UNIVERSAL INFERENCE FOR INCOMPLETE MODELS

HIROAKI KAIDO
Department of Economics, Boston University

YI ZHANG
Institute Economic and Social Research (IESR), Jinan University

This paper develops a robust inference method with finite-sample validity for general discrete choice models. The procedure's appeal is its simplicity and versatility; it compares a novel likelihood-ratio statistic to a fixed critical value and can be used in models with set-valued predictions and nuisance parameters. It does not require moment selection tuning parameters or resampling. The key is to construct a likelihood that captures the worst-case scenario for controlling the test's size. We show that the proposed test is valid even if unknown selection mechanisms are incidental parameters that may vary arbitrarily across units.

KEYWORDS: Universal inference, Incomplete models, Incidental parameter problem.

Hiroaki Kaido: hkaido@bu.edu
Yi Zhang: yzhangjnu@outlook.com

This paper builds on an earlier work by the authors titled "Robust Likelihood-ratio Tests in Incomplete Economic Models", which evolved into two papers. One is this paper that focuses on universal inference. The other ongoing work focuses on the optimality properties of likelihood ratio tests.

We are grateful to Tim Armstrong, Federico Bugni, Shuowen Chen, Andrew Chesher, Larry Epstein, Jean-Jacques Forneron, Marc Henry, Hidehiko Ichimura, Toru Kitagawa, Michal Kolesar, Xun Lu, Francesca Molinari, Ulrich Müller, Whitney Newey, and Adam Rosen for their helpful comments. We thank, for their comments, the seminar, lecture, and conference participants at the following: BU, Cornell, Columbia, Hitotsubashi, HKUST, Kobe, Kyoto, Michigan State, Princeton, Rutgers, Texas A&M, UCSD, U. Tokyo (CREPE Lectures), UWO, CEMMAP Workshops: Advances in Microeconometrics (Doshisha U., 2016, Vanderbilt, 2019), CEME, JEA Fall Meeting 2017, BU-BC joint econometrics workshop, ASSA Meeting 2018, NASMES 2018, New Frontiers in Econometrics (UConn), and Conference on Incomplete Models (Northwestern/Cemmap). We thank Undral Byambadalai and Junwen Lu for their excellent research assistance. Financial support from NSF grants SES-1824344 and SES-2018498 is gratefully acknowledged. Kaido thanks Microsoft Research (New England), where part of the research was conducted. Financial Support from the National Natural Science Foundation of China (NSFC) grant No.72203077 is gratefully acknowledged.

1. INTRODUCTION

Economic models often make *set-valued predictions* when the researcher works with weak assumptions. We consider models in which, given observable and unobservable exogenous variables (X, U), an outcome variable Y is known to take values in a discrete set $G(U|X;\theta)$ with probability 1. This general discrete choice model nests widely-used models such as binary, multinomial, and ordered choice models, where G contains a single value of Y. If G contains multiple values, the model is silent about how the observed outcome gets selected from G. A growing number of empirical models exhibit such set-valued predictions. The examples that fall into this class include but are not limited to discrete games (Tamer, 2003, Ciliberto and Tamer, 2009), discrete choice models with heterogeneous choice sets (Barseghyan et al., 2021), discrete choice models with endogeneity (Chesher and Rosen, 2017), English auctions (Haile and Tamer, 2003), and product offerings (Eizenberg, 2014). Following Tamer (2003), we call them *incomplete models*.

We propose a novel, universally valid inference method for this class of models. In particular, we consider testing the composite hypotheses

$$H_0 : \theta \in \Theta_0, \quad \text{v.s.} \quad H_0 : \theta \in \Theta_1, \tag{1.1}$$

for disjoint subsets Θ_0, Θ_1 of a parameter space. The universality means the proposed test has a size control property in any finite samples without complex regularity conditions (Wasserman et al., 2020). Our test compares a tailor-made likelihood-ratio statistic to a fixed critical value. As such, it does not require the user to choose any moment selection or other regularization tuning parameters commonly used in the literature. Inverting the test yields confidence regions for the entire parameter and their functionals, such as parameter subcomponents and counterfactual probabilities.

A key insight is that any incomplete model has a structure that allows us to calculate densities corresponding to the least-favorable scenarios for a minimax testing problem. We use this insight to construct a test statistic. Specifically, we start with a simple problem of distinguishing a pair of parameters $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$. For each parameter value, the model admits a set of densities $\mathfrak{q}_\theta$. Heuristically, it suffices to find a pair of densities $(q_{\theta_0}, q_{\theta_1})$ such that q_{θ_0} is compatible with θ_0 and least favorable for size control, while q_{θ_1} is least favorable for maximizing power. Building on Huber and Strassen (1973), we ensure that such a *least-favorable pair (LFP)* of densities exists and is computable. Forming a likelihood-ratio statistic from such pairs is the key to universality. We combine it with the cross-fitting technique proposed by Wasserman et al. (2020) and show that the test controls its size in finite samples regardless of our lack of understanding of the selection mechanism.

The proposed test also provides robustness against *incidental parameters*. The true data-generating process may be characterized by selection mechanisms that are heterogeneous across cross-sectional units.[1] In an extreme case, each cross-sectional unit may have its own selection mechanism. However, empirical studies often assume the same selection mechanism generated the observed data across cross-sectional units or impose regularity conditions for classic limit theorems that implicitly limit heterogeneity of selections. Our proposed test remains valid without such assumptions.

The proposed procedure has two distinct features that set it apart from the textbook LR test. However, implementing these additional steps is straightforward. The first feature is that it uses a custom likelihood function that can be constructed by solving a convex program. We provide a numerical algorithm and also demonstrate that leading examples have closed-form likelihoods. The second feature is that the test uses cross-fitting. This means that we estimate model parameters in a subsample, use the remaining subsample to evaluate the likelihood-ratio, and aggregate statistics after swapping the roles of the two subsamples. This enables us to apply a conditioning argument and a Chernoff-style bound.

The method proposed here offers a tractable way for practitioners to perform inference on general discrete choice models without relying on ad-hoc assumptions. It is particularly effective in settings where the outcome is relatively low-dimensional, involves various

[1] The selection mechanisms may represent different objects depending on contexts. In discrete games, they represent equilibrium selection mechanisms (Jovanovic, 1989). In panel dynamic discrete choice models, they represent unobserved initial conditions (Honoré and Tamer, 2006). In discrete choice models with heterogeneous choice sets, they represent unknown choice-set formation processes (Barseghyan et al., 2021). They may vary across cross-sectional units such as markets and individuals.

types of covariates, and θ may contain nuisance parameters. The method can be applied to both small and large samples. For the theoretical front, this paper provides a new likelihood-based test with finite sample validity for incomplete models. To our knowledge, this is the first such procedure.

1.1. *Relation to the literature*

The study of incomplete structural models has a long history dating back to the work of Wald (1950). Economic models with multiple equilibria are well-known examples of incomplete models. Jovanovic (1989) developed a formal framework to examine the empirical content of such models. It is worthwhile to note that the class of incomplete models considered here also covers a wide range of empirical models beyond games with multiple equilibria. Systematic ways to derive partially identifying restrictions for such models have been developed (Tamer, 2003, Galichon and Henry, 2011, Beresteanu et al., 2011, Chesher and Rosen, 2017). Building on this line of work, we use the *sharp identifying restrictions* to incorporate all information in the original structural model to construct a likelihood function.

We contribute to the literature on inference by providing a novel test that has the following properties: (i) it has finite sample validity without complex regularity conditions (universality); (ii) it is robust to the incidental parameter problem; (iii) it is applicable to models with mixed data types (e.g., continuous and discrete covariates); and (iv) one can construct confidence regions for the entire parameter or its functionals.[2] We do so by using a likelihood-ratio statistic. Likelihood-based methods are also considered by Chen et al. (2011, 2018), Kaido and Molinari (2024), but they aim at asymptotically uniformly valid

[2] Each of the problems above is studied somewhat separately. For (i), Horowitz and Lee (2023) develop a method with finite sample validity for models characterized as an optimization problem; For (ii), Hahn and Moon (2010) tackles the problem using panel data. Epstein et al. (2016) develop a central limit theorem robust to the incidental parameter problem; (iii) For discrete covariates one can use unconditional moment inequalities (see Canay and Shaikh, 2017, and references therein). For continuous covariates, one needs to work with a continuum (or increasing number) of moment inequalities (Andrews and Shi, 2013, Chernozhukov et al., 2013). See also Kaido and Molinari (2024) for this point who develop a method related to this paper's approach; (iv) Subvector inference is studied, for example, by Bugni et al. (2017), Kaido et al. (2019), Cox and Shi (2022), Andrews et al. (2023).

inference. This paper builds on an earlier work by the authors (Kaido and Zhang, 2019), which evolved into two separate papers, including this one. The other ongoing work focuses on the finite-sample and asymptotic optimality properties of the related likelihood-ratio tests.

Likelihood-based inference is commonly used in discrete choice models. However, as Wasserman et al. (2020) notes "The (limiting) null distribution of the classical likelihood-ratio statistic is often intractable when used to test composite null hypotheses in irregular statistical models." Their universal inference technique is simple and does not require complex regularity conditions. Hence, it is also attractive in the context of discrete choice models. Nonetheless, we cannot directly apply their framework since they make two key assumptions. First, they assume the model yields a unique likelihood function. Second, they assume random sampling as a baseline.[3] Neither of these assumptions is guaranteed in incomplete models. For each value of exogenous variables and a structural parameter, the model implies multiple (typically infinitely many) likelihoods since the model does not specify an unknown selection mechanism. Furthermore, the unspecified selection mechanism can vary arbitrarily across experiments. Each selection mechanism may depend on the observable and unobservable variables. It can introduce unknown heterogeneity to the sampling distribution across experiments. We address these issues by constructing a likelihood robust to the worst case scenario.

2. SET-UP

Let $Y \in \mathcal{Y} \subseteq \mathbb{R}^{d_Y}$ and $X \in \mathcal{X} \subseteq \mathbb{R}^{d_X}$ denote, respectively, observable endogenous and exogenous variables, and $U \in \mathcal{U} \subseteq \mathbb{R}^{d_U}$ denote latent variables. For a metric space $\mathcal{Z}$, let $\Delta(\mathcal{Z})$ denote the set of all Borel probability measures on $(\mathcal{Z}, \Sigma_{\mathcal{Z}})$, where $\Sigma_{\mathcal{Z}}$ is the Borel σ-algebra on $\mathcal{Z}$. For $\mathcal{Z} = \mathcal{Y} \times \mathcal{X}$, we let $\Sigma_{\mathcal{Z}}$ equal the product σ-algebra $\Sigma_Y \times \Sigma_X$. Let Θ denote a parameter space.[4] Let $F_\theta(\cdot|x)$ denote a family of conditional distributions of U

[3] To be precise, they assume the knowledge of the conditional likelihood based on a subsample D_0 given another subsample D_1, which is not guaranteed in incomplete models. See Section 2.1.

[4] The theoretical framework below accommodates both finite-dimensional and infinite-dimensional parameters.

given $X = x$. We let $F = \{F_\theta, \theta \in \Theta\}$ be the collection of the conditional laws. For any vector of random elements, we let $Z^n = (Z_1, \ldots, Z_n)$.

For each $\theta \in \Theta$, let $G(\cdot|\cdot;\theta) : \mathcal{U} \times \mathcal{X} \twoheadrightarrow \mathcal{Y}$ be a weakly measurable correspondence. This map summarizes the model prediction. It maps the exogenous variables $(x, u) \in \mathcal{X} \times \mathcal{U}$ to permissible outcomes. For example, in discrete games, $G(U|X;\theta)$ may collect all pure-strategy Nash equilibrium outcomes given observable and unobservable payoff shifters (X, U) (see Example 1). The observable outcome Y is a measurable selection of the prediction satisfying

$$Y \in G(U|X;\theta), \ a.s. \tag{2.1}$$

The model does not impose any restrictions on how Y is selected. This structure nests models with *complete predictions* as a special case. A model makes a complete prediction if there is a function $g(\cdot|\cdot;\theta) : \mathcal{X} \times \mathcal{U} \to \mathcal{Y}$ such that $Y = g(U|X;\theta)$.

Let $\mathcal{C}$ be the collection of all closed subsets of $\mathcal{Y}$. Define the *containtment functional* of G by

$$\nu_\theta(A|x) \equiv \int 1\{G(u|x;\theta) \subseteq A\} dF_\theta(u|x), \ A \in \mathcal{C}. \tag{2.2}$$

This functional characterizes *all* conditional distributions of the measurable selections of $G(U|x;\theta)$ by the following set (Artstein, 1983, Theorem 2.1):

$$\text{core}\,(\nu_\theta(\cdot|x)) \equiv \{Q \in \mathcal{M}(\Sigma_Y, \mathcal{X}) : Q(A|x) \geq \nu_\theta(A|x), \ A \in \mathcal{C}\}, \tag{2.3}$$

where $\mathcal{M}(\Sigma_Y, \mathcal{X})$ is the collection of laws of random variables supported on $\mathcal{Y}$ conditional on X. The moment inequalities $Q(\cdot|x) \geq \nu_\theta(\cdot|x)$ characterizing $\text{core}\,(\nu_\theta(\cdot|x))$ are known as the *sharp identifying restrictions* as they can partially identify θ without losing distributional information (Galichon and Henry, 2011, Molchanov and Molinari, 2018). Our method applies to any model characterized by such restrictions.

2.1. *An Overview of the Main Results*

Consider testing

$$H_0 : \theta \in \Theta_0, \quad \text{v.s.} \quad H_0 : \theta \in \Theta_1 \tag{2.4}$$

for disjoint subsets Θ_0, Θ_1 of the parameter space. The proposed procedure combines a likelihood ratio test (LRT) with a cross-fitting technique. In a nutshell, it takes the following steps.

Step 1: Split samples into D_0 and D_1. Let a likelihood function based on D_0 be

$$\mathcal{L}_0(\theta) = \prod_{i \in D_0} q_\theta(Y_i|X_i), \tag{2.5}$$

where $\theta \mapsto q_\theta(y|x)$ is a "tailor-made" likelihood function;

Step 2: Let $\hat{\theta}_1$ be any estimator of θ constructed from sample D_1. Let $\hat{\theta}_0$ be the *restricted maximum likelihood estimator (RMLE)* based on D_0:

$$\hat{\theta}_0 \in \arg\max_{\theta \in \Theta_0} \mathcal{L}_0(\theta); \tag{2.6}$$

Step 3: Let

$$T_n = \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)}. \tag{2.7}$$

The *cross-fit likelihood-ratio (LR) statistic* is

$$S_n = \frac{T_n + T_n^{\text{swap}}}{2}, \tag{2.8}$$

where T_n^{swap} is calculated in the same way as T_n after swapping the roles of D_0 and D_1.

Step 4: Reject H_0 if $S_n > 1/\alpha$. Do not reject H_0 otherwise.

One can also construct a confidence region for functionals $\varphi(\theta)$ of the parameter by inverting the test (see Section 3).

