen so that $E_{Q_0^n}[\phi_n(s^n)] = \alpha$.

The LFP consisting of the product measures in Theorem 3.1 provides an important link through which we may connect the incomplete model to standard frameworks. Below, we demonstrate this by studying the large-sample approximations and asymptotic local power of the tests. While both

---

[25]This result does not follow from Corollary 4.2 of HS, who assumed that a sample is independently distributed (p. 258).

issues can be analyzed without assuming identically distributed latent variables, we maintain this assumption to simplify the notation and our analysis.[26]

Gaussian Approximation

One of the consequences of the product structure is that the upper rejection probability of ϕ_n admits a Gaussian approximation in large samples. For ease of exposition, we assume that the u_i's are i.i.d., which in turn implies that Q_0^n is an i.i.d. law from Corollary 3.1. Hence, properly normalized sample averages follow classical limit theorems under this law. We use this insight to obtain an asymptotically valid critical value. Since Q_0^n is the least favorable, the asymptotic size is controlled under any distribution under the null hypothesis.

Let z_α be the $1 - \alpha$ quantile of the standard normal distribution and let $\Lambda \in dQ_1/dQ_0$. For each n, let

$$C_n^* \equiv \exp\left(n\mu_{Q_0} + \sqrt{n}z_\alpha\sigma_{Q_0}\right), \tag{33}$$

where $\mu_{Q_0} \equiv E_{Q_0}[\ln \Lambda(s)]$, and $\sigma_{Q_0}^2 \equiv \mathrm{Var}_{Q_0}(\ln \Lambda(s))$. Observe that μ_{Q_0} and σ_{Q_0} depend only on the LFP but not on the unknown DGP. Once the LFP is found, computing μ_{Q_0} and σ_{Q_0} is straightforward because Q_0 is a discrete distribution and Λ is known. The critical value in (33) is constructed in such a way that the following convergence holds:

$$\sup_{P^n \in \mathcal{P}_{\theta_0}^n} P^n\left(\Lambda_n(s^n) > C_n^*\right) = Q_0^n\left(\Lambda_n(s^n) > C_n^*\right) \to \Pr\left(Z > z_\alpha\right) = \alpha, \tag{34}$$

where Z is a standard normal random variable. This critical value is computed without any resampling or simulation and therefore can be done so easily. Despite its simplicity, it has the advantage of being asymptotically valid even if the true DGP is highly heterogeneous and dependent.

Let ϕ_n^* be a test that rejects the null hypothesis if and only if $\Lambda_n > C_n^*$. The following proposition then follows.

Proposition 3.1. *Suppose Assumption 3.1 holds and that $0 \leq \sigma_{Q_0}^2 < \infty$. Then, the test controls the asymptotic size:*

$$\limsup_{n \to \infty} \sup_{P \in \mathcal{P}_{\theta_0}^n} E_P[\phi_n^*(s^n)] \leq \alpha. \tag{35}$$

Furthermore, (35) holds with equality when $\sigma_{Q_0}^2 > 0$.

---

[26]If u_i is not identically distributed, one should invoke a central limit theorem for independent and not identically distributed (i.n.i.d.) sequences under Q_0^n (e.g., White, 2001) to obtian a Gaussian approximation. Similarly, for local approximations of experiments with an i.n.i.d. sequence, Rieder (1994) (Section 2.3) provides a general framework, which can be applied to the sequence $\{Q_{\theta_n,\xi,h}^n\}$ defined below.

3.4 Asymptotic Local Power

Building on Theorem 3.1, we analyze the asymptotic local power of the tests. In what follows, suppose that Θ is a subset of Euclidean space $\mathbb{R}^d$ and let $\varphi : \Theta \to \mathbb{R}$ be a continuously differentiable function with gradient $\dot{\varphi}_\theta : \mathbb{R}^d \to \mathbb{R}$. Consider the following hypotheses:

$$H_0 : \varphi(\theta) \leq 0, \ v.s. \ H_1 : \varphi(\theta) > 0. \tag{36}$$

Various hypotheses of empirical interest can be formulated in this way. Our goal here is to characterize the upper envelope of the lower power of the tests for (36) in an asymptotic framework. Theorem 3.1 serves as a building block for this purpose, as it allows us to embed our problem into a more standard one. In this section, we assume that $u_i, i = 1, \ldots, n$ are i.i.d. throughout.

We consider localized experiments. Let $\theta_0 \in \Theta$ be a parameter such that $\varphi(\theta_0) = 0$ and let $\{\theta_{n,\xi,h}\}, (\xi, h) \in \mathbb{R}^d \times \mathbb{R}^d$ be a sequence of alternative parameter values, which we specify below. We call ξ a *fixed shift* and h a *local parameter*. The sequence of parameters induces a sequence of belief functions $\nu^n_{\theta_{n,\xi,h}}$. Suppose that for each n and (ξ, h), the conditions of Theorem 3.1 hold for $\nu^n_{\theta_0}$ and $\nu^n_{\theta_{n,\xi,h}}$. Then, there exists an LFP $(Q^n_0, Q^n_1) \in \mathcal{P}^n_{\theta_0} \times \mathcal{P}^n_{\theta_{n,\xi,h}}$. If there exists model $\theta \mapsto \mathsf{Q}_\theta$ indexed by θ defined on a neighborhood of θ_0 such that $Q^n_0 = \mathsf{Q}^n_{\theta_0}$ and $Q^n_1 = \mathsf{Q}^n_{\theta_{n,\xi,h}}$ for all n, one may consider the following sequence of experiments:

$$\mathcal{E}_n = \left(S^n, \Sigma_{S^n}, \mathsf{Q}^n_{\theta_{n,\xi,h}} : h \in \mathbb{R}^d \right). \tag{37}$$

In this hypothetical environment, observations are generated using a sequence of probability laws that are the least favorable for testing θ_0 against $\theta_{n,\xi,h}$. The fact that the experiments are characterized by probabilities instead of capacities allows us to employ standard asymptotic tools. In particular, we employ the *limits of experiments* argument in the style of Le Cam (1972, 1986). Heuristically, if one wants to obtain an asymptotic upper envelope of $\pi_{\theta_{n,\xi,h}}$, one may consider the least favorable sequence of DGPs $\{\mathsf{Q}^n_{\theta_{n,\xi,h}}\}$ for power maximization. It turns out that the lower power of any test can be matched by a power function of a limit experiment, which is often more straightforward to analyze. The limit experiment can then be used to derive an asymptotic power envelope.

While the argument above suggests that we may use the standard limits of experiments framework, a few non-standard features arise. First, the underlying model may not satisfy the differentiability in quadratic mean condition, which is sufficient for the LAN of the experiments. Instead, the model is typically directionally differentiable (in the L^2 sense) and satisfies the LAN property separately on a finite number of convex cones that partition the local parameter space, which requires us to consider sub-experiments of (37). Second, some alternatives are not robustly testable. Hence, to conduct a meaningful power analysis, one needs to construct local alternatives with care.[27]

---

[27]We modify the definition of the local alternative so that the sequences of LFPs $\{Q^n_{\theta_0}\}$ and $\{Q^n_{\theta_{n,\xi,h}}\}$ are contiguous.

3.4.1 Asymptotic Power against Robustly Testable Local Alternatives

Recall that in Example 1, setting the strategic interaction effects to $\theta_1 < 0$ made $\mathcal{P}_{\theta_0}$ and $\mathcal{P}_{\theta_1}$ disjoint no matter how small the deviation from the null $\theta_0 = 0$ was. Below, we start with a relatively simple setting in which $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_0 + h/\sqrt{n}$ induce sets of distributions that are disjoint for any n. In this setting, it suffices to index local alternatives using h only, and hence we use $\theta_{n,h}$ instead of $\theta_{n,\xi,h}$.[28]

We define the notion of differentiability below. For this, let L_P^2 denote the set of square integrable functions on S with respect to measure P. For each $f, g \in L_P^2$, we then let $\langle f, g \rangle_{L_P^2}$ and $\|f\|_{L_P^2}$ denote the L^2 inner product and L^2 norm, respectively. Consider a parametric family of distributions $\{P_\theta, \theta \in V\}$, where V is an open subset of Θ.

Definition 3.1. *Let $\theta \mapsto P_\theta$ be a model such that P_θ is absolutely continuous with respect to a σ-finite measure μ on S. The model is said to be L^2 differentiable at $\theta \in V$ tangentially to set $\mathcal{T} \subset \mathbb{R}^d$ if there exists a square integrable (w.r.t. P_θ) function $\dot{\ell}_\theta : S \to \mathbb{R}^d$ such that, for every $h \in \mathcal{T}$,*

$$\left\| p_{\theta+\tau h}^{1/2} - p_\theta^{1/2} \left(1 + \frac{1}{2} \tau h' \dot{\ell}_\theta \right) \right\|_{L_\mu^2} = o(\tau), \tag{38}$$

as $\tau \downarrow 0$, where $p_\theta = dP_\theta/d\mu$ for all θ.

This is the L^2 differentiability of the square root density commonly used in the literature. If one could embed the LFPs into an L^2 differentiable model $\theta \mapsto \mathsf{Q}_\theta$ with $\mathcal{T}$ being a suitable limit of the local parameter space $\{h \in \sqrt{n}(\Theta - \theta_0) : \varphi(\theta_0 + h/\sqrt{n}) > 0\}$, asymptotic local power can be analyzed in a standard way. However, as we show below, the model is often only directionally differentiable. That is, the form of the derivative, $\dot{\ell}_\theta$, varies across the subsets of $\mathcal{T}$ (see the discussions below). The following high-level assumption states this formally.[29] For this, let $\overset{P^n}{\leadsto}$ denote weak convergence under the sequence $\{P^n\}$ of distributions. Let $\mathbb{C}(0, \epsilon)$ denote an open cube centered on the origin with edges of length 2ϵ. A set $\Gamma \subset \mathbb{R}^d$ is said to be locally equal to set $\Upsilon \subset \mathbb{R}^d$ if $\Gamma \cap \mathbb{C}(0, \epsilon) = \Upsilon \cap \mathbb{C}(0, \epsilon)$ for some $\epsilon > 0$ (Andrews, 1999).

Assumption 3.2 (Local parameter cones and L^2 directional differentiability)**.** *(i) Set $\{\xi \in \Theta - \theta_0 : \varphi(\theta_0 + \xi) > 0\}$ is locally equal to convex cone $\mathcal{T}(\theta_0)$; (ii) There exists set $\mathbb{J}$ and a collection of convex cones (containing 0) $\{\mathcal{T}_j(\theta_0), j \in \mathbb{J}\}$ such that $\mathcal{T}_j(\theta_0) \cap \mathcal{T}_{j'}(\theta_0) = \{0\}, \forall j \neq j'$ and $\bigcup_j \mathcal{T}_j(\theta_0) = \mathcal{T}(\theta_0)$; (iii) For each $j \in \mathbb{J}$, there exists model $\theta \mapsto \mathsf{Q}_{j,\theta}$ defined on a neighborhood of θ_0 such that an LFP $(Q_{0,j,\tau h}, Q_{1,j,\tau h}) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_0 + \tau h}$ satisfies*

$$Q_{0,j,\tau h} = \mathsf{Q}_{j,\theta_0}, \quad and \quad Q_{1,j,\tau h} = \mathsf{Q}_{j,\theta_0+\tau h}, \tag{39}$$

---

[28]We elaborate on the role of ξ in the next section.

[29]For the results that follow, it suffices that a model satisfies Assumption 3.2 for an LFP. The LFPs are unique up to the Radon–Nikodym derivative (HS, 1973), and thus they all lead to the same quadratic expansion of the log-LR process. While one could alternatively take (40) as a high-level condition, we do not do so because Assumption 3.2 is often easier to check. A similar comment applies to Assumption 3.3.

for all $\tau \in (0, \bar{\tau}]$ for some $\bar{\tau} > 0$, and $\theta \mapsto Q_{j,\theta}$ is L^2 differentiable at θ_0 tangentially to $\mathcal{T}_j(\theta_0)$.

The assumption above imposes a regularity condition on the LFP for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_0 + \tau h$ for $\tau > 0$. While θ_0 is fixed, both the least favorable distributions (under H_0 and H_1) may depend on deviation τh. Hence, we index the LFP using τh. The restrictions in (39) require that the least favorable distribution $Q_{0,j,\tau h}$ under H_0 remains the same for all sufficiently small τ, and this is given by point Q_{j,θ_0} in the L^2 differentiable model $\theta \mapsto Q_{j,\theta}$. As we demonstrate below through examples, Assumption 3.2 (and similarly Assumption 3.3) can be checked by analyzing the LFP. In all our examples, the cardinality of $\mathbb{J}$ is finite. Appendix E also provides the primitive conditions that ensure the key condition (L^2 differentiability) (see Assumption E.1, Proposition E.1, and Corollary E.1).

Below, we let $\dot{\ell}_{j,\theta_0}$ denote the L^2 derivative defined for $h \in \mathcal{T}_j(\theta_0)$. Under Assumption 3.2, one may expand the log-LR for every $h \in \mathcal{T}_j(\theta_0)$ as follows:

$$\ln \frac{dQ^n_{j,\theta_0+h/\sqrt{n}}}{dQ^n_{j,\theta_0}} = h'\Delta_{j,n} - \frac{1}{2}h'C_j h + o_{Q^n_{j,\theta_0}}(1), \tag{40}$$

where $\Delta_{j,n} \overset{Q^n_{j,\theta_0}}{\rightsquigarrow} \Delta_j \sim N(0, C_j)$ and $C_j = E_{Q_{j,\theta_0}}[\dot{\ell}_{j,\theta_0}\dot{\ell}'_{j,\theta_0}]$. In what follows, we call $\Delta_{j,n}$ the central sequence (or normalized score) and C_j the information matrix. Consider the following sub-experiments:

$$\mathcal{E}_{j,n} \equiv \left(S^n, \Sigma_{S^n}, Q^n_{j,\theta_0+h/\sqrt{n}} : h \in \mathcal{T}_j(\theta_0) \right), \ j \in \mathbb{J}. \tag{41}$$

When Assumption 3.2 holds, for each j, the limit of the experiments (as $n \to \infty$) is

$$\mathcal{E}_j = \left(\mathbb{R}^d, \Sigma_{\mathbb{R}^d}, N(C_j h, C_j) : h \in \mathcal{T}_j(\theta_0) \right), \tag{42}$$

which is also equivalent to $(\mathbb{R}^d, \Sigma_{\mathbb{R}^d}, N(h, C_j^{-1}) : h \in \mathcal{T}_j(\theta_0))$ if C_j is non-singular (see Van der Vaart, 2000, Ch. 9). In other words, the experiment is equivalent to the one in which the researcher observes a single normal random vector whose mean and variance are $h \in \mathcal{T}_j(\theta_0)$ and C_j^{-1}, respectively. The asymptotic local (lower) power of a test is then bounded from above by the corresponding power in the limit experiment. The power envelope can be derived by considering the highest possible power for testing $H_0 : \dot{\varphi}_{\theta_0} h \leq 0$ against $H_1 : \dot{\varphi}_{\theta_0} h > 0$ at level-α.

As is well known, these limit experiments are Gaussian shift experiments defined on suitable subsets of $\mathbb{R}^d$. If Assumption 3.2 holds with $\mathbb{J} = \{I\}$ and $\mathcal{T}_{I,\theta_0} = \mathbb{R}^d$, we obtain the LAN (Le Cam, 1986). Assumption 3.2 slightly extends the LAN, and this allows us to consider experiments defined separately on the subsets (cones) of the local parameter space. This extension is motivated by the fact that central sequence $\Delta_{j,n}$ and information matrix C_j may differ across the local parameter cones. This is because as h varies (and hence $\nu_{\theta_{n,h}}$ varies), the LFP defined through the convex

program in (18) may change in a non-differentiable (but directionally differentiable) way.[30] This may lead to distinct central sequence and information matrix pairs across the cones. Owing to this non-standard feature, our results below concern asymptotically optimal statistical decisions when the underlying model may only be directionally differentiable (in the L^2 sense). This complements the recent developments on statistical inference and decisions in non-standard models in which the parameters of interest are only directionally differentiable, whereas the underlying model is regular (Hirano and Porter, 2012; Song, 2014; Fang, 2014; Fang and Santos, 2018; Hong and Li, 2018).

To characterize the power envelope, we define the tangent cone of the score functions and efficient influence function of φ. For each $j \in \mathbb{J}$, let

$$\mathcal{G}_{j,\theta_0} \equiv \{g \in L^2_{\mathsf{Q}_{j,\theta_0}}(S) : g = h'\dot{\ell}_{j,\theta_0}, \ h \in \mathcal{T}_j(\theta_0)\}. \tag{43}$$

We call the set above the *tangent cone* of the model. The influence curve $\varrho_j \in L^2_{\mathsf{Q}_{j,\theta_0}}(S)$ of φ is such that, for any $g \in \mathcal{G}_{j,\theta_0}$,

$$\left|\varphi(\theta_0 + \tau h) - \varphi(\theta_0) - \tau\langle\varrho_j, g\rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}}\right| = o(\tau), \tag{44}$$

as $\tau \downarrow 0$. The *efficient influence function* (or canonical gradient) $\tilde{\varrho}_j$ of φ is then defined as the projection of ϱ_j on the closure of $\mathcal{G}_{j,\theta_0}$ (which is often called the *tangent set*).

The following theorem characterizes the asymptotic upper bound of the lower power and provides a test that achieves the bound (for a given cone).

Theorem 3.2. *Suppose Assumption 3.2 holds. Let $j \in \mathbb{J}$. Suppose that φ is such that $\varphi(\theta_0) = 0$ and has influence curve ϱ_j. Let ϕ_n be a level-α test for $H_0 : \varphi(\theta) \leq 0$ against $H_1 : \varphi(\theta) > 0$ and $\pi_{n,\theta}(\phi_n)$ be its lower power under ν_θ^n. Then, for any $h \in \mathcal{T}_j(\theta_0)$,*

$$\limsup_{n\to\infty} \pi_{n,\theta_0+h/\sqrt{n}}(\phi_n) \leq 1 - \Phi\left(z_\alpha - \frac{\langle\varrho_j, h'\dot{\ell}_{j,\theta_0}\rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}}}{\|\tilde{\varrho}_j\|_{L^2_{\mathsf{Q}_{j,\theta_0}}}}\right). \tag{45}$$

Let $T_{j,n}$ be a statistic such that

$$T_{j,n} = \frac{n^{-1/2}\sum_{i=1}^n \tilde{\varrho}_j(s_i)}{\|\tilde{\varrho}_j\|_{L^2_{\mathsf{Q}_{j,\theta_0}}}} + o_{\mathsf{Q}_{j,\theta_0}^n}(1). \tag{46}$$

Let $\phi_{j,n}^*$ be a test that rejects the null hypothesis iff $T_{j,n} \geq z_\alpha$. Then, the test is of asymptotically

---

[30]See Shapiro (1988), Dempe (1993), and the references therein for the directional differentiability of solutions to parametric convex programs. Our primitive conditions (see Appendix E.1) for Assumptions 3.2 and 3.3 are based on Shapiro (1988).

level-α and, for any $h \in \mathcal{T}_j(\theta_0)$,

$$\lim_{n \to \infty} \pi_{n, \theta_0 + h/\sqrt{n}}(\phi_{j,n}^*) = 1 - \Phi\left(z_\alpha - \frac{\langle \varrho_j, h'\dot{\ell}_{j,\theta_0}\rangle_{L^2_{Q_{j,\theta_0}}}}{\|\tilde{\varrho}_j\|_{L^2_{Q_{j,\theta_0}}}}\right). \tag{47}$$

The asymptotic power envelope in (45) coincides with that for the one-sided test $\dot{\varphi}_{\theta_0} h \leq 0$ against $\dot{\varphi}_{\theta_0} h > 0$ in the Gaussian shift experiment. The theorem also implies that the test based on the (rescaled) efficient influence function achieves the power envelope toward alternatives with $h \in \mathcal{T}_j(\theta_0)$.[31] Therefore, the key factor is to find the efficient influence function.

We next revisit Example 1.

Example 1 (Binary response game (continued)). Let $\Theta = \{\theta \in \mathbb{R}^2 : \theta^{(1)} \leq 0, \theta^{(2)} \leq 0\}$. Consider testing the hypothesis as in (36) with $\varphi(\theta) = p'\theta = -\theta^{(1)} - \theta^{(2)}$, where $p = (-1, -1)'$. To localize the experiment at $\theta_0 = (0, 0)'$, we start by describing the LFPs (and minimax tests) for all the parameter values under consideration. Let $\Theta_1 \equiv \{\theta \in \Theta : \varphi(\theta) > 0\}$ be the set of parameter values under the alternative.

Let $\mathbb{J} = \{\mathrm{I}, \mathrm{II}, \mathrm{III}\}$. Then, define the following parameter sets:

$$\Theta_{\mathrm{I}} \equiv \left\{\theta_1 \in \Theta_1 : \Phi(\theta_1^{(1)})(1 - \Phi(\theta_1^{(2)})) \leq \frac{1}{4}, \Phi(\theta_1^{(2)})(1 - \Phi(\theta_1^{(1)})) \leq \frac{1}{4}\right\} \tag{48}$$

$$\Theta_{\mathrm{II}} \equiv \left\{\theta_1 \in \Theta_1 : \Phi(\theta_1^{(1)})(1 - \Phi(\theta_1^{(2)})) > \frac{1}{4}\right\} \tag{49}$$

$$\Theta_{\mathrm{III}} \equiv \left\{\theta_1 \in \Theta_1 : \Phi(\theta_1^{(2)})(1 - \Phi(\theta_1^{(1)})) > \frac{1}{4}\right\}. \tag{50}$$

Figure 2 (left) shows these sets. In addition to the case in which $\theta_1^{(1)}$ and $\theta_1^{(2)}$ are both strictly negative and comparable in magnitude (as discussed in Section 3.2), we consider two other cases here. Across all subcases, the density of Q_0 is $(q_0(0,0), q_0(1,1), q_0(1,0), q_0(0,1)) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. However, Q_1 and the minimax test vary (Table 1).[32] If $\theta_1^{(1)}$ is substantially smaller than $\theta_1^{(2)}$ (i.e., $\theta_1 \in \Theta_{\mathrm{II}}$), a larger mass moves to the region over which $s = (1, 0)$ is predicted than to the region over which $s = (0, 1)$ is predicted (recall Figure 1). The minimax test then rejects H_0 with a positive probability ($\gamma = 4\alpha$) only when $s = (1, 0)$ is observed. A similar comment applies to the case in which $\theta \in \Theta_{\mathrm{III}}$.

Consider a local alternative $\theta_{n,h} = \theta_0 + h/\sqrt{n}$ such that $\varphi(\theta_{n,h}) > 0$. Suppose that $\theta_{n,h} \in \Theta_j$ for some $j \in \mathbb{J}$ for all sufficiently large n. Then, the local parameter must belong to one of the

---

[31]The power envelope can be expressed using $\dot{\varphi} h$ and C_j as in Theorem 15.4 in Van der Vaart (2000) if C_j is non-singular and $\mathcal{G}_{j,\theta_0}$ is a linear subspace. The description above is slightly more general to handle cases in which C_j may be singular and $\mathcal{G}_{j,\theta_0}$ is a convex cone (Rieder, 2014).

[32]Proposition D.1 in Appendix D provide these results formally.

Table 1: Q_1 and minimax tests

Sets	$q_1(0,0)$	$q_1(1,1)$	$q_1(1,0)$	$q_1(0,1)$		Minimax test
			Q_1			
Θ_{I}	$\frac{1}{4}$	$\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})$	$\frac{3}{8} - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})/2$	$\frac{3}{8} - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})/2$	$\phi(s) = \begin{cases} 2\alpha & s \in \{(1,0),(0,1)\} \\ 0 & \text{otherwise} \end{cases}$	
Θ_{II}	$\frac{1}{4}$	$\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})$	$\frac{1}{4} - \Phi(\theta_1^{(1)})(\Phi(\theta_1^{(2)}) - \frac{1}{2})$	$\frac{1}{2}(1 - \Phi(\theta_1^{(1)}))$	$\phi(s) = \begin{cases} 4\alpha & s = (1,0) \\ 0 & \text{otherwise} \end{cases}$	
Θ_{III}	$\frac{1}{4}$	$\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})$	$\frac{1}{2}(1 - \Phi(\theta_1^{(2)}))$	$\frac{1}{4} - \Phi(\theta_1^{(2)})(\Phi(\theta_1^{(1)}) - \frac{1}{2})$	$\phi(s) = \begin{cases} 4\alpha & s = (0,1) \\ 0 & \text{otherwise} \end{cases}$	

Figure 2: Subcases in Proposition 1 and local parameter cones
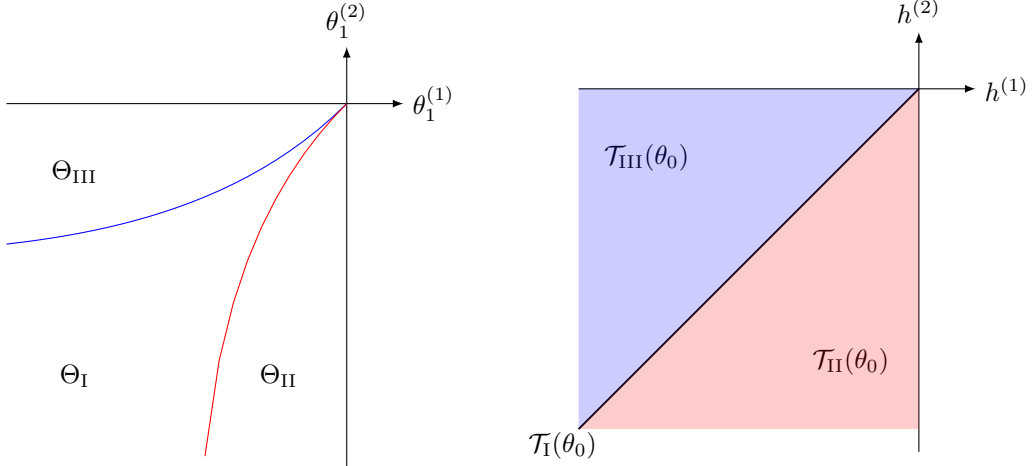
Left panel: Parameter space and subcases;
Right panel: Local parameter cones at $\theta_0 = (0,0)'$: $\mathcal{T}_{\mathrm{I},\theta_0}$ (solid line), $\mathcal{T}_{\mathrm{II}}(\theta_0)$ (red), and $\mathcal{T}_{\mathrm{III},\theta_0}$ (blue).

following cones:

$$\mathcal{T}_{\mathrm{I}}(\theta_0) = \{h \in \mathbb{R}^2 : h = (\bar{h}, \bar{h})', \bar{h} \in (-\infty, 0]\} \tag{51}$$

$$\mathcal{T}_{\mathrm{II}}(\theta_0) = \{h \in \mathbb{R}^2 : h = (h^{(1)}, h^{(2)})', -\infty < h^{(2)} < h^{(1)} \leq 0\} \tag{52}$$

$$\mathcal{T}_{\mathrm{III}}(\theta_0) = \{h \in \mathbb{R}^2 : h = (h^{(1)}, h^{(2)})', -\infty < h^{(1)} < h^{(2)} \leq 0\}. \tag{53}$$

These cones are localized versions of the parameter subsets (Θ_{I}-Θ_{III}), as shown in Figure 2 (right).