We focus on key properties of the proposed test for now and defer discussion of how to construct $\hat{\theta}_1$ and $\mathcal{L}_0$ to Sections 2.2-2.3. The cross-fit LR test controls its size in any finite samples. That is, for a wide class $\mathcal{P}_0^n$ of data-generating processes (DGPs) compatible with the null hypothesis,

$$\sup_{P^n \in \mathcal{P}_0^n} P^n\big(S_n > \frac{1}{\alpha}\big) \leq \alpha, \tag{2.9}$$

and this size control result holds for any n.

Following Wasserman et al. (2020), we call the procedure above *universal inference* (or *universal hypothesis test*). The idea is that the inference applies universally to any model described by (2.1) without further regularity conditions. This point is important because many of the existing inference methods require additional regularity conditions, some of which are hard to verify. [5]

Our test builds on the recent work of Wasserman et al. (2020), which demonstrated the universality of their statistical test for a model consisting of probability distributions, denoted as $\{P_\theta, \theta \in \Theta\}$, over an arbitrary set Θ. Our statistical model is $\{\mathcal{P}_\theta, \theta \in \Theta\}$, where $\mathcal{P}_\theta$ itself is a set of probability distributions. Hence, we cannot directly apply their framework to our setting. To extend the scope of universal inference, this paper introduces two innovations. First, they assumed the availability of a unique likelihood function. Incomplete models admit infinitely many likelihoods. We address this issue by constructing a tailor-made likelihood q_θ. Second, they assumed the researcher could calculate the likelihood of D_0 conditional on D_1 and vice versa. Their baseline assumption is that (Y_i, X_i) are drawn independently from an identical distribution. In our setting, the joint distribution of outcomes across units (and hence across D_0 and D_1) is not uniquely determined (see (3.1) below). Hence, one cannot calculate the conditional likelihood required by their framework. We address this issue by constructing a product likelihood from a least favorable density. That is, pretending as if data were generated from q_θ for some $\theta \in \Theta_0$ independently across $i \in D_0$ conditional on D_1 allows us to control the test's size.

[5] See discussion in Kaido et al. (2022)

Finally, we note that the selection mechanism can be heterogeneous across experiments because the model does not restrict how it varies across i.[6] To see this, write the conditional law of $Y_i|X_i = x$ as

$$P_i(\cdot|x) = \int_{\mathcal{U}} \eta_i(\cdot|u,x)dF_\theta(u|x),\ \eta_i(\cdot|u,x) \in \Delta(G(u|x;\theta)), \tag{2.10}$$

where the conditional distribution η_i over the predicted set of outcome values $G(u|x;\theta)$ represents the unknown selection mechanism. It is allowed to vary across experiments arbitrarily. In an extreme case, each experiment can have its own selection mechanism. As in the fixed-effects approach, we aim to make inferences about the common parameter θ and its functionals, while staying agostic about how the *incidental parameters* $\{\eta_i\}$ are related to the observables. This feature does not appear in Wasserman et al.'s (2020) framework. In short, we develop a universal inference procedure that is robust to the model incompleteness and incidental parameter problem.

REMARK 2.1: *Most of the existing inference methods that apply laws of large numbers and central limit theorems implicitly impose assumptions on the heterogeneity of η_i. A leading example is to assume the identity of the selection mechanism $\eta_i = \eta$ for all i. An alternative assumption is samples are stationary and strongly mixing (Chernozhukov et al., 2007, Andrews and Soares, 2010), which imposes implicit restrictions on the selection mechanisms. An exception is the central limit theorem of Epstein et al. (2016), which allowed arbitrary heterogeneity and dependence.*

The result above is also related to the *asymptotic uniform validity* results shown for many of the existing proposals (see Canay and Shaikh, 2017). Relative to them, Theorem 1 is novel in the following respects. First, it establishes the finite-sample validity of the proposed method instead of asymptotic validity. To our knowledge, such results have not been available outside a class of models formulated as optimization problems (Horowitz

[6]More generally, the outcomes can be arbitrarily correlated across experiments. With a slight modification, our procedure remains valid in this setting (see Appendix B).

and Lee, 2023). Second, it does not require tuning parameters for moment selection and additional regularization commonly used to ensure the asymptotic uniform validity.[7]

In the next few sections, we provide details on how to construct the LR statistic. The key is to construct $\mathcal{L}_0$ from a "least favorable" parametric model $\{q_\theta, \theta \in \Theta_0\}$ against a "representative" alternative hypothesis $q_{\hat{\theta}_1}$.

2.2. *Choice of* $\hat{\theta}_1$

Any existing estimator of θ based on sample D_1 can be used as $\hat{\theta}_1$. A leading choice is an extremum estimator $\hat{\theta}_1$ that minimizes a sample criterion function $\theta \mapsto \hat{\mathsf{Q}}_1(\theta)$. For example, $\hat{\mathsf{Q}}_1$ can be based on sample moment inequalities (Chernozhukov et al., 2007, 2013)

$$\hat{\mathsf{Q}}_1(\theta) = \sup_{j,x} \frac{\nu_\theta(A_j|x) - \hat{P}_1(A_j|x)}{\hat{s}_{\theta,1}(A_j|x)}, \tag{2.11}$$

where $\hat{P}_1(A_j|x)$ is an estimator of the conditional probability $P(A_j|x)$ based on D_1, and $\hat{s}_{\theta,1}$ is an estimator of the standard error of $\hat{P}_1$. Another possibility is to use a negative log-likelihood:

$$\hat{\mathsf{Q}}_1(\theta) = \sum_{i \in D_1} \ln p_\theta(Y_i|X_i; \hat{p}_n), \tag{2.12}$$

where $p_\theta(\cdot|\cdot;\hat{p}_n)$ is the Kullback-Leibler projection of the empirical distribution (Kaido and Molinari, 2024).

2.3. *How to Construct* $\mathcal{L}_0$

Below, we condition on D_1, assume $\hat{\theta}_1$ is already computed, and treat it as fixed. We first discuss why we construct a certain likelihood. For each $x \in \mathcal{X}$, let

$$\mathfrak{q}_{\theta,x} \equiv \{q : q(\cdot|x) = dQ(\cdot|x)/d\mu,\ Q \in \operatorname{core}(\nu_\theta(\cdot|x))\}, \tag{2.13}$$

[7] For confidence regions for the entire parameter or identified sets, tuning parameters for selecting moments are commonly used (Chernozhukov et al., 2007, Andrews and Soares, 2010). For subvector inference, additional regularization parameters may be needed (Kaido et al., 2019).

and let $\mathfrak{q}_\theta \equiv \{\mathfrak{q}_{\theta,x}, x \in \mathcal{X}\}$. Given $\hat{\theta}_1$, let $p(\cdot|x) \in \mathfrak{q}_{\hat{\theta}_1,x}$ be a conditional density that is compatible with $\hat{\theta}_1$. Now consider testing whether the data in D_0 is compatible with the null hypothesis or the "representative" alternative p. For each $\theta \in \Theta_0$, consider forming a likelihood ratio between an element of $\mathfrak{q}_{\theta,x}$ and $p(\cdot|x) \in \mathfrak{q}_{\hat{\theta}_1,x}$. Some elements of $\mathfrak{q}_{\theta,x}$ may be very easy to distinguish from p. Using such densities for the denominator of the LR statistic can lead to overrejection if the true DGP differs from it.

To address this issue, we construct a family of *least favorable* densities for distinguishing each $\theta \in \Theta_0$ from $\hat{\theta}_1$. Define the Kullback-Leibler (KL) divergence by

$$I(f(\cdot|x)||f'(\cdot|x)) \equiv \int_{S_x} \ln \frac{f(y|x)}{f'(y|x)} f(y|x) d\mu, \tag{2.14}$$

where $S_x = \{y \in \mathcal{Y} : f(y|x) > 0\}$. We construct a parametric model as follows.

DEFINITION 2.1—LFP-based parametric model: *A family of densities* $\{q_\theta, \theta \in \Theta_0 \cup \{\hat{\theta}_1\}\}$ *is an* LFP-based parametric model *if (i) for each* $\theta \in \Theta_0$ *and* $x \in \mathcal{X}$,

$$q_\theta(\cdot|x) = \underset{q(\cdot|x) \in \mathfrak{q}_{\theta,x}}{\arg\min}\; I(q(\cdot|x) + p(\cdot|x)||q(\cdot|x)) \tag{2.15}$$

for $p(\cdot|x) \in \mathfrak{q}_{\hat{\theta}_1,x}$*; and (ii)* $q_{\hat{\theta}_1}(\cdot|x) = p(\cdot|x)$.

Given the representative density $q_{\hat{\theta}_1} = p$ under the unrestricted model, the challenge is to identify the density that is compatible with $\theta \in \Theta_0$ and is appropriate for $\mathcal{L}_0$. Using the theory for minmax tests, it can be shown that q_θ in Definition 2.1 is the least-favorable density for distinguishing $\mathfrak{q}_\theta$ from $q_{\hat{\theta}_1}$ (see Proposition 5.1 below). We compute the LFP-based parametric model as follows. First, $p \in \mathfrak{q}_{\hat{\theta}_1}$ can be found by solving the following linear feasibility problem:

$$\text{Find } p(\cdot|x) \in \Delta^{\mathcal{Y}} \tag{2.16}$$
$$s.t. \sum_{y \in A} p(y|x) \geq \nu_{\hat{\theta}_1}(A|x),\; A \in \mathcal{C}.$$

Since we condition on D_1, we also view p as fixed. Next, given p and $\theta \in \Theta_0$, and $x \in \mathcal{X}$, solve the following convex program:

$$q_\theta(\cdot|x) = \underset{q(\cdot|x)\in\Delta^{\mathcal{Y}}}{\arg\min} \sum_{y\in\mathcal{Y}} \ln\Big(\frac{q(y|x)+p(y|x)}{q(y|x)}\Big)(q(y|x)+p(y|x)) \tag{2.17}$$
$$s.t. \sum_{y\in A} q(y|x) \geq \nu_\theta(A|x;\theta),\ A\in\mathcal{C}, x\in X.$$

As we will show, q_θ can be derived analytically in leading examples. Also, we note that numerically solving (2.17) is straightforward as long as the cardinality of $\mathcal{Y}$ is moderate (see Remark 2.2). We then compute MLE $\hat{\theta}_0$ in (2.6). Finally, compute

$$T_n = \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} = \frac{\prod_{i\in D_0} q_{\hat{\theta}_1}(Y_i|X_i)}{\prod_{i\in D_0} q_{\hat{\theta}_0}(Y_i|X_i)}, \tag{2.18}$$

and let S_n be defined as in (2.8).[8]

REMARK 2.2: *The computational cost of solving* (2.17) *(or* (2.16)*) remains low as long as $\mathcal{Y}$ does not contain many elements. For example, for an entry game with two players, one may characterize the core of ν_θ by two inequality and two equality restrictions. Our procedure is most effective in such settings. Also, there exists a sub-collection $\mathcal{A}\subset\mathcal{C}$, known as the* minimal core determining class*, which characterizes the core of ν_θ and has the smallest cardinality (Luo and Wang, 2017, Ponomarev, 2022). It can be identified by an algorithm that does not require any data (Ponomarev, 2022). Thus, for $\mathcal{Y}$ with high cardinality, it is useful to first reduce the number of constraints by focusing on $\mathcal{A}$.*

3. MAIN THEORETICAL RESULTS

We make the following assumption on sampling.

[8]To compute T_n, we only need q_θ to be defined over a smaller parameter space $\Theta_0 \cup \{\hat{\theta}_1\}$, which is a non-random domain conditional on D_1.

ASSUMPTION 1: *(i)* $(Y_i, X_i, U_i), i = 1, \dots, n$ *are independently distributed across* i*; (ii) for each* i*,* $U_i | X_i = x \sim F_\theta(\cdot|x)$*.*

Independently distributed (X_i, U_i) is often assumed. Y_i's independence across i is subtle as it rules out the cross-sectional dependence of outcomes through selection.[9] However, this assumption does not rule out the possibility that the selection mechanism η_i differs across experiments. Hence, the selection can be arbitrarily heterogeneous across i. We also provide a modified version of the procedure that relaxes Y_i's independence across i (see Appendix B).

For each $\theta \in \Theta$, let

$$\mathcal{P}^n_\theta \equiv \Big\{ P^n \in \Delta(\mathcal{Z}^n) : P^n = \bigotimes_{i=1}^n P_i, P_i(\cdot|x) = \int_{\mathcal{U}} \eta_i(\cdot|u,x) dF_\theta(u|x), \\ \eta_i(\cdot|u,x) \in \Delta(G(u|x;\theta)),\ a.s. \Big\}. \quad (3.1)$$

Let $\mathcal{P}^n_0 \equiv \{P^n \in \mathcal{P}^n_\theta : \theta \in \Theta_0\}$ be the set of data generating processes (DGPs) compatible with H_0. The following theorem establishes the universal validity of the proposed test.

THEOREM 1: *Suppose Assumption 1 holds. Then, for any* $n \in \mathbb{N}$*,*

$$\sup_{P^n \in \mathcal{P}^n_0} P^n\big(S_n > \frac{1}{\alpha}\big) \le \alpha. \quad (3.2)$$

We build on this result to construct a finite-sample valid confidence region for functions of θ. Let $\varphi : \Theta \to \mathbb{R}^{d_\varphi}$. For each $\varphi^* \in \mathbb{R}^{d_\varphi}$, let $\Theta_0(\varphi^*) \equiv \{\theta \in \Theta : \varphi(\theta) = \varphi^*\}$ and $\Theta_1(\varphi^*) \equiv \{\theta \in \Theta : \varphi(\theta) \neq \varphi^*\}$. For example, a leading example of φ is the counterfactual choice probability.[10] Let $T_n(\varphi^*)$ be defined as in (2.18) for this hypothesis, and let $S_n(\varphi^*)$

---

[9] See Epstein et al. (2016) for further discussion on the matter and its consequences on limit theorems.

[10] One can also construct confidence regions for the entire parameter by taking φ to be the identity map and letting $\Theta_0(\theta^*) = \{\theta \in \Theta : \varphi(\theta) = \theta^*\} = \{\theta^*\}$ or a subvector of θ by taking φ to be the projection map to suitable coordinates.

be defined similarly. Define

$$CS_n \equiv \left\{\varphi^* \in \mathbb{R}^{d_\varphi} : S_n(\varphi^*) \leq \frac{1}{\alpha}\right\}. \tag{3.3}$$

Finally, define $\mathcal{F}_0^n \equiv \left\{(\varphi^*, P^n) : P^n \in \mathcal{P}_\theta^n,\ \varphi(\theta) = \varphi^*,\ \text{for some } \theta \in \Theta\right\}$. Then, the following coverage result holds.

COROLLARY 1: *Suppose Assumption 1 holds. Then, for any* $n \in \mathbb{N}$,

$$\inf_{(\varphi^*, P^n) \in \mathcal{F}_0^n} P^n\left(\varphi^* \in CS_n\right) \geq 1 - \alpha. \tag{3.4}$$

4. EXAMPLES

We illustrate the main theorem through examples.

Example 1 (Tests of Strategic Interaction)**.** Consider a two-player static game of complete information (Bresnahan and Reiss, 1990, 1991). Each player may either choose $y^{(j)} = 0$ or $y^{(j)} = 1$. The payoff of player j is

$$\pi^{(j)} = y^{(j)}\left(x^{(j)\prime}\delta^{(j)} + \beta^{(j)} y^{(-j)} + u^{(j)}\right),\ j = 1, 2 \tag{4.1}$$

where $y^{(-j)} \in \{0, 1\}$ is the opponent's action, $x^{(j)}$ is player j's observable characteristics, and $u^{(j)}$ is an unobservable payoff shifter. Suppose the players play a pure strategy Nash equilibrium (PSNE). Let $\theta = (\beta', \delta')'$, and assume $\beta \leq 0$. For each θ, the following

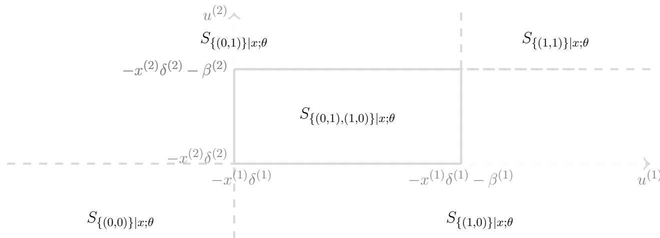

FIGURE 1.—Level sets of $G(\cdot|x;\theta)$ with $\beta^{(j)} < 0, j = 1, 2$.

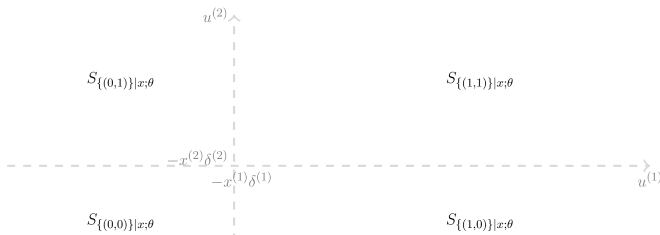

FIGURE 2.—Level sets of $G(\cdot|x;\theta)$ with $\beta^{(j)} = 0, j = 1, 2$.

correspondence gives the set of equilibria: (Beresteanu et al., 2011, Proposition 3.1):

$$G(u|x;\theta) = \begin{cases} \{(0,0)\} & u \in S_{\{(0,0)\}|x;\theta} \equiv \{u : u^{(j)} < -x^{(j)\prime}\delta^{(j)}, j = 1, 2\}, \\ \{(0,1)\} & u \in S_{\{(0,1)\}|x;\theta} \equiv \{u^{(1)} < -x^{(1)\prime}\delta^{(1)}, u^{(2)} > -x^{(2)\prime}\delta^{(2)}\} \\ & \quad \cup \{-x^{(1)\prime}\delta^{(1)} < u^{(1)} < -x^{(1)\prime}\delta^{(1)} - \beta^{(1)}, u^{(2)} > x^{(2)\prime}\delta^{(2)} - \beta^{(2)}\}, \\ \{(1,0)\} & u \in S_{\{(1,0)\}|x;\theta} \equiv \{u^{(1)} > -x^{(1)\prime}\delta^{(1)} - \beta^{(1)}, u^{(2)} < -x^{(2)\prime}\delta^{(2)} - \beta^{(2)}\} \\ & \quad \cup \{-x^{(1)\prime}\delta^{(1)} < u^{(1)} < -x^{(1)\prime}\delta^{(1)} - \beta^{(1)}, u^{(2)} < -x^{(2)\prime}\delta^{(2)}\}, \\ \{(1,1)\} & u \in S_{\{(1,1)\}|x;\theta} \equiv \{u : u^{(j)} > -x^{(j)\prime}\delta^{(j)} - \beta^{(j)}, j = 1, 2\}, \\ \{(1,0),(0,1)\} & u \in S_{\{(0,1),(1,0)\}|x;\theta} \equiv \{u : -x^{(j)\prime}\delta^{(j)} < u^{(j)} < -x^{(j)\prime}\delta^{(j)} - \beta^{(j)}, j = 1, 2\}. \end{cases} \tag{4.2}$$

The model admits multiple equilibria $\{(1,0),(0,1)\}$ when $\beta^{(j)} < 0$ (see Figure 1).

Let $F_\theta(\cdot|x)$ be the conditional distribution of $U|X$. The set $\mathfrak{q}_{\theta,x}$ of densities compatible with θ is

$$\mathfrak{q}_\theta = \Big\{ q \in \Delta : \ q((0,0)|x) = F_\theta(S_{\{(0,0)\}|x;\theta}|x); \qquad q((1,1)|x) = F_\theta(S_{\{(1,1)\}|x;\theta}|x);$$
$$F_\theta(S_{\{(1,0)\}|x;\theta}|x) \le q((1,0)|x) \le F_\theta(S_{\{(1,0)\}|x;\theta}) + F_\theta(S_{\{(0,1),(1,0)\}|x;\theta}|x), \quad x \in \mathcal{X} \Big\}. \tag{4.3}$$

Consider testing

$$H_0 : \beta^{(j)} = 0, \ j = 1, 2 \quad v.s. \quad H_1 : \beta^{(j)} < 0, \ \text{for some } j. \tag{4.4}$$

Let $\hat{\theta}_1$ be an estimator of θ based on D_1. For example, it can be the minimizer of a sample criterion function based on the moment restrictions in (4.3). For any $p(\cdot|x)$ solving (2.16), the least-favorable density solving (2.17) is unique. This is because the model implies a unique density under H_0 (see Figure 2).[11] For example, suppose $U = (U^{(1)}, U^{(2)})$ follows the bivariate standard normal distribution. The following is the LFP-based likelihood:

$$\begin{aligned} q_\theta(y|x) &= \prod_{\bar{y} \in \mathcal{Y}} F_\theta(S_{\{\bar{y}\}|x;\theta}|x)^{1\{y=\bar{y}\}} \\ &= (1 - \Phi(x^{(1)\prime}\delta^{(1)}))(1 - \Phi(x^{(2)\prime}\delta^{(2)}))^{1\{y=(0,0)\}} \\ &\quad \times (1 - \Phi(x^{(1)\prime}\delta^{(1)}))\Phi(x^{(2)\prime}\delta^{(2)})^{1\{y=(0,1)\}} \times \Phi(x^{(1)\prime}\delta^{(1)})(1 - \Phi(x^{(2)\prime}\delta^{(2)}))^{1\{y=(1,0)\}} \\ &\quad \times \Phi(x^{(1)\prime}\delta^{(1)})\Phi(x^{(2)\prime}\delta^{(2)})^{1\{y=(1,1)\}}. \end{aligned}$$

Let $\hat{\theta}_0$ be the restricted maximum likelihood estimator (RMLE) from D_0:

$$\hat{\theta}_0 \in \arg\max_{\theta \in \Theta_0} \mathcal{L}_0(\theta), \tag{4.5}$$

which consists of $\beta^{(j)} = 0, j = 1, 2$, and a maximizer of $\delta \mapsto \prod_{i \in D_0} q_{(0,\delta)}(Y_i|X_i)$. Now, one can compute the test statistic S_n via (2.7)-(2.8).

[11] Completeness under the null hypothesis also occurs in other contexts such as dynamic discrete choice models (Chen and Kaido, 2023). Our framework does not need this requirement. In general, the solution to (2.17) can be derived analytically even if the model is incomplete under H_0 (see Proposition 4.1).

The next example is on the counterfactual probabilities.

Example 2 (Counterfactual Probabilities)**.** Consider the same example. Define the potential entry of player j by

$$Y^{(j)}(x^{(j)},y^{(-j)})=1\{x^{(j)\prime}\delta^{(j)}+\beta^{(j)}y^{(-j)}+U^{(j)}\geq 0\}. \tag{4.6}$$

This is the counterfactual entry decision by player j when the covariates and the other player's action are set to $(x^{(j)},y^{(-j)})$. Define the *counterfactual entry probability* by

$$\varphi(\theta)\equiv P(Y^{(j)}(x^{(j)},y^{(-j)})=1)=F_\theta(\{u:x^{(j)\prime}\delta^{(j)}+\beta^{(j)}y^{(-j)}\geq -u^{(j)}\}), \tag{4.7}$$

where F_θ is the marginal distribution of U. Consider testing,

$$H_0:\varphi=\varphi_0,\ \ v.s.\ \ H_1:\varphi\neq\varphi_0. \tag{4.8}$$

We derive the LFP-based density in closed form. Let us introduce the following objects:

$$\eta_1(\theta;x)=1-F_\theta(S_{\{(0,0)\}|x;\theta}|x)-F_\theta(S_{\{(1,1)\}|x;\theta}|x), \tag{4.9}$$

$$\eta_2(\theta;x)=F_\theta(S_{\{(1,0)\}|x;\theta}|x)+F_\theta(S_{\{(0,1),(1,0)\};\theta}|x), \tag{4.10}$$

$$\eta_3(\theta;x)=F_\theta(S_{\{(1,0)\}|x;\theta}|x). \tag{4.11}$$

Here, $\eta_1(\theta;x)$ is the probability allocated by the model to either $(1,0)$ or $(0,1)$ occurring as an equilibrium outcoe of the game; $\eta_2(\theta;x)$ ($\eta_3(\theta;x)$) is the upper (lower) bound implied by the model on the probability that $(1,0)$ is the equilibrium outcome of the game.

PROPOSITION 4.1: *For any p and θ such that $p\notin\mathfrak{q}_\theta$, the LFP-based density q_θ has the following closed-form:*

$$q_\theta((0,0)|x)=F_\theta(S_{\{(0,0)\}|x;\theta}|x) \tag{4.12}$$

$$q_\theta((1,1)|x)=F_\theta(S_{\{(1,1)\}|x;\theta}|x) \tag{4.13}$$

$$q_\theta((0,1)|x) = \begin{cases} \frac{p((0,1)|x)}{p((1,0)|x)+p((0,1)|x)}\eta_1(\theta;x) & \theta \in \Theta_1(x,p) \\ \eta_1(\theta;x) - \eta_2(\theta;x) & \theta \in \Theta_2(x,p) \\ \eta_1(\theta;x) - \eta_3(\theta;x) & \theta \in \Theta_3(x,p) \end{cases} \tag{4.14}$$

$$q_\theta((1,0)|x) = \begin{cases} \frac{p((1,0)|x)}{p((1,0)|x)+p((0,1)|x)}\eta_1(\theta;x) & \theta \in \Theta_1(x,p) \\ \eta_2(\theta;x) & \theta \in \Theta_2(x,p) \\ \eta_3(\theta;x) & \theta \in \Theta_3(x,p) \end{cases}, \tag{4.15}$$

where

$$\Theta_1(x,p) \equiv \left\{\theta \in \Theta : \eta_3(\theta;x) \le \frac{p((1,0)|x)}{p((1,0)|x)+p((0,1)|x)}\eta_1(\theta;x) \le \eta_2(\theta;x)\right\} \tag{4.16}$$

$$\Theta_2(x,p) \equiv \left\{\theta \in \Theta : \frac{p((1,0)|x)}{p((1,0)|x)+p((0,1)|x)}\eta_1(\theta;x) > \eta_2(\theta;x)\right\} \tag{4.17}$$

$$\Theta_3(x,p) \equiv \left\{\theta \in \Theta : \frac{p((1,0)|x)}{p((1,0)|x)+p((0,1)|x)}\eta_1(\theta;x) < \eta_3(\theta;x)\right\}. \tag{4.18}$$

The test statistic $S_n(\varphi_0)$ can be calculated at any null counterfactual probability φ_0. Corollary 1 ensures inverting this test yields a universally valid confidence interval for the counterfactual probability.

In the analysis above, we considered the potential entry of a player when we set the other player's action to a certain value. Instead, one may be interested in the *potential (or counterfactual) equilibrium outcome* $Y(x)$, which is an equilibrium outcome that would realize when we set X to x. It is a measurable selection of $G(U|x;\theta)$. In general, we cannot uniquely determine the *counterfactual equilibrium probability* $P(Y(x) = y), y \in \mathcal{Y}$. However, we can express its bounds as functions of θ. For example, the sharp upper bound on the counterfactual probability of $Y(x) = (1,0)$ is

$$\begin{aligned} \varphi(\theta) &\equiv \sup_{P \in \mathcal{P}_\theta} P(Y(x) = (1,0)) \\ &= F_\theta(G(U|x;\theta) \cap \{(1,0)\} \neq \emptyset) = F_\theta(S_{\{(1,0)\}|x;\theta}) + F_\theta(S_{\{(0,1),(1,0)\};\theta}), \end{aligned} \tag{4.19}$$

where F_θ is the marginal distribution of U. The analysis for this object is essentially the same as before.

Example 3 (Set-Valued Control Functions)**.** Consider a triangular model

$$Y = 1\{\alpha D + W'\beta + \varepsilon > 0\}, \tag{4.20}$$

$$D = 1\{Z'\delta + V > 0\}, \tag{4.21}$$

where Y is a binary outcome, D is a binary treatment, and W is a vector of observable control variables. Suppose that ε is independent of D conditional on (W, V). This means (W, V) serve as a set of control variables. With a binary endogenous variable, we cannot recover V as a function of other observables. However, we can construct the following *set-valued control function*:

$$\boldsymbol{V}(D, Z; \delta) = \begin{cases} [-Z'\delta, \infty) & \text{if} \quad D = 1 \\ (-\infty, -Z'\delta] & \text{if} \quad D = 0. \end{cases} \tag{4.22}$$

Suppose the conditional distribution of $\varepsilon|V$ belongs to a location family and one can write $\varepsilon = \eta V - U$, where $U \perp D|W, V$. Let $x = (l, w, z)$. Without further assumptions, the model's prediction is

$$G(u|x;\theta) = \Big\{y \in \{1, 0\} : y = 1\{\alpha d + w'\beta + \eta v - u > 0\},\ v \in \boldsymbol{V}(d, z; \delta)\Big\}, \tag{4.23}$$

One can simplify G with additional sign restrictions. For example, if $\eta < 0$,[12]

$$G(u|x;\theta) = \begin{cases} \{1\} & \text{if } d = 0,\ u < w'\beta - \eta z'\delta \\ \{0\} & \text{if } d = 1,\ u \geq \alpha + w'\beta - \eta z'\delta \\ \{0,1\} & \text{otherwise.} \end{cases} \tag{4.24}$$

[12] See Lemma A.1 for a full description of G.

The following proposition characterizes the LFP-based density.