Below, as an example, we consider the case $\theta_{n,h} \in \Theta_{\mathrm{II}}$ for all sufficiently large n. Since Q_1 is as shown in Table 1, one may embed the LFP into model $\theta \mapsto \mathsf{Q}_{\mathrm{II},\theta}$ whose density is

$$(\mathsf{q}_{\mathrm{II},\theta}(0,0), \mathsf{q}_{\mathrm{II},\theta}(1,1), \mathsf{q}_{\mathrm{II},\theta}(1,0), \mathsf{q}_{\mathrm{II},\theta}(0,1))$$
$$= \left(\frac{1}{4}, \Phi(\theta^{(1)})\Phi(\theta^{(2)}), \frac{1}{4} - \Phi(\theta^{(1)})(\Phi(\theta^{(2)}) - \frac{1}{2}), \frac{1}{2}(1 - \Phi(\theta^{(1)})) \right). \tag{54}$$

Then, the L^2 derivative of the model for $h \in \mathcal{T}_{\mathrm{II}}(\theta_0)$ is

$$\dot{\ell}_{\mathrm{II},\theta_0}(s) = 1\{s = (1,1)\} \begin{pmatrix} \frac{2}{\sqrt{2\pi}} \\ \frac{2}{\sqrt{2\pi}} \end{pmatrix} + 1\{s = (1,0)\} \begin{pmatrix} 0 \\ \frac{-2}{\sqrt{2\pi}} \end{pmatrix} + 1\{s = (0,1)\} \begin{pmatrix} \frac{-2}{\sqrt{2\pi}} \\ 0 \end{pmatrix}. \tag{55}$$

The log-likelihood function can be expanded as in (40) with $\Delta_{\mathrm{II},n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\ell}_{\mathrm{II},\theta_0}$ and the information matrix: $C_{\mathrm{II}} = \begin{pmatrix} \frac{1}{\pi} & \frac{1}{2\pi} \\ \frac{1}{2\pi} & \frac{1}{\pi} \end{pmatrix}$. Therefore, the limit experiment is

$$\mathcal{E}_{\mathrm{II}} = \left(\mathbb{R}^2, \Sigma_{\mathbb{R}^2}, N(h, C_{\mathrm{II}}^{-1}) : h \in \mathcal{T}_{\mathrm{II}}(\theta_0) \right), \tag{56}$$

in which one observes a single random vector $Z \sim N(h, C_{\mathrm{II}}^{-1})$. From Theorem 3.2, it then suffices to consider the power for testing $H_0 : p'h = 0$ against $p'h > 0$ in this simple experiment. With $p = (-1, -1)'$, it can be shown that the power envelope is

$$\limsup_{n \to \infty} \pi_{n,\theta_0 + h/\sqrt{n}}(\phi_n) \leq 1 - \Phi\left(z_\alpha - \frac{-h^{(1)} - h^{(2)}}{\sqrt{4\pi/3}} \right). \tag{57}$$

This bound can be achieved using a test that rejects the null when the following statistic exceeds z_α:

$$T_{\mathrm{II},n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[-\frac{4}{3}\sqrt{\frac{3}{2}} 1\{s_i = (1,1)\} + \frac{2}{3}\sqrt{\frac{3}{2}} (1\{s_i = (1,0)\} + 1\{s_i = (0,1)\}) \right]. \tag{58}$$

Heuristically, this means that the test rejects H_0 when one observes $(1,0)$ or $(0,1)$ frequently relative to $(1,1)$. In general, the power envelope and optimal test depend on local cone $\mathcal{T}_j(\theta_0)$ and the functional of interest. Appendix D describes the efficient influence functions for this example.

Remark 3.3. The optimal test above compares the relative frequencies of two events $\{(1,1)\}$ and $\{(1,0),(0,1)\}$. It therefore only uses the information on the *number of entrants* (i.e., duopoly

v.s. monopoly) in each market, which is the feature (or transformation) of the outcome s used in Bresnahan and Reiss (1990, 1991) and Berry (1992). For testing competing hypotheses on $\varphi(\theta) = -\theta^{(1)} - \theta^{(2)}$, using such a transformed outcome indeed leads to an optimal test. However, our theory suggests that the choice of transformation depends, in general, on the functional of interest (φ) and the direction of alternatives ($\mathcal{T}_j(\theta_0)$). For example, with $\varphi(\theta) = -\theta^{(1)} - 2\theta^{(2)}$ and $h \in \mathcal{T}_{\text{II}}(\theta_0)$, the optimal test compares the relative frequencies of $\{(1, 1)\}$ and $\{(1, 0)\}$.

Remark 3.4. Theorem 3.2 applies to each cone $\mathcal{T}_j(\theta_0)$ and hence can be used to obtain a test that asymptotically achieves the power envelope against the alternatives in $\mathcal{T}_j(\theta_0)$. The optimal test, however, may differ by cone.[33]

3.4.2 Asymptotic Power against Shifted Local Alternatives

As discussed earlier, not all alternatives are robustly testable. Consider the following example.

Example 2 (Roy model (continued)). Suppose that the researcher wants to know if the share of individuals who have higher economic prospects in sector 0 is above a certain percentage. The parameter of interest is $\theta^{(1,0)} = m_\theta((Y_0, Y_1) = (1, 0))$, and the null and alternative hypotheses can be expressed as $H_0 : p'\theta = \theta^{(1,0)} \leq c$ and $H_1 : p'\theta > c$ for some $c \in [0, 1]$ with $p = (0, 0, 1)$. Mourifie, Henry, and Meango (2018) showed that the sharp identifying restrictions are

$$\theta^{(1,0)} \leq P(\{(1, 0)\}) \tag{59}$$

$$\theta^{(0,1)} \leq P(\{(1, 1)\}) \tag{60}$$

$$\theta^{(0,0)} = P(\{(0, 0)\}) + P(\{(0, 1)\}). \tag{61}$$

For simplicity, suppose that $\theta^{(0,0)}$ is known to be $1/6$. This implies $P(\{(1, 0)\}) + P(\{(1, 1)\}) = 5/6$, which in turn simplifies the restrictions to

$$\theta^{(1,0)} \leq P(\{(1, 0)\}) \leq \frac{5}{6} - \theta^{(0,1)} \tag{62}$$

$$\frac{1}{6} = P(\{(0, 0)\}) + P(\{(0, 1)\}). \tag{63}$$

Now, consider testing θ_0 against the alternative $\theta_1 = \theta_0 + \xi$, where $\xi = (0, \xi^{(0,1)}, \xi^{(1,0)})'$ with $\xi^{(1,0)} > 0$. As shown in Figure 3 (Alt. 1), the interval $[\theta_0^{(1,0)} + \xi^{(1,0)}, \frac{5}{6} - \theta^{(0,1)} - \xi^{(0,1)}]$ to which $P(\{(1, 0)\})$ belongs under this alternative has a non-empty intersection with the interval under the null until $\xi^{(1,0)}$ becomes sufficiently large. Indeed, $\mathcal{P}_{\theta_0 + \xi}$ becomes disjoint from $\mathcal{P}_{\theta_0}$ only when $\xi^{(1,0)} > \frac{5}{6} - \theta_0^{(0,1)} - \theta_0^{(1,0)}$ (Alt. 2).[34] This means that against any local alternative of the form

---

[33]When a functional satisfies $\dot{\varphi}_{\theta_0} = c \times (-1, -1)$ for some $c > 0$ in Example 1, the optimal test is common across the cones because the projection of the influence function onto $cl(\mathcal{G}_{j,\theta_0})$ is common across j. However, this does not hold with the other functionals.

[34]Another way to make $\mathcal{P}_{\theta_0}$ and $\mathcal{P}_{\theta_0 + \xi}$ disjoint is to shift the interval to the left in Figure 3 by a sufficiently large amount. For this, one needs $5/6 - \theta_0^{(0,1)} - \xi^{(0,1)} < \theta_0^{(1,0)}$. However, $\theta_0 + \xi$ does not satisfy $\varphi(\theta_0 + \xi) > 0$ in this case.

$\theta_0 + h/\sqrt{n}$, the lower power of any level-α test is eventually dominated (weakly) by α, which does not lead to useful comparisons.

Figure 3: Bounds on $P(\{(1,0)\})$
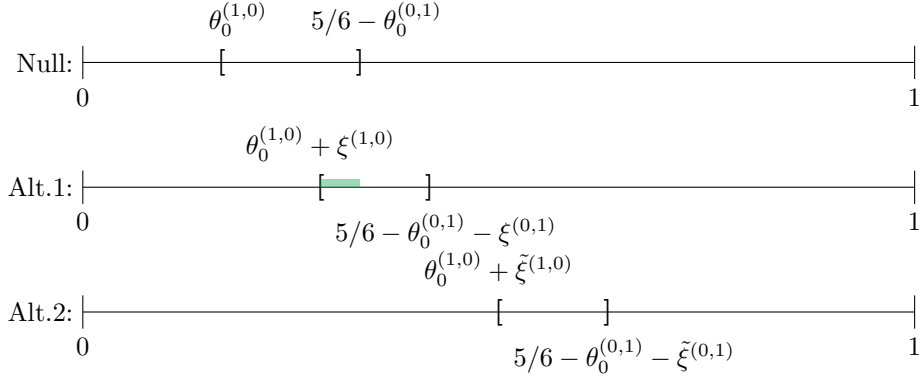
Note: The interval under Alt.1 has a non-empty intersection (in green) with the interval under the null. When $\tilde{\xi}^{(1,0)} > \frac{5}{6} - \theta_0^{(0,1)} - \theta_0^{(1,0)}$ (Alt.2), the two intervals are disjoint, i.e., $\mathcal{P}_{\theta_0+\tilde{\xi}} \cap \mathcal{P}_{\theta_0} = \emptyset$.

Given the challenge above, we conduct a local power analysis as follows. First, we *shift* θ_0 by vector ξ, which does not depend on n. Hence, at $\theta_0 + \xi$, certain local deviations can be robustly detectable. We then analyze the limit of experiments constructed from a sequence of LFPs induced by such alternatives. Specifically, the *shifted local alternative* is

$$\theta_{n,\xi,h} = \theta_0 + \xi + h/\sqrt{n}, \tag{64}$$

where $\theta_0 \in \Theta$ is such that $\varphi(\theta_0) = 0$, and ξ and h take values in the following sets:

$$\Xi_{\theta_0} \equiv \{\xi \in \Theta - \theta_0 : \varphi(\theta_0 + \xi) > 0, \mathcal{P}_{\theta_0} \cap \mathcal{P}_{\theta_0+\xi} \neq \emptyset\} \tag{65}$$

$$\mathcal{T}_n(\theta_0, \xi) \equiv \{h \in \sqrt{n}(\Theta - \theta_0 - \xi) : \mathcal{P}_{\theta_0} \cap \mathcal{P}_{\theta_0+\xi+h/\sqrt{n}} = \emptyset\}. \tag{66}$$

In other words, Ξ_{θ_0} collects the deviations for which $\theta_0 + \xi$ are not robustly testable. Set $\mathcal{T}_n(\theta_0, \xi)$ then collects local deviations that make $\theta_{n,\xi,h}$ robustly testable. Among the points in Ξ_{θ_0}, we focus on those for which $\mathcal{T}_n(\theta_0, \xi)$ is non-empty.

Under this construction, we may apply Theorem 3.1 for any $\tau > 0$ to obtain the LFP $(Q_{0,j,\tau h}, Q_{1,j,\tau h}) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_0+\xi+\tau h}$. Suppose that the following assumption holds for $\theta_0 \in \Theta$ and $\xi \in \Xi_{\theta_0}$.

Assumption 3.3 (Local parameter cones and L^2 directional differentiability). *(i) Set $\{h \in \Theta - \theta_0 - \xi : \mathcal{P}_{\theta_0} \cap \mathcal{P}_{\theta_0+\xi+h} = \emptyset\}$ is locally equal to convex cone $\mathcal{T}(\theta_0, \xi)$; (ii) There exists set $\mathbb{J}$ and a collection of convex cones (containing 0) $\{\mathcal{T}_j(\theta_0, \xi), j \in \mathbb{J}\}$ such that $\mathcal{T}_j(\theta_0, \xi) \cap \mathcal{T}_{j'}(\theta_0, \xi) = \{0\}, \forall j \neq j'$ and $\bigcup_j \mathcal{T}_j(\theta_0, \xi) = \mathcal{T}(\theta_0, \xi)$; (iii) For each $j \in \mathbb{J}$, there exists model $\vartheta \mapsto Q_{j,\vartheta}$ defined on a neighborhood of $\vartheta = 0$ such that the LFP $(Q_{0,j,\tau h}, Q_{1,j,\tau h}) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_0+\xi+\tau h}$ satisfies*

$$Q_{0,j,\tau h} = Q_{j,0}, \quad and \quad Q_{1,j,\tau h} = Q_{j,\tau h}, \tag{67}$$

28

for all $\tau \in (0, \bar{\tau}]$ for some $\bar{\tau} > 0$, and $\vartheta \mapsto \mathsf{Q}_{j,\vartheta}$ is L^2 differentiable at 0 tangentially to $\mathcal{T}_j(\theta_0, \xi)$.

In what follows, we let $\dot{\ell}_{j,0}, j \in \mathbb{J}$ denote the L^2 derivatives and focus on testing the linear hypotheses, namely $\varphi(\theta) = p'\theta - c$. Define the influence curve ϱ_j of $\varphi(\theta) = p'\theta - c$ as a square integrable function ϱ_j that satisfies, for every $h \in \mathcal{T}_j(\theta_0, \xi)$, $p'h = \langle \varrho_j, g \rangle_{L^2_{\mathsf{Q}_{j,0}}}$, where $g = h'\dot{\ell}_{j,0}$. Let $\mathcal{G}_{j,\theta_0} = \{g \in L^2_{\mathsf{Q}_{j,0}} : g = h'\dot{\ell}_{j,0}, h \in \mathcal{T}_j(\theta_0, \xi)\}$. Let $\tilde{\varrho}_j$ be the efficient influence function, which is the projection of ϱ_j to the closure of $\mathcal{G}_{j,0}$. We then obtain the following result.

Theorem 3.3. *Suppose Assumption 3.3 holds. Let $j \in \mathbb{J}$. Let $\theta_0 \in \Theta$ such that $\varphi(\theta_0) = 0$. Let ϕ_n be a level-α test for $H_0 : \varphi(\theta) \leq 0$ against $H_1 : \varphi(\theta) > 0$ and $\pi_{n,\theta}(\phi_n)$ be its lower power under ν_θ^n. Then, for any $h \in \mathcal{T}_j(\theta_0, \xi)$,*

$$\limsup_{n \to \infty} \pi_{n,\theta_{n,\xi,h}}(\phi_n) \leq 1 - \Phi\left(z_\alpha - \frac{\langle \varrho_j, h'\dot{\ell}_{j,0} \rangle_{L^2_{\mathsf{Q}_{j,0}}}}{\|\tilde{\varrho}_j\|_{L^2_{\mathsf{Q}_{j,0}}}} \right). \tag{68}$$

Let $T_{j,n}$ be a statistic such that

$$T_{j,n} = \frac{n^{-1/2} \sum_{i=1}^n \tilde{\varrho}_j(s_i)}{\|\tilde{\varrho}_j\|_{L^2_{\mathsf{Q}_{j,0}}}} + o_{\mathsf{Q}_{j,0}^n}(1). \tag{69}$$

Let $\phi_{j,n}^*$ be a test that rejects the null hypothesis iff $T_{j,n} \geq z_\alpha$. Then, the test is of level-α and, for any $h \in \mathcal{T}_{j,\theta_0}$,

$$\liminf_{n \to \infty} \pi_{n,\theta_{n,\xi,h}}(\phi_{j,n}^*) = 1 - \Phi\left(z_\alpha - \frac{\langle \varrho_j, h'\dot{\ell}_{j,0} \rangle_{L^2_{\mathsf{Q}_{j,0}}}}{\|\tilde{\varrho}_j\|_{L^2_{\mathsf{Q}_{j,0}}}} \right). \tag{70}$$

Below, we again use Example 2 to illustrate the local power analysis.

Example 2 (Roy model (continued))**.** Figure 4 shows Ξ_{θ_0} and the cones in Assumption 3.3. Since $\theta^{(0,0)} = 1/6$ is known, we can plot the parameters in a two-dimensional simplex $\{(\theta^{(1,0)}, \theta^{(0,1)}) \in [0,1]^2 : \theta^{(0,1)} + \theta^{(1,0)} \leq 5/6\}$. Here, we take $\theta_0^{(1,0)}, = 1/6, \theta^{(0,1)} = 1/2$, and test $H_0 : \theta^{(1,0)} \leq 1/6$ against $H_1 : \theta^{(1,0)} > 1/6$. This configuration implies that the alternative $\theta_0 + \xi$ is not robustly testable unless $\xi^{(1,0)} > 1/6$. The green region in Figure 4 shows the set of not robustly testable alternatives.[35]

One can see that the local parameter space $\mathcal{T}(\theta_0, \xi)$ is non-empty when ξ is a boundary point of Ξ_{θ_0} toward p. For example, at $\theta_0 + \xi_A$, the local parameter space $\mathcal{T}(\theta_0, \xi_A)$ is given by a half space $\{h = (h^{(0,0)}, h^{(0,1)}, h^{(1,0)}) : h^{(0,0)} = 0, h^{(1,0)} > 0\}$. At $\theta_0 + \xi_B$, the local parameter space is $\{h = (h^{(0,0)}, h^{(0,1)}, h^{(1,0)}) : h^{(0,0)} = 0, h^{(1,0)} > 0, h^{(0,1)} + h^{(1,0)} = 0\}$. At each boundary point, one can then conduct a local power analysis.

As an illustration, take $\xi = \xi_A \equiv (0, 1/6, -1/6)'$. It turns out that in this setting, it suffices to consider a single convex cone $\mathcal{T}_{\mathrm{I}}(\theta_0, \xi_A) \equiv \mathcal{T}(\theta_0, \xi_A)$. Let $\theta_{n,\xi_A,h} = \theta_0 + \xi_A + h/\sqrt{n}$ with $h \in \mathcal{T}_{\mathrm{I}}(\theta_0, \xi_A)$

---

[35]Once translated by $-\theta_0$, this region represents the set Ξ_{θ_0} of fixed shifts.

Figure 4: Set of not robustly testable alternatives and local parameter spaces
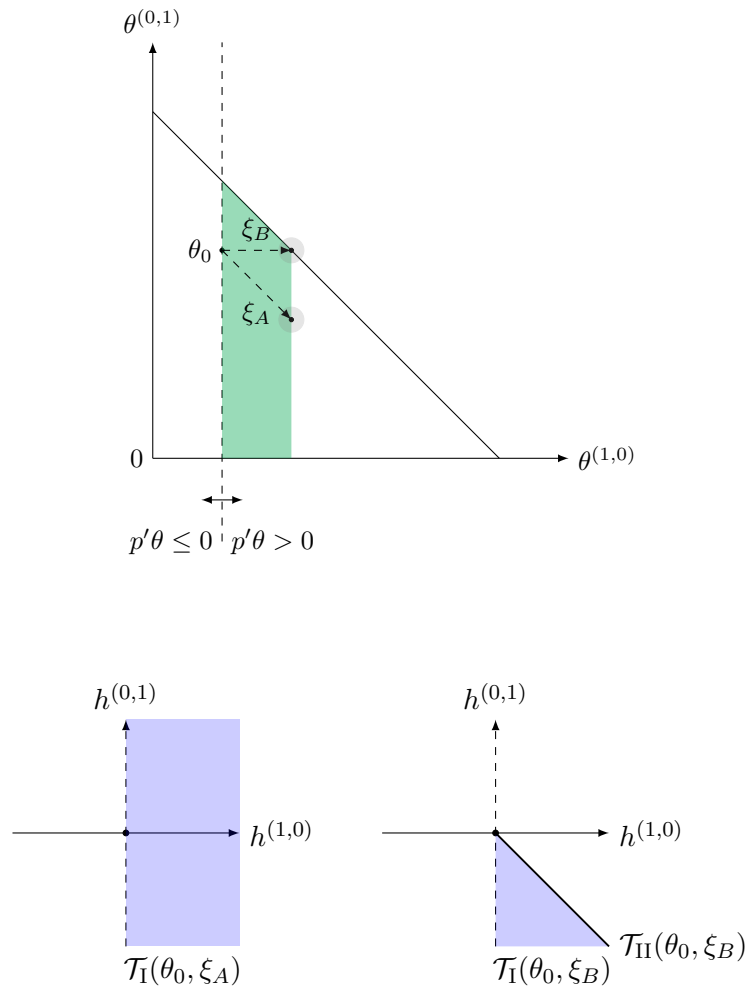
Note: $\mathcal{T}_{\mathrm{I}}(\theta_0, \xi_A)$ (half space) coincides with the local parameter space at $\theta_0 + \xi_A$.
$\mathcal{T}_{\mathrm{I}}(\theta_0, \xi_B)$ (shaded area not including the solid line) and $\mathcal{T}_{\mathrm{II}}(\theta_0, \xi_B)$ (solid line) are tangent cones at $\theta_0 + \xi_B$.

be our shifted local alternative.[36] Hence, the LFP for $\mathcal{P}_{\theta_0}$ and $\mathcal{P}_{\theta_{n,\xi_A},h}$ has the densities

$$(q_0(0,0), q_0(0,1), q_0(1,0), q_0(1,1)) = (\frac{1}{12}, \frac{1}{12}, \frac{1}{3}, \frac{1}{2}) \tag{71}$$

$$(q_1(0,0), q_1(1,1), q_1(1,0), q_1(1,1)) = (\frac{1}{12}, \frac{1}{12}, \frac{1}{3} + \frac{h^{(1,0)}}{\sqrt{n}}, \frac{1}{2} - \frac{h^{(1,0)}}{\sqrt{n}}). \tag{72}$$

These LFPs can be embedded into model $\vartheta \mapsto \mathsf{Q}_{\mathrm{I},\vartheta}$ whose density $\mathsf{q}_{\mathrm{I},\vartheta}$ is given by

$$(\mathsf{q}_{\mathrm{I},\vartheta}(0,0), \mathsf{q}_{\mathrm{I},\vartheta}(1,1), \mathsf{q}_{\mathrm{I},\vartheta}(1,0), \mathsf{q}_{\mathrm{I},\vartheta}(1,1)) = (\frac{1}{12}, \frac{1}{12}, \frac{1}{3} + \vartheta^{(1,0)}, \frac{1}{2} - \vartheta^{(1,0)}). \tag{73}$$

This model is L^2 directionally differentiable (at 0) tangentially to $\mathcal{T}_{\mathrm{I}}(\theta_0, \xi_A)$ with the following directional derivative:

$$\dot{\ell}_{\mathrm{I},0}(s) = 1\{s_i = (1,0)\} \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} + 1\{s_i = (1,1)\} \begin{pmatrix} 0 \\ 0 \\ -2 \end{pmatrix}. \tag{74}$$

The log-likelihood function can then be expanded as in (40) with $\Delta_{\mathrm{I},n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_{\mathrm{I},0}$ and information matrix $C_{\mathrm{I}} = \begin{pmatrix} \mathbf{0}_2 & 0 \\ 0 & 5 \end{pmatrix}$, which is singular, where $\mathbf{0}_2$ is a two-by-two matrix of zeros. The limit experiment is

$$\mathcal{E}_{\mathrm{I}} = \Big(\mathbb{R}^3, \Sigma_{\mathbb{R}^3}, N(C_{\mathrm{I}}h, C_{\mathrm{I}}) : h \in \mathcal{T}_{\mathrm{I}}(\theta_0, \xi_A) \Big), \tag{75}$$

in which one observes a single random vector $Z \sim N(C_{\mathrm{I}}h, C_{\mathrm{I}})$. Here, the information matrix is not full rank. This experiment essentially involves a single normal random variable with mean $5h^{(1,0)}$ and variance 5. With $p = (0,0,1)'$, the efficient influence function is

$$\tilde{\varrho}_{\mathrm{I}}(s) = \frac{3}{5} 1\{s = (1,0)\} - \frac{2}{5} 1\{s = (1,1)\}. \tag{76}$$

Theorem 3.2 then implies, for any level-α test ϕ_n,

$$\limsup_{n \to \infty} \pi_{n,\theta_{n,\xi_A},h}(\phi_n) \leq 1 - \Phi\Big(z_\alpha - \sqrt{5}h^{(1,0)} \Big). \tag{77}$$

This bound can be achieved using a test that rejects the null when the following statistic exceeds z_α:

$$T_{\mathrm{I},n} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big[\frac{3}{\sqrt{5}} 1\{s_i = (1,0)\} - \frac{2}{\sqrt{5}} 1\{s_i = (1,1)\} \Big]. \tag{78}$$

Heuristically, the test rejects H_0 when one observes $(1,0)$ sufficiently more frequently than $(1,1)$, which is the robust prediction of the model when $\theta^{(1,0)}$ is sufficiently large.[37]

---

[36] At $\theta_0 + \xi_A$, one needs to consider a single cone that coincides with the local parameter space.

[37] The case with $\xi_B = (0, 1/6, 0)'$ can be analyzed similarly. For $\theta_{n,\xi_B,h} = \theta_0 + \xi_B + h/\sqrt{n}$, one needs to consider

4 Monte Carlo Experiments

We conduct Monte Carlo experiments to examine the performance of our tests.

4.1 Tests of Strategic Interaction Effects

The first set of experiments evaluates the size and power of the tests on the strategic interaction effects. The design of the experiment is based on Example 1, in which we generate $u_i \overset{i.i.d.}{\sim} N(0, I_2)$ for $i = 1, \cdots, n$. Whenever multiple equilibria exist, we select an outcome using one of the three selection mechanisms below. The first one is an i.i.d. selection mechanism, which selects $(1, 0)$ out of $G(u|\theta) = \{(1, 0), (0, 1)\}$ if an i.i.d. Bernoulli random variable v_i takes 1. Otherwise, $(0, 1)$ is selected. The second mechanism selects $(1, 0)$ when another Bernoulli random variable $\tilde{v}_i$ takes 1, where $\{\tilde{v}_i\}$ is a i.n.i.d. sequence. Let N_k^* be an increasing sequence of integers.[38] For each i, let $h(i) = N_k^*$, where $N_{k-1}^* < i \le N_k^*$. We define

$$
\tilde{v}_i = \begin{cases} 1 & \Psi_{h(i)}^G(u^\infty) > \frac{1/4 + \Phi(\theta^{(1)})/2 - \Phi(\theta^{(1)})\Phi(\theta^{(2)})}{1/2 + (\Phi(\theta^{(1)}) + \Phi(\theta^{(2)}))/2 - 2\Phi(\theta^{(1)})\Phi(\theta^{(2)})} \\ 0 & \Psi_{h(i)}^G(u^\infty) \le \frac{1/4 + \Phi(\theta^{(1)})/2 - \Phi(\theta^{(1)})\Phi(\theta^{(2)})}{1/2 + (\Phi(\theta^{(1)}) + \Phi(\theta^{(2)}))/2 - 2\Phi(\theta^{(1)})\Phi(\theta^{(2)})}, \end{cases} \tag{79}
$$

where $\Psi_n^G(u^\infty) = \frac{\sum_{i=1}^n I[G(u_i|\theta) = \{(1,0)\}]}{\sum_{i=1}^n I[G(u_i|\theta) = \{(1,0)\} \, or \, \{(0,1)\}]}$. One may view observations for which $N_{k-1}^* < i \le N_k^*$ as members of a cluster. Under this selection mechanism, the outcomes are dependent on each cluster, which in turn makes the outcome sequence heterogeneous (and non-ergodic). The third mechanism generates data from the LFP, which draws an outcome sequence from Q_0^n when $\theta = \theta_0$ and from Q_1^n when $\theta = \theta_1$.

We evaluate the size and power of the test in Section 3.4.1 on $H_0 : p'\theta = 0$ against $H_1 : p'\theta > 0$ with $p = (-1, -1)'$. The test based on the statistic in (58) achieves the power envelope (see Appendix D.1.2). We therefore evaluate the size and power of this test. To make a comparison, we consider another test, namely the Wald–Wolfowitz runs test, which non-parametrically tests the i.i.d.-ness of the outcome sequence (see Wald and Wolfowitz, 1940; Cho and White, 2011). Since the model is complete under H_0 and u_i is i.i.d., the resulting outcome sequence is i.i.d. under the null hypothesis. The test therefore should have power only when the selection introduces heterogeneity and/or dependence.