PROPOSITION 4.2: *Let $F_\theta(\cdot|w)$ be the conditional CDF of $U|W$.*[13] *For any p and θ such that $p \notin \mathfrak{q}_\theta$, the LFP-based density q_θ has the following closed-form:*

$$q_\theta(1|x) = \begin{cases} F_\theta(w'\beta - \eta z'\delta|w) & \text{if } \eta < 0,\ d = 0,\ p(1|x) \le F_\theta(w'\beta - \eta z'\delta|w), \\ 1 - F_\theta(\alpha + w'\beta - \eta z'\delta|w) & \text{if } \eta < 0,\ d = 1,\ p(1|x) \ge 1 - F_\theta(\alpha + w'\beta - \eta z'\delta|w), \\ F_\theta(\alpha d + w'\beta|w) & \text{if } \eta = 0, \\ F_\theta(w'\beta - \eta z'\delta|w) & \text{if } \eta > 0,\ d = 0,\ p(1|x) \ge F_\theta(w'\beta - \eta z'\delta|w), \\ F_\theta(\alpha + w'\beta - \eta z'\delta|w) & \text{if } \eta > 0,\ d = 1,\ p(1|x) \le F_\theta(\alpha + w'\beta - \eta z'\delta|w), \\ p(1|x) & \text{otherwise.} \end{cases}$$

$$q_\theta(0|x) = 1 - q_\theta(1|x). \tag{4.25}$$

We may examine the effect of D by testing $H_0 : \alpha = 0$. Furthermore, we may examine the exogeneity of D by testing if $H_0 : \eta = 0$ is true. In addition, bounds on structural estimands can be obtained. For example, the lower bound on the average structural function $\text{ASF}(d) = E[1\{\alpha d + W'\beta + \varepsilon > 0\}]$ studied in Blundell and Powell (2003, 2004) is

$$\varphi(\theta) = E\Big[\inf_{v \in \boldsymbol{V}(D,Z;\delta)} F_{U|W}(\alpha d + W'\beta + \eta v|W)\Big],$$

where expectation is taken with respect to (D, Z, W, η).[14] We may conduct a one-sided hypothesis test on $\varphi(\theta)$ and invert it to obtain a confidence interval.

5. THEORY BEHIND UNIVERSAL INFERENCE FOR INCOMPLETE MODELS

This section outlines the machinery behind Theorem 1.

[13] Note that U is independent of L (a function of Z) conditional on (W, V). Hence U's distribution only depends only on W.

[14] Similarly, the lower bound on the average treatment effect is

$$\varphi(\theta) = E\Big[\inf_{v \in \boldsymbol{V}(D,Z;\delta)} \Big(F_{U|W}(\alpha + W'\beta + \eta v|W) - F_{U|W}(W'\beta + \eta v|W)\Big)\Big].$$

5.1. *Universal inference & Least Favorable Pair*

Wasserman et al. (2020) consider a probabilistic model $\{P_\theta, \theta \in \Theta\}$, where $P_\theta \in \Delta(\mathcal{Z})$. Let p_θ be the density of P_θ. Their split LR statistic is

$$T_n = \frac{\mathcal{L}_0(\hat{\theta}_1)}{\mathcal{L}_0(\hat{\theta}_0)} = \frac{\prod_{i \in D_0} p_{\hat{\theta}_1}(Z_i)}{\prod_{i \in D_0} p_{\hat{\theta}_0}(Z_i)}, \tag{5.1}$$

where $\hat{\theta}_1$ is an estimator computed from D_1, $\hat{\theta}_0 = \arg\max_{\theta \in \Theta_0} \mathcal{L}_0(\theta)$. They apply a Chernoff-type bound to the log-likelihood ratio to show the universal validity of their procedure.[15] For their argument to go through, the likelihood $\mathcal{L}_0(\cdot)$ must represent the conditional distribution of $\{Z_i, i \in D_0\}$ given D_1 when evaluated at the true parameter value. In our setting, the form of this distribution is unknown due to the selection mechanism. Instead of applying their argument directly, we let $\mathcal{L}_0(\cdot)$ represent the least-favorable distributions for a certain testing problem. Specifically, given $\hat{\theta}_1$ and a distribution $q_{\hat{\theta}_1}(\cdot|x) = p(\cdot|x)$ computed in (2.16), we consider distinguishing $\{\mathfrak{q}_\theta, \theta \in \Theta_0\}$ against a "representative" alternative $q_{\hat{\theta}_1}$. For each $\theta \in \Theta_0$, we compute the least favorable distribution $q_\theta \in \mathfrak{q}_\theta$, which is most difficult to distinguish from $q_{\hat{\theta}_1}$. We outline below the formal argument.

In what follows, let $Z = (Y, X)$ denote the vector of observable variables, and let $\mathcal{Z} = \mathcal{Y} \times \mathcal{X}$. Let P be the joint distribution of (Y, X), $P_{Y|X}$ be the conditional distribution of $Y|X$, and P_X be the marginal distribution of X, and we assume P_X is known. For each $\theta \in \Theta$, the following is the set of distributions of Z that are compatible with the model assumption:

$$\mathcal{P}_\theta \equiv \Big\{ P \in \Delta(\mathcal{Z}) : P_{Y|X}(\cdot|x) = \int_{\mathcal{U}} \eta(\cdot|u, x) dF_\theta(u|x), \quad \text{for some } \eta(\cdot|u, x) \in \Delta(G(u|x; \theta)) \Big\}, \tag{5.2}$$

[15] To be precise, they apply Markov's inequality to T_n, which can be viewed as bounding the tail probability using the moment generating function (MGF) of the log-likelihood. The key is to use a simple yet nontrivial result $E_{P_\theta}[e^{\ln T_n}] \leq 1$. They call this approach "poor man's Chernoff bound". See their discussion on page 16882.

where η is a conditional law of Y supported on $G(u|x;\theta)$.

Our starting point is an analog of the Neyman–Pearson framework. 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, $H_0 : \theta = \theta_0$, against a simple alternative hypothesis, $H_1 : \theta = \theta_1$. In *complete* models, a well-defined reduced form induces a unique likelihood function. In such settings, an optimal test is an LR test by the Neyman–Pearson lemma. In incomplete models, however, the model generally admits a (non-singleton) set $\mathcal{P}_\theta$ of distributions, which prevents us from directly applying the Neyman–Pearson lemma.

Hence, we consider *minimax tests* building on Huber and Strassen (1973, 1974).[16] Let $\phi : \mathcal{Z} \mapsto [0, 1]$ be a (possibly randomized) test. For each P on $(\mathcal{Z}, \Sigma_\mathcal{Z})$, the rejection probability of ϕ is $E_P[\phi(Z)] = \int \phi(z)dP$. Let $\pi_\theta(\phi) \equiv \inf_{P\in\mathcal{P}_\theta} E_P[\phi(Z)]$ be the *power guarantee* of ϕ under θ. This is the power value certain to be obtained regardless of the unknown selection mechanism. We call ϕ a *level-α minimax test* if it satisfies the following conditions:

$$\sup_{P\in\mathcal{P}_{\theta_0}} E_P[\phi(Z)] \le \alpha \;, \tag{5.3}$$

and

$$\pi_{\theta_1}(\phi) \ge \pi_{\theta_1}(\tilde{\phi}),\ \forall \tilde{\phi} \text{ satisfying (5.3)}. \tag{5.4}$$

Here, (5.3) imposes a uniform size control requirement. In (5.4), tests are ranked in terms of their power guarantee.

The lower envelope of $\mathcal{P}_\theta$

$$\nu_\theta(A) \equiv \inf_{P\in\mathcal{P}_\theta} P(A) \tag{5.5}$$

[16] See Ch. 8 of Lehmann and Romano (2006) for a general treatment of the topic.

is a non-additive set function called *belief function* (or totally monotone capacity).[17] We also define the upper envelope $\nu_\theta^*(A) \equiv \sup_{P\in\mathcal{P}_\theta} P(A)$, which satisfies the conjugacy relationship $\nu_\theta^*(A) = 1 - \nu_\theta(A^c)$ to the belief function.[18]

We summarize the model's prediction by the correspondence $\Gamma : \mathcal{X} \times \mathcal{U} \times \Theta \to \mathcal{Z}$:

$$\Gamma(x,u;\theta) \equiv \{(y,x) \in \mathcal{Z} : y \in G(u|x;\theta)\}. \tag{5.6}$$

This map collects all values of the observable variables compatible with θ. By Choquet's theorem (e.g., Choquet, 1954, Philippe et al., 1999, Molchanov, 2006), we may represent ν_θ and ν_θ^* using the distribution of the random set $\Gamma(X,U;\theta)$:

$$\nu_\theta(A) = \int_{\mathcal{X}} \int_{\mathcal{U}} 1\{\Gamma(x,u;\theta) \subset A\} dF_\theta(u|x) dP_X(x)$$
$$\nu_\theta^*(A) = \int_{\mathcal{X}} \int_{\mathcal{U}} 1\{\Gamma(x,u;\theta) \cap A \neq \emptyset\} dF_\theta(u|x) dP_X(x).$$

A belief function is a special case of two-monotone capacities whose properties have proven powerful for conducting robust inference (Huber, 1981).[19] In particular, for models whose lower envelopes are two-monotone, Huber and Strassen (1973) showed that the rejection region of a minimax test takes the form $\{z : \Lambda(z) > t\}$ for a measurable function Λ. Further, the following results follow from their Theorems 4.1 and 6.1. For this, let $\frac{dQ_1}{dQ_0} = \{\frac{q_1}{q_0} : q_j \in \frac{dQ_j}{dv},\ q_j \geq 0,\ j = 0,1,\ q_0 + q_1 > 0\}$ be the Radon-Nikodym derivative, where v is a measure that dominates $Q_j, j = 0,1$. Below, we take v to be the counting measure.

[17] It is not a probability measure due to the lack of additivity. When the model is complete (i.e., $\mathcal{P}_\theta$ is a singleton) ν_θ is a probability measure.

[18] The total monotonicity of ν_θ follows from Philippe et al. (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.

[19] See Appendix A.1

PROPOSITION 5.1: *Suppose Assumption 1 holds. Let $\mathcal{P}_{\theta_0}$ be defined as in (5.2). Then, (i) a* least-favorable pair (LFP) $(Q_0, Q_1) \in \mathcal{P}_{\theta_0} \times \mathcal{P}_{\theta_1}$ *exists such that for all $t \in \mathbb{R}$,*

$$\nu^*_{\theta_0}(\Lambda > t) = Q_0(\Lambda > t) \tag{5.7}$$

$$\nu_{\theta_1}(\Lambda > t) = Q_1(\Lambda > t), \tag{5.8}$$

where Λ is a version of dQ_1/dQ_0.
(ii) The LFP densities $q_j = dQ_j/d\upsilon, j = 0, 1$ solve the following convex program:

$$(q_0(\cdot|x), q_1(\cdot|x)) = \underset{p_0(\cdot|x), p_1(\cdot|x)}{\arg\min} \sum_{y \in \mathcal{Y}} \ln\left(\frac{p_0(y|x) + p_1(y|x)}{p_0(y|x)}\right)(p_0(y|x) + p_1(y|x)) \tag{5.9}$$

$$s.t.\ \nu_{\theta_0}(B|x) \le \sum_{s \in B} p_0(y|x),\ B \subset \mathcal{Y},\ x \in X$$

$$\nu_{\theta_1}(B|x) \le \sum_{s \in B} p_1(y|x),\ B \subset \mathcal{Y},\ x \in X;$$

Heuristically, 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.[20]

We use Proposition 5.1 to construct the LFP-based likelihood. Namely, for each $\theta \in \Theta_0$, we form the LFP $(Q_\theta, Q_{\hat{\theta}_1}) \in \mathcal{P}_\theta \times \{Q_{\hat{\theta}_1}\}$, where $Q_{\hat{\theta}_1}/d\upsilon = q_{\hat{\theta}_1}$, and we set $q_\theta = dQ_\theta/d\upsilon$. The key is that, for each $\theta \in \Theta_0$, we set q_θ to be the density (within $\mathfrak{q}_\theta$) that is most difficult to distinguish from $q_{\hat{\theta}_1}$. Hence, we observe a large value of T_n only if there is strong evidence against the null hypothesis. This construction allows us to control the test's size even if the nature generates data from the least-favorable scenario for the size control (i.e., generating data from q_θ for some $\theta \in \Theta_0$).

[20] An extension of the classic Neyman–Pearson lemma is known to hold (Huber and Strassen, 1973). That is, an LR-test of the form

$$\phi(z) = \begin{cases} 1 \text{ if } & \Lambda(z) > C \\ \gamma \text{ if } & \Lambda(z) = C \\ 0 \text{ if } & \Lambda(z) < C, \end{cases}$$

is a level-α minmax test.

6. AN EMPIRICAL APPLICATION

As an illustration, we apply the proposed tests to the model of Lambert (2019) who studied the effects of lobbying on regulatory actions taken by the U.S. financial regulators.[21] One of the primary tools used by financial regulators to ensure the stability of the financial system is regulatory enforcement actions. These actions can take different forms, including fines, penalties, cease-and-desist orders, and other measures.

Our goals here are two-fold. First, revisiting the empirical question studied by Lambert (2019), we aim to test whether the banks' lobbying actions have any effects on the regulatory enforcement actions. The banks' lobbying behavior can be explained by the bank's rent-seeking behavior for receiving preferential treatment from the regulator (Stigler, 2021, Peltzman, 1976) or their voluntary information provision to reduce the chance of facing severe enforcement actions. This suggests the bank's lobbying decision is likely endogenous. In particular, their unobserved characteristics (e.g., unmeasured credit risk exposure), may affect their lobbying decisions and the regulator's enforcement decisions. Hence, we use the set-valued control function approach as in Example 3 to account for self-selection. Second, we also formally test whether the lobbying decisions are made endogenously.

Let Y_i be a binary outcome that takes 1 if the regulatory agency issues a severe enforcement action on bank i at time t. Suppose

$$Y_i = \mathbf{1}\{\alpha D_i + W_i'\beta + \varepsilon_i > 0\}, \tag{6.1}$$

$$D_i = \mathbf{1}\{Z_i'\delta + V_i > 0\}, \tag{6.2}$$

where $D_i = 1$ if bank i engaged in a lobbying activity, otherwise $D_i = 0$, and W_i represents additional control variables. The latent variable ε_i in the outcome equation may be related to the unobservable characteristics V_i, which are relevant to the banks lobbying decision. Hence, we assume $\varepsilon_i = \eta V_i + U_i$, where U_i satisfies $U_i \perp D_i | W_i, V_i$.

[21] We consider the Office of the Comptroller of the Currency (OCC), the Federal Deposit Insurance Corporation (FDIC), and the Federal Reserve System (Fed).