Figures 5 and 6 show the power of the optimal test and runs test, respectively. The power of our test changes little across the selection mechanisms. This can be explained as follows. Our test statistic in (58) treats the two outcomes, $(0, 1)$ and $(1, 0)$, symmetrically. While the selection mechanism affects the relative frequencies of the two outcomes, what matters for the statistic is the frequency of $\{(0, 1), (1, 0)\}$ (relative to $(1, 1)$), and hence its power curve is insensitive to the

---

two tangent cones, $\mathcal{T}_j(\theta_0, \xi_B), j \in \{\mathrm{I}, \mathrm{II}\}$, because the form of the LFP changes depending on the cone h to which belongs (see Appendix D.2). Interestingly, the optimal test statistic is still given by (78) in both cases.

[38]In our simulations, we set $N_k^* = 2^{2^k}$.

selection mechanism.[39] The Wald–Wolfowitz test has non-zero power when the selection is i.n.i.d.; it becomes noticeable only when h is very large and its lower power is significantly below that of the optimal test. As expected, it does not have any power when the selection mechanism is i.i.d. or the LFP.

4.2 Tests on the Distribution of Potential Outcomes

The second set of experiments is based on Example 2, which we use to evaluate the performance of the tests when the model is incomplete under H_0. The model may be complete under certain alternatives. As described in Section 3.4, we consider testing $H_0 : p'\theta \leq c$ against $H_1 : p'\theta > c$, where $p = (0, 0, 1)'$ and $c = 1/6$. As before, we assume that $\theta^{(0,0)} = 1/6$ is known and localize the experiment at $\theta_0 = (\theta_0^{(0,0)}, \theta_0^{(0,1)}, \theta_0^{(1,0)})' = (1/6, 1/2, 1/6)'$.

The set of selection mechanisms is the same as the one in Section 4.1. One difference is that the model predicts multiple outcomes when $u = (0, 0)$ or $u = (1, 1)$. In both cases, an i.i.d. selection mechanism selects one of the outcomes ($s = (0, 0)$ when $G(u) = \{(0, 0), (0, 1)\}$ and $s = (1, 0)$ when $G(u) = \{(1, 0), (1, 1)\}$) when an i.i.d. Bernoulli random variable v_i is 1. Similarly, an i.n.i.d. selection mechanism selects one of the predicted outcomes when $\tilde{v}_i$ in (79) is 1. The LFP selection mechanism is defined in the same way as before.

We consider the following shifted local alternatives:

$$\text{Alt.A:} \qquad \theta_{n,\xi_A,h} = \theta_0 + \xi_A + (0, 0, \bar{h})'/\sqrt{n}, \ \bar{h} > 0 \tag{80}$$

$$\text{Alt.B-I:} \qquad \theta_{n,\xi_B,h} = \theta_0 + \xi_B + (0, -\bar{h}, \bar{h}/2)'/\sqrt{n}, \ \bar{h} > 0 \tag{81}$$

$$\text{Alt.B-II:} \qquad \theta_{n,\xi_B,h} = \theta_0 + \xi_B + (0, -\bar{h}, \bar{h})'/\sqrt{n}, \ \bar{h} > 0. \tag{82}$$

Under Alternative A, we consider a sequence of the parameters that tends to $\theta_0 + \xi_A$, where $h \in \mathcal{T}_I(\theta_0, \xi_A)$. Under Alternatives B-I and B-II, we consider the parameters that tend to $\theta_0 + \xi_B$, where $h \in \mathcal{T}_I(\theta_0, \xi_B)$ and $h \in \mathcal{T}_{II}(\theta_0, \xi_B)$, respectively.

Figures 7 and 9 report the results with $n = 1000$ and $S = 2000$, respectively. The right panel of Figure 7 shows the power curves of the optimal test against Alternative A. Because of the model incompleteness, its performance varies significantly across the selection mechanisms. As predicted by the theory, the power of the test under the LFP selection mechanism essentially coincides with the power envelope. Both the i.i.d. and the i.n.i.d. selection mechanisms are considerably more favorable; that is, the actual power of the test under these mechanisms is much higher than under the LFP selection. In particular, the power of the test under the i.i.d. selection mechanism is already 1 even when $\bar{h} = 0$ because the test can detect deviations from the null under this selection mechanism even for $\theta = \theta_0 + \xi$ that are not robustly testable. The left panel of Figure 7 shows the power curves against alternatives of the form $\theta_0 + w\xi_A$ with $w \in [0, 1]$. The figure shows that the

---

[39]This insensitivity is not a generic feature of the optimal test. See the next example.

test may have non-trivial power even for such alternatives when the selection mechanisms are i.i.d. or i.n.i.d.[40]

Figure 8 shows the power curves against Alternative B-I. When $\bar{h} = 0$, the model is indeed complete.[41] However, as $\bar{h}$ increases, the region of incompleteness enlarges, leading to differences in the local power across the selection mechanisms. Under Alternative B-II, the model stays complete under the local alternatives. Therefore, the power of the test is essentially the same across the selection mechanisms.

5 Extensions

5.1 Tests in the Presence of Nuisance Components

We now consider testing the hypotheses on subcomponents of θ. Let $\theta = (\beta', \delta')' \in \Theta_\beta \times \Theta_\delta$, where β is a $k \times 1$-sub-vector of interest and δ is a $(d-k) \times 1$ vector of nuisance parameters. Consider the following hypotheses:

$$H_0 : \beta = \beta_0, \ \delta \in \Theta_\delta, \quad \text{v.s.} \quad H_1 : \beta \neq \beta_0, \ \delta \in \Theta_\delta. \tag{83}$$

This problem can be recast as a special case of (1) with $\varphi(\theta) = \beta$, $K_0 = \{\beta_0\}$, and $K_1 = \{\beta \in \mathbb{R}^k : \beta \neq \beta_0\}$.[42]

In this general setting, both hypotheses are composite in terms of the structural parameters. Therefore, Lemma 3.1 is not directly applicable. However, the result is still useful for constructing tests that have desirable properties. To this end, we partition the parameter space into two sets, namely Θ_0 and Θ_1, where $\Theta_0 = \{\beta_0\} \times \Theta_\delta$ and $\Theta_1 = \{\beta : \beta \neq \beta_0\} \times \Theta_\delta$. We focus on this setting, whereas the theory below applies more generally to the hypotheses of the form in (1) by taking $\Theta_0 = \{\theta : \varphi(\theta) \in K_0\}$ and $\Theta_1 = \{\theta : \varphi(\theta) \in K_1\}$.

Throughout, the researcher's action is binary, that is $a = 1$ (reject) or $a = 0$ (accept). For each $\theta \in \Theta$ and action $a \in \{0, 1\}$, define a loss function $\mathsf{L} : \Theta \times \{0, 1\} \to \mathbb{R}_+$ by

$$\mathsf{L}(\theta, a) \equiv a I_{\Theta_0}(\theta) + \zeta(1-a) I_{\Theta_1}(\theta),$$

where $\zeta > 0$. The loss from the Type-I error is normalized to 1. The trade-off between the Type-I and Type-II errors is determined by parameter ζ.

---

[40]Under the LFP selection, the data are drawn from $Q_{\theta_0} \in \mathcal{P}_{\theta_0 + w\xi_A}$, and hence the power of the test does not exceed the nominal level.

[41]The model remains incomplete regarding selection from $G(u) = \{(0,0), (0,1)\}$ when $u = (0,0)$, however, these outcomes are not used by the test and hence do not affect the power.

[42]Sub-vector inference has been actively studied in the context of partially identified models, particularly moment inequality models (Romano and Shaikh, 2008; Bugni, Canay, and Shi, 2017; Kaido, Molinari, and Stoye, 2019). Here, we focus on hypothesis tests on the sub-vectors of the structural parameters.

For each test ϕ and $\theta \in \Theta$, define the *upper risk* by

$$R(\theta, \phi) = \max_{P \in \mathcal{P}_\theta} \int \phi(s) I_{\Theta_0}(\theta) + \zeta(1 - \phi(s)) I_{\Theta_1}(\theta) dP(s)$$

$$= \int \phi(s) d\nu_\theta^*(s) I_{\Theta_0}(\theta) + \zeta(1 - \int \phi(s) d\nu_\theta(s)) I_{\Theta_1}(\theta), \tag{84}$$

where the integrals in (84) are Choquet integrals (see Appendix A). The upper risk determines the trade-off between the size ($R_0(\theta, \phi) \equiv \sup_{P \in \mathcal{P}_\theta} \int \phi dP = \int \phi d\nu_\theta^*$ for $\theta \in \Theta_0$) and the lower power ($\inf_{P \in \mathcal{P}_\theta} \int \phi dP = \int \phi d\nu_\theta$ for $\theta \in \Theta_1$). What remains is to incorporate parameter uncertainty. For this, let μ be a (prior) probability distribution over Θ. We write μ as $\mu = \tau \mu_0 + (1 - \tau)\mu_1$, where $\tau \in (0, 1)$ and μ_0, μ_1 are suitable probability measures supported on Θ_0 and Θ_1, respectively. Define

$$r(\mu, \phi) \equiv \int_\Theta R(\theta, \phi) d\mu(\theta)$$

$$= \tau \int \phi(s) d\kappa_0^*(s) + (1 - \tau)\zeta(1 - \int \phi(s) d\kappa_1(s)), \tag{85}$$

where $\kappa_0^* = \int_{\Theta_0} \nu_\theta^* d\mu_0$ and $\kappa_1 = \int_{\Theta_1} \nu_\theta d\mu_1$.[43] This risk function uses the prior probability to reflect parameter uncertainty, while it uses the belief function (and its conjugate) to incorporate the decision maker's willingness to be robust against incompleteness. In what follows, we call r the *Bayes–Dempster–Shafer (BDS) risk*. We then call ϕ a *BDS test* if it minimizes the BDS risk.[44] One of the components of the BDS risk is $\pi_{\kappa_1}(\phi) = \int \phi d\kappa_1$. We call this object the *weighted average lower power* (WALP). The interpretation of $\pi_{\kappa_1}(\phi)$ is similar to that of the standard weighted average power (Andrews and Ploberger, 1994, 1995). Choosing a suitable μ_1, one may direct the power of a test toward certain alternatives. However, π_{κ_1} takes the average of the guaranteed power value instead of the actual unknown power.

The following theorem characterizes the BDS test. For this, let $core(\kappa) \equiv \{P \in \Delta(S) : P(A) \geq \kappa(A), \forall A \subset S\}$. In what follows, we assume that $core(\kappa_0) \cap core(\kappa_1) = \emptyset$.[45]

Lemma 5.1. *Let the BDS risk be defined as in (85). Then, there exists a BDS test such that, for any $\zeta > 0$,*

$$\phi(s) = \begin{cases} 1 & if \quad \Lambda(s) > C \\ \gamma & if \quad \Lambda(s) = C \\ 0 & if \quad \Lambda(s) < C, \end{cases} \tag{86}$$

where $C = \tau/\zeta(1 - \tau)$, and Λ is a version of dQ_1/dQ_0 for the LFP $(Q_0, Q_1) \in core(\kappa_0) \times core(\kappa_1)$

---

[43]The second equality in (85) is established in the proof of Theorem 5.1.

[44]The axiomatic foundations for this type of preference (when S is the payoff-relevant state space) is given in Gul and Pesendorfer (2014) and Epstein and Seo (2015) (in the context of repeated experiments).

[45]To ensure this condition, it is sufficient to have at least one $\bar{A} \subset S$ such that $P_0(\bar{A}) < P_1(\bar{A})$ (or $P_0(\bar{A}) > P_1(\bar{A})$) for all $(P_0, P_1) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_1}$ and $(\theta_0, \theta_1) \in \Theta_0 \times \Theta_1$. In Example 1, one may take $\bar{A} = \{(1, 1)\}$.

such that, for all $t \in \mathbb{R}_+$,

$$Q_0(\Lambda > t) = \kappa_0^*(\Lambda > t), \quad \text{and} \quad Q_1(\Lambda > t) = \kappa_1(\Lambda > t). \tag{87}$$

One may view this as an analog of Lemma 3.1. A key difference is that the LFP belongs to the product of the cores of the capacities κ_0 and κ_1. Hence, κ_0 and κ_1 are both belief functions, which in turn allows us to compute the LFP in a tractable way.

The analysis above is useful for constructing optimal tests for minimizing risk. However, the BDS tests are not designed to control size uniformly over Θ_0. Therefore, we consider a test that controls size and maximizes the WALP. For this, we fix μ_1 throughout and follow the developments on the tests in the presence of nuisance parameters (Chamberlain, 2000; Elliott, Müller, and Watson, 2015; Moreira and Moreira, 2013).

The following minimax theorem characterizes the minimax test as a BDS test for the least favorable prior (if it exists). For this, let $\mathcal{M}(\mu_1) \equiv \{\mu : \mu = \tau\mu_0 + (1-\tau)\mu_1, \mu_0 \in \Delta(\Theta_0), \tau \in [0,1]\}$, where μ_1 is fixed. In what follows, we drop μ_1 from the argument of $\mathcal{M}$, but its dependence should be understood. We then let $\boldsymbol{\Phi}$ be the set of randomized tests.

Theorem 5.1. *Let the upper risk R be defined as in (84). Suppose that Θ is compact. Then,*

$$\sup_{\mu \in \mathcal{M}} \inf_{\phi \in \boldsymbol{\Phi}} \int_{\Theta} R(\theta, \phi) d\mu(\theta) = \inf_{\phi \in \boldsymbol{\Phi}} (\sup_{\theta \in \Theta_0} R_0(\theta, \phi) \vee R_1(\phi)), \tag{88}$$

where $R_0(\theta, \phi) = \int \phi(s) d\nu_\theta^*(s) I_{\Theta_0}(\theta)$ and $R_1(\phi) = \zeta(1 - \pi_{\kappa_1}(\phi))$. Furthermore, there exists $\phi^\dagger$ that achieves equality in (88).

Remark 5.1. Suppose that ζ is chosen so that the maximum BDS risk (left-hand side of (88)) equals α. Then, $\phi^\dagger$ is a level-α test that maximizes the WALP. The theorem suggests that such a test can be approximated (in terms of risk) by the sequence of tests $\{\phi_\ell, \ell = 1, 2, \dots\}$ such that ϕ_ℓ is a BDS test for some prior μ_ℓ, and $\int_{\Theta} R(\theta, \phi_\ell) d\mu_\ell(\theta) \to \sup_{\mu \in \mathcal{M}} \int_{\Theta} R(\theta, \phi^\dagger) d\mu(\theta)$.

5.2 Covariates

This section extends the base framework to incorporate observable covariates. Each individual experiment is described by $(S, X, U, G, \Theta; \upsilon, m)$, where S, U, G, Θ are defined as before. We let X denote the finite set of covariate values and $\{\upsilon_\theta, \theta \in \Theta\}$ be a family of distributions on X. Throughout, we assume that each $\upsilon_\theta \in \Delta(X)$ has full support on X. Measure $m_\theta(\cdot | x)$ then determines the conditional law of u given x. The prediction of the model is then summarized by a weakly measurable correspondence $(u, x) \mapsto F(u, x | \theta) \equiv \{(s, x) : s \in G(u | \theta, x)\} \subset S \times X$ for each $\theta \in \Theta$. As before, this correspondence induces a belief function on $S \times X$

$$\nu_\theta(A) = \int 1\{F(u, x | \theta) \subset A\} dm_\theta(u | x) d\upsilon_\theta(x), \; A \subset S \times X. \tag{89}$$

If A is a rectangle $A = A_s \times A_x$ for some $A_s \subset S$ and $A_x \subset X$, one may write it as

$$\nu_\theta(A) = \int_{A_x} m_\theta(G(u|\theta,x) \subset A_s|x) d\upsilon_\theta(x) = \int_{A_x} \nu_\theta(A_u|x) d\upsilon_\theta(x), \tag{90}$$

which can be viewed as the mean of the conditional belief function $\nu_\theta(\cdot|x)$. The subsequent analysis, starting with the Neyman–Pearson lemma, is then essentially the same as before.

A simplification occurs when υ does not depend on θ. Consider $\theta_0, \theta_1 \in \Theta$ with $\theta_0 \neq \theta_1$. Because of the additivity of υ, it suffices to consider sets of the form $B \times \{x\}$, where $B \subset S$ and $x \in X$. Then, the program that determines the LFP is

$$(Q_0, Q_1) = \underset{P_0, P_1 \in \Delta(S \times X)}{\arg\min} \int H\Big(\frac{dP_0}{d(P_0 + P_1)}\Big) d(P_0 + P_1) \tag{91}$$
$$s.t. \ \nu_{\theta_0}(B \times \{x\}) \leq P_0(B \times \{x\}), \ B \subset S, x \in X$$
$$\nu_{\theta_1}(B \times \{x\}) \leq P_1(B \times \{x\}), \ B \subset S, x \in X.$$

Observe that the constraints simplify to

$$\nu_{\theta_j}(B \times \{x\}) = \nu_{\theta_j}(B|x)\upsilon(x) \leq \sum_{s \in B} p_j(s|x)\upsilon_j(x), \ j = 0, 1. \tag{92}$$

Taking $B = S$, one obtains $\upsilon(x) \leq \upsilon_j(x)$ for all $x \in X$ with $j = 0, 1$. Since υ is a measure, this implies $\upsilon_0 = \upsilon_1 = \upsilon$, and the resulting LR statistic does not depend on υ. The LR statistic $dQ_1/dQ_0 = q_1(s|x)/q_0(s|x)$ can then be calculated by solving, for each x,

$$\min_{p_0(\cdot|x), p_1(\cdot|x)} \sum_{s \in S} \ln\Big(\frac{p_0(s|x) + p_1(s|x)}{p_0(s|x)}\Big)(p_0(s|x) + p_1(s|x)) \tag{93}$$
$$s.t. \ \nu_{\theta_0}(B|\theta,x) \leq \sum_{s \in B} p_j(s|x), \ B \subset S, x \in X$$
$$\nu_{\theta_1}(B|\theta,x) \leq \sum_{s \in B} p_j(s|x), \ B \subset S, x \in X.$$

6 Concluding Remarks

This study explored robust likelihood-based inference methods for incomplete economic models. A key result is the existence of an LFP consisting of product measures, through which we may connect incomplete models to standard frameworks and thus obtain asymptotic approximations and analyze optimality properties while remaining agnostic about the selection. However, some problems need further work. They include methods of inference on the subcomponents of θ, especially when the dimension of the parameter is moderately high and a framework that can handle solution concepts involving some form of mixing (e.g., mixed Nash, Bayes correlated equilibria) which are undertaken in ongoing work. Another important avenue for future research is an extension of the current results

to statistical inference or decision problems outside hypothesis testing.

References

ALIPRANTIS, C. D., AND K. BORDER (2006): *Infinite dimensional analysis: a hitchhiker's guide.* Springer Science & Business Media.

ANDREWS, D. W., AND P. J. BARWICK (2012): "Inference for parameters defined by moment inequalities: A recommended moment selection procedure," *Econometrica*, 80(6), 2805–2826.

ANDREWS, D. W., AND W. PLOBERGER (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica: Journal of the Econometric Society*, pp. 1383–1414.

——— (1995): "Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative," *The Annals of Statistics*, pp. 1609–1629.

ANDREWS, D. W., AND G. SOARES (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78(1), 119–157.

ANDREWS, D. W. K. (1999): "Estimation When a Parameter is on a Boundary," *Econometrica*, 67(6), 1341–1383.

ARADILLAS-LOPEZ, A., AND E. TAMER (2008): "The Identification Power of Equilibrium in Simple Games," *Journal of Business & Economic Statistics*, 26(3), 261–283.

ARMSTRONG, T. B. (2014): "Weighted KS statistics for inference on conditional moment inequalities," *Journal of Econometrics*, 181(2), 92 – 116.

——— (2018): "On the choice of test statistic for conditional moment inequalities," *Journal of Econometrics*, 203(2), 241 – 255.

BEDNARSKI, T. (1982): "Binary experiments, minimax tests and 2-alternating capacities," *The Annals of Statistics*, pp. 226–232.

BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2011): "Sharp identification regions in models with convex moment predictions," *Econometrica*, 79(6), 1785–1821.

BERESTEANU, A., AND F. MOLINARI (2008): "Asymptotic properties for a class of partially identified models," *Econometrica*, 76(4), 763–814.

BERRY, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica: Journal of the Econometric Society*, pp. 889–917.

BOYD, S., AND L. VANDENBERGHE (2004): *Convex optimization.* Cambridge university press.

BRESNAHAN, T. F., AND P. C. REISS (1990): "Entry in monopoly market," *The Review of Economic Studies*, 57(4), 531–553.

——— (1991): "Empirical models of discrete games," *Journal of Econometrics*, 48(1), 57–81.

BUGNI, F., I. CANAY, AND X. SHI (2017): "Inference for functions of partially identified parameters in moment inequality models," *Quantitative Economics*, 8, 1–38.

BUGNI, F. A. (2010): "Bootstrap inference in partially identified models defined by moment inequalities: Coverage of the identified set," *Econometrica*, 78(2), 735–753.

CANAY, I. A. (2010): "EL inference for partially identified models: Large deviations optimality and bootstrap validity," *Journal of Econometrics*, 156(2), 408 – 425.

CHAMBERLAIN, G. (2000): "Econometric applications of maxmin expected utility," *Journal of Applied Econometrics*, pp. 625–644.

CHEN, X., E. T. TAMER, AND A. TORGOVITSKY (2011): "Sensitivity Analysis in Semiparametric Likelihood Models," *Cowles Foundation Discussion Paper No. 1836.*

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and confidence regions for parameter sets in econometric models1," *Econometrica*, 75(5), 1243–1284.

CHESHER, A., AND A. ROSEN (2017): "Generalized instrumental variable models," *Econometrica*, 85(3), 959–989.

CHO, J. S., AND H. WHITE (2011): "Generalized runs tests for the IID hypothesis," *Journal of econometrics*, 162(2), 326–344.

CHOQUET, G. (1954): "Theory of capacities," in *Annales de l'institut Fourier*, vol. 5, pp. 131–295.

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

DE PAULA, Á., AND X. TANG (2012): "Inference of signs of interaction effects in simultaneous games with incomplete information," *Econometrica*, pp. 143–172.

DEMPE, S. (1993): "Directional differentiability of optimal solutions under Slater's condition," *Math Program*, 59(1-3), 49–69.

DEMPSTER, A. P. (1967): "Upper and lower probabilities induced by a multivalued mapping," *The annals of mathematical statistics*, pp. 325–339.

EIZENBERG, A. (2014): "Upstream Innovation and Product Variety in the U.S. Home PC Market," *The Review of Economic Studies*, 81(3 (288)), 1003–1045.

ELLIOTT, G., U. K. MÜLLER, AND M. W. WATSON (2015): "Nearly optimal tests when a nuisance parameter is present under the null hypothesis," *Econometrica*, 83(2), 771–811.

EPSTEIN, L. G., H. KAIDO, AND K. SEO (2016): "Robust Confidence Regions for Incomplete Models," *Econometrica*, 84(5), 1799–1838.

EPSTEIN, L. G., AND K. SEO (2015): "Exchangeable capacities, parameters and incomplete theories," *Journal of Economic Theory*, 157, 879–917.

FANG, Z. (2014): "Optimal plug-in estimators of directionally differentiable functionals," *Working Paper, Texas A&M.*

FANG, Z., AND A. SANTOS (2018): "Inference on directionally differentiable functions," *The Review of Economic Studies*, 86(1), 377–412.

GALICHON, A., AND M. HENRY (2006): "Inference in incomplete models," *Available at SSRN 886907.*

———— (2009): "A test of non-identifying restrictions and confidence regions for partially identified parameters," *Journal of Econometrics*, 152(2), 186–196.

———— (2011): "Set identification in models with multiple equilibria," *The Review of Economic Studies*, p. rdr008.

———— (2013): "Dilation bootstrap," *Journal of Econometrics*, 177(1), 109–115.

GAUVIN, J. (1977): "A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming," *Mathematical Programming*, 12(1), 136–138.

GUL, F., AND W. PESENDORFER (2014): "Expected uncertain utility theory," *Econometrica*, 82(1), 1–39.

HAILE, P. A., AND E. TAMER (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111(1).

HÄUSLER, E., AND H. LUSCHGY (2015): *Stable Convergence and Stable Limit Theorems*, vol. 74. Springer.

HECKMAN, J. J., AND B. E. HONORÉ (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58(5), 1121–1149.

HIRANO, K., AND J. R. PORTER (2012): "Impossibility Results for Nondifferentiable Functionals," *Econometrica*, 80(4), 1769–1790.

HONG, H., AND J. LI (2018): "The numerical delta method," *Journal of Econometrics*, 206(2), 379–394.

HUBER, P. J. (1981): *Robust statistics.* Wiley, New York.

HUBER, P. J., AND V. STRASSEN (1973): "Minimax tests and the Neyman-Pearson lemma for capacities," *The Annals of Statistics*, pp. 251–263.

———— (1974): "Note: Correction to minimax tests and the Neyman-Pearson lemma for capacities," *The Annals of Statistics*, 2(1), 223–224.

JOVANOVIC, B. (1989): "Observable implications of models with multiple equilibria," *Econometrica: Journal of the Econometric Society*, pp. 1431–1437.

KAIDO, H., F. MOLINARI, AND J. STOYE (2019): "Confidence intervals for projections of partially identified parameters," *Econometrica*, 87(4), 1397–1432.

KAIDO, H., AND A. SANTOS (2014): "Asymptotically Efficient Estimation of Models Defined by Convex Moment Inequalities," *Econometrica*, 82(1), 387–413.

KAWAI, K., AND Y. WATANABE (2013): "Inferring Strategic Voting," *The American Economic Review*, 103(2), 624–662.

LE CAM, L. (1972): "Limits of experiments," in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics*, pp. 245–261, Berkeley, Calif. University of California Press.

LE CAM, L. (1986): "Asymptotic Methods in Statistical Decision Theory," *Springer Series in Statistics*.

LEHMANN, E. L., AND J. P. ROMANO (2006): *Testing statistical hypotheses*. Springer Science & Business Media.

LUO, Y., AND H. WANG (2017a): "Core determining class and inequality selection," *American Economic Review, Papers & Proceedings*.

LUO, Y., AND H. WANG (2017b): "Core Determining Class: Construction, Approximation and Inference," Working Paper, University of Florida.

MACCHERONI, F., AND M. MARINACCI (2005): "A strong law of large numbers for capacities," *The Annals of Probability*, 33(3), 1171–1178.

MIYAUCHI, Y. (2016): "Structural estimation of pairwise stable networks with nonnegative externality," *Journal of Econometrics*, 195(2), 224 – 235.

MOLCHANOV, I. (2006): *Theory of random sets*. Springer Science & Business Media.

MOLCHANOV, I., AND F. MOLINARI (2018): *Random sets in econometrics*, vol. 60 of *Econometric Society Monographs*. Cambridge University Press, Cambridge.

MOREIRA, H. A., AND M. J. MOREIRA (2013): "Contributions to the theory of optimal tests," Working Paper.

MOURIFIE, I., M. HENRY, AND R. MEANGO (2018): "Sharp Bounds and Testability of a Roy Model of STEM Major Choices," SSRN Electronic Journal.

PHILIPPE, F., G. DEBS, AND J.-Y. JAFFRAY (1999): "Decision making with monotone lower probabilities of infinite order," *Mathematics of Operations Research*, 24(3), 767–784.

RIEDER, H. (1994): "Robust Asymptotic Statistics," *Springer Series in Statistics*.