With a binary endogenous variable, we cannot uniquely recover V_i. We construct the following set-valued control function

$$\boldsymbol{V}(D_i, Z_i'\delta) = \begin{cases} [-Z_i'\delta, \infty) & \text{if} \quad D_i = 1 \\ (-\infty, Z_i'\delta] & \text{if} \quad D_i = 0 \end{cases} \tag{6.3}$$

Following Example 3, we construct the model's set-valued prediction as follows

$$G(U_i|X_i;\theta) \equiv \Big\{ y \in \{0,1\} : y = \mathbf{1}[\alpha D_i + W_i'\beta + \eta V_i + U_i > 0],\ V_i \in \operatorname{Sel} \boldsymbol{V}(D_i, Z_i, \delta) \Big\}\ , \tag{6.4}$$

where $\operatorname{Sel}\boldsymbol{V}$ is the set of all *measurable selections* of $\boldsymbol{V}$.[22]

We examine the effect of lobbying on the regulator's actions by testing

$$H_0 : \alpha = 0 \quad v.s. \quad H_1 : \alpha \neq 0.$$

For this, we follow the construction of $\mathcal{L}_0$ in Section 2.3. Specifically, we first compute $\hat{\theta}_1$, an estimator of $\theta = (\alpha, \beta', \delta', \eta)'$ by minimizing the sample criterion function as in (2.11) with

$$\nu_\theta(\{1\}|d = 0, w, z) = 1 - F_\theta(w'\beta - \eta z'\gamma|w), \ \text{ for } \theta \text{ such that } \eta < 0, \tag{6.5}$$

$$\nu_\theta(\{0\}|d = 1, w, z) = F_\theta(\alpha + w'\beta - \eta z'\delta|w), \ \text{ for } \theta \text{ such that } \eta < 0, \tag{6.6}$$

$$\nu_\theta(\{1\}|d = 1, w, z) = F_\theta(\alpha + w'\beta - \eta z'\delta|w), \ \text{ for } \theta \text{ such that } \eta \geq 0, \tag{6.7}$$

$$\nu_\theta(\{0\}|d = 0, w, z) = 1 - F_\theta(w'\beta - \eta z'\delta|w), \ \text{ for } \theta \text{ such that } \eta \geq 0. \tag{6.8}$$

We then obtain $q_{\hat{\theta}_1}$ by solving (2.16). Next, using Proposition 4.2, we maximize $\mathcal{L}_0$ subject to the constraint $\alpha = 0$, which yields $\hat{\theta}_0$. We obtain the split-LR statistic T_n in (2.18). Re-

[22] Lambert (2019) assumed the latent variables ε_i and v_i were bivariate normal distributions with mean zero and unit variance independent of Z_i. Here, we work with an assumption on the conditional law of ε_i only.

peating them while swapping the roles of D_0 and D_1 and aggregating the resulting statistics yields the cross-fit LR statistic S_n. Finally, we compare S_n to $1/\alpha$.

Furthermore, we test the endogeneity of lobbying decisions by considering

$$H_0 : \eta = 0 \quad v.s. \quad H_1 : \eta \neq 0.$$

This can be done by taking the same steps as above but replacing the constraint $\alpha = 0$ with $\eta = 0$ upon maximizing $\theta \mapsto \mathcal{L}_0(\theta)$.

6.1. *Data*

Our primary data consists of annual records of regulatory enforcement actions, lobbying activities, and other variables from 2008 to 2019, obtained from three sources. The first source is the S&P Capital IQ Pro database, which includes enforcement actions, financial activities, and demographic information.[23] The second source is the Center for Responsive Politics (CRP), which provides information on the lobbying activities of commercial banks. The third source is the Bureau of Economic Analysis, which provides county-level personal income, which serves as a regional economic control variable (see Lambert, 2019, for details). We have merged the data obtained from CRP and the Capital IQ Pro database to obtain the full set of banks' lobbying activities used for our analysis.

6.1.1. *Regulatory Enforcement Actions*

The Capital IQ Pro database provides information on the year and type of regulatory enforcement actions. Various actions may be taken if a regulator identifies financial weaknesses, managerial issues, or violations of banking regulations during the examination. These actions either request banks to adopt a resolution or require them to sign agreements to address the problem. When the bank and regulator sign on a bilateral written agreement about necessary actions, we treat it as a form of enforcement action. We also treat cease

[23] Since the original data (SNL Financial Database) used by Lambert (2019) is not publicly available, we use data from Capital IQ Pro (also called S&P Global Market Intelligence) established in 2016, which builds on datasets that used to be part of S&P Capital IQ and SNL Financial Database.

and desist orders, prompt corrective action directives, and deposit insurance threats as enforcement actions.[24] We let $Y_i = 1$ if a federal agency (OCC, FDIC, or Fed) issues a severe enforcement action against a given bank in the year the action becomes effective and zero otherwise.

6.1.2. *Bank Lobbying*

We use lobbying disclosure reports of commercial banks to determine each bank's lobbying activity in a given year.[25] All lobbying activities are recorded at the parent financial institution level rather than at the individual bank (subsidiary) level. A bank's lobbying status D_i is 1 if the parent institution is active in lobbying in a given year, otherwise $D_i = 0$. We identify 196 banks that were active in lobbying in any of the years from 2008 to 2019. This corresponds to 1489 lobbying bank-year observations.[26] In our analysis below, we use year 2010, which contained the largest number of lobbying activities between 2008 and 2012.[27]

6.1.3. *Instrumental variables and Control variables*

Following Lambert (2019), we use two instrumental variables (IV). The first IV is the initial market size, measured by a bank's total assets in 1998 relative to its within-state peers' total assets. The second IV is the distance (in kilometers) between the bank's headquarters and Washington, DC. The control variables include various covariates capturing the banks' financial characteristics and regional economic conditions (see Section D in the Online Supplement). The control variables we consider in the model are six proxies variables for CAMELS ratings. They are capital adequacy, asset quality, management quality, earnings, liquidity, and sensitivity to market risk. Additionally, we also control for a set of

[24] These are considered "severe" actions. There are also less severe actions, which we do not consider. They are usually issued against individuals associated with banks. See Section D.3 in the Online Supplement for details.

[25] The Lobbying Disclosure Act (LDA) of 1995 requires banks to report information on their lobbying activities to the Senate Office of Public Records (SOPR).

[26] Lobbying made by coalitions or associations is not considered in our sample, as they do not disclose membership information. This may result in an underestimation of some banks' lobbying activities.

[27] Table C.I presents the time distribution of lobbying banks. The proportion of lobbying banks was relatively high from 2008 to 2012 but has decreased since 2013 and remained fairly low from 2013 to 2019.

financial and demographic variables, including the deposit to asset ratio, total core deposits, size of total assets, bank ages and real personal income growth at the county level.

6.2. *Results*

In an ongoing work, we conduct tests of the the hypotheses discussed above. The results will be added to a revised version of the paper.

7. MONTE CARLO EXPERIMENTS

We examine the performance of the proposed test through simulations. First, we use the two-player entry game example with the following payoff:

$$\pi^{(j)} = y^{(j)}\big(\theta^{(j)}y^{(-j)} + u^{(j)}\big),\ j = 1, 2. \tag{7.1}$$

We then test $H_0 : \theta^{(j)} = 0, j = 1, 2$ against $H_1 : \theta^{(j)} < 0$ for some j. As discussed in Example 1, the model is complete under the null hypothesis, which determines $\mathcal{L}_0$. Hence, one only needs to determine $\hat{\theta}_1$. We consider two options. Both are extreme estimators. The first one is a minimizer of a sample criterion function based on the sample analog of the sharp identifying restrictions as in (2.11). We call this estimator a moment-based estimator. The second one maximizes the information-based objective function, which uses the Kullback-Leibler projection $p_\theta(Y_i; \hat{p}_n)$ of the empirical distribution $\hat{p}_n$ to the incomplete model (Kaido and Molinari, 2024). For each $\theta \in \Theta$, The estimator $\hat{\theta}_1$ is a maximizer of $\hat{\mathsf{Q}}_1(\theta) \equiv \sum_{i\in D_1} \ln p_\theta(Y_i; \hat{p}_n)$. We call this estimator an MLE.

We set the sample size to $n = 100$ and 200. We calculate the rejection probability of the test at the alternatives with $\theta^{(j)} = -h$ with $h \geq 0$. Under each alternative, the outcome $Y_i = (1, 0)$ is selected with probability 0.5 whenever the model admits multiple equilibria. Figure 3 plots the power of the two cross-fit tests and the power envelope. The two tests have similar power curves, suggesting the choice of the initial estimator $\hat{\theta}_1$ does not seem to matter, at least for this example. Overall, the proposed tests are conservative. Their rejection probabilities at $h = 0$ are nearly 0. However, they could detect mild strategic interaction effects (e.g., $h = 0.5$) with reasonable rejection probabilities (approximately 0.7 for $n = 100$ and 0.98 for $n = 200$) even in small samples.

In the second experiment, we use the specification of payoff functions in (4.1). For each j, $X^{(j)}$ is a covariate that takes $K = 5$ discrete values $\{-2, -1, 0, 1, 2\}$, and the two covariates are generated independently. When multiple equilibria are predicted, one of them gets selected with probability 0.5. We test $H_0 : \delta^{(j)} = 0, j = 1, 2$ against $H_1 : \delta^{(j)} \neq 0$ for some j. A difference from the previous specification is that the model is incomplete under the null hypothesis. While the model remains incomplete under H_0, carrying out the test is straightforward because $\mathcal{L}_0$ can be derived analytically (Proposition 4.1). We evaluate the power of the test against alternatives with $\delta^{(j)} = h, h > 0$ for samples of size $n = 100$ and 200. For this experiment, we use the moment-based estimator as $\hat{\theta}_1$.

Figure 4 summarizes the result. As in the previous example, the test tends to be conservative. Nonetheless, it has meaningful power against alternatives, even in small samples, and it exhibits monotonically increasing power curves. We also compare our test to the test developed by Bugni et al. (2017). Their test uses a statistic based on the sample analog of moment inequalities. We hence call it a moment-based test. Note that the previous designs use small sample sizes. The covariates $(X^{(1)}, X^{(2)})$ take 25 different values with equal probabilities, which makes the essential sample size in each bin very small. This feature was not an issue for the cross-fit LR test. However, it caused computational issues for implementing Bugni et al.'s (2017) test.[28] Hence, we compare the performance of the two tests with $n = 5000$. Figure 5 reports the power curves of the two tests. Both tests show monotonically increasing power curves, and neither of them is uniformly dominant. Finally, Table I reports the computation time required to implement the tests.[29] To mimic a realistic scenario, we parallelized Bugni et al.'s (2017) bootstrap repetitions across multiple

[28]For example, when evaluating their objective functions in the sample or with bootstrap samples, the moments based on empty bins caused computational issues. To avoid adding modifications not considered in their original paper, we work in an environment in which their procedure is reliable. With $n = 5000$, we can ensure that we have 200 observations in each bin on average.

[29]They were computed using Boston University's Shared Computing Cluster (SCC) nodes equipped with 2.6 GHz Intel Xeon processor (E5-2650v2) and 128GB memory (per node).

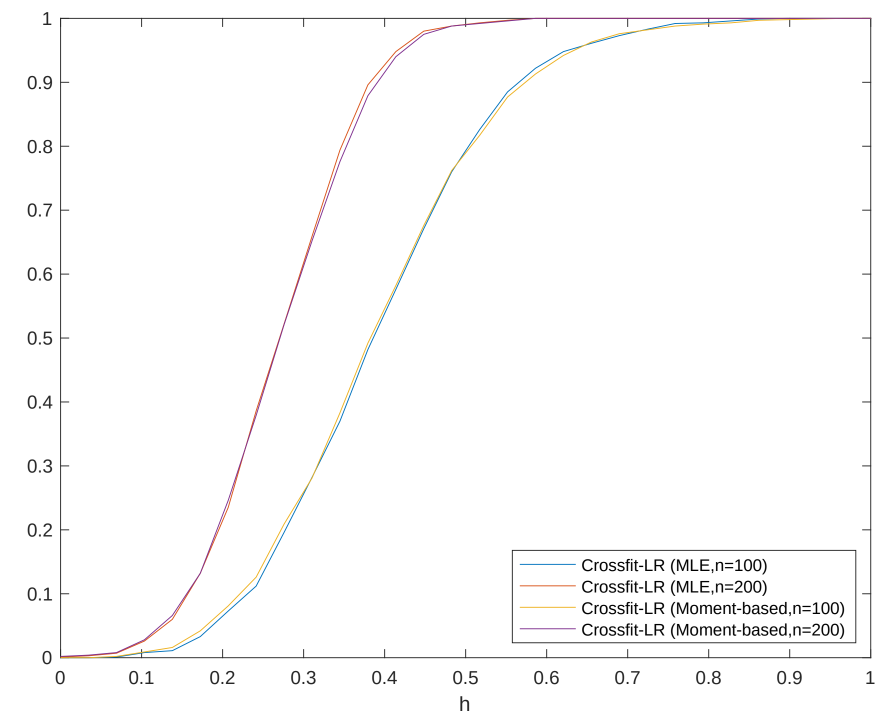

FIGURE 3.—Power of the Cross-Fit Tests for testing $H_0 : \beta^{(j)} = 0, j = 1, 2$.

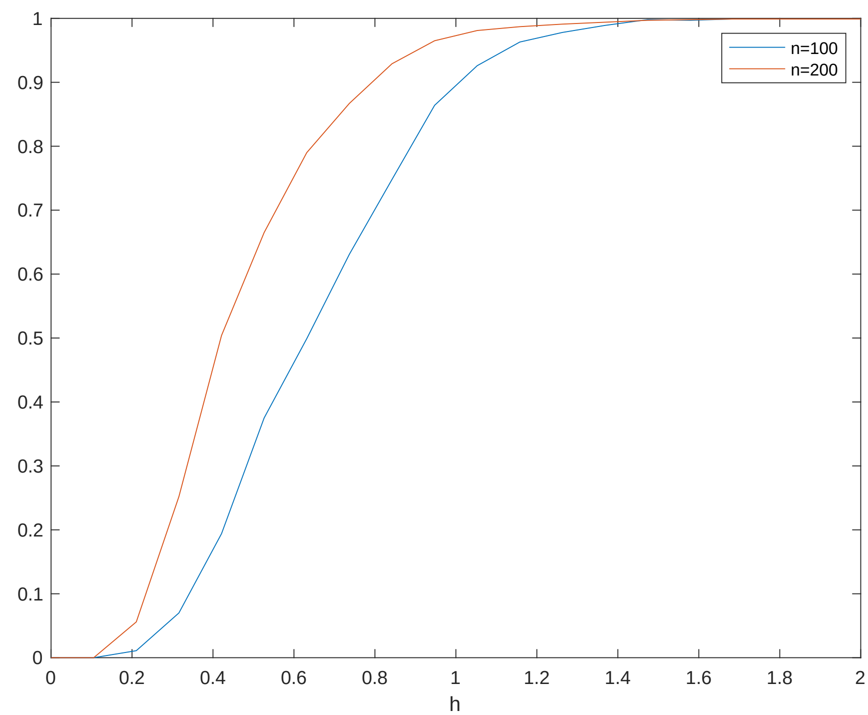

FIGURE 4.—Power of the Cross-Fit Tests for testing $H_0 : \delta^{(j)} = 0, j = 1, 2$. ($S = 1000$ replications)

Cross-fit LR test	Moment-based test		
	4 cores	8 cores	16 cores
13.75	111.65	56.64	41.84

TABLE I

COMPUTATION TIME (IN SECONDS)

Note: The median computation time is calculated based on $S = 1000$ simulations for the Cross-fit LR test. For the moment-based test of Bugni et al. (2017), we parallelized bootstrap replications with 4, 8, and 16 cores. The median computation time is calculated based on $S = 100$ simulation repetitions.

processors. The table shows that the cross-fit LR test takes about 14 seconds, which is significantly below the computation time (about 42 seconds) required for the moment-based test with 16 cores.

In sum, the simulation results show that the cross-fit LR test has non-trivial power even in small samples. In large samples for which existing tests are applicable, the proposed test has power properties comparable to a well-established test.