——— (2014): "One-sided confidence about functionals over tangent cones," *arXiv preprint arXiv:1412.1701*.

ROCKAFELLAR, R. T. (1972): *Convex analysis*. Princeton university press.

ROMANO, J. P., AND A. M. SHAIKH (2008): "Inference for identifiable parameters in partially identified econometric models," *Journal of Statistical Planning and Inference*, 138(9), 2786–2807.

SHAFER, G. (1982): "Belief functions and parametric models," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 322–352.

SHAPIRO, A. (1988): "Sensitivity Analysis of Nonlinear Programs and Differentiability Properties of Metric Projections," *SIAM Journal on Control and Optimization*, 26(3), 628–645.

SONG, K. (2014): "Local asymptotic minimax estimation of nonregular parameters with translation-scale equivariant maps," *Journal of Multivariate Analysis*, 125, 136 – 158.

STRASSER, H. (1985): *Mathematical theory of statistics: statistical experiments and asymptotic decision theory*, vol. 7. Walter de Gruyter.

TAMER, E. (2003): "Incomplete simultaneous discrete response model with multiple equilibria," *The Review of Economic Studies*, 70(1), 147–165.

VAN DER VAART, A. W. (2000): *Asymptotic statistics*. Cambridge university press.

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes: with Applications to Statistics.* Springer-Verlag, New York.

WALD, A. (1950): "Remarks on the estimation of unknown parameters in incomplete systems of equations," in *Statistical Inference in Dynamic Economic Models,*, ed. by T. Koopmans, pp. 305–310. John Wiley, New York.

WALD, A., AND J. WOLFOWITZ (1940): "On a test whether two samples are from the same population," *The Annals of Mathematical Statistics*, 11(2), 147–162.

WHITE, H. (2001): *Asymptotic theory for econometricians.* Academic Press, San Diego, rev. ed edn.
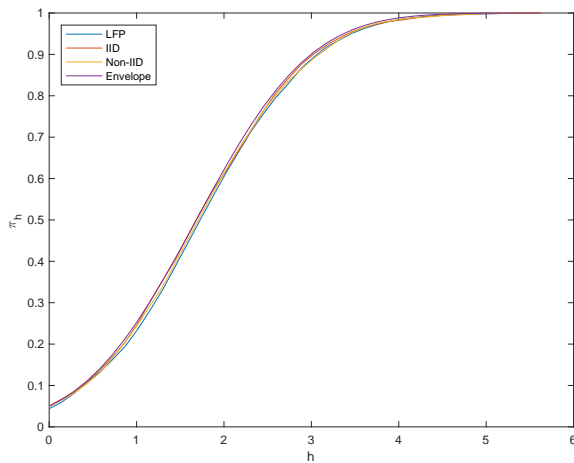
Figures

Figure 5: Local power of the robust test: $n = 1000$, $S = 5000$, $\theta_{n,h} = (-\bar{h}/\sqrt{n}, -\bar{h}/\sqrt{n})$
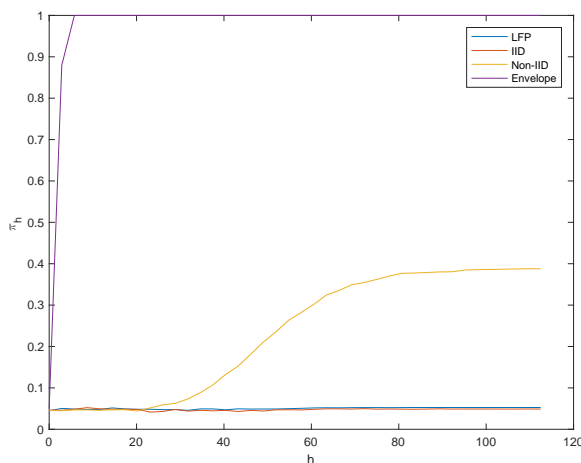

Figure 6: Local power of the Wald-Wolfowitz test: $n = 1000$, $S = 5000$, $\theta_{n,h} = (-\bar{h}/\sqrt{n}, -\bar{h}/\sqrt{n})$
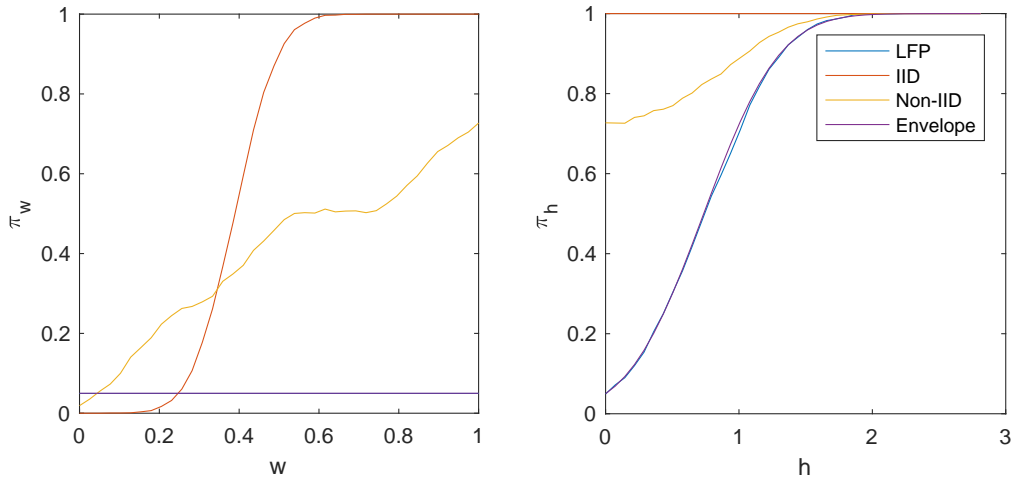
Figure 7: Power against $\theta_0 + w\xi_A$ (left) and $\theta_0 + \xi_A + h/\sqrt{n}$ (right) with $h = (0, 0, \bar{h})' \in \mathcal{T}(\theta_0, \xi_B)$
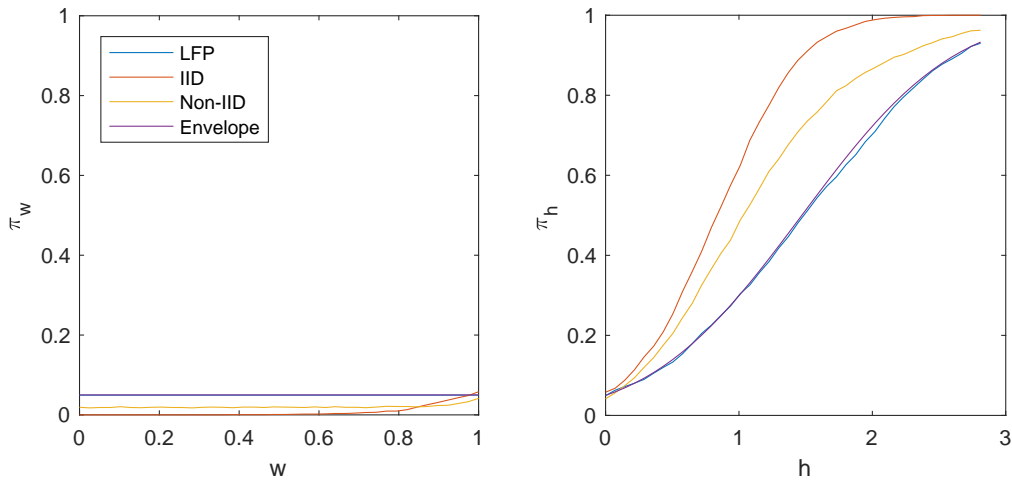

Figure 8: Power against $\theta_0 + w\xi_B$ (left) and $\theta_0 + \xi_B + h/\sqrt{n}$ (right) with $h = (0, -\bar{h}, \bar{h}/2)' \in \mathcal{T}_{\mathrm{I}}(\theta_0, \xi_B)$
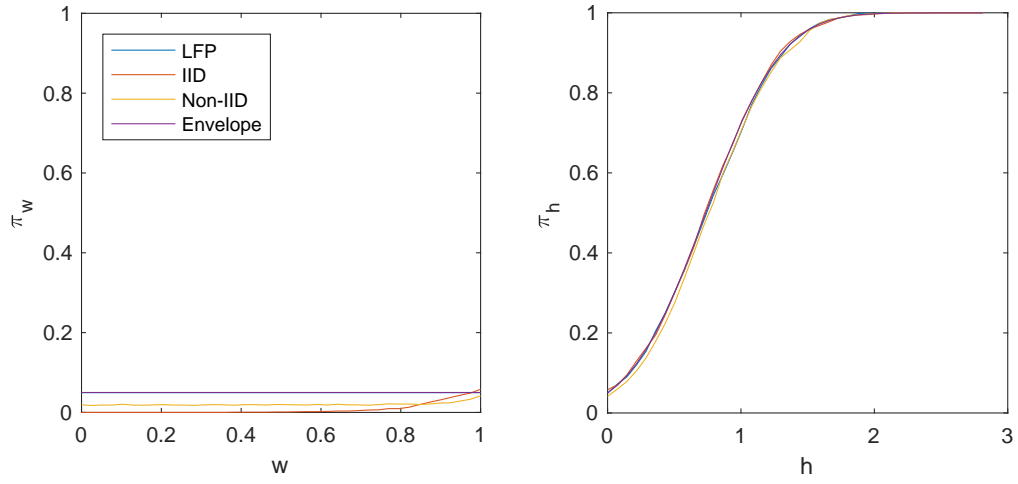
Figure 9: Power against $\theta_0 + w\xi_B$ (left) $\theta_0 + \xi_B + h/\sqrt{n}$ (right) with $h = (0, -\bar{h}, \bar{h})' \in \mathcal{T}_{\mathrm{II}}(\theta_0, \xi_B)$

Supplementary Material for "Robust Likelihood Ratio Tests for Incomplete Economic Models"

A Capacities

Let Ω be a compact metric space and let Σ_Ω denote its Borel σ-algebra. Let $\mathcal{K}(\Omega)$ be the set of compact subsets of Ω endowed with the Hausdorff metric. Let $\mathcal{C}(\Omega)$ be the set of continuous functions on Ω. Let $\Delta(\Omega)$ be the set of Borel probability measures on Ω endowed with the weak topology.

A set function ν^* is said to be a *capacity* if ν^* satisfies the following conditions:

(i) $\nu^*(\emptyset) = 0, \nu^*(\Omega) = 1$,

(ii) $A \subset B \Rightarrow \nu^*(A) \le \nu^*(B)$, for all $A, B \in \Sigma_\Omega$.

(iii) $A_n \uparrow A \Rightarrow \nu^*(A_n) \uparrow \nu^*(A)$, for all $\{A_n, n \ge 1\} \subset \Sigma_\Omega$ and $A \in \Sigma_\Omega$.

(iv) $F_n \downarrow F, F_n$ closed $\Rightarrow \nu^*(F_n) \downarrow \nu^*(F)$.

One may define integral operations with respect to capacities as follows. Let $f : \Omega \to \mathbb{R}$ be a measurable function. The *Choquet integral* of f with respect to ν is defined by

$$\int f d\nu \equiv \int_{-\infty}^0 (\nu(\{\omega : f(\omega) \ge t\}) - \nu(\Omega))dt + \int_0^\infty \nu(\{\omega : f(\omega) \ge t\})dt, \tag{94}$$

where the integrals on the right hand side are Riemann integrals.

The following result due to Choquet follows from Theorems 1-3 in Philippe, Debs, and Jaffray (1999).

Lemma A.1. *Let Ω be a Polish space. Let M be a probability measure on $\mathcal{K}(\Omega)$. Let $\mathcal{P} = \{P \in \Delta(\Omega) : P = \int P_K dM(K), P_K \in \Delta(K)\}$. Then, $\nu(\cdot) = \inf_{P \in \mathcal{P}} P(\cdot)$ is a belief function and satisfies*

$$\nu(A) = M(\{K \subset A\}). \tag{95}$$

In each experiment characterized by the tuple $(S, U, \Theta, G; m)$, one may apply the lemma above with $\mathcal{P} = \mathcal{P}_\theta$, $\nu = \nu_\theta$, $K = G(u|\theta)$, and M is the law of K induced by correspondence G and m_θ.

B Proofs

B.1 Proof of Lemma 3.1

Proof of Lemma 3.1. It is straightforward to show ν_θ^* is a capacity satisfying conditions (i)-(iv) in Appendix A. Since $G(\cdot|\theta)$ is weakly measurable, the map $u \mapsto G(u|\theta)$ defines a measurable map from U to $\mathcal{K}(S)$. Let $\tilde{m}$ be the induced measure of m_θ on $\mathcal{K}(S)$ by this map. Then, by Lemma A.1, $P \in \mathcal{P}_\theta$ is equivalent to

$P \in core(\nu)$ for an infinitely monotone capacity ν such that $\nu(A) = \tilde{m}(K \subset A)$ for all $A \in \Sigma_\Omega$. By Lemma 2.5 in HS,

$$\nu(A) = \inf_{P \in \mathcal{P}_\theta} P(A) = \nu_\theta(A), \quad \text{for all } A \in \Sigma_S, \tag{96}$$

and hence ν_θ is infinitely monotone. By the previous step, $\nu^*_{\theta_1}$ and $\nu^*_{\theta_0}$ are also 2-alternating and 2-monotone capacities respectively. Let Λ be the Radon-Nikodym derivative of $\nu^*_{\theta_1}$ and $\nu^*_{\theta_0}$ in the sense of HS (Section 3). Then, by their Theorem 4.1, the conclusion of the lemma follows. ∎

B.2 Proof of Theorem 3.1, Corollary 3.1 and Auxiliary Lemmas

We use Theorem 8.1.1 in Lehmann and Romano (2006) to show Theorem 3.1. For ease of reference, we copy their theorem below (with a slight change of notation to avoid conflicts). For this, let E, E' be measurable spaces and let $\mathcal{P} = \{P_\eta \in \Delta(S) : \eta \in E \cup E'\}$ be families of probability distributions on S with densities $p_\eta = dP_\eta/dv$ parameterized by $\eta \in E \cup E'$. Throughout, we assume that the map $(s, \eta) \mapsto p_\eta(s)$ is jointly measurable.

Theorem B.1 (Theorem 8.1.1. of Lehmann and Romano (2006)). *For any distributions μ, μ' over Σ_E and $\Sigma_{E'}$, let $\phi_{\mu,\mu'}$ be the most powerful test for testing*

$$f(s) = \int_E p_\eta(s) d\mu(\eta) \tag{97}$$

at level α against

$$f'(s) = \int_{E'} p_\eta(s) d\mu(\eta) \tag{98}$$

and let $\beta_{\mu,\mu'}$ be its power against the alternative f'. If there exist μ and μ' such that

$$\sup_{\eta \in E} E_{P_\eta}[\phi_{\mu,\mu'}(s)] \le \alpha \tag{99}$$

$$\inf_{\eta \in E'} E_{P'_\eta}[\phi_{\mu,\mu'}(s)] = \beta_{\mu,\mu'}, \tag{100}$$

then:

(i) $\phi_{\mu,\mu'}$ maximizes $\inf_{\eta \in E'} E_{P'_\eta}[\phi_{\mu,\mu'}(s)]$ among all level-α tests of the hypothesis $H : \eta \in E$ and is the unique test with this property if it is the unique most powerful level-α test for testing f against f'.

(ii) The pair of distributions μ, μ' is least favorable in the sense that for any other pair $\tilde{\mu}, \tilde{\mu}'$ we have

$$\beta_{\mu,\mu'} \le \beta_{\tilde{\mu},\tilde{\mu}'}. \tag{101}$$

Lemma B.1. *Let ν_θ be defined as in (7), and let ν^*_θ be its conjugate. Let $f : S \to \mathbb{R}$ be a measurable function. Similarly, for each $i \in \mathbb{N}$, let $f_i : S \to \mathbb{R}$ be a measurable function. Then,*

(i) There exists a minimizing measure $Q \in \Delta(S)$ and a maximizing measure $Q^* \in \Delta(S)$ such that, for

2

any $t \in \mathbb{R}$,

$$\nu_\theta(f(s) > t) = Q(f(s) > t), \tag{102}$$

and

$$\nu_\theta^*(f(s) > t) = Q^*(f(s) > t). \tag{103}$$

(ii) If, for each $i \in \mathbb{N}$ and each measurable function f_i, Q_i, Q_i^* are the minimizing and maximizing measures in the sense of (102)-(103), it follows that

$$\nu_\theta^n \Big(\sum_{i=1}^n f_i(s_i) > t \Big) = Q^n \Big(\sum_{i=1}^n f_i(s_i) > t \Big), \tag{104}$$

and

$$\nu_\theta^{*,n} \Big(\sum_{i=1}^n f_i(s_i) > t \Big) = Q^{*n} \Big(\sum_{i=1}^n f_i(s_i) > t \Big), \tag{105}$$

for all $t \in \mathbb{R}$, where $Q^n = \bigotimes_{i=1}^n Q_i$ and $Q^{*n} = \bigotimes_{i=1}^n Q_i^* \in \Delta(S^n)$.

Proof. (i) As shown in the proof of Lemma 3.1, ν_θ^* is a 2-alternating capacity. Since S is finite, any function on S is upper semi-continuous by the continuity of f. By Lemma 2.4 in HS, there exists a probability measure $p^* \in \Delta(S)$ such that for all $t \in \mathbb{R}$, $\nu_\theta^*(f(s) > t) = p^*(f(s) > t)$. This ensures (103). Similarly, let $g = -f$ and note that g is again upper semicontinuous. Applying Lemma 2.4 in HS to the event $\{g \geq -t\}$, there exists $p \in \Delta(s)$ such that, for any $t \in \mathbb{R}$,

$$
\begin{aligned}
\nu_\theta^*(g \geq -t) = p(g \geq t) \quad &\Leftrightarrow \quad 1 - \nu_\theta^*(g < -t) = 1 - p(g < t) \\
&\Leftrightarrow \quad \nu_\theta(f > t) = p(f > t).
\end{aligned}
\tag{106}
$$

This therefore establishes (102).

(ii) For each i, let $Y_i \equiv \min_{s_i \in G(u_i|\theta)} f_i(s_i)$ and $Z_i \equiv f_i(s_i)$. Note that Y_i is a function of u_i, and hence we use $m_{\theta,i}$ for the law of Y_i induced by u_i. For each experiment, we have

$$G(u|\theta) \subseteq \{s \in S : f_i(s) > t\} \Leftrightarrow \min_{s \in G(u|\theta)} f_i(s) > t. \tag{107}$$

Therefore, by Lemma A.1,

$$\nu_{\theta,i}(f_i(s_i) > t) = m_{\theta,i}(\min_{s_i \in G(u_i|\theta)} f_i(s_i) > t) = m_{\theta,i}(Y_i > t), \ \forall t \in \mathbb{R}. \tag{108}$$

By (i), there is $Q_i \in \Delta(S)$ such that

$$\nu_{\theta,i}(f_i(s_i) > t) = Q_i(Z_i > t), \ \forall t \in \mathbb{R}. \tag{109}$$

Hence, by (108)-(109), $Y_i \stackrel{d}{=} Z_i$ for all i.

Let $\mathcal{P}_\theta^n$ be defined as in (26) and let $\nu_\theta^n, \nu_\theta^{*n}$ be the lower and upper probabilities of $\mathcal{P}_\theta^n$ respectively. By

3

Lemma A.1, ν_θ^n is a belief function and ν_θ^{*n} is its conjugate. Therefore,

$$\nu_\theta^n\Big(\sum_{i=1}^n f_i(s_i) > t\Big) = m_\theta^n\Big(u^n \in U^n : G^n(u^n|\theta) \subseteq \{\sum_{i=1}^n f_i(s_i) > t\}\}\Big). \tag{110}$$

Since $G^n(u^n|\theta) = \prod_{i=1}^n G(u_i|\theta)$, inside the parenthesis we have:

$$\prod_{i=1}^n G(u_i|\theta) \subseteq \{s^n : \sum_{i=1}^n f_i(s_i) > t\}$$

$$\Leftrightarrow \min_{s^n \in \prod_{i=1}^n G(u_i|\theta)} \sum_{i=1}^n f_i(s_i) > t \Leftrightarrow \sum_{i=1}^n \min_{s_i \in G(u_i|\theta)} f_i(s_i) > t. \tag{111}$$

By (110)-(111) and recalling that $Y_i = \min_{s_i \in G(u_i|\theta)} f_i(s_i)$, we have

$$\nu_\theta^n\Big(\sum_{i=1}^n f_i(s_i) > t\Big) = m_\theta^n\Big(\sum_{i=1}^n \min_{s_i \in G(u_i|\theta)} f_i(s_i) > t\Big) = m_\theta^n\Big(\sum_{i=1}^n Y_i > t\Big). \tag{112}$$

Let $\{Y_1, Y_2, \cdots, Y_i, \cdots\}$ be independently distributed according to m_θ^n, and let $\{Z_1, Z_2, \cdots, Z_i \cdots\}$ be independently distributed according to Q^n. Then, $\sum_{i=1}^n Y_i \overset{d}{=} \sum_{i=1}^n Z_i$ because $(Y_1, \ldots, Y_n) \overset{d}{=} (Z_1, \ldots, Z_n)$. Therefore, for all $t \in \mathbb{R}$,

$$m_\theta^n\Big(\sum_{i=1}^n Y_i > t\Big) = Q^n(\sum_{i=1}^n Z_i > t). \tag{113}$$

By (112)-(113) and ν_θ^n being the lower probability of $\mathcal{P}_\theta^n$, we have

$$\min_{P \in \mathcal{P}_\theta^n} P(\sum_{i=1}^n f_i(s_i) > t) = \nu_\theta^n\Big(\sum_{i=1}^n f_i(s_i) > t\Big) = Q^n(\sum_{i=1}^n f_i(s_i) > t). \tag{114}$$

This establishes (104). One may show (105) by a similar argument. ∎

Proof of Theorem 3.1. Recall that $Q_0^n = \otimes_{i=1}^n Q_{0,i}$, $Q_1^n = \otimes_{i=1}^n Q_{1,i}$, and Λ_n is a version of the Radon-Nikodym derivative of them. We follow Section 8.3 in Lehmann and Romano (2006) and show the following statements:

(a) When s^n is distributed according to a distribution in $\mathcal{P}_{\theta_0}^n$, the probability of the event $\{s^n : \Lambda_n > t\}$ is largest (for any t), i.e. Λ_n is stochastically largest, when the distribution of s^n is $Q_0^n = \otimes_{i=1}^n Q_{0,i}$.

(b) When s^n is distributed according to a distribution in $\mathcal{P}_{\theta_1}^n$, the probability of the event $\{s^n : \Lambda_n > t\}$ is smallest (for any t), i.e. Λ_n is stochastically smallest, when the distribution of s^n is $Q_1^n = \otimes_{i=1}^n Q_{1,i}$.

(c) Λ_n is stochastically larger when the distribution of s is Q_1^n than when it is Q_0^n.

These statements are summarized by

$$Q_0^{n,\prime}(\Lambda_n > t) \overset{(a)}{\leq} Q_0^n(\Lambda_n > t) \overset{(c)}{\leq} Q_1^n(\Lambda_n > t) \overset{(b)}{\leq} Q_1^{n,\prime}(\Lambda_n > t), \tag{115}$$

for all t, $Q_0^{n,\prime} \in \mathcal{P}_{\theta_0}^n$, and $Q_1^{n,\prime} \in \mathcal{P}_{\theta_1}^n$.

Below, we invoke Lemma B.1. For this, let $f_i(\cdot) = \ln \Lambda_i(\cdot)$, where $\Lambda_i \in dQ_{1,i}/dQ_{0,i}$. Let (Q^{*n}, Q^n) be the product measures in Lemma B.1 with $f_i = \ln \Lambda_i$ for $i = 1, \ldots, n$. Note that $\Lambda_n > t$ is equivalent to $\sum_{i=1}^n f_i(s_i) > \ln t$. By Lemma B.1 with $t' = \ln t$, it then follows that

$$\nu_{\theta_0}^{*n}(\Lambda_n > t) = \nu_{\theta_0}^{*n}\Big(\sum_{i=1}^n f_i(s_i) > t'\Big) = Q^{*n}\Big(\sum_{i=1}^n f_i(s_i) > t'\Big) = Q^{*n}(\Lambda_n > t), \tag{116}$$

where $Q^{*n} = Q_0^n$. Recall that $\nu_{\theta_0}^{*n}(\Lambda_n > t) = \sup_{Q_0^{n,'} \in \mathcal{P}_{\theta_0}^n} Q_0^{n,'}(\Lambda_n > t)$. This therefore means Q_0^n makes Λ_n stochastically largest among all distributions in $\mathcal{P}_{\theta_0}^n$ and hence ensures inequality (a) in (115).

Similarly, again by Lemma B.1,

$$\nu_\theta^n(\Lambda_n > t) = \nu_\theta^n\Big(\sum_{i=1}^n f_i(s_i) > t'\Big) = Q^n\Big(\sum_{i=1}^n f_i(s_i) > t'\Big) = Q^n(\Lambda_n > t), . \tag{117}$$

where $Q^n = Q_1^n$. Therefore, Q^n makes Λ_n stochastically smallest and hence ensures inequality (c) in (115).

The middle inequality in (115) follows from Corollary 3.2.1 in Lehmann and Romano (2006) and the Neyman-Pearson lemma.