8. CONCLUSION

This paper develops a novel likelihood-based test and confidence regions for incomplete models. They apply to a wide range of discrete choice models involving set-valued predictions. Yet, they are simple to implement because they do not require any resampling or moment selection. To retain simplicity, this paper uses a simple two-fold cross-fitting method. An avenue for further research is to examine whether alternative sample-splitting schemes (e.g., K-fold cross fitting) can improve the statistical properties of the proposed test.

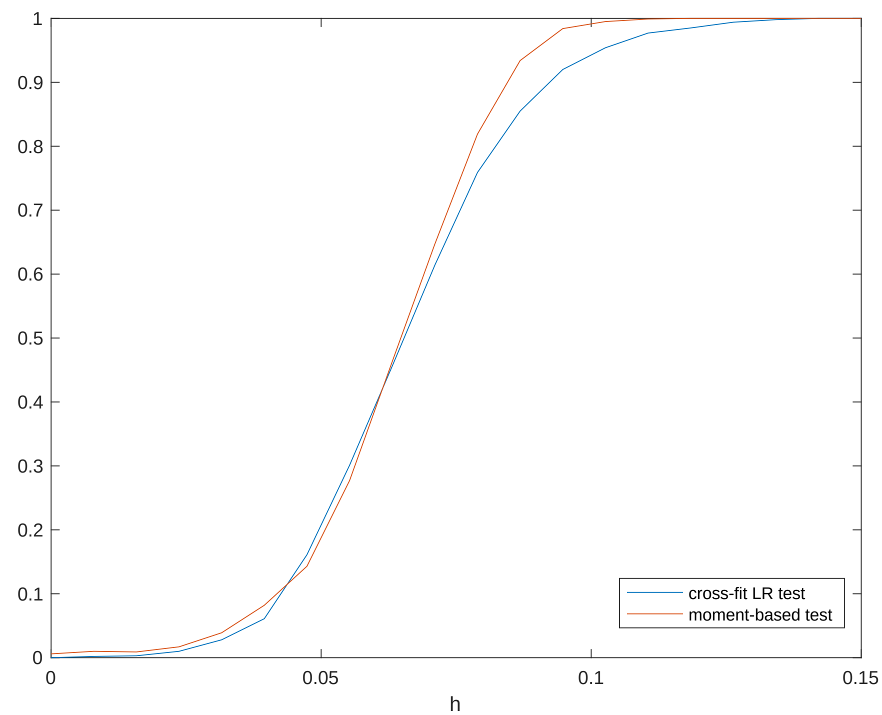

FIGURE 5.—Power of the Cross-fir LR and Moment-based Tests: ($n = 5000$, $S = 1000$ replications)

APPENDIX A: PROOFS

A.1. *Preliminaries*

Below, we introduce capacities and their basic properties. We refer to Denneberg (1994) for technical treatments.

Let $\Sigma_{\mathcal{Z}}$ be the Borel σ-algebra. A function $\nu : \Sigma_{\mathcal{Z}} \to \mathbb{R}$ with $v(\emptyset) = 0$ is a *capacity*. Throughout, we assume $\nu(A) \geq 0,\ \forall A \in \Sigma_{\mathcal{Z}}$, $\nu(\mathcal{Z}) = 1$ (i.e. normalized). We also assume ν is monotone. That is, for any $A, B \in \Sigma_{\mathcal{Z}}$, $A \subseteq B \Rightarrow \nu(A) \leq \nu(B)$. Capacity ν is said to be *monotone of order* k or, for short, *k-monotone* if for any $A_i \subset S, i = 1 \cdots, k$,

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

If the property holds for any k, it is called a *totally monotone* capacity. The conjugate $\nu^*(A) = 1 - \nu(A^c)$ of a k-monotone capacity is called a k-*alternating* capacity. For any capacity ν and a real-valued function f on $\mathcal{Z}$, the *Choquet integral* of f with respect to ν is defined by

$$\int f d\nu \equiv \int_{-\infty}^{0} (\nu(\{s : f(s) \geq t\}) - \nu(\mathcal{Z}))dt + \int_0^\infty \nu(\{s : f(s) \geq t\})dt. \tag{A.2}$$

Let $A(D_1) \in \Sigma_{\mathcal{Z}}$ be a measurable set, which is allowed to depend on subsample D_1. For each $\theta \in \Theta$, let

$$\nu_\theta^*(A(D_1)|D_1) \equiv \int 1\{G(u|X;\theta) \cap A(D_1) \neq \emptyset\} dF_\theta(u). \tag{A.3}$$

The set function $\nu_\theta^*(\cdot|D_1)$ is then a totally-alternating capacity (Philippe et al., 1999).

A.2. *Proof of Theorems*

PROOF OF THEOREM 1: We present a version of the proof for the split-sample statistic T_n first. Let $\theta \in \Theta_0$ and let $P^n \in \mathcal{P}_\theta^n$. Let P^{D_j} be the marginal distribution of P^n on D_j, and let $P^n(\cdot|D_1)$ be the conditional distribution given $\{(Y_i, X_i), i \in D_1\}$. By Markov's inequality,

$$P^n(T_n > \frac{1}{\alpha}) \leq \alpha E_{P^n}[T_n] = \alpha E_{P^{D_1}}[E_{P^n}[T_n|D_1]]. \quad \text{(A.4)}$$

For each $i \in D_0$, let $\Lambda_i = q_{\hat{\theta}_1}(Z_i)/q_\theta(Z_i)$ and define $T_n^*(\theta) = \mathcal{L}_0(\hat{\theta}_1)/\mathcal{L}_0(\theta) = \prod_{i \in D_0} \Lambda_i$. We may bound $E_{P^n}[T_n|D_1]$ as follows.

$$\begin{aligned} E_{P^n}[T_n|D_1] &\leq \sup_{\tilde{P}^n \in \mathcal{P}_\theta^n} E_{\tilde{P}^n}[T_n|D_1] \\ &\overset{(1)}{\leq} \sup_{\tilde{P}^n \in \mathcal{P}_\theta^n} E_{\tilde{P}^n}[T_n^*(\theta)|D_1] \\ &\overset{(2)}{=} \prod_{i \in D_0} \sup_{\tilde{P} \in \mathcal{P}_\theta} E_{\tilde{P}}[\Lambda_i|D_1] \\ &\overset{(3)}{=} \prod_{i \in D_0} \int \Lambda_i d\nu_\theta^*(\cdot|D_1), \end{aligned} \quad \text{(A.5)}$$

where (1) is due to $T_n \leq T_n^*(\theta)$ for any $\theta \in \Theta_0$ due to $\hat{\theta}_0$ being the maximizer of $\mathcal{L}_0$, (2) follows because of Assumption 1 (i) and the definition of $\mathcal{P}_\theta^n$ in (3.1), and (3) follows from $\sup_{\tilde{P} \in \mathcal{P}_\theta} E_{\tilde{P}}[\Lambda_i|D_1] = \int \Lambda_i d\nu_\theta^*(\cdot|D_1)$, a property of the Choquet integral for two-alternating capacities (Schmeidler, 1986). Observe that

$$\int \Lambda_i d\nu^*(\cdot|D_1) = \int_0^\infty \nu_\theta^*(\Lambda_i \geq t|D_1)dt = \int_0^\infty \nu_\theta^*(\Lambda_i > t|D_1)dt. \quad \text{(A.6)}$$

By Proposition 5.1 (i), there exists $Q_\theta \in \mathcal{P}_\theta$ such that, for all t,

$$\nu_\theta^*(\Lambda_i > t|D_1) = Q_\theta(\Lambda_i > t|D_1), \quad \text{(A.7)}$$

and Q_θ's conditional density q_θ solves (2.17) by Proposition 5.1 (ii). Let A be the support of q_θ. Then,

$$\int_0^\infty \nu_\theta^*(\Lambda_i > t|D_1)dt = \int_0^\infty Q_\theta(\Lambda_i > t|D_1)dt$$

$$= E_{Q_\theta}\Big[\frac{q_{\hat\theta_1}(Z_i)}{q_\theta(Z_i)}\Big|D_1\Big] = \int_A \frac{q_{\hat\theta_1}(z)}{q_\theta(z)} q_\theta(z)dz \le \int q_{\hat\theta_1}(z)dz = 1. \quad \text{(A.8)}$$

Combining (A.5)-(A.8) yields $E_{P^{D_1}}[T_n|D_1] \le 1$. Conclude that $P^n(T_n > \frac{1}{\alpha}) \le \alpha E_{P^n}[T_n] \le 1$ from (A.4) and observe that the bound applies uniformly across $P^n \in \mathcal{P}_0^n$.

For the cross-fit test, arguing as in (A.4),

$$P^n(S_n > \frac{1}{\alpha}) \le \alpha E_{P^n}[S_n] = \alpha E_{P^n}\Big[\frac{T_n + T_n^{\text{swap}}}{2}\Big] = \frac{\alpha}{2}(E_{P^n}[T_n] + E_{P^n}[T_n^{\text{swap}}]). \quad \text{(A.9)}$$

The rest of the proof is essentially the same. *Q.E.D.*

PROOF OF COROLLARY 1: The argument is by the standard test inversion. By Theorem 1, for any n,

$$P^n(\varphi^* \notin CS_n) = P^n(S_n(\varphi^*) > \frac{1}{\alpha}) \le \alpha,$$

implying $P^n(\varphi^* \in CS_n) \ge 1 - \alpha$ uniformly in $(\varphi^*, P^n) \in \mathcal{F}_0^n$. *Q.E.D.*

A.3. *Proof of the Propositions*

PROOF OF PROPOSITION 4.1: By (4.3), it suffices solve the following program

$$\begin{aligned}
&\operatorname*{arg\,min}_{q(\cdot|x)\in\Delta^{\mathcal{Y}}} \sum_{y\in\{(0,0),(1,1),(1,0),(0,1)\}} \ln\Big(\frac{q(y|x)+p(y|x)}{q(y|x)}\Big)(q(y|x)+p(y|x))\\
&\quad s.t.\ q((0,0)|x) = F_\theta(S_{\{(0,0)\}|x;\theta}|x)\\
&\qquad q((1,1)|x) = F_\theta(S_{\{(1,1)\}|x;\theta}|x)\\
&\qquad F_\theta(S_{\{(1,0)\}|x;\theta}|x) \le q((1,0)|x) \le F_\theta(S_{\{(1,0)\}|x;\theta}) + F_\theta(S_{\{(0,1),(1,0)\}|x;\theta}|x).
\end{aligned}$$

Note that $q_\theta((0,0)|x)$ and $q((1,1)|x)$ are determined by the equality constraints, which gives (4.12)-(4.13). Therefore, it suffices to solve the problem above for $q((1,0)|x)$. Let $z =$

$q((1,0)|x)$ and note that one can express $q((0,1)|x)$ as $q((0,1)|x) = \eta_1(\theta;x) - z$. Consider

$$\min_{z\in[0,1]} \ln\Big(\frac{z+p((1,0)|x)}{z}\Big)(z+p((1,0)|x)) + \ln\Big(\frac{\eta_1(\theta;x)-z+p((0,1)|x)}{\eta_1(\theta;x)-z}\Big)(\eta_1(\theta;x)-z+p((0,1)|x))$$
$$s.t. F_\theta(S_{\{(1,0)\}|x;\theta}|x) \le z \le F_\theta(S_{\{(1,0)\}|x;\theta}) + F_\theta(S_{\{(0,1),(1,0)\}|x;\theta}|x)$$

If $F_\theta(S_{\{(0,1),(1,0)\}|x;\theta}|x) > 0$, Slater's condition is satisfied. Solving the Karush-Kuhn-Tucker (KKT) condition for this problem yields (4.15). (4.14) follows from $q((0,1)|x) = \eta_1(\theta;x) - z$. If $F_\theta(S_{\{(0,1),(1,0)\}|x;\theta}|x) = 0$, the model is complete, and the solution reduces to $q_\theta(y|x) = F_\theta(S_{\{y\}|x;\theta}|x)$ for all y, which is a special case of (4.12)-(4.15). *Q.E.D.*

LEMMA A.1: *Suppose the model is characterized by* (4.20)-(4.21). *Then,* $G(u|x)$ *takes the following form.*

If $\eta < 0$,

$$G(u|x;\theta) = \begin{cases} \{1\} & \text{if } d=0,\ u < w'\beta - \eta z'\delta \\ \{0\} & \text{if } d=1,\ u \ge \alpha + w'\beta - \eta z'\delta \\ \{0,1\} & \text{otherwise}. \end{cases} \tag{A.10}$$

If $\eta > 0$,

$$G(u|x;\theta) = \begin{cases} \{1\} & \text{if } d=1,\ u < \alpha + w'\beta - \eta z'\delta \\ \{0\} & \text{if } d=0,\ u \ge w'\beta - \eta z'\delta \\ \{0,1\} & \text{otherwise.} \end{cases} \tag{A.11}$$

Finally, if $\eta = 0$,

$$G(u|x;\theta) = \begin{cases} \{1\} & u < \alpha d + w'\beta \\ \{0\} & u \ge \alpha d + w'\beta. \end{cases} \tag{A.12}$$

PROOF: We first derive (A.10). Suppose $\eta < 0$ and $d = 1$ first. By (4.20) and $\varepsilon = \eta v - u$, $y = 1$ if

$$u < \alpha + w'\beta + \eta v, \quad \text{for some } v \in [-z'\delta, \infty). \tag{A.13}$$

Then, by $\eta < 0$, we can write this event as $u \in \bigcup_{v \in [-z'\delta,\infty)} (-\infty, \alpha + w'\beta + \eta v) = (-\infty, \alpha + w'\beta - \eta z'\delta)$. Similarly, $y = 0$ if

$$u \geq \alpha + w'\beta + \eta v, \quad \text{for some } v \in [-z'\delta, \infty). \tag{A.14}$$

This means $y = 0$ is always consistent with the model because ηv is unbounded from below. Hence,

$$G(u|1, w, z; \theta) = \begin{cases} \{0, 1\} & u < \alpha + w'\beta - \eta z'\delta \\ \{0\} & u \geq \alpha + w'\beta - \eta z'\delta. \end{cases} \tag{A.15}$$

Now consider the case $d = 0$. By the same argument, $y = 1$ if

$$u < w'\beta + \eta v, \quad \text{for some } v \in (-\infty, -z'\delta]. \tag{A.16}$$

This means $y = 1$ is always consistent with the model because ηv is unbounded from above. Similarly, $y = 0$ if

$$u \geq w'\beta + \eta v, \quad \text{for some } v \in (-\infty, -z'\delta], \tag{A.17}$$

which is equivalent to $u \geq w'\beta - \eta z'\delta$. Therefore,

$$G(u|0, w, z; \theta) = \begin{cases} \{1\} & u < w'\beta - \eta z'\delta \\ \{0, 1\} & u \geq w'\beta - \eta z'\delta. \end{cases} \tag{A.18}$$

Combining (A.15)-(A.18) yields (A.10).