Let $E = \mathcal{P}_{\theta_0}^n, E' = \mathcal{P}_{\theta_1}^n$. Let $\mu, \mu' \in \Delta(E)$ be distributions, each assigning probability 1 to a single distribution, μ to $Q_0^n \in \mathcal{P}_{\theta_0}^n$ and μ' to $Q_1^n \in \mathcal{P}_{\theta_0}^n$. Let (C_n, γ_n) be chosen so that $E_{Q_0^n}[\phi_n(s^n)] = \alpha$, where ϕ_n is the likelihood-ratio test defined as in (31). The argument above shows that μ, μ' satisfy (99)-(100). The conclusion of the theorem then follows from applying Theorem B.1 to the present setting. ∎

Proof of Corollary 3.1. By Theorem 3.1, the LFP (Q_0^n, Q_i^n) exists, and they are product measures. Note that, in the application of Lemma B.1, Q_i (and Q_i^*) is identical across i because ν_θ (and ν_θ^*) is identical across i. The conclusion of the Corollary then follows by arguing as in the proof of Theorem 3.1. ∎

B.3   Proof of Proposition 3.1 and Theorems 3.2-3.3

Proof of Proposition 3.1. Note that

$$\begin{aligned}
\sup_{P \in \mathcal{P}_{\theta_0}^n} E_P[\phi_n^*(s^n)] &= \sup_{P \in \mathcal{P}_{\theta_0}^n} P(\Lambda_n(s^n) > C_n^*) \\
&= \nu_{\theta_0}^{*n}(\Lambda_n(s^n) > C_n^*) \\
&= Q_0^n\big(\Lambda_n(s^n) > C_n^*\big) \\
&= Q_0^n\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \ln \frac{dQ_1}{dQ_0}(s_i) - E_{Q_0}[\ln \frac{dQ_1}{dQ_0}(s_i)] > \sigma_{Q_0} z_\alpha\Big),
\end{aligned} \tag{118}$$

where the second equality follows from ν^{*n} being the upper probability of $\mathcal{P}_{\theta_0}^n$, and the third equality follows from Q_0^n being the least favorable null distribution by Theorem 3.1. For each i, let $Z_i \equiv \ln \frac{dQ_1}{dQ_0}(s_i)$. Under Q_0^n, $\{Z_i\}_{i=1}^n$ is an i.i.d. sequence with a finite variance due to $\sigma_{Q_0}^2 < \infty$. Hence, if $\sigma_{Q_0} > 0$, by the

CLT for i.i.d. random variables, one obtains

$$\lim_{n\to\infty} Q_0^n\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\ln\frac{dQ_1}{dQ_0}(s_i) - E_{Q_0}[\ln\frac{dQ_1}{dQ_0}(s_i)]}{\sigma_{Q_0}} > z_\alpha\right) = \Pr(Z > z_\alpha) = \alpha, \tag{119}$$

where $Z \sim N(0,1)$. If $\sigma_{Q_0} = 0$, the summand in (118) is identically 0 and hence the probability of the event is zero and hence $\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_{\theta_0}^n} E_P[\phi_n^*(s^n)] \leq \alpha$. ■

Proof of Theorem 3.2. Let ϕ_n be a level-α test for $H_0 : \varphi(\theta) \leq 0$ against $H_1 : \varphi(\theta) > 0$. Since $\varphi(\theta_0) \leq 0$, ϕ_n is necessarily a level-α test for testing $\theta = \theta_0$ against $\theta_1 = \theta_0 + h/\sqrt{n}$. For any n, the lower power of ϕ_n is then bounded from above by that of the minimax test in Theorem 3.1, which we denote by ϕ_n^* below. Thus,

$$\pi_{n,\theta_{n,h}}(\phi_n) = \inf_{P^n\in\mathcal{P}_{\theta_{n,h}}^n} E_{P^n}[\phi_n] \leq \inf_{P^n\in\mathcal{P}_{\theta_{n,h}}^n} E_{P^n}[\phi_n^*] = \pi_{\theta_{n,h}}(\phi_n^*). \tag{120}$$

Let $j \in \mathbb{J}$ and let $h \in \mathcal{T}_j(\theta_0)$. By Assumption 3.2, the LFP $(Q_{0,j,\tau h}, Q_{1,j,\tau h}) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_0+\tau h}$ satisfies

$$Q_{0,j,\tau h} = \mathsf{Q}_{j,\theta_0}, \quad \text{and} \quad Q_{1,j,\tau h} = \mathsf{Q}_{j,\theta_0+\tau h}, \quad \text{for all } 0 < \tau \leq \bar{\tau}. \tag{121}$$

This and Theorem 3.1 imply, for each $j \in \mathbb{J}$ and $h \in \mathcal{T}_j(\theta_0)$,

$$\pi_{n,\theta_0+h/\sqrt{n}}(\phi_n^*) = \int \phi_n^* d\nu_{\theta_0+h/\sqrt{n}}^n = \int \phi_n^* d\mathsf{Q}_{j,\theta_0+h/\sqrt{n}}^n, \tag{122}$$

for n sufficiently large. Hence, it suffices to analyze the asymptotic power under $\{\mathsf{Q}_{j,\theta_0+h/\sqrt{n}}^n\}$.

The underlying model $\theta \mapsto \mathsf{Q}_{j,\theta}$ is L^2 differentiable tangentially to $\mathcal{T}_j(\theta_0)$. By Lemma 25.14 in Van der Vaart (2000), the log-likelihood ratio of the LFP can be expanded as

$$L_n = \ln\frac{d\mathsf{Q}_{j,\theta_0+h/\sqrt{n}}^n}{d\mathsf{Q}_{j,\theta_0}^n} = h'\Delta_{j,n} - \frac{1}{2}h'C_j h + o_{\mathsf{Q}_{j,\theta_0}^n}(1), \tag{123}$$

where $\Delta_{j,n} = n^{-1/2}\sum_{i=1}^n \dot{\ell}_{j,\theta_0}(s_i)$ and hence $L_n \overset{\mathsf{Q}_{j,\theta_0}^n}{\rightsquigarrow} N(-\frac{\sigma^2}{2}, \sigma^2)$ with $\sigma^2 = h'C_j h$. By Theorem 9.4 of Van der Vaart (2000), the sequence $\mathcal{E}_{j,n}$ of localized experiments in (41) then converges to the Gaussian limit experiment $\mathcal{E}_j$ in (42). This ensures that, for any $j \in \mathbb{J}$ and $h \in \mathcal{T}_j(\theta_0)$, there is a subsequence along which $\pi_{n,\theta_0+h/\sqrt{n}}(\phi_n^*) \to \pi_h$, where π_h is a power function in the Gaussian limit experiment (see the proof of Theorem 15.1 in Van der Vaart, 2000). For any $h \in \mathcal{T}_j(\theta_0)$ with $\dot{\varphi}_{\theta_0}h < 0$, we have $\varphi(\theta_0 + h/\sqrt{n}) < 0$ for all n sufficiently large. Hence, by $\theta_0 + h/\sqrt{n}$ satisfying the null for all n sufficiently large and ϕ_n^* being level-α,

$$\pi_h \leq \limsup_{n\to\infty} \pi_{n,\theta_0+h/\sqrt{n}}(\phi_n^*) \leq \alpha. \tag{124}$$

By continuity, $\pi_h \leq \alpha$ for all h such that $\dot{\varphi}_{\theta_0}h \leq 0$. This, in turn, implies that π_h is a power function of a level-α test for testing $H_0 : \dot{\varphi}_{\theta_0}h \leq 0$ against $H_1 : \dot{\varphi}_{\theta_0}h > 0$ in $\mathcal{E}_j$, and hence it is bounded by the power of the uniformly most powerful test. The rest of the proof parallels the proof of Theorem 15.4 in Van der Vaart (2000) if C_j is non-singular and the tangent set is a linear subspace. The first claim of the theorem

6

then holds with $\tilde{\varrho}_j = \dot{\varphi}_{\theta_0} C_j^{-1} \dot{\ell}_{j,\theta_0}$.

In case C_j is singular or the tangent set is a convex cone (not a linear subspace), we follow the argument in the proof of Theorem 2.1 in Rieder (2014). Let $h \in \mathcal{T}_j(\theta_0)$ be a vector such that $\dot{\varphi}_{\theta_0} h = c > 0$. We rewrite it as $h = \tau a$, where $\tau > 0$ and a is a unit vector. We then let $g = a' \dot{\ell}_{j,\theta_0} \in \mathcal{G}_{j,\theta_0}$. Note that, by the definition of ϱ_j,

$$\dot{\varphi}_{\theta_0} a = \langle \varrho_j, g \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}}, \tag{125}$$

and hence $\tau = c / \langle \varrho_j, g \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}}$. One may now rewrite (123) as

$$L_n = \ln \frac{d\mathsf{Q}^n_{j,\theta_0+h/\sqrt{n}}}{d\mathsf{Q}^n_{j,\theta_0}} = \tau \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(s_i) - \frac{\tau^2}{2} \|g\|_{L^2_{\mathsf{Q}_{j,\theta_0}}} + o_{\mathsf{Q}^n_{j,\theta_0}}(1). \tag{126}$$

By Corollary 3.4.2 in Rieder (1994), the asymptotic power of any test satisfying (124) is then dominated by $1 - \Phi(z_\alpha - \tau \|g\|_{L^2_{\mathsf{Q}_{j,\theta_0}}})$. Now let $g \to \tilde{\varrho}_j$ in $L^2_{\mathsf{Q}_{j,\theta_0}}(S)$. Then,

$$\tau \|g\|_{L^2_{\mathsf{Q}_{j,\theta_0}}} = \frac{c \|g\|_{L^2_{\mathsf{Q}_{j,\theta_0}}}}{\langle \varrho_j, g \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}}} \to \frac{c \|\tilde{\varrho}_j\|_{L^2_{\mathsf{Q}_{j,\theta_0}}}}{\langle \varrho_j, \tilde{\varrho}_j \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}}} = \frac{c}{\|\tilde{\varrho}_j\|_{L^2_{\mathsf{Q}_{j,\theta_0}}}}, \tag{127}$$

where the last equality follows from $\langle \varrho_j, \tilde{\varrho}_j \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}} = \|\tilde{\varrho}_j\|^2_{L^2_{\mathsf{Q}_{j,\theta_0}}}$ due to $\tilde{\varrho}_j$ being the projection of ϱ_j to $cl(\mathcal{G}_{j,\theta_0})$. Therefore, the power bound is obtained as the following limit

$$\lim_{g \to \tilde{\varrho}_j} 1 - \Phi\left(z_\alpha - \tau \|g\|_{L^2_{\mathsf{Q}_{j,\theta_0}}}\right) = 1 - \Phi\left(z_\alpha - \frac{c}{\|\tilde{\varrho}_j\|_{L^2_{\mathsf{Q}_{j,\theta_0}}}}\right). \tag{128}$$

The first claim of the theorem then follows from noting that $c = \dot{\varphi}_{\theta_0} h = \langle \varrho_j, h' \dot{\ell}_{j,\theta_0} \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}}$.

For the second claim, let $h \in \mathcal{T}_j(\theta_0)$. Then, by Le Cam's third lemma,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\varrho}_j(s_i) \overset{\mathsf{Q}^n_{j,\theta_{n,h}}}{\rightsquigarrow} N\left(\langle \tilde{\varrho}_j, h' \dot{\ell}_{j,\theta_0} \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}}, \|\tilde{\varrho}_j\|^2_{L^2_{\mathsf{Q}_{j,\theta_0}}}\right). \tag{129}$$

Therefore,

$$\lim_{n \to \infty} \pi_{n,\theta_0+h/\sqrt{n}}(\phi^*_{j,n}) = 1 - \Phi\left(z_\alpha - \frac{\langle \tilde{\varrho}_j, h' \dot{\ell}_{\theta_0} \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}}}{\|\tilde{\varrho}_j\|_{L^2_{\mathsf{Q}_{j,\theta_0}}}}\right) \geq 1 - \Phi\left(z_\alpha - \frac{\langle \varrho_j, h' \dot{\ell}_{\theta_0} \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}}}{\|\tilde{\varrho}_j\|_{L^2_{\mathsf{Q}_{j,\theta_0}}}}\right), \tag{130}$$

where the inequality follows from $\langle \tilde{\varrho}_j, h' \dot{\ell}_{\theta_0} \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}} \geq \langle \varrho_j, h' \dot{\ell}_{\theta_0} \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}}$ by $\langle \tilde{\varrho}_j, g \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}} \geq \langle \varrho_j, g \rangle_{L^2_{\mathsf{Q}_{j,\theta_0}}}$ for any $g \in \mathcal{G}_{j,\theta_0}$ due to $\tilde{\varrho}_j$ being the projection of ϱ_j to $cl(\mathcal{G}_{j,\theta_0})$. This establishes the claim of the theorem. ∎

Proof of Theorem 3.3. Let $j \in \mathbb{J}$. Consider $h \in \mathcal{T}_j(\theta_0, \xi)$. By Assumption 3.3, for any $\tau \in (0, \bar{\tau}]$, the LFP $(Q_{0,j,\tau h}, Q_{1,j,\tau h}) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_0 + \xi + \tau h}$ satisfies

$$Q_{0,j,\tau h} = \mathsf{Q}_{j,0}, \quad \text{and} \quad Q_{1,j,\tau h} = \mathsf{Q}_{j,\tau h}, \tag{131}$$

where $\vartheta \mapsto Q_{j,\vartheta}$ is L^2 differentiable tangentially to $\mathcal{T}_j(\theta_0, \xi)$. The rest of the argument then parallels that of Proof of Theorem 3.2. ∎

B.4 Proof of Theorems in Section 5

In what follows, we repeatedly use the fact that, for any nonnegative measurable function g on S, belief function ν and its conjugate ν^*, one has

$$\int g(s)d\nu^*(s) = \int \max_{s \in K} g(s) dM_\nu \tag{132}$$

$$\int g(s)d\nu(s) = \int \min_{s \in K} g(s) dM_\nu, \tag{133}$$

where M_ν is the probability measure on $\mathcal{K}(S)$ associated with ν (see Lemma A.1).

Proof of Lemma 5.1. We first start with showing (84) and (85). For this, observe that

$$R(\theta, \phi) = \max_{P \in \mathcal{P}_\theta} \int \phi(s)I_{\Theta_0}(\theta) + \zeta(1 - \phi(s))I_{\Theta_1}(\theta)dP(s)$$

$$= \max_{P \in \mathcal{P}_\theta}(I_{\Theta_0}(\theta) - \zeta I_{\Theta_1}(\theta)) \int \phi(s)dP(s) + \zeta I_{\Theta_1}(\theta)$$

$$= I_{\Theta_0}(\theta) \max_{P \in \mathcal{P}_\theta} \int \phi(s)dP(s) - \zeta I_{\Theta_1}(\theta) \min_{P \in \mathcal{P}_\theta} \int \phi(s)dP(s) + \zeta I_{\Theta_1}(\theta)$$

$$= \int \phi(s)d\nu_\theta^*(s)I_{\Theta_0}(\theta) + \zeta(1 - \int \phi(s)d\nu_\theta(s))I_{\Theta_1}(\theta), \tag{134}$$

where the third equality follows from the fact that $I_{\Theta_0}(\theta) - \zeta I_{\Theta_1}(\theta) > 0$ if and only if $I_{\Theta_0}(\theta) = 1$ (and $I_{\Theta_0}(\theta) - \zeta I_{\Theta_1}(\theta) \leq 0$ if and only if $I_{\Theta_1}(\theta) = 1$). The last equality follows from $core(\nu_\theta) = \mathcal{P}_\theta$ by Theorem 3 in Philippe, Debs, and Jaffray (1999) and the fact that, for any nonnegative bounded function g on S, $\int gd\nu \leq \int gdP \leq \int gd\nu^*$ for all $P \in core(\nu)$. Using this, write

$$r(\mu, \phi) \equiv \int_\Theta R(\theta, \phi)d\mu(\theta)$$

$$= \tau \int \int \phi(s)I_{\Theta_0}(\theta)d\nu_\theta^*(s)d\mu_0(\theta) + \zeta(1 - \tau)(1 - \int \int \phi(s)I_{\Theta_1}(\theta)d\nu_\theta(s)d\mu_1(\theta)). \tag{135}$$

By Lemma A.1, for each ν_θ, there is a unique Borel probability measure M_θ on $\mathcal{K}(S)$ such that

$$\nu_\theta(A) = M_\theta(K \subset A), \ \forall A \subset S. \tag{136}$$

Let q be a σ-finite measure on $\mathcal{K}(S)$ such that $M_\theta \ll q$. Let M_{κ_1} be a Borel probability measure on $\mathcal{K}(S)$

such that $dM_{\kappa_1}/dq = \int_{\Theta_1} \frac{dM_\theta}{dq} d\mu_1(\theta)$. For any $A \subset S$, it then follows that

$$
\begin{aligned}
\int_{\Theta_1} \nu_\theta(A) d\mu_1(\theta) &= \int_{\Theta_1} M_\theta(K \subset A) d\mu_1(\theta) \\
&= \int_{\Theta_1} \int_{\mathcal{K}(S)} 1\{K \subset A\} dM_\theta d\mu_1(\theta) \\
&= \int_{\mathcal{K}(S)} 1\{K \subset A\} \int_{\Theta_1} \frac{dM_\theta}{dq} d\mu_1(\theta) dq(K) \\
&= \int_{\mathcal{K}(S)} 1\{K \subset A\} dM_{\kappa_1}(K) \\
&= \int_S 1\{s \in A\} d\kappa_1(s) \\
&= \kappa_1(A),
\end{aligned}
\tag{137}
$$

where the third equality follows from Fubini's theorem. The existence and uniqueness of κ_1 follows again from Choquet's theorem (Lemma A.1). Note that by $\phi \geq 0$ and the definition of the Choquet integral, one can show by the same argument

$$
\begin{aligned}
\int \int \phi(s) I_{\Theta_1}(\theta) d\nu_\theta(s) d\mu_1(\theta) &= \int_{\Theta_1} \int \inf_{s \in K} \phi(s) dM_\theta(K) d\mu_1(\theta) \\
&= \int \inf_{s \in K} \phi(s) \int_{\Theta_1} \frac{dM_\theta}{dq} d\mu_1(\theta) dq(K) \\
&= \int \phi(s) d\kappa_1(s),
\end{aligned}
\tag{138}
$$

where the second equality follows from Fubini's theorem. Similarly, it follows that

$$
\int \int \phi(s) I_{\Theta_0}(\theta) d\nu_\theta^*(s) d\mu_0(\theta) = \int \phi(s) d\kappa_0^*(s).
\tag{139}
$$

By (135), (138), and (139), we have

$$
r(\mu, \phi) = \tau \int \phi(s) d\kappa_0^*(s) + \zeta(1-\tau)\left(1 - \int \phi(s) d\kappa_1(s)\right).
\tag{140}
$$

Therefore, (85) holds. Minimizing the BDS risk is then equivalent to minimizing

$$
\tilde{r}_t(\mu, \phi) = t \int \phi(s) d\kappa_0^*(s) - \int \phi(s) d\kappa_1(s),
\tag{141}
$$

where $t = \tau/\zeta(1-\tau) > 0$. Let $A \equiv \{s : \phi(s) > 0\}$. Minimizing the risk function above with respect to ϕ is then equivalent to minimizing the 2-alternating function $w_t(A) \equiv t\kappa_0^*(A) - \kappa_1(A)$ with respect to $A \subset S$. By Lemmas 3.1 and 3.2 in HS, for each $t \in [0, \infty]$, there exists a set $A_t \subset S$ such that

$$
w_t(A_t) = \inf_{A \subset S} w_t(A),
\tag{142}
$$

and $\{A_t, t \geq 0\}$ forms an increasing family of sets. Now define $\Lambda(s) \equiv \inf\{t | s \in A_t\}$. By Theorem 4.1 in HS, the conclusion of the theorem then follows. ∎

Proof of Theorem 5.1. Note that μ_1 is fixed throughout and $\mathcal{M} = \{\mu : \mu = \tau\mu_0 + (1-\tau)\mu_1, \mu_0 \in \Delta(\Theta_0), \tau \in$

$[0, 1]\}$. In what follows, we therefore redefine R in (84) as

$$R(\theta, \phi) = \int \phi(s) d\nu_\theta^*(s) I_{\Theta_0}(\theta) + \zeta(1 - \int \phi(s) d\kappa_1(s)) I_{\Theta_1}(\theta)$$
$$= R_0(\theta, \phi) I_{\Theta_0}(\theta) + R_1(\phi) I_{\Theta_1}(\theta), \tag{143}$$

where $\kappa_1 = \int_{\Theta_1} \nu_\theta d\mu_1(\theta)$. First, we show $\sup_{\mu \in \mathcal{M}} \inf_{\phi \in \mathbf{\Phi}} \int R(\theta, \phi) d\mu \leq \inf_{\phi \in \mathbf{\Phi}} \sup_{\theta \in \Theta} R(\theta, \phi)$. This follows because for any (θ, ϕ), one has $\inf_{\phi'} R(\theta, \phi') \leq R(\theta, \phi) \leq \sup_{\theta'} R(\theta', \phi)$, and hence

$$\sup_\theta \inf_{\phi'} R(\theta, \phi') \leq \inf_\phi \sup_{\theta'} R(\theta', \phi).$$

Note that $\sup_\theta \inf_{\phi'} R(\theta, \phi') \geq \inf_{\phi \in \mathbf{\Phi}} \int R(\theta, \phi) d\mu$ for any μ, and hence, the first claim follows.

The other direction follows from Lemma C.1. To see this, let

$$\beta \equiv \sup_{\mu \in \mathcal{M}} \inf_{\phi \in \mathbf{\Phi}} \int R(\theta, \phi) d\mu. \tag{144}$$

If $\beta = \infty$, the result is trivial. If $\beta < \infty$, set $f(\theta) = \beta$ for all $\theta \in \Theta$. By construction, $f(\theta) \geq \inf_{\phi \in \mathbf{\Phi}} \int R(\theta, \phi) d\mu$ for every μ. By Lemma C.1, this is equivalent to the existence of $\phi^\dagger \in \mathbf{\Phi}$ such that $\beta \geq R(\theta, \phi^\dagger)$, $\forall \theta \in \Theta$. This implies

$$\beta \geq \sup_{\theta \in \Theta} R(\theta, \phi^\dagger) \geq \inf_{\phi \in \mathbf{\Phi}} \sup_{\theta \in \Theta} R(\theta, \phi). \tag{145}$$

Finally, observe that by (143), $\sup_{\theta \in \Theta} R(\theta, \phi) = \sup_{\theta \in \Theta_0} R_0(\theta, \phi) \vee R_1(\phi)$. This completes the proof. ∎

C   Auxiliary Lemmas

Below, we identify each decision function (randomized test) ϕ with a Markov kernel $\phi : S \times \mathcal{B}_{\{0,1\}} \to [0, 1]$ and let $\mathbf{\Phi}$ be the set of all decision functions. We then equip $\mathbf{\Phi}$ with the weak topology (see Häusler and Luschgy, 2015, Definition 2.2). The following lemma is an analog of Lemma 46.1 in Strasser (1985) for the BDS risk. We state it as a lemma because the BDS risk R is defined through Choquet integrals with respect to capacities (instead of measures) and hence Lemma 46.1 in Strasser (1985) is not directly applicable.[46]

Lemma C.1. *Suppose S is finite and Θ is compact. Let R be defined as in (84). For every $f : \Theta \to \mathbb{R}$, the following assertions are equivalent.*

(i) There exists $\phi \in \mathbf{\Phi}$ such that $f(\theta) \geq R(\theta, \phi)$ for every $\theta \in \Theta$.

(ii) $\int f d\mu \geq \inf_{\phi \in \mathbf{\Phi}} \int R(\theta, \phi) d\mu(\theta)$ for every $\mu \in \Delta(\Theta)$.

Proof. The implication $(i) \Rightarrow (ii)$ is obvious. We therefore prove the other implication. Consider the

---

[46]They also show their results to generalized decision functions, which we do not pursue here.

following sets of functions:

$$M_1 = \{f\}, \quad M_2 = \{h \in \mathcal{C}(\Theta) : h(\cdot) = R(\cdot, \phi), \phi \in \mathbf{\Phi}\}.$$

We mimic the proof of Lemma 46.1 in Strasser (1985) while replacing M_2 with the set above. For this, let $M \subseteq \mathcal{C}(\Theta)$ be an arbitrary set. For any $m \in \Delta(\Theta)$, the lower envelope of M is defined as

$$\psi_M(m) \equiv \inf\{\int f dm, f \in M\}. \tag{146}$$

Also define $\alpha(M) \equiv \bigcup_{f \in M}\{g \in \mathcal{C}(\Theta) : f \le g\}$. This is the set of continuous functions that dominate some function in M. Since M_2 is compact by Lemma C.3, $\alpha(M_2)$ is closed and hence coincides with its closure $\overline{\alpha(M_2)}$ (Strasser, 1985, Remark 45.4).

By (ii), $\psi_{M_2}(m) \le \psi_{M_1}(m)$ for all $m \in \Delta(\Theta)$. By Lemma C.2, M_2 is subconvex. Then, by Theorem 45.6 in Strasser (1985), for every $f \in M_1$, there is $g \in \alpha(M_2) = \overline{\alpha(M_2)}$ such that $g \le f$. By the construction of $\alpha(M_2)$, this means there exists $\phi \in \mathbf{\Phi}$ such that

$$R(\cdot, \phi) \le g(\cdot) \le f(\cdot). \tag{147}$$

This completes the proof. ∎

Below, a set M is said to be *subconvex*, if for any $\alpha \in (0, 1)$ and $h_1, h_2 \in M$, there exists $h_3 \in M$ such that $h_3 \le \alpha h_1 + (1 - \alpha)h_2$.

Lemma C.2. M_2 *is subconvex.*

Proof. Let $h_1, h_2 \in M_2$. Then, there exist $\phi_1, \phi_2 \in \mathbf{\Phi}$ such that $h_j(\cdot) = R(\cdot, \phi_j), j = 1, 2$. Therefore, for any $\alpha \in (0, 1)$,

$$
\begin{aligned}
\alpha h_1(\theta) + (1 - \alpha)h_2(\theta) =& \alpha R(\theta, \phi_1) + (1 - \alpha)R(\theta, \phi_2) \\
=& \zeta\left(\alpha \int \phi_1(s)d\nu_\theta^*(s) + (1 - \alpha)\int \phi_2(s)d\nu_\theta^*(s)\right)I_{\Theta_0}(\theta) \\
& + \left(\alpha(1 - \int \phi_1(s)d\nu_\theta(s)) + (1 - \alpha)(1 - \int \phi_2(s)d\nu_\theta(s))\right)I_{\Theta_1}(\theta). \tag{148}
\end{aligned}
$$

Note that ν_θ^* is 2-alternating. Therefore, the Choquet integral with respect to ν_θ^* is subadditive. The Choquet integral is also positively homogeneous. Therefore,

$$
\begin{aligned}
\alpha \int \phi_1(s)d\nu_\theta^*(s) + (1 - \alpha)\int \phi_2(s)d\nu_\theta^*(s) =& \int \alpha\phi_1(s)d\nu_\theta^*(s) + \int (1 - \alpha)\phi_2(s)d\nu_\theta^*(s) \\
\ge& \int \alpha\phi_1(s) + (1 - \alpha)\phi_2(s)d\nu_\theta^*(s). \tag{149}
\end{aligned}
$$

Similarly, by the 2-monotonicity of ν_θ, the Choquet integral with respect to it is superadditive and positively homogeneous. Therefore,

$$\alpha(1 - \int \phi_1(s)d\nu_\theta(s)) + (1 - \alpha)(1 - \int \phi_2(s)d\nu_\theta(s)) \ge 1 - \int \alpha\phi_1(s) + (1 - \alpha)\phi_2(s)d\nu_\theta(s). \tag{150}$$

Combining (148)-(150), we obtain $\alpha h_1(\cdot) + (1 - \alpha)h_2(\cdot) \geq h_3(\cdot)$, where $h_3(\cdot) = R(\cdot, \phi_3)$ with $\phi_3 = \alpha\phi_1 + (1 - \alpha)\phi_2$. Hence, $h_3 \in M_2$. Conclude that M_2 is subconvex. $\blacksquare$

Lemma C.3. *Suppose that S is finite and Θ is compact. Then, M_2 is weakly compact.*

Proof. Equip M_2 with the weak topology. Let ϱ be the counting measure. We let $\boldsymbol{\Phi} = K^1(\varrho)$ denote the set of Markov kernels equipped with the weak topology. It is the coarsest topology that makes any functional of the following form continuous:

$$T(\phi) = \int_S \int_{\{0,1\}} h(a)\phi(s, da)f(s)d\varrho(s), \ f \in L^1_\varrho(S), h \in \mathcal{C}_b(\{0,1\}). \tag{151}$$

By Theorem 2.7 in Häusler and Luschgy (2015), $\boldsymbol{\Phi}$ is compact if the set $\varrho\boldsymbol{\Phi} \equiv \{\varrho\phi : \varrho\phi(\cdot) = \int_S \phi(s, \cdot)d\varrho(s), \phi \in \boldsymbol{\Phi}\}$ is relatively compact in $\Delta(\{0, 1\})$. Note that $\{0, 1\}$ is compact. Hence, $\varrho\boldsymbol{\Phi}$ is uniformly tight, which implies that $\varrho\boldsymbol{\Phi}$ is relatively compact by Prohorov's theorem (van der Vaart and Wellner, 1996, Problem 1.12.1). This ensures the compactness of $\boldsymbol{\Phi}$.

Below, let $\mathcal{K}(S)$ the set of all nonempty (and necessarily closed) subsets of S. By Choquet's theorem, a belief function ν_θ can be expressed by its canonical representation $(\mathcal{K}(S), K, \hat{m}_\theta)$, where K is a random set following a probability measure $\hat{m}_\theta$ on $\mathcal{K}(S)$ such that $\nu_\theta(A) = \hat{m}_\theta(K \subset A)$ for all $A \in \mathcal{K}(S)$. Below, we adopt this canonical representation and also denote the measure on $\mathcal{K}(S)$ by m_θ rather than $\hat{m}_\theta$. We also note that we write $\phi(s) = \int_{\{0,1\}} \phi(s, da)$ in what follows.

Define $g : \boldsymbol{\Phi} \to \mathcal{C}(\Theta)$ pointwise by $\phi \mapsto R(\cdot, \phi)$. We argue that this map is continuous, where we equip $\boldsymbol{\Phi}$ and $\mathcal{C}(\Theta)$ with weak topologies. Let $\phi_n, n = 1, 2, \cdots$ be a sequence such that $\phi_n \to \phi \in \boldsymbol{\Phi}$ weakly.

$$\int_S \int_{\{0,1\}} \phi_n(s, da)d\nu^*_\theta(s) = \int_S \phi_n(s)d\nu^*_\theta(s) = \int_S \max_{s \in K} \phi_n(s)dm_\theta(K) \tag{152}$$

Since S is finite and $1\{s = s'\} \in L^1_\varrho(S)$ for any $s' \in S$, ϕ_n converging weakly implies

$$\phi_n(s') = \sum_{s \in S} \phi_n(s)1\{s = s'\} \to \sum_{s \in S} \phi(s)1\{s = s'\} = \phi(s'), \ \forall s' \in S. \tag{153}$$

Therefore, ϕ_n converges pointwise to ϕ.

Fix $K \in \mathcal{K}(S)$. Note that $(s, n) \mapsto \phi_n(s)$ is continuous with respect to the discrete topology. By Berge's maximum theorem, $\max_{s \in K} \phi_n(s) \to \max_{s \in K} \phi(s)$. Hence,

$$\begin{aligned}
\lim_{n\to\infty} \int_S \int_{\{0,1\}} \phi_n(s, da)d\nu^*_\theta(s) &= \lim_{n\to\infty} \int_S \max_{s \in K} \phi_n(s)dm_\theta(K) \\
&= \int_S \lim_{n\to\infty} \max_{s \in K} \phi_n(s)dm_\theta(K) \\
&= \int_S \max_{s \in K} \phi(s)dm_\theta(K) \\
&= \int_S \phi(s)d\nu^*_\theta(s) \\
&= \int_S \int_{\{0,1\}} \phi(s, a)d\nu^*_\theta(s). \tag{154}
\end{aligned}$$

12

where the second equality follows from the convergence of $\max_{s \in K} \phi_n(s)$ and the dominated convergence theorem.

Consider any finite Borel measure μ on Θ. The result above and the dominated convergence theorem imply

$$\lim_{n \to \infty} \int_\Theta \int_S \int_{\{0,1\}} \phi_n(s, da) d\nu_\theta^*(s) I_{\Theta_0}(\theta) d\mu(\theta) = \int_\Theta \lim_{n \to \infty} \int_S \int_{\{0,1\}} \phi_n(s, da) d\nu_\theta^*(s) I_{\Theta_0}(\theta) d\mu(\theta) \qquad (155)$$

$$= \int_\Theta \int_S \int_{\{0,1\}} \phi(s, da) d\nu_\theta^* I_{\Theta_0}(\theta) d\mu(\theta). \qquad (156)$$

By a similar argument, one can also show

$$\lim_{n \to \infty} \int_\Theta \int_S \int_{\{0,1\}} \phi_n(s, da) d\nu_\theta(s) I_{\Theta_1}(\theta) d\mu(\theta) = \int_\Theta \lim_{n \to \infty} \int_S \int_{\{0,1\}} \phi_n(s, da) d\nu_\theta(s) I_{\Theta_1}(\theta) d\mu(\theta) \qquad (157)$$

$$= \int_\Theta \int_S \int_{\{0,1\}} \phi(s, da) d\nu_\theta I_{\Theta_1}(\theta) d\mu(\theta). \qquad (158)$$

Note that Θ is a compact set in a metric space. Corollary 14.15 in Aliprantis and Border (2006) then ensures that the dual space of $\mathcal{C}(\Theta)$ is the set of finite Borel measures on Θ. Combining these results and noting that $R(\theta, \phi) = \zeta \int_S \int_{\{0,1\}} \phi(s, da) d\nu_\theta^* I_{\Theta_0}(\theta) + (1 - \int_S \int_{\{0,1\}} \phi(s, da) d\nu_\theta) I_{\Theta_1}(\theta)$, it follows that $R(\cdot, \phi_n) \to R(\cdot, \phi)$ in $\mathcal{C}(\Theta)$ with respect to the weak topology. This establishes that g is continuous. Hence, M_2 is the continuous image of a compact set. Conclude that M_2 is weakly compact. ■

D   Examples

D.1   Example 1: Binary response game

In this section, we provide details on Example 1 including the computation of the belief function, least favorable pair, and minimax test. Recall that $S = \{(0,0), (1,1), (1,0), (0,1)\}$. There exist 14 subsets to be considered (without considering the empty set and S). One can then compute the the lower and upper bounds of the probability of each event by mimicking the calculation in (20). The results are summarized in Table 2.

D.1.1   LFP

The following proposition characterizes the LFP and minimax tests.

Proposition D.1. *Let (S, U, Θ, G) be as defined in Example 1. Suppose that u follows the bivariate standard normal distribution. Let $\alpha \in (0, 1/4)$, $\theta_0 = (0,0)'$, and $\theta_1 < 0$. Then, for any $j \in \mathbb{J} = \{I, II, III\}$, the density of Q_0 is $(q_0(0,0), q_0(1,1), q_0(1,0), q_0(0,1)) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. The densities of Q_1 for $\theta_1 \in \Theta_j, j \in \mathbb{J}$ and minimax tests are as in Table 1.*

Proof of Proposition D.1. First, observe that the upper and lower probabilities of $A_1, \cdots, A_4$ fully charac-

Table 2: The lower and upper probability bounds

A	$\nu_\theta(A) = \min P(A)$	$\nu_\theta^* = \max P(A)$
$A_1 = \{(0,0)\}$	$\frac{1}{4}$	$\frac{1}{4}$
$A_2 = \{(1,1)\}$	$\Phi(\theta^{(1)})\Phi(\theta^{(2)})$	$\Phi(\theta^{(1)})\Phi(\theta^{(2)})$
$A_3 = \{(0,1)\}$	$\frac{1}{4} - \Phi(\theta^{(1)})\Phi(\theta^{(2)}) + \frac{\Phi(\theta^{(2)})}{2}$	$\frac{1}{2}(1 - \Phi(\theta^{(1)}))$
$A_4 = \{(1,0)\}$	$\frac{1}{4} - \Phi(\theta^{(1)})\Phi(\theta^{(2)}) + \frac{\Phi(\theta^{(1)})}{2}$	$\frac{1}{2}(1 - \Phi(\theta^{(2)}))$
$A_5 = \{(0,0),(1,1)\}$	$\frac{1}{4} + \Phi(\theta^{(1)})\Phi(\theta^{(2)})$	$\frac{1}{4} + \Phi(\theta^{(1)})\Phi(\theta^{(2)})$
$A_6 = \{(0,0),(0,1)\}$	$\frac{1}{2} - \Phi(\theta^{(1)})\Phi(\theta^{(2)}) + \frac{\Phi(\theta^{(2)})}{2}$	$\frac{3}{4} - \frac{\Phi(\theta^{(1)})}{2}$
$A_7 = \{(0,0),(1,0)\}$	$\frac{1}{2} - \Phi(\theta^{(1)})\Phi(\theta^{(2)}) + \frac{\Phi(\theta^{(1)})}{2}$	$\frac{3}{4} - \frac{\Phi(\theta^{(2)})}{2}$
$A_8 = A_6^c = \{(1,1),(1,0)\}$	$\frac{1}{4} + \frac{\Phi(\theta^{(1)})}{2}$	$\frac{1}{2}(1 - \Phi(\theta^{(2)})) + \Phi(\theta^{(1)})\Phi(\theta^{(2)})$
$A_9 = A_7^c = \{(1,1),(0,1)\}$	$\frac{1}{4} + \frac{\Phi(\theta^{(2)})}{2}$	$\frac{1}{2}(1 - \Phi(\theta^{(1)})) + \Phi(\theta^{(1)})\Phi(\theta^{(2)})$
$A_{10} = A_5^c = \{(1,0),(0,1)\}$	$\frac{3}{4} - \Phi(\theta^{(1)})\Phi(\theta^{(2)})$	$\frac{3}{4} - \Phi(\theta^{(1)})\Phi(\theta^{(2)})$
$A_{11} = A_1^c = \{(1,1),(1,0),(0,1)\}$	$\frac{3}{4}$	$\frac{3}{4}$
$A_{12} = A_2^c = \{(0,0),(1,0),(0,1)\}$	$1 - \Phi(\theta^{(1)})\Phi(\theta^{(2)})$	$1 - \Phi(\theta^{(1)})\Phi(\theta^{(2)})$
$A_{13} = A_3^c = \{(0,0),(1,1),(1,0)\}$	$\frac{1}{2} + \frac{\Phi(\theta^{(1)})}{2}$	$\frac{3}{4} - \frac{\Phi(\theta^{(2)})}{2} + \Phi(\theta^{(1)})\Phi(\theta^{(2)})$
$A_{14} = A_4^c = \{(0,0),(1,1),(0,1)\}$	$\frac{1}{2} + \frac{\Phi(\theta^{(2)})}{2}$	$\frac{3}{4} - \frac{\Phi(\theta^{(1)})}{2} + \Phi(\theta^{(1)})\Phi(\theta^{(2)})$
$S = \{(0,0),(1,1),(1,0),(0,1)\}$	1	1

terize the constraints in the convex program. To see this, observe that, for example,

$$\nu_\theta(A_5) = m_\theta(G(u|\theta) \subset \{(0,0),(1,1)\})$$

$$= m_\theta(G(u|\theta) = \{(0,0)\}) + m_\theta(G(u|\theta) = \{(1,1)\}) = \nu_\theta(A_1) + \nu_\theta(A_2) = \frac{1}{4} + \Phi(\theta^{(1)})\Phi(\theta^{(2)}), \quad (159)$$

where we note that the additivity of ν_θ for this event is due to the form of the correspondence in (2) and does not hold in general. One can compute $\nu_\theta(A_6)$ and $\nu_\theta(A_7)$ similarly. The upper bounds $\nu_\theta^*(A_j)$ for $j = 8, \cdots, 14$ can then be computed using the conjugacy of ν_θ and ν_θ^*. Similarly, the upper bounds $\nu_\theta^*(A_j)$ for $j = 1, \cdots, 7$ imply the lower bounds $\nu_\theta(A_j)$ for $j = 8, \cdots, 14$.

In sum, it suffices to impose the constraints that arise from the upper and lower probabilities of $A_1, \cdots, A_4$. Further, $q_0(0,0) = q_1(0,0) = 1/4$ regardless of the parameter value. This allows to simplify the convex program as

$$\min_{(q_0,q_1)} - \ln\left(\frac{q_0(1,1)}{q_0(1,1) + q_1(1,1)}\right)(q_0(1,1) + q_1(1,1)) - \ln\left(\frac{q_0(1,0)}{q_0(1,0) + q_1(1,0)}\right)(q_0(1,0) + q_1(1,0))$$

$$- \ln\left(\frac{q_0(0,1)}{q_0(0,1) + q_1(0,1)}\right)(q_0(0,1) + q_1(0,1))$$

$$\text{s.t.} \quad \frac{1}{4} - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) + \frac{\Phi(\theta_1^{(2)})}{2} \leq q_1(0,1) \leq \frac{1}{2}(1 - \Phi(\theta_1^{(1)})) \quad (160)$$

$$\frac{1}{4} - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) + \frac{\Phi(\theta_1^{(1)})}{2} \leq q_1(1,0) \leq \frac{1}{2}(1 - \Phi(\theta_1^{(2)})) \quad (161)$$

$$q_1(1,1) = \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) \quad (162)$$

$$q_1(1,1) + q_1(1,0) + q_1(0,1) = \frac{3}{4} \quad (163)$$

$$q_0(1,1) = q_0(1,0) = q_0(0,1) = \frac{1}{4}. \quad (164)$$

Note that (162)-(164) imply that the values of $q_0(1,1)$, $q_0(1,0)$, $q_0(0,1)$, and $q_1(1,1)$ are determined uniquely. Hence, it remains to optimize the problem with respect to $q_1(1,0)$ and $q_1(0,1)$. For this, let $y = q_1(1,0)$. Then, one may write $q_1(0,1) = 3/4 - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) - y$ due to (163) and (164). Hence, the problem reduces to an optimization problem with a single control variable. Using this, define the Lagrangian by

$$\mathcal{L}(y,\lambda) \equiv -\ln\left(\frac{1/4}{1/4+y}\right)\left(\frac{1}{4}+y\right) - \ln\left(\frac{1/4}{\frac{1}{4}+\frac{3}{4}-\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})-y}\right)\left(\frac{1}{4}+\frac{3}{4}-\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})-y\right)$$
$$- \lambda_1\left(\frac{1}{2}-\frac{\Phi(\theta_1^{(2)})}{2}-y\right) - \lambda_2\left(y-\frac{1}{4}-\frac{\Phi(\theta_1^{(1)})}{2}+\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})\right). \quad (165)$$

By Theorem 28.3 in Rockafellar (1972), the saddle point of the Lagrangian characterizes the optimal solution of the original problem. The Karush-Kuhn-Tucker (KKT) conditions are as follows:

$$1 - \ln\left(\frac{1/4}{1/4+y}\right) - 1 + \ln\left(\frac{1/4}{1-\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})-y}\right) + \lambda_1 - \lambda_2 = 0 \quad (166)$$

$$\lambda_1\left(\frac{1}{2}-\frac{\Phi(\theta_1^{(2)})}{2}-y\right) = 0 \quad (167)$$

$$\lambda_2\left(y-\frac{1}{4}-\frac{\Phi(\theta_1^{(1)})}{2}+\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})\right) = 0 \quad (168)$$

$$\lambda_1, \lambda_2 \geq 0. \quad (169)$$

Below, we consider three cases based on the value of the Lagrange multipliers.

Case 1 ($\lambda_1 = \lambda_2 = 0$): Suppose that $\lambda_1 = 0$ and $\lambda_2 = 0$. Then, the solution from (166) is

$$y = \frac{3}{8} - \frac{\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})}{2}. \quad (170)$$

Substituting this into the complementary slackness conditions (167) and (168) yields

$$\frac{1}{2} - \frac{\Phi(\theta_1^{(2)})}{2} - \frac{3}{8} + \frac{\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})}{2} \geq 0, \quad \frac{3}{8} - \frac{\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})}{2} - \frac{1}{4} - \frac{\Phi(\theta_1^{(1)})}{2} + \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) \geq 0, \quad (171)$$

which can be simplified as

$$\Phi(\theta_1^{(2)})(1-\Phi(\theta_1^{(1)})) \leq \frac{1}{4}, \qquad \Phi(\theta_1^{(1)})(1-\Phi(\theta_1^{(2)})) \leq \frac{1}{4}. \quad (172)$$

By $y = q_1(1,0)$, $q_1(0,1) = 3/4 - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) - y$, and (170), the LFP is

$$(q_0(0,0), q_0(1,1), q_0(1,0), q_0(0,1)) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \quad (173)$$

$$(q_1(0,0), q_1(1,1), q_1(1,0), q_1(0,1)) = \left(\frac{1}{4}, \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}), \frac{3-4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})}{8}, \frac{3-4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})}{8}\right). \quad (174)$$

15

Hence, the likelihood-ratio statistic is given by

$$
\Lambda(s) = \begin{cases}
1 & s = (0,0) \\
4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) & s = (1,1) \\
\frac{3 - 4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})}{2} & s = (1,0) \\
\frac{3 - 4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)})}{2} & s = (0,1).
\end{cases}
\tag{175}
$$

Under Q_0, $\Lambda(s)$ is supported on $\{4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}), 1, (3-4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}))/2\}$ with probabilities $(1/4, 1/4, 1/2)$. For $\theta^{(j)} < 0, j = 1, 2$, one has $4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) < 1 < (3-4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}))/2$. The largest value of the support is therefore $(3 - 4\Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}))/2$. Setting C to this value and solving

$$
\alpha = E_{Q_0}[\phi(s)] = \gamma Q_0(\pi(s) \geq C) = \gamma Q_0\big(s = (1,0) \cup s = (0,1)\big) = \frac{\gamma}{2},
\tag{176}
$$

one obtains $\gamma = 2\alpha$. This gives the level-α minimax test.

Case 2 ($\lambda_1 = 0$ and $\lambda_2 > 0$): Suppose that $\lambda_1 = 0$ and $\lambda_2 > 0$. By the complementary slackness condition (168), the solution is obtained at the lower bound $y = \frac{1}{4} - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) + \frac{\Phi(\theta_1^{(1)})}{2}$. Substituting this into (166) and noting that $\lambda_1 = 0$, we have

$$
\lambda_2 = \ln\left(\frac{\frac{1}{2} - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) + \frac{\Phi(\theta_1^{(1)})}{2}}{\frac{3}{4} - \frac{\Phi(\theta_1^{(1)})}{2}} \right).
\tag{177}
$$

The difference between the numerator and denominator in the logarithm above is

$$
\frac{1}{2} - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) + \frac{\Phi(\theta_1^{(1)})}{2} - \big(\frac{3}{4} - \frac{\Phi(\theta_1^{(1)})}{2}\big) = \Phi(\theta_1^{(1)})(1 - \Phi(\theta_1^{(2)}) - \frac{1}{4}.
\tag{178}
$$

Therefore, $\lambda_2 > 0$ if and only if $\Phi(\theta_1^{(1)})(1 - \Phi(\theta_1^{(2)}) > 1/4$. Similarly, the complementary slackness condition (167) is satisfied with $\lambda_1 = 0$ and $y = \frac{1}{4} - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) + \frac{\Phi(\theta_1^{(1)})}{2}$. Note that the constraint $\frac{1}{2} - \frac{\Phi(\theta_1^{(2)})}{2} - y \geq 0$ is trivially satisfied because

$$
\frac{1}{2} - \frac{\Phi(\theta_1^{(2)})}{2} - y = \frac{1}{2} - \frac{\Phi(\theta_1^{(2)})}{2} - \left(\frac{1}{4} - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) + \frac{\Phi(\theta_1^{(1)})}{2} \right)
\tag{179}
$$

$$
= (\frac{1}{2} - \Phi(\theta_1^{(1)}))(\frac{1}{2} - \Phi(\theta_1^{(2)})) \geq 0,
\tag{180}
$$

where the last inequality follows from $\theta_1^{(j)} \leq 0$ for $j = 1, 2$. Hence, if $\Phi(\theta_1^{(1)})(1 - \Phi(\theta_1^{(2)})) > \frac{1}{4}$, the least favorable pair is

$$
(q_0(0,0), q_0(1,1), q_0(1,0), q_0(0,1)) = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right)
\tag{181}
$$

$$
(q_1(0,0), q_1(1,1), q_1(1,0), q_1(0,1)) = \left(\frac{1}{4}, \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}), \frac{1}{4} - \Phi(\theta_1^{(1)})(\Phi(\theta_1^{(2)}) - \frac{1}{2}), \frac{1}{2}(1 - \Phi(\theta_1^{(1)})) \right),
\tag{182}
$$

where we used $y = q_1(1,0)$, $q_1(0,1) = 3/4 - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) - y$, and $y = \frac{1}{4} - \Phi(\theta_1^{(1)})\Phi(\theta_1^{(2)}) + \frac{\Phi(\theta_1^{(1)})}{2}$. The rest of the analysis is similar to Case 1.

Case 3 ($\lambda_1 > 0$ and $\lambda_2 = 0$): The argument is similar to the one for Case 2. Hence, we omit the proof. ∎

D.1.2   Efficient influence function and optimal tests

Proposition D.2. *Suppose that the conditions of Proposition D.1 hold. Let $p = (p^{(1)}, p^{(2)})'$ with $p^{(j)} < 0$ for $j = 1, 2$ and let $\varphi(\theta) = p'\theta$. Then, Assumption 3.2 holds.*

Proof of Proposition D.2. (i) Observe that $\{\xi \in \Theta - \theta_0 : \varphi(\theta_0 + \xi) > 0\} = \{\xi \in (-\infty, 0]^2 : \xi^{(1)} < 0, \text{ or } \xi^{(2)} < 0\}$, which is indeed a convex cone; (ii) Let $\mathbb{J} = \{I, II, III\}$ and let $\{\mathcal{T}_j(\theta_0), j \in \mathbb{J}\}$ be defined as in (51)-(53). These cones satisfy the requirements in Assumption 3.2 (ii); (iii) Based on Table 1, let $\mathsf{Q}_{I,\theta}$ be a model with the following density

$$(\mathsf{q}_{I,\theta}(0,0), \mathsf{q}_{I,\theta}(1,1), \mathsf{q}_{I,\theta}(1,0), \mathsf{q}_{I,\theta}(0,1))$$
$$= \left(\frac{1}{4}, \Phi(\theta^{(1)})\Phi(\theta^{(2)}), \frac{3 - 4\Phi(\theta^{(1)})\Phi(\theta^{(2)})}{8}, \frac{3 - 4\Phi(\theta^{(1)})\Phi(\theta^{(2)})}{8} \right). \quad (183)$$

Similarly, for $j = II$ and III, let q_j be defined accordingly based on Table 1. By Proposition D.1, for each $j \in \mathbb{J}$, the LFP $(Q_{0,j,\tau h}, Q_{1,j,\tau h}) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_0 + \tau h}$ satisfies

$$Q_{0,j,\tau h} = \mathsf{Q}_{j,\theta_0}, \quad \text{and} \quad Q_{1,j,\tau h} = \mathsf{Q}_{j,\theta_0 + \tau h}, \quad (184)$$

for all $\tau \in (0, \bar{\tau}]$ for some $\bar{\tau} > 0$. Furthermore, $\theta \mapsto \mathsf{Q}_{j,\theta}$ is L^2 differentiable at θ_0 tangentially to $\mathcal{T}_j(\theta_0)$ by Proposition D.3. ∎

Proposition D.3. *Suppose that the conditions of Proposition D.1 hold. Let $p = (p^{(1)}, p^{(2)})'$ with $p^{(j)} < 0$ for $j = 1, 2$ and let $\varphi(\theta) = p'\theta$. Then, for each $j \in \{I, II, III\}$, the model $\theta \mapsto \mathsf{Q}_{j,\theta}$ is L^2 differentiable at θ_0 tangentially to $\mathcal{T}_j(\theta_0)$. Furthermore, the efficient influence functions are*

$$\tilde{\varrho}_I(s) = \frac{2\sqrt{2\pi}}{3}(p^{(1)} + p^{(2)})1\{s = (1,1)\} - \frac{\sqrt{2\pi}}{3}(p^{(1)} + p^{(2)})(1\{s = (1,0)\} + 1\{s = (0,1)\}) \quad (185)$$

$$\tilde{\varrho}_{II}(s) = (b_{II}^{(1)}(p) + b_{II}^{(2)}(p))1\{s = (1,1)\} - b_{II}^{(2)}(p)1\{s = (1,0)\} - b_{II}^{(1)}(p)1\{s = (0,1)\} \quad (186)$$

$$\tilde{\varrho}_{III}(s) = (b_{III}^{(1)}(p) + b_{III}^{(2)}(p))1\{s = (1,1)\} - b_{III}^{(1)}(p)1\{s = (1,0)\} - b_{III}^{(2)}(p)1\{s = (0,1)\}, \quad (187)$$

where

$$b_{II}(p) = \underset{b^{(2)} \leq b^{(1)} \leq 0}{\arg \min} \frac{1}{4}[(\sqrt{2\pi}p - b)'1]^2 + \frac{1}{4}(\sqrt{2\pi}p^{(2)} - b^{(2)})^2 + \frac{1}{4}(\sqrt{2\pi}p^{(1)} - b^{(1)})^2 \quad (188)$$

$$b_{III}(p) = \underset{b^{(1)} \leq b^{(2)} \leq 0}{\arg \min} \frac{1}{4}[(\sqrt{2\pi}p - b)'1]^2 + \frac{1}{4}(\sqrt{2\pi}p^{(1)} - b^{(1)})^2 + \frac{1}{4}(\sqrt{2\pi}p^{(2)} - b^{(2)})^2. \quad (189)$$

Proof of Proposition D.3. Case I: First, we consider the case in which $h \in \mathcal{T}_I(\theta_0)$. Let

$$\dot{\ell}_{I,\theta_0} \equiv 1\{s_i = (1,1)\} \begin{pmatrix} \frac{2}{\sqrt{2\pi}} \\ \frac{2}{\sqrt{2\pi}} \end{pmatrix} + (1\{s_i = (1,0)\} + 1\{s_i = (0,1)\}) \begin{pmatrix} \frac{-1}{\sqrt{2\pi}} \\ \frac{-1}{\sqrt{2\pi}} \end{pmatrix}. \quad (190)$$

17

Let $\mathsf{q}_{I,\theta}$ denote the density of $\mathsf{Q}_{I,\theta}$. Then, by Proposition D.1 (Case I), for $h = (\bar{h}, \bar{h})' \in \mathcal{T}_I(\theta_0)$,

$$\mathsf{q}_{I,\theta_0+\tau h}^{1/2} - \mathsf{q}_{I,\theta_0}^{1/2}$$

$$= \left(\frac{1}{2} 1\{s = (0,0)\} + \Phi(\tau\bar{h}) 1\{s = (1,1)\} + \left(\frac{3 - 4\Phi(\tau\bar{h})^2}{8} \right)^{\frac{1}{2}} (1\{s = (1,0)\} + 1\{s = (0,1)\}) \right)$$

$$\quad - \frac{1}{2} \left(1\{s = (0,0)\} + 1\{s = (1,1)\} + 1\{s = (1,0)\} + 1\{s = (0,1)\} \right)$$

$$= (\Phi(\tau\bar{h}) - 1/2) 1\{s = (1,1)\} + \left(\left(\frac{3 - 4\Phi(\tau\bar{h})^2}{8} \right)^{\frac{1}{2}} - \frac{1}{2} \right) \left(1\{s = (1,0)\} + 1\{s = (0,1)\} \right)$$

$$= (\Phi(0) + \Phi'(0)\tau\bar{h} + o(\tau) - 1/2) 1\{s = (1,1)\}$$

$$\quad + \left(\left(\frac{3 - 4\Phi(0)^2}{8} \right)^{\frac{1}{2}} - \frac{1}{2} \left(\frac{3 - 4\Phi(0)^2}{8} \right)^{-\frac{1}{2}} \Phi(0)\Phi'(0)\tau\bar{h} + o(\tau) - \frac{1}{2} \right) \left(1\{s = (1,0)\} + 1\{s = (0,1)\} \right)$$

$$= (\frac{1}{\sqrt{2\pi}}\tau\bar{h} + o(\tau)) 1\{s = (1,1)\} - (\frac{1}{2\sqrt{2\pi}}\tau\bar{h} + o(\tau)) \left(1\{s = (1,0)\} + 1\{s = (0,1)\} \right), \tag{191}$$

where the third equality follows from taking a Taylor expansion of $\mathsf{q}_{I,\theta_0+\tau h}^{1/2}$ with respect to τ at 0.