Now suppose $\eta > 0$. Consider the case $d = 1$. $y = 1$ if (A.13) holds. Since ηv is unbounded from above, $y = 1$ is consistent with the model regardless of the value of u. Similarly, $y = 0$ if (A.14) holds, and this event can be written as $u \geq \alpha + w'\beta - z'\delta$. Hence,

$$G(u|1,w,z;\theta) = \begin{cases} \{1\} & u < \alpha + w'\beta - \eta z'\delta \\ \{0,1\} & u \geq \alpha + w'\beta - \eta z'\delta. \end{cases} \tag{A.19}$$

Suppose $d = 0$. Again, $y = 1$ if (A.16) holds, which is equivalent to $u < w'\beta - \eta z'\delta$. Similarly, $y = 0$ if (A.17) holds, which means $y = 0$ is consistent with the model regardless of the value of u in this case. Hence,

$$G(u|1,w,z;\theta) = \begin{cases} \{0,1\} & u < w'\beta - \eta z'\delta \\ \{0\} & u \geq w'\beta - \eta z'\delta. \end{cases} \tag{A.20}$$

Combining (A.19)-(A.20) yields

$$G(u|x;\theta) = \begin{cases} \{1\} & \text{if } d = 1,\ u < \alpha + w'\beta - \eta z'\delta \\ \{0\} & \text{if } d = 0,\ u \geq w'\beta - \eta z'\delta \\ \{0,1\} & \text{otherwise.} \end{cases} \tag{A.21}$$

Finally, consider the case $\eta = 0$. By (4.21) and $\varepsilon = \eta v - u$,

$$G(u|x;\theta) = \begin{cases} \{1\} & u < \alpha d + w'\beta \\ \{0\} & u \geq \alpha d + w'\beta. \end{cases} \tag{A.22}$$

Q.E.D.

PROOF OF PROPOSITION 4.2: Let

$$\mathfrak{q}_{\theta,-} \equiv \Big\{ q(\cdot|x) \in \Delta : q(1|d=0,w,z) \geq F_\theta(w'\beta - \eta z'\delta|w),$$

$$q(0|d=1,w,z) \geq F_\theta(\alpha + w'\beta - \eta z'\delta|w), x \in \mathcal{X}\Big\},$$

$$\mathfrak{q}_{\theta,+} \equiv \Big\{q(\cdot|x) \in \Delta : q(1|d=1,w,z) \geq F_\theta(\alpha + w'\beta - \eta z'\delta|w),$$
$$q(0|d=0,w,z) \geq 1 - F_\theta(w'\beta - \eta z'\delta|w), x \in \mathcal{X}\Big\}.$$

Suppose for the moment $\eta < 0$. By Lemma A.1 (with $\eta < 0$),

$$\begin{aligned}\nu_\theta(\{1\}|d=0,w,z) &= \int 1\{G(u|x;\theta) \subseteq \{1\}\}dF_\theta(u|w)\\ &= \int 1\{u < w'\beta - \eta z'\delta\}dF_\theta(u|w) = F_\theta(w'\beta - \eta z'\delta|w).\end{aligned}$$

Similarly,

$$\begin{aligned}\nu_\theta(\{0\}|d=1,w,z) &= \int 1\{G(u|x;\theta) \subseteq \{0\}\}dF_\theta(u|w)\\ &= \int 1\{u \geq \alpha + w'\beta - \eta z'\delta\}dF_\theta(u|w) = 1 - F_\theta(\alpha + w'\beta - \eta z'\delta|w).\end{aligned}$$

Recall that $2^{\mathcal{Y}} = \{\emptyset, \{0\}, \{1\}, \{0,1\}\}$, $\nu_\theta(\emptyset|x) = 0$, and $\nu_\theta(\{0,1\}|x) = 1$. Hence, the only nontrivial restrictions are the ones above. Therefore, $\mathfrak{q}_\theta = \mathfrak{q}_{\theta,-}$ if $\eta < 0$. Similarly, $\mathfrak{q}_\theta = \mathfrak{q}_{\theta,+}$ if $\eta \geq 0$.

We continue to focus on the case with $\eta < 0$. For any p obtained from (2.16), the LFP-based density solves (2.17). For example, consider x such that $l = 0$, and let $r = q(1|x)$. Then,

$$\begin{aligned}q_\theta(1|x) = \arg\min_r \ln\Big(\frac{r + p(1|x)}{r}\Big)(r + p(1|x)) + \ln\Big(\frac{1 - r + p(0|x)}{1 - r}\Big)(1 - r + p(0|x))\\ s.t.\ r \geq F_\theta(w'\beta - \eta z'\gamma|w).\end{aligned}$$

The KKT conditions for this problem are

$$-\frac{p(1|x)}{r} + \ln\frac{p(1|x) + r}{r} + \frac{p(0|x)}{1 - r} - \ln\frac{1 - r + p(0|x)}{1 - r} - \lambda = 0 \tag{A.23}$$

$$r \geq F_\theta(w'\beta - \eta z'\gamma|w) \tag{A.24}$$

$$\lambda(F_\theta(w'\beta - \eta z'\gamma|w) - r) = 0 \tag{A.25}$$

$$\lambda \geq 0. \tag{A.26}$$

One one hand, if $r > F_\theta(w'\beta - \eta z'\gamma|w)$, it must be the case that $\lambda = 0$ by (A.25)-(A.26). Then, $r = p(1|x)$ solves (A.23). On the other hand, if $r = F_\theta(w'\beta - \eta z'\gamma|w)$, (A.23) and some algebra can show $\lambda \geq 0$ holds if and only if $p(1|x) \leq F_\theta(w'\beta - \eta z'\gamma|w)$. Solving the system above yields the following result:

$$q_\theta(1|x) = \begin{cases} F_\theta(w'\beta - \eta z'\gamma|w) & \text{if } d = 0, p(1|x) \leq F_\theta(w'\beta - \eta z'\gamma|w), \\ p(1|x) & \text{if } d = 0, p(1|x) > F_\theta(w'\beta - \eta z'\gamma|w), 0. \end{cases} \tag{A.27}$$

Repeating the same type of analysis for $d = 1$ gives

$$q_\theta(1|x) = \begin{cases} 1 - F_\theta(\alpha + w'\beta - \eta z'\gamma|w) & \text{if } d = 1, p(1|x) \geq 1 - F_\theta(\alpha + w'\beta - \eta z'\gamma|w), \\ p(1|x) & \text{if } d = 1, p(1|x) < 1 - F_\theta(\alpha + w'\beta - \eta z'\gamma|w). \end{cases} \tag{A.28}$$

The analysis for the case with $\eta \geq 0$ is similar and is therefore omitted. When $\eta = 0$, the model is complete, and hence $q_\theta(1|x) = F_\theta(\alpha d + w'\beta)$. Collecting all results yields (4.25). *Q.E.D.*

PROOF OF PROPOSITION 5.1: Note that a belief function is a totally monotone capacity. Hence, it is also 2-monotone, implying $\nu^*_{\theta_j}$ is 2-alternating for $j = 0, 1$. By Theorem 4.1 of Huber and Strassen (1973). The first claim follows. The second claim follows from taking $\Phi(\cdot) = -\ln(\cdot)$ in their Theorem 6.1 and noting the fact that the distribution of X is not restricted by the hypotheses. *Q.E.D.*

APPENDIX B: GENERAL DGPS WITH UNKNOWN DEPENDENCE

We consider more general data-generating processes than $\mathcal{F}_0$. For each i, let $G_i : \mathcal{U} \times \mathcal{X} \times \Theta \rightarrow \mathcal{Y}$ be a weakly measurable correspondence. Let

$$G^n(u^n|x^n;\theta) \equiv \prod_{i=1}^{n} G_i(u_i|X_i;\theta). \tag{B.1}$$

Let

$$\tilde{\mathcal{P}}^n_\theta \equiv \Big\{P^n \in \Delta(\mathcal{Z}^n) : P^n(\cdot|x^n) = \int_{\mathcal{U}^n} \eta(A|u^n, x^n) dF^n_\theta(u^n),\ \forall A \in \Sigma_{\mathcal{Y}^n}, \\ \eta(\cdot|u^n, x^n) \in \Delta(G^n(u^n|x^n;\theta)),\ a.s.\Big\}. \tag{B.2}$$

This set allows arbitrary dependence of the outcome sequence $Y^n = (Y_1, \ldots, Y_n)$ through the selection mechanism across n units. The following result is from Kaido and Zhang (2019).

THEOREM 2: *Suppose $\{(U_i, X_i)\}$ are independently distributed across i. Suppose, for each i, $\mathcal{P}_{\theta_0,i}$ and $\mathcal{P}_{\theta_1,i}$ are disjoint. Then, LFP $(Q^n_0, Q^n_1) \in \tilde{\mathcal{P}}^n_{\theta_0} \times \tilde{\mathcal{P}}^n_{\theta_1}$ exists such that for all $t \in \mathbb{R}_+$,*

$$\nu^{*,n}_{\theta_0}(\Lambda_n > t) = Q^n_0(\Lambda_n > t) \tag{B.3}$$

$$\nu^{n}_{\theta_1}(\Lambda_n > t) = Q^n_1(\Lambda_n > t), \tag{B.4}$$

where $\Lambda_n = dQ^n_1/dQ^n_0$. The LFP consists of the product measures:

$$Q^n_0 = \bigotimes_{i=1}^{n} Q_{0,i}, \quad \textit{and} \quad Q^n_1 = \bigotimes_{i=1}^{n} Q_{1,i}, \tag{B.5}$$

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;

The result above states for distinguishing θ_0 against θ_1, the least favorable pair consists of product measures. If (X_i, U_i) are identically distributed across i, the theorem implies the LFP consists of i.i.d. laws. This allows us to generalize the result in the text.

The i.i.d. sampling assumption in Wasserman et al. (2020) can be relaxed as long as one can compute a likelihood for D_0 conditional on D_1. This requirement is crucial for the sample-splitting (and cross-fitting) to work, but recall that the distribution of outcomes can be heterogeneous and dependent across i in unknown ways in our setting. This feature makes it hard to define a conditional likelihood and apply their argument. For the conditioning argument to work, we construct $\hat{\theta}_1$ using outcomes that can be uniquely determined by (X_i, U_i) but not by the selection. Using these insights, Theorem 3 below generalizes the universal inference result to a wider class of distributions.

We now explain how to construct an estimator $\hat{\theta}_1$ of θ under the general model.

DEFINITION B.1—Initial estimator: *$\hat{\theta}_1$ is an* initial estimator *of θ such that (i) it is constructed from $\{W_i = \varphi(Y_i, X_i), i \in D_1\}$ for a measurable function $\varphi : \mathcal{Y} \times \mathcal{X} \to \mathcal{W}$; (ii) Under H_0, one can wirte W_i as $W_i = f(X_i, U_i)$ for some measurable function $f : \mathcal{X} \times \mathcal{U} \to \mathcal{W}$.*

Definition B.1 states $\hat{\theta}_1$ is a function of observable variables from D_1, which can also be expressed as a function of the exogenous variables (X_i, U_i) at least under the null hypothesis. In Example 1, the model is complete under H_0. Hence, there is a unique reduced form $Y_i = g(U_i|X_i; \theta)$. This ensures we can construct $\hat{\theta}_1$ using directly $W_i = Y_i, i \in D_1$, i.e., φ is the identity map. For example, one can use an extremum estimator

$$\hat{\theta}_1 \in \arg\min_{\theta\in\Theta} \hat{\mathsf{Q}}_1(\theta), \tag{B.6}$$

where $\hat{\mathsf{Q}}_1$ can be an objective function based on moment inequalities as discussed earlier.

In Example 2, the model is incomplete under H_0. Hence, some outcomes cannot be expressed as a function of the exogenous variables (e.g., $Y_i = (1, 0)$ gets selected from multiple equilibria). Nonetheless, as shown by Bresnahan and Reiss (1990), Berry (1992), the number of entrants

$$W_i = \varphi(Y_i) = \sum_j 1\{Y_i^{(j)} = 1\} \tag{B.7}$$

is uniquely determined as a function of (X_i, U_i). For a two-player game, a natural candidate is

$$\hat{\theta}_1 \in \arg\max_{\theta \in \Theta} \sum_{i \in D_1} \sum_{k=0}^{2} 1\{W_i = k\} \ln F_\theta(S_{k,X_i}). \tag{B.8}$$

With this change, the rest of the procedure remains the same. For each i, let $q_{\hat{\theta}_1,i}$ be a solution to (2.16). Solve (2.17) to find $q_{\theta,i}, \theta \in \Theta_0$. Finally, form the cross-fit LR statistic as in (2.8). One can use the product LR statistic due to Theorem 2. The following theorem establishes the validity of the procedure. Let $\tilde{\mathcal{P}}_0^n \equiv \{P^n \in \tilde{\mathcal{P}}_\theta^n : \theta \in \Theta_0\}$.

THEOREM 3: *Suppose $(X_i, U_i), i = 1, \dots, n$ are independently distributed across i, and $U_i | X_i = x \sim F_{\theta,i}(\cdot | x)$.*

$$\sup_{P^n \in \tilde{\mathcal{P}}_0^n} P^n\Big(S_n > \frac{1}{\alpha}\Big) \leq \alpha. \tag{B.9}$$

PROOF: Let $\theta \in \Theta_0$ and $\hat{\theta}_1$ be an estimator of θ described in Definition B.1. Note that $\hat{\theta}_1$ is a function of $(X_i, U_i), i \in D_1$ only, and (X_i, U_i) are independently distributed across i. Therefore, conditioning on $(X_i, U_i), i \in D_1$ does not provide additional information on the observations from D_0 through selection across the two subsamples. Below, we condition on D_1 and treat $\hat{\theta}_1$ as fixed. Let $Q^n_{\hat{\theta}_1} \in \tilde{\mathcal{P}}^n_{\hat{\theta}_1}$. Consider a minimax testing problem between $\tilde{P}_\theta^n$ and $\{Q^n_{\hat{\theta}_1}\}$. By Theorem 2, there is a product LFP $(Q_\theta^n, Q^n_{\hat{\theta}_1}) \in \tilde{P}_\theta^n \times \{Q^n_{\hat{\theta}_1}\}$. We use Q_θ^n below.