Note that, by (190), $h = (\bar{h}, \bar{h})'$, and $\mathsf{q}_{I,\theta_0}^{1/2} = \frac{1}{2}(1\{s = (0,0)\} + 1\{s = (1,1)\} + 1\{s = (1,0)\} + 1\{s = (0,1)\})$ by Proposition D.1, it follows that

$$\frac{1}{2}\tau h' \dot{\ell}_{I,\theta_0} \mathsf{q}_{I,\theta_0}^{1/2} = \frac{1}{\sqrt{2\pi}}\tau\bar{h} 1\{s = (1,1)\} - \frac{1}{2\sqrt{2\pi}}\tau\bar{h} \left(1\{s = (1,0)\} + 1\{s = (0,1)\} \right) \tag{192}$$

By (191), (192), and the triangle and Cauchy-Schwarz inequalities, for any given $h \in \mathcal{T}_I(\theta_0)$,

$$\left\| \mathsf{q}_{I,\theta_0+\tau h}^{1/2} - \mathsf{q}_{I,\theta_0}^{1/2}(1 + \frac{1}{2}\tau h' \dot{\ell}_{I,\theta_0}) \right\|_{L^2_\mu} = o(\tau). \tag{193}$$

This establishes that $\dot{\ell}_{I,\theta_0}$ in (190) is the L^2 derivative for $h \in \mathcal{T}_I(\theta_0)$. Therefore, the tangent cone is

$$\mathcal{G}_{I,\theta_0} = \left\{ g \in L^2_{\mathsf{Q}_{\theta_0}} : g = \frac{4\bar{h}}{\sqrt{2\pi}} 1\{s = (1,1)\} - \frac{2\bar{h}}{\sqrt{2\pi}}(1\{s = (1,0)\} + 1\{s = (0,1)\}), \bar{h} \leq 0 \right\}$$

$$= \left\{ g \in L^2_{\mathsf{Q}_{\theta_0}} : g = 2b 1\{s = (1,1)\} - b(1\{s = (1,0)\} + 1\{s = (0,1)\}), b \leq 0 \right\}. \tag{194}$$

Let $p = (p^{(1)}, p^{(2)})'$ with $p^{(j)} < 0$ for $j = 1,2$. The influence curve ϱ_I must satisfy

$$p'h = E_{\mathsf{Q}_{I,\theta_0}} \left[\varrho_I h' \ell_{\theta_0} \right]. \tag{195}$$

Let $b < 0$ and let $\varrho_I = 2b 1\{s = (1,1)\} - b(1\{s = (1,0)\} + 1\{s = (0,1)\})$. Then,

$$E_{\mathsf{Q}_{I,\theta_0}} \left[\varrho_I h' \ell_{\theta_0} \right] = \frac{8b\bar{h}}{\sqrt{2\pi}} E_{\mathsf{Q}_{I,\theta_0}}[1\{s = (1,1)\}] + \frac{2b\bar{h}}{\sqrt{2\pi}}(E_{\mathsf{Q}_{I,\theta_0}}[1\{s = (1,0)\}] + E_{\mathsf{Q}_{I,\theta_0}}[1\{s = (0,1)\}]) \tag{196}$$

$$= \frac{2b\bar{h}}{\sqrt{2\pi}} + \frac{b\bar{h}}{\sqrt{2\pi}} = \frac{3b\bar{h}}{\sqrt{2\pi}}. \tag{197}$$

Note that $p'h = (p^{(1)} + p^{(2)})\bar{h}$, and hence setting $b = \frac{\sqrt{2\pi}}{3}(p^{(1)} + p^{(2)})$ gives the following efficient influence function as a projection of the influence curve on the closure of $\mathcal{G}_{I,\theta_0}$:

$$\tilde{\varrho}_I = \frac{2\sqrt{2\pi}}{3}(p^{(1)} + p^{(2)}) 1\{s = (1,1)\} - \frac{\sqrt{2\pi}}{3}(p^{(1)} + p^{(2)})(1\{s = (1,0)\} + 1\{s = (0,1)\}).$$

18

Case II: Suppose $h \in \mathcal{T}_{\text{II}}(\theta_0)$. Arguing as in Case I, it can be shown that

$$\dot{\ell}_{\text{II},\theta_0}(s) = 1\{s = (1,1)\}\begin{pmatrix}\frac{2}{\sqrt{2\pi}} \\ \frac{2}{\sqrt{2\pi}}\end{pmatrix} + 1\{s = (1,0)\}\begin{pmatrix}0 \\ \frac{-2}{\sqrt{2\pi}}\end{pmatrix} + 1\{s = (0,1)\}\begin{pmatrix}\frac{-2}{\sqrt{2\pi}} \\ 0\end{pmatrix} \tag{198}$$

is the L^2 derivative when $h \in \mathcal{T}_{\text{II}}(\theta_0)$. Therefore, the tangent cone can be written

$$\mathcal{G}_{\text{II},\theta_0} = \Big\{ g \in L^2_{\mathbf{Q}_{\theta_0}} : g = (b^{(1)} + b^{(2)})1\{s = (1,1)\} - b^{(2)}1\{s = (1,0)\} - b^{(1)}1\{s = (0,1)\},$$
$$-\infty < b^{(2)} < b^{(1)} \leq 0 \Big\}. \tag{199}$$

Let $\varrho_{\text{II}}(s) \equiv (\sqrt{2\pi}p^{(1)} + \sqrt{2\pi}p^{(2)})1\{s = (1,1)\} - \sqrt{2\pi}p^{(2)}1\{s = (1,0)\} - \sqrt{2\pi}p^{(1)}1\{s = (0,1)\}$. It is straightforward to show ϱ_{II} is an influence curve for φ. The efficient influence function is then given by the projection of ϱ_{II} onto $cl(\mathcal{G}_{\text{II},\theta_0})$, which is $\tilde{\varrho}_{\text{II}} = \arg\min_{g \in cl(\mathcal{G}_{\text{II},\theta_0})} \|\varrho_{\text{II}} - g\|^2_{L^2_{\mathbf{Q}_{\text{II},\theta_0}}}$. Note that $b \mapsto b'\dot{\ell}_{\text{II},\theta_0}$ is a continuous map from $\mathcal{T}_{\text{II}}(\theta_0)$ to $\mathcal{G}_{\text{II},\theta_0}$ by $\dot{\ell}_{\text{II},\theta_0}$ being square integrable. Hence, the efficient influence function $\tilde{\varrho}_{\text{II}}$ is as given in (186) with

$$b_{\text{II}}(p) = \arg\min_{b \in cl(\mathcal{T}_{\text{II}}(\theta_0))} E_{\mathbf{Q}_{\theta_0}}\Big[(\varrho_{\text{II}} - ((b^{(1)} + b^{(2)})1\{s = (1,1)\} - b^{(2)}1\{s = (1,0)\} - b^{(1)}1\{s = (0,1)\}))^2\Big]$$
$$= \arg\min_{b^{(2)} \leq b^{(1)} \leq 0} \frac{1}{4}[(\sqrt{2\pi}p - b)'1]^2 + \frac{1}{4}(\sqrt{2\pi}p^{(2)} - b^{(2)})^2 + \frac{1}{4}(\sqrt{2\pi}p^{(1)} - b^{(1)})^2.$$

The analysis of Case III ($h \in \mathcal{T}_{\text{III}}(\theta_0)$) is similar to the one above. Therefore, we omit its proof. ∎

D.2 Example 2: Roy model

In this section, we provide details on Example 2. Recall that $S = \{(0,0),(0,1),(1,0),(1,1)\}$, and the sharp identifying restrictions are given as (59)-(61).

D.2.1 LFP

We start with the following characterization of the LFP for the hypotheses considered in the text. For this, let $\bar{c} > 0$ be a known constant.

Proposition D.4. *Let (S, U, G, Θ) be defined as in Example 2. Let m_θ be a discrete distribution on S whose distribution is uniquely determined by $\theta = (\theta^{(0,0)}, \theta^{(0,1)}, \theta^{(1,0)})'$. Suppose (i) $\theta_0^{(0,0)} = \theta_1^{(0,0)} = \bar{c} > 0$ and $\theta_0^{(1,0)} < 1 - \bar{c} - \theta_0^{(0,1)}$; and (ii) $\theta_1^{(1,0)} > 1 - \bar{c} - \theta_0^{(0,1)}$.*

Case 1: If $\theta_1^{(1,0)} < 1 - \bar{c} - \theta_1^{(0,1)}$, the densities of the LFP $(Q_0, Q_1) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_1}$ are:

$$(q_0(0,0), q_0(0,1), q_0(1,0), q_0(1,1)) = (\frac{\bar{c}}{2}, \frac{\bar{c}}{2}, 1 - \bar{c} - \theta_0^{(0,1)}, \theta_0^{(0,1)}) \tag{200}$$

$$(q_1(0,0), q_1(0,1), q_1(1,0), q_1(1,1)) = (\frac{\bar{c}}{2}, \frac{\bar{c}}{2}, \theta_1^{(1,0)}, 1 - \bar{c} - \theta_1^{(1,0)}). \tag{201}$$

Case 2: If $\theta_1^{(1,0)} = 1 - \bar{c} - \theta_1^{(0,1)}$, *the densities of the LFP* $(Q_0, Q_1) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_1}$ *are:*

$$(q_0(0,0), q_0(0,1), q_0(1,0), q_0(1,1)) = (\frac{\bar{c}}{2}, \frac{\bar{c}}{2}, 1 - \bar{c} - \theta_0^{(0,1)}, \theta_0^{(0,1)}) \tag{202}$$

$$(q_1(0,0), q_1(0,1), q_1(1,0), q_1(1,1)) = (\frac{\bar{c}}{2}, \frac{\bar{c}}{2}, \theta_1^{(1,0)}, \theta_1^{(0,1)}). \tag{203}$$

Before giving the proof of the claim above, a few remarks are in order. To simplify the analysis, we assume that $\theta_0^{(0,0)} = \theta_1^{(0,0)}$ are known. Recall that the sharp identifying restrictions in (59)-(61) imply

$$\theta_0^{(1,0)} \le P_0(\{(1,0)\}) \le 1 - \bar{c} - \theta_0^{(0,1)}. \tag{204}$$

The assumption $\theta_0^{(1,0)} < 1 - \bar{c} - \theta_0^{(0,1)}$ therefore makes the model incomplete at θ_0. Finally, the assumption $\theta_1^{(1,0)} > 1 - \bar{c} - \theta_0^{(0,1)}$ is equivalent to $\nu_{\theta_1}(\{(1,0)\}) > \nu_{\theta_0}(\{(1,0)\})$, which ensures that $\mathcal{P}_{\theta_0} \cap \mathcal{P}_{\theta_1} \ne \emptyset$.

Proof of Proposition D.4. The following convex program characterizes the LFP:

$$\min_{(p_0,p_1) \in \Delta^3 \times \Delta^3} \sum_{s \in \{(0,0),(0,1),(1,0),(1,1)\}} - \ln \left(\frac{p_0(s)}{p_0(s) + p_1(s)} \right) (p_0(s) + p_1(s)) \tag{205}$$

$$s.t. \ \theta_j^{(1,0)} \le p_j(1,0), \ j = 0, 1, \tag{206}$$

$$\theta_j^{(0,1)} \le p_j(1,1), \ j = 0, 1, \tag{207}$$

$$\bar{c} = p_j(0,0) + p_j(0,1), \ j = 0, 1. \tag{208}$$

First, we concentrate out $p_0(0,0), p_1(0,0), p_0(0,1), p_1(0,1)$ from the problem. The subset of the KKT conditions that involves these components are, for $s \in \{(0,0), (0,1)\}$

$$-\frac{p_0(s) + p_1(s)}{p_0(s)} \frac{p_0(s) + p_1(s) - p_0(s)}{(p_0(s) + p_1(s))^2} (p_0(s) + p_1(s)) - \ln \left(\frac{p_0(s)}{p_0(s) + p_1(s)} \right) - \lambda_1 = 0 \tag{209}$$

$$-\frac{p_0(s) + p_1(s)}{p_0(s)} \frac{-p_0(s)}{(p_0(s) + p_1(s))^2} (p_0(s) + p_1(s)) - \ln \left(\frac{p_0(s)}{p_0(s) + p_1(s)} \right) - \lambda_2 = 0 \tag{210}$$

$$p_j(0,0) + p_j(0,1) = \bar{c}, \ j = 0, 1. \tag{211}$$

for some $\lambda_1 \ne 0$ and $\lambda_2 \ne 0$. The first two conditions can be simplified as

$$\frac{p_1(s)}{p_0(s)} = 1 + \lambda_2 - \lambda_1, \ s \in \{(0,0), (0,1)\}, \tag{212}$$

which implies $\frac{p_1(0,0)}{p_0(0,0)} = \frac{p_1(0,1)}{p_0(0,1)}$, and hence, if $p_0(0,0) \in (0, \bar{c})$, $\frac{p_1(0,0)}{p_1(0,1)} = \frac{p_0(0,0)}{p_0(0,1)} = \beta$ for some $0 < \beta < \infty$. This, together with (211) yields

$$(p_j(0,0), p_j(0,1)) = (\frac{\beta \bar{c}}{1 + \beta}, \frac{\bar{c}}{1 + \beta}), \ j = 0, 1. \tag{213}$$

For example, one can take the following as a solution $(q_j(0,0), q_j(0,1)) = (\frac{\bar{c}}{2}, \frac{\bar{c}}{2})$. After concentrating out $p_0(0,0), p_1(0,0), p_0(0,1), p_1(0,1)$, the problem becomes maximizing

$$\min_{(p_0,p_1) \in \Delta^3 \times \Delta^3} \sum_{s \in \{(1,0),(1,1)\}} - \ln \left(\frac{p_0(s)}{p_0(s) + p_1(s)} \right) (p_0(s) + p_1(s)) \tag{214}$$

20

subject to the constraints in (206)-(208). The KKT conditions are

$$-\frac{p_1(1,0)}{p_0(1,0)} - \ln\left(\frac{p_0(1,0)}{p_0(1,0) + p_1(1,0)}\right) - \chi_1 - \chi_5 = 0 \tag{215}$$

$$-\frac{p_1(1,1)}{p_0(1,1)} - \ln\left(\frac{p_0(1,1)}{p_0(1,1) + p_1(1,1)}\right) - \chi_3 - \chi_5 = 0 \tag{216}$$

$$1 - \ln\left(\frac{p_0(1,0)}{p_0(1,0) + p_1(1,0)}\right) - \chi_2 - \chi_6 = 0 \tag{217}$$

$$1 - \ln\left(\frac{p_0(1,1)}{p_0(1,1) + p_1(1,1)}\right) - \chi_4 - \chi_6 = 0 \tag{218}$$

$$\chi_1(p_0(1,0) - \theta_0^{(1,0)}) = 0 \tag{219}$$

$$\chi_2(p_1(1,0) - \theta_1^{(1,0)}) = 0 \tag{220}$$

$$\chi_3(p_0(1,1) - \theta_0^{(0,1)}) = 0 \tag{221}$$

$$\chi_4(p_1(1,1) - \theta_1^{(0,1)}) = 0. \tag{222}$$

where $\chi_j \geq 0$ for $j = 1, \ldots, 4$, and the original inequality and equality constraints are also imposed.

Case 1: $(\chi_1 = 0, \chi_2 > 0, \chi_3 > 0, \chi_4 = 0)$

Suppose $\chi_3 > 0$. Then, $p_0(1,1) = \theta_0^{(0,1)}$ by the complementary slackness condition (221). It also implies $p_0(1,0) = 1 - \bar{c} - \theta_0^{(0,1)}$ by $p_0(1,0) + p_0(1,1) = 1 - \bar{c}$. Further, $\chi_1 = 0$ because the constraints associated with χ_1 and χ_3 cannot bind simultaneously due to the assumption that $\theta_0^{(1,0)} < 1 - \bar{c} - \theta_0^{(0,1)}$. Suppose further that $\chi_2 > 0$. Then, $p_1(1,0) = \theta_1^{(1,0)}$ and $p_1(1,1) = 1 - \bar{c} - \theta_1^{(1,0)}$. We also assume that $\chi_4 = 0$.

Now note that (215)-(220) reduce to

$$-\frac{\theta_1^{(1,0)}}{1 - \bar{c} - \theta_0^{(0,1)}} - \ln\left(\frac{1 - \bar{c} - \theta_0^{(0,1)}}{1 - \bar{c} - \theta_0^{(0,1)} + \theta_1^{(1,0)}}\right) - \chi_5 = 0 \tag{223}$$

$$-\frac{1 - \bar{c} - \theta_1^{(1,0)}}{\theta_0^{(0,1)}} - \ln\left(\frac{\theta_0^{(0,1)}}{\theta_0^{(0,1)} + 1 - \bar{c} - \theta_1^{(1,0)}}\right) - \chi_3 - \chi_5 = 0 \tag{224}$$

$$1 - \ln\left(\frac{1 - \bar{c} - \theta_0^{(0,1)}}{1 - \bar{c} - \theta_0^{(0,1)} + \theta_1^{(1,0)}}\right) - \chi_2 - \chi_6 = 0 \tag{225}$$

$$1 - \ln\left(\frac{\theta_0^{(0,1)}}{\theta_0^{(0,1)} + 1 - \bar{c} - \theta_1^{(1,0)}}\right) - \chi_6 = 0 \tag{226}$$

It can be shown that, when $\theta_1^{(1,0)} > 1 - \bar{c} - \theta_0^{(0,1)}$, the system can be solved for $(\chi_2, \chi_3, \chi_5, \chi_6)$ that satisfies $\chi_j > 0$ for $j = 2, 3$, and hence the solution (the remaining components of LFP) is

$$(q_0(1,0), q_0(1,1)) = (1 - \bar{c} - \theta_0^{(0,1)}, \theta_0^{(0,1)}) \tag{227}$$

$$(q_1(1,0), q_1(1,1)) = (\theta_1^{(1,0)}, 1 - \bar{c} - \theta_1^{(1,0)}). \tag{228}$$

Case 2: $(\chi_1 = 0, \chi_2 > 0, \chi_3 > 0, \chi_4 > 0)$

The difference from Case 1 is that $\chi_4 > 0$ is assumed. By the complementary slackness condition, this implies $p_1(1,1) = \theta_1^{(0,1)}$, which also equals $1 - \bar{c} - \theta_1^{(1,0)}$ due to $\chi_2 > 0$ and $p_1(1,0) + p_1(1,1) = 1 - \bar{c}$.

Therefore, the solutions are

$$(q_0(1,0), q_0(1,1)) = (1 - \bar{c} - \theta_0^{(0,1)}, \theta_0^{(0,1)}) \tag{229}$$

$$(q_1(1,0), q_1(1,1)) = (\theta_1^{(1,0)}, \theta_1^{(0,1)}). \tag{230}$$

Similar to Case 1, one may show that there exists $(\chi_2, \ldots, \chi_6)$ that solves the KKT conditions with $\chi_j > 0$ for $j = 2, 3, 4$. $\blacksquare$

D.2.2 Efficient influence function, power envelope, and optimal tests

As in Section 3.4.2, we consider testing $H_0 : p'\theta = \theta^{(1,0)} \leq c$ against $H_1 : p'\theta > c$ for some $c \in (0,1)$ with $p = (0,0,1)'$. Consider the following three configurations analyzed in the text (see also Figure 4) with $\bar{c} = \frac{1}{6}$. In each configuration, the null and alternative parameter values are

$$\theta_0 = (\theta_0^{(0,0)}, \theta_0^{(0,1)}, \theta_0^{(1,0)}) = (\frac{1}{6}, \frac{1}{2}, \frac{1}{6}) \tag{231}$$

$$\theta_1 = (\theta_1^{(0,0)}, \theta_1^{(0,1)}, \theta_1^{(1,0)}) = \Big(\frac{1}{6}, \frac{1}{2} + \xi^{(0,1)} + \frac{h^{(0,1)}}{\sqrt{n}}, \frac{1}{6} + \xi^{(1,0)} + \frac{h^{(1,0)}}{\sqrt{n}}\Big), \tag{232}$$

where ξ and h are taken from one of the following specifications:

Case A-I : $\xi_A = (0, \xi^{(0,1)}, \xi^{(1,0)}) = (0, \frac{-1}{6}, \frac{1}{6})$,

$\mathcal{T}_\mathrm{I}(\theta_0, \xi_A) = \{h = (h^{(0,0)}, h^{(0,1)}, h^{(1,0)}) : h^{(0,0)} = 0, h^{(1,0)} > 0\}$;

Case B-I : $\xi_B = (0, \xi^{(0,1)}, \xi^{(1,0)}) = (0, 0, \frac{1}{6})$,

$\mathcal{T}_\mathrm{I}(\theta_0, \xi_B) = \{h = (h^{(0,0)}, h^{(0,1)}, h^{(1,0)}) : h^{(0,0)} = 0, h^{(1,0)} > 0, h^{(0,1)} + h^{(1,0)} < 0\}$;

Case B-II : $\xi_B = (0, \xi^{(0,1)}, \xi^{(1,0)}) = (0, 0, \frac{1}{6})$,

$\mathcal{T}_\mathrm{II}(\theta_0, \xi_B) = \{h = (h^{(0,0)}, h^{(0,1)}, h^{(1,0)}) : h^{(0,0)} = 0, h^{(1,0)} > 0, h^{(0,1)} + h^{(1,0)} = 0\}$.

Proposition D.5. *Suppose that the conditions of Proposition D.4 hold. Let $p = (0,0,1)'$ and let $\varphi(\theta) = p'\theta$. Then, Assumption 3.3 holds for each of the cases above. Furthermore, for all cases, the efficient influence function is*

$$\tilde{\varrho}(s) = \frac{3}{5}1\{s = (1,0)\} - \frac{2}{5}\{s = (1,1)\}), \tag{233}$$

and the asymptotic power envelope $1 - \Phi\Big(z_\alpha - \sqrt{5}h^{(1,0)}\Big)$ is achieved by a test that rejects H_0 when the following statistic exceeds z_α:

$$T_{\mathrm{I},n} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\frac{3}{\sqrt{5}}1\{s_i = (1,0)\} - \frac{2}{\sqrt{5}}1\{s_i = (1,1)\}\Big]. \tag{234}$$

Proof of Proposition D.5. We analyze the three cases separately.

Case A-I: The alternative parameter configuration satisfies the assumptions of Case 1 in Proposition D.4.

Therefore, the LFPs are

$$(q_0(0,0), q_0(0,1), q_0(1,0), q_0(1,1)) = (\frac{\bar{c}}{2}, \frac{\bar{c}}{2}, 1 - \bar{c} - \theta_0^{(0,1)}, \theta_0^{(0,1)}) = (\frac{1}{12}, \frac{1}{12}, \frac{1}{3}, \frac{1}{2}) \tag{235}$$

$$(q_1(0,0), q_1(1,1), q_1(1,0), q_1(1,1)) = (\frac{\bar{c}}{2}, \frac{\bar{c}}{2}, \theta_1^{(1,0)}, 1 - \bar{c} - \theta_1^{(1,0)}) = (\frac{1}{12}, \frac{1}{12}, \frac{1}{3} + \frac{h^{(1,0)}}{\sqrt{n}}, \frac{1}{2} - \frac{h^{(1,0)}}{\sqrt{n}}). \tag{236}$$

Let $\vartheta \mapsto \mathsf{Q}_{\mathrm{I},\vartheta}$ be a model defined on a neighborhood of $\vartheta = 0$ for which the density $\mathsf{q}_{\mathrm{I},\vartheta}$ is given by

$$(\mathsf{q}_{\mathrm{I},\vartheta}(0,0), \mathsf{q}_{\mathrm{I},\vartheta}(1,1), \mathsf{q}_{\mathrm{I},\vartheta}(1,0), \mathsf{q}_{\mathrm{I},\vartheta}(1,1)) = (\frac{1}{12}, \frac{1}{12}, \frac{1}{3} + \vartheta^{(1,0)}, \frac{1}{2} - \vartheta^{(1,0)}). \tag{237}$$

Due to the form of the LFP, Assumption 3.3 is satisfied with the underlying model $\vartheta \mapsto \mathsf{Q}_{\mathrm{I},\vartheta}$. Calculations similar to the ones in (191) can ensure that the L^2-derivative is

$$\dot{\ell}_{\mathrm{I},0}(s) = 1\{s = (1,0)\} \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} - 1\{s = (1,1)\} \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}. \tag{238}$$

Hence, the tangent cone of the model is

$$\mathcal{G}_{\mathrm{I},0} = \{g \in L_{\mathsf{Q}_{\mathrm{I},0}}^2 : g = 3h^{(1,0)}1\{s = (1,0)\} - 2h^{(1,0)}1\{s = (1,1)\}), h^{(1,0)} > 0\}. \tag{239}$$

The influence curve of φ must satisfy $p'h = \langle \varrho_{\mathrm{I}}, g \rangle_{L_{\mathsf{Q}_{\mathrm{I},0}}^2}$ for $g \in \mathcal{G}_{\mathrm{I},\theta_0+\xi}$. Since $p = (0,0,1)'$, $\varrho_{\mathrm{I}} = \frac{3}{5}1\{s = (1,0)\} - \frac{2}{5}\{s = (1,1)\})$ satisfies the requirement. Observe also that ϱ_{I} is in $\mathcal{G}_{\mathrm{I},0}$ and hence in its closure. Therefore, it is the efficient influence function, i.e. $\varrho_{\mathrm{I}} = \tilde{\varrho}_{\mathrm{I}}$, which in turn implies $\|\tilde{\varrho}_{\mathrm{I}}\|_{L_{\mathsf{Q}_{\mathrm{I},0}}^2} = 1/\sqrt{5}$. By Theorem 3.3, for any level-α test ϕ_n,

$$\limsup_{n \to \infty} \pi_{n,\theta_n,\xi_A,h}(\phi_n) \le 1 - \Phi\left(z_\alpha - \sqrt{5}h^{(1,0)}\right). \tag{240}$$

Again, by Theorem 3.3, this bound can be achieved by a test that rejects the null when the statistic in (234) exceeds z_α.

Case B-I: The alternative parameter configuration again satisfies the assumptions of Case 1 in Proposition D.4. Therefore, the LFPs are given as in (235)-(236). The rest of the analysis parallels Case A-I and is omitted.

Case B-II: The alternative parameter configuration satisfies the assumptions of Case 2 in Proposition D.4. Hence, the LFP is

$$(q_0(0,0), q_0(0,1), q_0(1,0), q_0(1,1)) = (\frac{\bar{c}}{2}, \frac{\bar{c}}{2}, 1 - \bar{c} - \theta_0^{(0,1)}, \theta_0^{(0,1)}) = (\frac{1}{12}, \frac{1}{12}, \frac{1}{3}, \frac{1}{2}), \tag{241}$$

$$(q_1(0,0), q_1(0,1), q_1(1,0), q_1(1,1)) = (\frac{\bar{c}}{2}, \frac{\bar{c}}{2}, \theta_1^{(1,0)}, \theta_1^{(0,1)}) = (\frac{1}{12}, \frac{1}{12}, \frac{1}{3} + \frac{h^{(1,0)}}{\sqrt{n}}, \frac{1}{2} + \frac{h^{(0,1)}}{\sqrt{n}}) \tag{242}$$

Let $\vartheta \mapsto \mathsf{Q}_{\mathrm{II},\vartheta}$ be a model defined on a neighborhood of $\vartheta = 0$ for which the density $\mathsf{Q}_{\mathrm{II},\vartheta}$ is given by

$$(\mathsf{q}_{\mathrm{II},\vartheta}(0,0), \mathsf{q}_{\mathrm{II},\vartheta}(1,1), \mathsf{q}_{\mathrm{II},\vartheta}(1,0), \mathsf{q}_{\mathrm{II},\vartheta}(1,1)) = (\frac{1}{12}, \frac{1}{12}, \frac{1}{3} + \vartheta^{(1,0)}, \frac{1}{2} + \vartheta^{(0,1)}). \tag{243}$$

23

Assumption 3.3 is satisfied with the underlying model $\vartheta \mapsto \mathsf{Q}_{\text{II},\vartheta}$. Calculations similar to the ones in (191) can ensure that the L^2-derivative is

$$\dot{\ell}_{\text{II},0}(s) = 1\{s = (1,0)\} \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} + 1\{s = (1,1)\} \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}. \tag{244}$$

Recall that the local parameter space is $\mathcal{T}_{\text{II}}(\theta_0, \xi_B) = \{h = (h^{(0,0)}, h^{(0,1)}, h^{(1,0)}) : h^{(0,0)} = 0, h^{(1,0)} > 0, h^{(0,1)} + h^{(1,0)} = 0\}$. Hence, the tangent cone of the model is

$$\mathcal{G}_{\text{II},0} = \left\{ g \in L^2_{\mathsf{Q}_{\text{I},0}} : g = 3h^{(1,0)} 1\{s = (1,0)\} - 2h^{(1,0)} 1\{s = (1,1)\}), h^{(1,0)} > 0 \right\}, \tag{245}$$

where we used $h^{(0,1)} = -h^{(1,0)}$. Observe that the tangent cone coincides with the one in (239). The rest of the analysis parallels Case A-I and is omitted. ∎

E Computing the L^2-derivative

In practice, the L^2 derivative, a key object for constructing the optimal tests, can be computed by an analytical method or a quadratic-programming method.[47] In what follows, let $\theta_1 = \theta_0 + \xi + \tau h$ with $\tau > 0$ and assume that $\mathcal{P}_{\theta_0} \cap \mathcal{P}_{\theta_1} = \emptyset$. A robustly testable alternative is a special case, and the argument below can be applied with $\xi = 0$.

Analytical method

Among the two, the analytical method is more straightforward and is recommended whenever possible. Let h be given and assume that Assumption 3.2 or 3.3 holds. For calculating the L^2 derivative, one can take the following steps.

Step 1: For $\tau > 0$, derive the LFP $(Q_{0,\tau h}, Q_{1,\tau h})$ in closed form by solving (18). Find a collection of models $\vartheta \mapsto \mathsf{Q}_{j,\vartheta}, j \in \mathbb{J}$ such that (67) (or (39)) holds.

Step 2: For each j, calculate $\dot{\ell}_{j,0}$ using the analytical form of $\mathsf{Q}_{j,\vartheta}$.

This is the approach taken in the analysis of Examples 1 and 2 (see Appendix D.1.2 and D.2.2). Also, in Step 1, one should check if (67) (or (39)) is satisfied for $Q_{0,\tau h}$. In Example 1, it is satisfied because $\mathcal{P}_{\theta_0} = \{Q_0\}$ is a singleton and by an inspection of the underlying model. In Example 2, the complementary slackness condition determines $Q_{0,\tau h}$ uniquely for all τ sufficiently small, and again the condition can be checked analytically.

---

[47]Another possibility is to use numerical differentiation, which is studied by Hong and Li (2018) in a related context.

Quadratic-Programming method

Under certain conditions, the L^2-derivative of the model can also be calculated as a solution to an auxiliary quadratic programming (QP) problem. Below, we discuss this alternative approach and provide primitive conditions for this method to work.

This approach is based on the stability analysis of a solution to a parametric convex program. Toward this end, we relate the problem in (18) to a parametric optimization problem studied in Shapiro (1988). Let L be the cardinality of S and let K be the cardinality of 2^S, the power set of S. Let $p_0, p_1 \in [0,1]^L$ denote vectors of probability mass functions on S. Let $x = (p'_0, p'_1)' \in [0,1]^{2L}$ denote a vector of control variables, and let $y = \theta_1$ denote a parameter vector in the optimization problem (defined below).

Define

$$f(x,y) \equiv \sum_{s \in S} H(\frac{p_0(s)}{p_0(s) + p_1(s)})[p_0(s) + p_1(s)], \tag{246}$$

and note that it does not depend on y. We impose constraints $\tilde{g}_\ell(x,y) \leq 0$ for $\ell = 1, \cdots, 2(K+L)$ and $\tilde{g}_\ell(x,y) = 0$ for $\ell = 2(K+L)+1, 2(K+L)+2$ with

$$\tilde{g}_\ell(x,y) = \nu_{\theta_0}(A_\ell) - \sum_{s \in A_\ell} p_0(s), \quad \ell = 1, \ldots, K \tag{247}$$

$$\tilde{g}_\ell(x,y) = \nu_{\theta_1}(A_{\ell-K}) - \sum_{s \in A_{\ell-K}} p_1(s), \quad \ell = K+1, \ldots, 2K \tag{248}$$

$$\tilde{g}_\ell(x,y) = -p_0(s_{\ell-2K}), \quad \ell = 2K+1, \ldots 2K+L, \tag{249}$$

$$\tilde{g}_\ell(x,y) = -p_1(s_{\ell-2K-L}), \quad \ell = 2K+L+1, \ldots 2K+2L, \tag{250}$$

$$\tilde{g}_\ell(x,y) = \sum_{s \in S} p_0(s) - 1, \quad \ell = 2K+2L+1 \tag{251}$$

$$\tilde{g}_\ell(x,y) = \sum_{s \in S} p_1(s) - 1, \quad \ell = 2K+2L+2. \tag{252}$$

These restrictions impose the sharp identifying restrictions together with the constraints that restrict $p_j, j = 0, 1$ in the probability simplex. Before proceeding, one can reduce the constraints by the following procedures.[48]

First, reduce the first $2K$ inequality constraints to the ones that are generated by a *core determining class* $\mathcal{C} \subset 2^S$ (Galichon and Henry, 2006, 2011).[49] Second, define

$$\mathcal{A} \equiv \{A \subset \mathcal{C} : \nu_{\theta_0}(A) = \nu^*_{\theta_0}(A) \text{ or } \nu_{\theta_0+\tau h}(A) = \nu^*_{\theta_0+\tau h}(A), \forall \tau \in (0, \bar{\tau}]\}, \tag{253}$$

which collects events that remain complete both under the null and local alternatives. For any $A \in \mathcal{C}$, combine the two inequality constraints, $\nu_{\theta_j}(A) - p_j(A) \leq 0$ and $\nu_{\theta_j}(A^c) - p_j(A^c) \leq 0$, and impose them as an equality constraint $\nu_{\theta_j}(A) = p_j(A)$. Finally, let $\mathcal{B} \equiv \mathcal{C} \setminus \mathcal{A}$, which collects the events associated with the

---

[48]This step is not strictly required, but it is recommended as the reduced system is more likely to satisfy Assumption E.1.

[49]Galichon and Henry (2011) provide a tractable characterization of the core determining class for incomplete models with G possessing a certain monotonicity property. Luo and Wang (2017a,b) provide algorithms to construct the core determining class for general incomplete models.

remaining inequality restrictions. We note that some of the inequality constraints associated with events in this collection may still bind when $\tau = 0$. Let M be the number constraints after reducing the restrictions and let the first $M_1 \leq M$ of the restrictions collect equality constraints.

The resulting convex program can then be written as

$$\min_x \ f(x, y) \tag{254}$$

$$s.t. \ g_\ell(x, y) = 0, \ \ell = 1, \dots, M_1 \tag{255}$$

$$g_\ell(x, y) \leq 0, \ \ell = M_1 + 1, \dots, M, \tag{256}$$

for some $g_\ell : \mathbb{R}^{2L} \times \Theta \to \mathbb{R}, \ell = 1, \dots, M$. This is the setting analyzed in Shapiro (1988).

Let $x_\tau = (q'_{0,\tau}, q'_{1,\tau})'$ be a solution to the problem above when y is set to $y_\tau = \theta_0 + \xi + \tau h$. Let $x_0 = (q'_{0,0}, q'_{1,0})'$ denote the solution when $\tau = 0$. We then construct two sets $\mathbf{A}$ and $\mathbf{B}$, which collect gradient vectors of the equality and binding inequality constraints at x_0 as follows.

1. Gradients associated with equality constraints ($\mathbf{A}$): For each $A \in \mathcal{A}$ and $j \in \{0, 1\}$, write the equality restriction $\nu_{\theta_j}(A) - \sum_{s \in A} q_{j,0}(s) = 0$ as $\nu_{\theta_j}(A) - a'x_0 = 0$ for a vector $a \in \{0, 1\}^{2L}$. Let $\mathbf{A}$ collect such vectors together with two additional vectors $(1'_L, 0'_L)'$ and $(0'_L, 1'_L)'$ that are associated with the equality constraints $\sum_{s \in S} q_{j,0}(s) - 1 = 0$ for $j = 0, 1$.

2. Gradients associated with inequality constraints ($\mathbf{B}$): Let $\mathcal{B}_0 \equiv \{A \in \mathcal{B} : \nu_{\theta_0}(A) = \sum_{s \in A} q_{0,0}(s) \text{ or } \nu_{\theta_1}(A) = \sum_{s \in A} q_{1,0}(s)\}$ be the collection of events for which an inequality constraint binds at $\tau = 0$. For each $A \in \mathcal{B}_0$ and $j \in \{0, 1\}$, write the active inequality restriction $\nu_{\theta_j}(A) - \sum_{s \in A} q_{j,0}(s) = 0$ as $\nu_{\theta_j}(A) - b'x_0 = 0$ for a vector $b \in \{0, 1\}^{2L}$. Let $\mathbf{B}$ collect all such vectors.

Recall that H is a twice continuously differentiable convex function. The following condition is sufficient for the directional differentiability of the model.

Assumption E.1. *(i) H is such that $H(z) > 0$, $H'(z) > 0$ (or $H'(z) < 0$), and $H''(z) > 0$ for all $z \in [0, 1]$; (ii) The solution $x_\tau = (q'_{0,\tau}, q'_{1,\tau})'$ is unique at $\tau = 0$; (iii) The vectors in $\mathbf{A}$ are linearly independent, and there is a vector $u \in \mathbb{R}^{2L}$ such that $u'a = 0$ for all $a \in \mathbf{A}$ and $u'b < 0$ for all $b \in \mathbf{B}$.*

Assumption E.1 (i) is a regularity condition on the objective function, which ensures that the solution of the problem satisfies a weak second-order condition. One can choose H that satisfies these conditions. Assumption E.1 (ii) requires that the solution, hence the LFP (not only its ratio), is unique when $\tau = 0$. This condition is satisfied in general when $\mathcal{P}_{\theta_0} \cap \mathcal{P}_{\theta_0 + \xi}$ is a singleton. A special case is that the local alternatives are robustly testable and the model is complete at $\tau = 0$, in which case $\mathcal{P}_{\theta_0}$ and $\mathcal{P}_{\theta_0 + \tau h}$ coincide at $\tau = 0$ and is a singleton. This occurs, for example, in Example 1. Assumption E.1 (iii) imposes the Mangasarian-Fromovitz constraint qualification (MFCQ). This condition ensures that the set of Lagrange multipliers is bounded for all $\tau \in (0, \bar{\tau}]$ (Gauvin, 1977; Shapiro, 1988), which is the key for the stability of the solution to local perturbations. Note that the constraints are linear and the gradient vectors in $\mathbf{A} \subset \{0, 1\}^{2L}$ and $\mathbf{B} \subset \{0, 1\}^{2L}$ do not need to be estimated. Hence, checking this condition is relatively straightforward.

Write the solution to (254)-(256) as $x(y)$ and denote its directional derivative with direction h by

$$Dx(y)[h] \equiv \lim_{\tau \downarrow 0} \frac{x(y + \tau h) - x(y)}{\tau}.$$

Let $\mathcal{X}_\tau \subset \mathbb{R}^M$ be the set of Lagrange multipliers under τ in a neighborhood of 0, which is, under MFCQ, bounded and the convex hull of a finite set E_τ of extreme points (Shapiro, 1988). For each $\chi \in \mathbb{R}^M$, let $J_+(\chi) = \{\ell : \chi^{(\ell)} > 0, \ell = M_1 + 1, \dots, M\}$, $J_0 = \{\ell : \chi^{(\ell)} = 0, \ell = M_1 + 1, \dots, M\}$, and $J(\chi) = \{1, \dots, M_1\} \cup J_+(\chi)$. Each element χ in E_0 is such that the gradient vectors $\{\nabla_x g_\ell(x_0, y_0), \ell \in J(\chi)\}$ are linearly independent. For each $\ell = 1, \dots, M$, define

$$\alpha_\ell(u, h) = u' \nabla_x g_\ell(x_0, y_0) + h' \nabla_y g_\ell(x_0, y_0),$$

and note that $\alpha_\ell(u, h) = u'a + h'\nabla_y \nu_{y_0}(A)$ for some $a \in \{0, 1\}^{2L}$ if g_ℓ originated from one of the constraints in (248) and $\alpha_\ell(u, h) = u'a$ otherwise because the other constraints (in (247) and (249)-(252)) do not involve θ_1. We then have the following directional differentiability result.

Proposition E.1. *Suppose Assumption E.1 holds. Then, $x(y)$ is directionally differentiable at $y = \theta_0 + \xi$ with the directional derivative*

$$Dx(\theta_0 + \xi)[h] = \arg\min_{u \in \bar{\Sigma}(h)} \bar{\zeta}(u), \tag{257}$$

where

$$\bar{\zeta}(u) = \sum_{\ell=1}^{L} H''\left(\frac{q_0(s^{(\ell)})}{q_0(s^{(\ell)}) + q_1(s^{(\ell)})}\right) \frac{(u_0^{(\ell)} q_1(s^{(\ell)}) - u_1^{(\ell)} q_0(s^{(\ell)}))^2}{(q_0(s^{(\ell)}) + q_1(s^{(\ell)}))^3} \tag{258}$$

$$\bar{\Sigma}(h) = \left\{u : \alpha_\ell(u, h) = 0, \ell \in J(\chi), \alpha_\ell(u, h) \leq 0, \ell \in J_0(\chi), \chi \in E_0\right\}. \tag{259}$$

Note that each α_ℓ is linear in u and the number of constraints is finite because E_0 is finite. Hence, $Dx(\theta_0 + \xi)[h]$ is a solution to a finite-dimensional (convex) QP.

To satisfy Assumption 3.3 (or Assumption 3.2), we also need to ensure (67) (or (39)). For this, among the inequality conditions in (256), let $J_{H_0} \subset \{M_1 + 1, \dots, M\}$ collect the indices associated with the constraints that restrict $q_{0,\tau}$, i.e. those of the form: $\nu_{\theta_0}(A_\ell) - \sum_{s \in A_\ell} p_0(s) \leq 0$. We then let $J_{+,H_0}(\chi) \equiv \{\chi^{(\ell)} > 0, \ell \in J_{H_0}\}$. The following condition is sufficient for (67) (or (39)).

Assumption E.2. *There is a path $\tau \mapsto \chi_\tau$ and $\bar{\tau} > 0$ such that (i) $\chi_\tau \in \mathcal{X}_\tau, \forall \tau \in (0, \bar{\tau}]$; (ii) $J_{+,H_0}(\chi_\tau) = J_{+,H_0}(\chi_0)$ for all $\tau \in (0, \bar{\tau}]$; and (iii) the solution $q_{0,0} \in \Delta(S)$ is uniquely determined by*

$$\nu_{\theta_0}(A_\ell) - \sum_{s \in A_\ell} q_{0,0}(s) = 0, \ \ell \in J_{+,H_0}(\chi_0). \tag{260}$$

In other words, there is a configuration of the Lagrange multipliers, under which the binding constraints under H_0 uniquely determines Q_0, which remains constant across $\tau \in (0, \bar{\tau}]$. As τ varies, the remaining components of the solution, $q_{1,\tau}, \chi_\tau$, are properly adjusted to satisfy the first order conditions (see (263)).

For each $j \in \mathbb{J}$ and $h \in \mathcal{T}_j(\xi, \theta_0)$, $Dx(\theta_0 + \xi)[h]$ is a $2L$-dimensional vector. Let us write it as

$(u_0^{*,(1)}, \ldots, u_0^{*,(L)}, u_1^{*,(1)}, \ldots, u_1^{*,(L)})'$. The following corollary gives a formula to calculate the L^2-derivative for $h \in \mathcal{T}_j(\xi, \theta_0)$.

Corollary E.1. *Suppose the conditions of Proposition E.1 hold. Suppose Assumption E.2 holds. Then, the L^2-derivative $\dot{\ell}_{j,0}$ satisfies*

$$h' \dot{\ell}_{j,0}(s) = \sum_{\ell=1}^{L} \frac{u_1^{*,(\ell)} 1\{s = s^{(\ell)}\}}{\mathsf{q}_{j,0}(s)}. \tag{261}$$

The proof of the results above are collected at the end of this appendix.

Proof of Proposition E.1 and Corollary E.1

Proof of Proposition E.1. The proof proceeds by showing the conditions of Theorem 5.1 in Shapiro (1988). First, for any $y \in \Theta$, the set of feasible solutions $\Omega(y) = \{x = (p_0', p_1')' : g_\ell(x, y) = 0, \ell = 1, \ldots, M_1, g_\ell(x, y) \leq 0, \ell = M_1 + 1, \ldots, M\}$ is a subset of a compact set in $\mathbb{R}^{2L}$ because p_j is in the probability simplex for $j = 0, 1$. This ensures Assumption 1 of Shapiro (1988). Assumption E.1 then ensures Assumption 2 in Shapiro (1988). Similarly, Assumption E.1 (iii) ensures Assumption 3 of Shapiro (1988). Finally, Assumption 4 in Shapiro (1988) is a second-order condition on the Lagrangian, which we show below.

Define

$$\mathcal{L}(x, y, \chi) = \sum_{s \in S} H\left(\frac{p_0(s)}{p_0(s) + p_1(s)}\right)[p_0(s) + p_1(s)] + \sum_{k=1}^{M} \chi^{(k)} g_k(x, y) \tag{262}$$

Note that $g_k(x, y)$ is affine in x for all k so that one may write $a_k' x - b_k(y)$ for some $a_k \in \{0, 1\}^{2L}$ and $b_k : \Theta \to \mathbb{R}$. Therefore, for $\ell = 1, \ldots, L$,

$$\frac{\partial}{\partial x^{(\ell)}} \mathcal{L}(x, y, \chi) = H'\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{p_1(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})} + H\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) + \sum_{k=1}^{M} \chi^{(k)} a_k^{(\ell)}, \tag{263}$$

and for $\ell = L + 1, \ldots, 2L$,

$$\frac{\partial}{\partial x^{(\ell)}} \mathcal{L}(x, y, \chi) = H'\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{-p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})} + H\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) + \sum_{k=1}^{M} \chi^{(k)} a_k^{(\ell)}. \tag{264}$$

Below, we derive the Hessian matrix of the Lagrangian. First, suppose $\ell \in \{1, \ldots, L\}$ and $m \in \{1, \ldots, L\}$

28

and $\ell = m$. Then,

$$
\begin{aligned}
\frac{\partial^2}{\partial x^{(\ell)} \partial x^{(m)}} \mathcal{L}(x, y, \chi) &= H''\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{p_1(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})} \times \frac{p_0(s^{(\ell)}) + p_1(s^{(\ell)}) - p_0(s^{(\ell)})}{(p_0(s^{(\ell)}) + p_1(s^{(\ell)}))^2} \\
&\quad + H'\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{-p_1(s^{(\ell)})}{(p_0(s^{(\ell)}) + p_1(s^{(\ell)}))^2} \\
&\quad + H'\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{p_0(s^{(\ell)}) + p_1(s^{(\ell)}) - p_0(s^{(\ell)})}{(p_0(s^{(\ell)}) + p_1(s^{(\ell)}))^2} \\
&= H''\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{p_1(s^{(\ell)})^2}{(p_0(s^{(\ell)}) + p_1(s^{(\ell)}))^3}
\end{aligned}
$$

Next, suppose $\ell \in \{1, \ldots, L\}$ and $m \in \{L+1, \ldots, 2L\}$ and $m = \ell + L$. Then, a similar calculation yields

$$
\frac{\partial^2}{\partial x^{(\ell)} \partial x^{(m)}} \mathcal{L}(x, y, \chi) = H''\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{-p_0(s^{(\ell)}) p_1(s^{(\ell)})}{(p_0(s^{(\ell)}) + p_1(s^{(\ell)}))^3}
$$

Finally, suppose $\ell \in \{L+1, \ldots, 2L\}$ and $m \in \{L+1, \ldots, 2L\}$ and $\ell = m$. Then,

$$
\frac{\partial^2}{\partial x^{(\ell)} \partial x^{(m)}} \mathcal{L}(x, y, \chi) = H''\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{p_0(s^{(\ell)})^2}{(p_0(s^{(\ell)}) + p_1(s^{(\ell)}))^3}.
$$

All other cases (with $\ell \neq m$) lead to 0 second order derivatives. Therefore,

$$
u' \nabla_{xx}^2 \mathcal{L}(x_0, y_0, \chi) u = \sum_{\ell=1}^{L} H''\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{(u_0^{(\ell)} p_1(s^{(\ell)}) - u_1^{(\ell)} p_0(s^{(\ell)}))^2}{(p_0(s^{(\ell)}) + p_1(s^{(\ell)}))^3} \geq 0. \tag{265}
$$

where $u = (u_0', u_1')'$. The object above is non-negative in general. For the second order condition in Shapiro (1988), we need it to be strictly positive for any non-zero vector u in the following critical cone C

$$
C = \{u : u' \nabla_x g_\ell(x_0, y_0) = 0, \ell = 1, \ldots, M_1; u' \nabla_x g_\ell(x_0, y_0) \leq 0, \ \ell = M_1 + 1, \ldots, M; u' \nabla f_x(x_0, y_0) \leq 0\}. \tag{266}
$$

The restriction, $u' \nabla f_x(x_0, y_0) \leq 0$, can also be written as

$$
\begin{aligned}
\sum_{\ell=1}^{L} H'\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{u_0^{(\ell)} p_1(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})} + u_0^{(\ell)} H\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) & \\
+ \sum_{\ell=1}^{L} H'\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{-u_1^{(\ell)} p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})} + u_1^{(\ell)} H\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) & \leq 0, \quad (267)
\end{aligned}
$$

and hence

$$
\sum_{\ell=1}^{L} H'\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{u_0^{(\ell)} p_1(s^{(\ell)}) - u_1^{(\ell)} p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})} \leq -\sum_{\ell=1}^{L} (u_0^{(\ell)} + u_1^{(\ell)}) H\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right). \tag{268}
$$

Now note that, by taking singleton events $A = \{s^{(\ell)}\}, \ell = 1, \ldots, L$, part of the restrictions $u' \nabla_x g_\ell(x_0, y_0) \leq 0, \ell = M_1 + 1, \ldots, M$ must include gradients of the form $\nabla_x g_\ell(x_0, y_0) = (-e_\ell', 0_L')', \ell = 1, \ldots, L$ where e_ℓ is a vector of zeros whose ℓ-th component is 1. This therefore requires $u_0^{(\ell)} \geq 0$ for all ℓ. Similarly, $u_1^{(\ell)} \geq 0$ for all ℓ. Further, Shapiro' second order condition requires $u \neq 0$, and hence, $u_j^{(\ell)} > 0$ for some ℓ and

$j \in \{0,1\}$. This implies that the RHS of (268) is strictly negative if we choose H to be strictly positive on $[0,1]$ as assumed in Assumption E.1. This in turn implies u must satisfy

$$\sum_{\ell=1}^{L} H'\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{u_0^{(\ell)} p_1(s^{(\ell)}) - u_1^{(\ell)} p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})} < 0. \tag{269}$$

Since H' is strictly positive (or strictly negative) for all $z \in [0,1]$, for at least one ℓ, one must have $u_0^{(\ell)} p_1(s^{(\ell)}) - u_1^{(\ell)} p_0(s^{(\ell)}) < 0$ (or > 0) to satisfy the inequality above. Therefore, for any $u \in C$ and $u \neq 0$, one must have

$$u' \nabla_{xx}^2 \mathcal{L}(x_0, y_0, \chi) u > 0 \tag{270}$$

This ensures Assumption 4 of Shapiro (1988). Note that $\bar{\zeta}$ in Proposition E.1 corresponds to $\zeta_{v,w}$ in Eq. (4.10) in Shapiro (1988), where $w = 0$ in our setting. To see this, observe that one of the terms in this function is ξ_λ defined in Eq. (3.3) in Shapiro (1988), which in our setting equals $u' \nabla_{xx}^2 \mathcal{L}(x_0, y_0, \chi) u$ because the components of $\nabla_{xy}^2 \mathcal{L}(x_0, y_0, \chi)$ and $\nabla_{yy}^2 \mathcal{L}(x_0, y_0, \chi)$ are 0 due to $\nabla_y \mathcal{L}(x, y, \chi)$ being a constant as g_k is separable in (x, y). Therefore, by (265) and this function being independent of χ, we obtain

$$\bar{\zeta}(u) = \sum_{\ell=1}^{L} H''\left(\frac{p_0(s^{(\ell)})}{p_0(s^{(\ell)}) + p_1(s^{(\ell)})}\right) \frac{(u_0^{(\ell)} p_1(s^{(\ell)}) - u_1^{(\ell)} p_0(s^{(\ell)}))^2}{(p_0(s^{(\ell)}) + p_1(s^{(\ell)}))^3}. \tag{271}$$

Also observe that $\bar{\Sigma}(h)$ in (259) corresponds to $\bar{\Sigma}(v)$ in Eq. (2.12) in Shapiro (1988). The claim of the proposition now follows from Theorem 5.1 of Shapiro (1988). ∎

Proof of Corollary E.1. Let $(q_{0,j,\tau h}, q_{1,j,\tau h})$ be the densities of the LFP between θ_0 and $\theta_0 + \xi + \tau h$ and recall that they are the solutions of a parametric convex program. By Assumption E.2, there is a model $\vartheta \mapsto \mathsf{Q}_{j,\vartheta}$ such that $q_{0,j,\tau h} = \mathsf{q}_{j,0}$ and $q_{1,j,\tau h} = \mathsf{q}_{j,\tau h}$ for all $\tau \in (0, \bar{\tau}]$ for some $\bar{\tau} > 0$, where $\mathsf{q}_{j,\vartheta}$ is the density of $\mathsf{Q}_{j,\vartheta}$. By Proposition E.1, the LFP density $\vartheta \mapsto \mathsf{q}_{j,\vartheta}$ is directionally differentiable with the directional derivative $u_1^{*,(\ell)} = \lim_{\tau \downarrow 0} \frac{\mathsf{q}_{j,\tau h}(s^{(\ell)}) - \mathsf{q}_{j,0}(s^{(\ell)})}{\tau}$. By the chain rule, this implies

$$\lim_{\tau \downarrow 0} \frac{\mathsf{q}_{j,\tau h}^{1/2}(s^{(\ell)}) - \mathsf{q}_{j,0}^{1/2}(s^{(\ell)})}{\tau} = \frac{u_1^{*,(\ell)}}{\mathsf{q}_{j,0}(s^{(\ell)})}. \tag{272}$$

Noting that $u_1^{*,(\ell)}$ is finite, and interchanging limits and sum, we obtain

$$\lim_{\tau \downarrow 0} \int_S \left(\mathsf{q}_{j,\tau h}^{1/2}(s) - \mathsf{q}_{j,0}^{1/2}(s) - \tau \sum_{\ell=1}^{L} \frac{u_1^{*,(\ell)} 1\{s = s^{(\ell)}\}}{\mathsf{q}_{j,0}(s)}\right)^2 \mathsf{q}_{j,0} d\mu(s)$$

$$= \lim_{\tau \downarrow 0} \sum_{\ell=1}^{L} \left(\mathsf{q}_{j,\tau h}^{1/2}(s^{(\ell)}) - \mathsf{q}_{j,0}^{1/2}(s^{(\ell)}) - \tau \frac{u_1^{*,(\ell)}}{\mathsf{q}_{j,0}(s^{(\ell)})}\right)^2 \mathsf{q}_{j,0}(s^{(\ell)})$$

$$= \sum_{\ell=1}^{L} \left(\lim_{\tau \downarrow 0} \left[\mathsf{q}_{j,\tau h}^{1/2}(s^{(\ell)}) - \mathsf{q}_{j,0}^{1/2}(s^{(\ell)}) - \tau \frac{u_1^{*,(\ell)}}{\mathsf{q}_{j,0}(s^{(\ell)})}\right]\right)^2 \mathsf{q}_{j,0}(s^{(\ell)}) = 0. \tag{273}$$

This ensures (261). ∎