The proof of the theorem is the same as the proof of Theorem 1 up to (A.4). We may bound $E_{P^n}[T_n|D_1]$ as follows.

$$\begin{aligned} E_{P^n}[T_n|D_1] &\le \sup_{\tilde{P}^n \in \tilde{\mathcal{P}}^n_\theta} E_{\tilde{P}^n}[T_n|D_1] \\ &\le \sup_{\tilde{P}^n \in \tilde{\mathcal{P}}^n_\theta} E_{\tilde{P}^n}[T^*_n(\theta)|D_1] \\ &= \sup_{\tilde{P}^n \in \tilde{\mathcal{P}}^n_\theta} \int_0^\infty \tilde{P}^n\big(T^*_n(\theta) > t \big| D_1\big) dt \\ &\le \int_0^\infty \sup_{\tilde{P}^n \in \tilde{\mathcal{P}}^n_\theta} \tilde{P}^n\big(T^*_n(\theta) > t \big| D_1\big) dt \\ &= \int_0^\infty Q^n_\theta\big(T^*_n(\theta) > t \big| D_1\big) dt, \end{aligned}$$

where the last equality is due to Q^n_θ being the least-favorable distribution in $\tilde{P}^n_\theta$. Let A^n be the support of Q^n_θ. Then,

$$\begin{aligned} &\int_0^\infty Q^n_\theta\big(T^*_n(\theta) > t \big| D_1\big) dt = E_{Q^n_\theta}\left[\frac{\prod_{i \in D_0} q_{\hat{\theta}_1}(Z_i)}{\prod_{i \in D_0} q_\theta(Z_i)}\right] \\ &= \int_{A^n} \frac{\prod_{i \in D_0} q_{\hat{\theta}_1}(z_i)}{\prod_{i \in D_0} q_\theta(z_i)} \prod_{i \in D_0} q_\theta(z_i) dz^{D_0} = \int_{A^n} \prod_{i \in D_0} q_{\hat{\theta}_1}(z_i) dz^{D_0} \le \prod_{i \in D_0} \int q_{\hat{\theta}_1}(z_i) dz_i = 1. \end{aligned} \tag{B.10}$$

Note that the second equality uses the fact $\prod_{i \in D_0} q_\theta$ is the density of Q^n_θ. The rest of the argument is the same as the proof of Theorem 1. *Q.E.D.*

APPENDIX C: TABLES AND FIGURES

[HK: I moved the definition of the variables to another file called supplement.tex]

Panel A: Lobbying Sample 2008-2012			
Year	Total number of banks	Number of lobbying banks	% of lobbying banks
2008	7,010	226	3.22
2009	6,813	160	2.35
2010	6,563	143	2.18
2011	6,366	133	2.09
2012	6,181	138	2.23
08-12	32,933	800	2.23
Panel B: Lobbying Sample 2013-2019			
2013	5,990	125	2.09
2014	5,782	102	1.76
2015	5,326	102	1.92
2016	5,131	93	1.81
2017	4,956	94	1.90
2018	4,634	94	2.03
2019	4,446	79	1.78
13-19	36,265	689	1.78
All years	69,198	1,489	2.15

TABLE C.I

TIME DISTRIBUTION OF THE LOBBYING SAMPLE RECORD

Panel A: 2008-2012 distribution of regulatory enforcement actions							
				Breakdown of severe actions			
Year	Any action	Severe actions	Less severe actions	Formal written agreements	Cease and desist orders	Prompt corrective action directives	Deposit insurance threats
2008	796	147	649	77	69	1	0
2009	1,331	352	979	149	190	13	0
2010	1,520	424	1096	177	227	20	0
2011	1,046	246	800	68	159	17	2
2012	784	164	620	47	106	11	0
Panel B: 2013-2019 distribution of regulatory enforcement actions							
2013	561	93	468	17	71	4	1
2014	463	51	412	5	41	5	0
2015	373	55	318	14	40	1	0
2016	370	35	335	8	25	2	0
2017	324	23	301	5	17	1	0
2018	351	30	321	5	22	3	0
2019	308	18	290	3	13	1	1

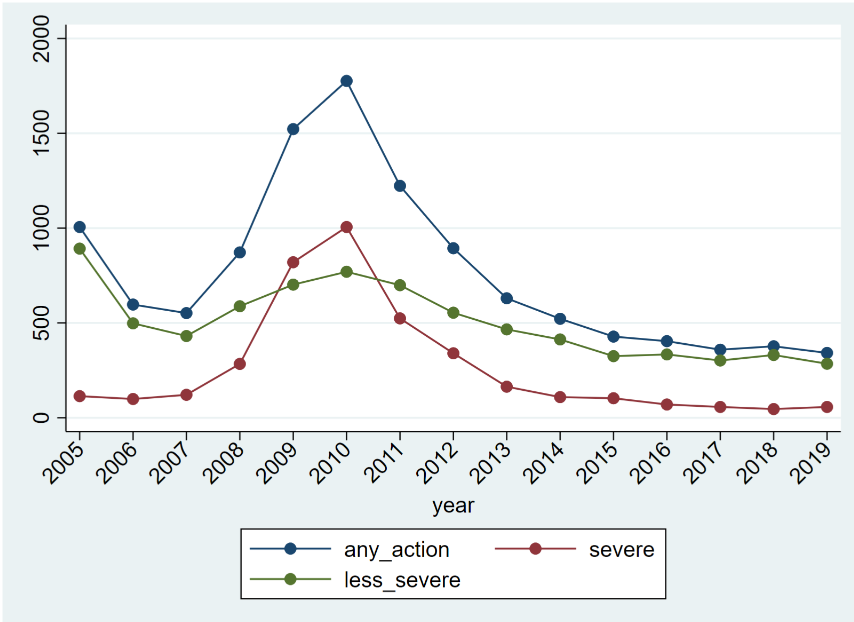

FIGURE C.1.—The enforcement action trend across different years

Panel A: Descriptive Statistics of the Sample Year 2010						
Variables	Mean	25th percentile	Median	75th percentile	Standard deviation	Number of observations
Dependent Variables (D_i and Y_i)						
Lobbying status	0.022	-	-	-	0.146	6,201
Severe action	0.067	-	-	-	0.250	6,201
Instrumental Variables (IVs)						
Distance to DC (km)	1,445,030	942.237	1383.535	1803.314	858.964	6,199
Initial Market Size(%)	0.077	0.003	0.009	0.028	0.486	6,201
Independent Variables (W_i)						
CAMELS proxy variables (also used in Z_i)						
Capital adequacy (%)	15.703	11.640	13.758	17.463	7.039	6,199
Asset quality (%)	-1.879	-2.240	-1.610	-1.230	0.994	6,056
Management quality	-0.024	0.000	0.000	0.000	0.106	6,201
Earnings (%)	0.035	0.031	0.035	0.039	0.007	6,198
Liquidity (%)	10.119	3.575	7.069	13.224	9.553	6,198
Sensitivity to market risk (%)	12.923	4.801	10.330	18.399	10.584	6,080
Financial and Demographic Variables						
Deposit-to-asset ratio (%)	83.814	80.754	85.169	88.542	6.545	6,201
Leverage (%)	9.931	8.220	9.290	11.14	3.115	6,200
Total core deposits ($)	714,086	59,586	136,095	358,584	2,859,852	6,201
Size ($)	1,216,969	77,450	176,557	512,906	5,587,400	6,201
Age	79.897	42.000	91.000	108.000	38.961	6,190
Personal income growth (%) (County-Level)	2.313	0.207	1.809	3.581	4.324	6,153

APPENDIX: REFERENCES

ANDREWS, D. W. AND X. SHI (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666. [4]

ANDREWS, D. W. AND G. SOARES (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78, 119–157. [9, 10]

ANDREWS, I., J. ROTH, AND A. PAKES (2023): "Inference for Linear Conditional Moment Inequalities," *The Review of Economic Studies*, 90, 2763–2791. [4]

ARTSTEIN, Z. (1983): "Distributions of random sets and random selections," *Israel Journal of Mathematics*, 46, 313–324. [6]

BARSEGHYAN, L., M. COUGHLIN, F. MOLINARI, AND J. C. TEITELBAUM (2021): "Heterogeneous Choice Sets and Preferences," *Econometrica*, 89, 2015–2048. [2, 3]

BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2011): "Sharp identification regions in models with convex moment predictions," *Econometrica*, 79, 1785–1821. [4, 15]

BERRY, S. T. (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica: Journal of the Econometric Society*, 889–917. [44]

BLUNDELL, R. AND J. L. POWELL (2003): "Endogeneity in nonparametric and semiparametric regression models," *Econometric society monographs*, 36, 312–357. [20]

BLUNDELL, R. W. AND J. L. POWELL (2004): "Endogeneity in Semiparametric Binary Response Models," *The Review of Economic Studies*, 71, 655–679. [20]

BRESNAHAN, T. F. AND P. C. REISS (1990): "Entry in monopoly market," *The Review of Economic Studies*, 57, 531–553. [14, 44]

——— (1991): "Empirical models of discrete games," *Journal of Econometrics*, 48, 57–81. [14]

BUGNI, F., I. CANAY, AND X. SHI (2017): "Inference for Subvectors and Other Functions of Partially Identified Parameters in Moment Inequality Models," *Quantitative Economics*, 8, 1–38. [4, 30, 33]

CANAY, I. A. AND A. M. SHAIKH (2017): *Practical and Theoretical Advances in Inference for Partially Identified Models*, Cambridge University Press, 271–306, Econometric Society Monographs. [4, 9]

CHEN, S. AND H. KAIDO (2023): "Robust Tests of Model Incompleteness in the Presence of Nuisance Parameters," . [16]

CHEN, X., T. M. CHRISTENSEN, AND E. TAMER (2018): "Monte Carlo Confidence Sets for Identified Sets," *Econometrica*, 86, 1965–2018. [4]

CHEN, X., E. T. TAMER, AND A. TORGOVITSKY (2011): "Sensitivity Analysis in Semiparametric Likelihood Models," *Cowles Foundation Discussion Paper No. 1836*. [4]

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and confidence regions for parameter sets in econometric models," *Econometrica*, 75, 1243–1284. [9, 10]

CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): "INTERSECTION BOUNDS: ESTIMATION AND INFERENCE," *Econometrica*, 81, 667–737. [4, 10]

CHESHER, A. AND A. M. ROSEN (2017): "Generalized instrumental variable models," *Econometrica*, 85, 959–989. [2, 4]

CHOQUET, G. (1954): "Theory of capacities," in *Annales de l'institut Fourier*, vol. 5, 131–295. [23]

CILIBERTO, F. AND E. TAMER (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828. [2]

COX, G. AND X. SHI (2022): "Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models," *The Review of Economic Studies*, 90, 201–228. [4]

DEMPSTER, A. P. (1967): "Upper and lower probabilities induced by a multivalued mapping," *The annals of mathematical statistics*, 325–339. [23]

DENNEBERG, D. (1994): *Non-additive measure and integral*, vol. 27, Springer Science & Business Media. [35]

EIZENBERG, A. (2014): "Upstream Innovation and Product Variety in the U.S. Home PC Market," *The Review of Economic Studies*, 81, 1003–1045. [2]

EPSTEIN, L. G., H. KAIDO, AND K. SEO (2016): "Robust Confidence Regions for Incomplete Models," *Econometrica*, 84, 1799–1838. [4, 9, 13]

EPSTEIN, L. G. AND K. SEO (2015): "Exchangeable capacities, parameters and incomplete theories," *Journal of Economic Theory*, 157, 879–917. [23]

GALICHON, A. AND M. HENRY (2011): "Set identification in models with multiple equilibria," *The Review of Economic Studies*, rdr008. [4, 6]

GUL, F. AND W. PESENDORFER (2014): "Expected uncertain utility theory," *Econometrica*, 82, 1–39. [23]

HAHN, J. AND H. R. MOON (2010): "PANEL DATA MODELS WITH FINITE NUMBER OF MULTIPLE EQUILIBRIA," *Econometric Theory*, 26, 863–881. [4]

HAILE, P. A. AND E. TAMER (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111. [2]

HONORÉ, B. E. AND E. TAMER (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–629. [3]

HOROWITZ, J. L. AND S. LEE (2023): "Inference in a Class of Optimization Problems: Confidence Regions and Finite Sample Bounds on Errors in Coverage Probabilities," *Journal of Business & Economic Statistics*, 41, 927–938. [4, 9]

HUBER, P. J. (1981): *Robust statistics.*, Wiley, New York. [23]

HUBER, P. J. AND V. STRASSEN (1973): "Minimax tests and the Neyman-Pearson lemma for capacities," *The Annals of Statistics*, 251–263. [3, 22, 23, 24, 42]

——— (1974): "Note: Correction to minimax tests and the Neyman-Pearson lemma for capacities," *The Annals of Statistics*, 2, 223–224. [22]

JOVANOVIC, B. (1989): "Observable implications of models with multiple equilibria," *Econometrica: Journal of the Econometric Society*, 1431–1437. [3, 4]

KAIDO, H. AND F. MOLINARI (2024): "Information Based Inference In Models With Set-Valued Predictions And Misspecification," ArXiv:2401.11046. [4, 10, 29]

KAIDO, H., F. MOLINARI, AND J. STOYE (2019): "Confidence Intervals for Projections of Partially Identified Parameters," *Econometrica*, 87, 1397–1432. [4, 10]

——— (2022): "Constraint Qualifications in Partial Identification," *Econometric Theory*, 38, 596–619. [8]

KAIDO, H. AND Y. ZHANG (2019): "Robust Likelihood Ratio Tests for Incomplete Economic Models," ArXiv:1910.04610. [5, 43]

LAMBERT, T. (2019): "Lobbying on regulatory enforcement actions: Evidence from US commercial and savings banks," *Management Science*, 65, 2545–2572. [25, 26, 27, 28]

LEHMANN, E. L. AND J. P. ROMANO (2006): *Testing statistical hypotheses*, Springer Science & Business Media. [22]

LUO, Y. AND H. WANG (2017): "Core determining class and inequality selection," *American Economic Review, Papers & Proceedings*. [12]

MOLCHANOV, I. (2006): *Theory of random sets*, Springer Science & Business Media. [23]

MOLCHANOV, I. AND F. MOLINARI (2018): *Random Sets in Econometrics*, Econometric Society Monograph Series, Cambridge University Press, Cambridge UK. [6]

PELTZMAN, S. (1976): "Toward a more general theory of regulation," *The Journal of Law and Economics*, 19, 211–240. [25]

PHILIPPE, F., G. DEBS, AND J.-Y. JAFFRAY (1999): "Decision making with monotone lower probabilities of infinite order," *Mathematics of Operations Research*, 24, 767–784. [23, 35]

PONOMAREV, K. (2022): "Essays in Econometrics," Ph.D. thesis, UCLA. [12]

SCHMEIDLER, D. (1986): "Integral representation without additivity," *Proceedings of the American Mathematical Society*, 97, 255–261. [36]

SHAFER, G. (1982): "Belief functions and parametric models," *Journal of the Royal Statistical Society. Series B (Methodological)*, 322–352. [23]

SRINIVAS, V., D. BYLER, R. WADHWANI, A. RANJAN, AND V. KRISHNA (2015): "Enforcement actions in the banking industry: Trends and lessons learned," *Deloitte Center for Financial Services Research Report. 12.12. 2018 tarihinde https://www2. deloitte. com/content/dam/insights/us/articles/bank-enforcement-actionstrends-in-banking-industry/DUP1372_EnforcementActionsBanking_120815. pdf adresinden erişildi*. []

STIGLER, G. J. (2021): "The theory of economic regulation," in *The political economy: Readings in the politics and economics of American public policy*, Routledge, 67–81. [25]

TAMER, E. (2003): "Incomplete simultaneous discrete response model with multiple equilibria," *The Review of Economic Studies*, 70, 147–165. [2, 4]

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, John Wiley, New York, 305–310. [4]

WASSERMAN, L., A. RAMDAS, AND S. BALAKRISHNAN (2020): "Universal inference," *Proceedings of the National Academy of Sciences*, 117, 16880–16890. [2, 3, 5, 8, 9, 21, 44]