Identification of Structural and Counterfactual Parameters in a Large Class of Structural Econometric Models *

Lixiong Li †

Johns Hopkins University

May 27, 2021

Abstract

This paper provides new identification results for partially identified models that impose moment restrictions on the latent variables. The new sharp identification results apply to *all* models in this class, including the models to which existing results do not apply. Examples include discrete choice with subjective expecations and censored regression models. This novel result builds on the new insight that restrictions on the support of the latent variables should be explicitly stated and should be treated differently from other moment restrictions. This paper also shows how to construct the identified set for counterfactual results with nonparametric distributional assumptions on the latent variables. This method is applicable to counterfactual analysis in a large class of complete and incomplete structural models, without the necessity to actually simulate the counterfactual results. Finally, I show that when the model is discrete, the identification condition can be simplified so that the inference can be efficiently implemented following recent developments in the literature on large-scale inference for linear systems.

*The first version is of December 5, 2018. I am grateful to Marc Henry, Keisuke Hirano and Joris Pinkse for their invaluable guidance and support throughout this project. I also thank Michael Gechter, Paul Grieco, Patrik Guggenberger, Ismael Mourifié, Charles Murry, Mark Roberts, Karl Schurter and Neil Wallace for helpful discussions and suggestions. All remaining errors are mine.

†Department of Economics, Johns Hopkins University. Email: lixiong.li@jhu.edu

1 Introduction

This paper studies the identification for the following model:

$$\mathbb{P}[(U,Z) \in \Gamma(\theta)] = 1 \text{ and } \mathbb{E}[r(U,Z;\theta)] = 0, \tag{1}$$

where θ is the parameter, U stands for the vector of unobservable variables, and Z is the vector of the observables. The parameter space Θ can be an arbitrary space, and the space of U and Z, denoted as $\mathcal{U}$ and $\mathcal{Z}$ are Polish spaces. $\Gamma(\theta)$ is a θ-dependent set of (U,Z), which restricts the support of (U,Z). Given any θ, $r(\cdot,\cdot;\theta)$ is a known function mapping $\mathcal{U}\times\mathcal{Z}$ to $\mathbb{R}^{\dim(r)}$ where $\dim(r) \geq 1$ is finite. In the following, I call the first restriction in (1) as the support restriction and the second restrictions as the moment restriction.

The goals of this paper are two-fold. Firstly, I derive the sharp identification results for models in (1) without imposing any other restrictions on (Γ, r) except for they being measurable and well-defined. Secondly, I explore how the framework in (1) can be used to conduct counterfactual analysis for structural models without specifying parametric distributional assumptions on the latent variables.

The first contribution of this paper is a new set of sharp identification results applicable to *all* models which fit in (1). This brings new identification results to models to which the existing results are not applicable. Examples include discrete choice with subjective expectations and censored regression models. This is done by revisiting the support-function-based approach as in Ekeland, Galichon and Henry (2010) and Beresteanu, Molchanov and Molinari (2011). After I the derive testable restrictions that are satisfied by all parameters in the identified set, I show that, when formulated properly, any model in (1) can be sharply characterized by these testable restrictions, in the sense that the identified set and the set of parameters satisfying the testable restrictions are indistinguishable in finite samples.

This novel identification result brings two new observations to the partial identification literature. The first observation is that the definition of the sharp identification typically adopted in the literature could be unnecessarily restrictive for models in (1). After I derive the testable restrictions which the true parameter must satisfy, I study the difference between the identified set and the set of parameters satisfying these testable restrictions. The sharp identification analysis in the literature so far focuses exclusively on the cases when there is no difference between these two sets. However, I find that, when these two sets differ, the difference is often so small that there does not exist a test which can distinguish these two sets in finite samples. This observation enables me to derive sharp identification results under much weaker regularity conditions than those in the related literature.

The second observation is that, all restrictions on the support of the latent variables should be explicitly stated and the support restrictions should be treated differently from the moment restrictions. If any moment restriction implictly restricts the support of the latent variable, one should make this restriction explicit and include it in $\Gamma(\theta)$. I show that following this principle alone is enough to ensure the sharpness of the identification. Reversely, failure to do so could lead to non-sharp identification results for models which can otherwise be sharply identified. This observation is not obvious, because the support restrictions can always be rewritten as an moment restriction. However, it turns out that this principle plays a key role in the identification analysis. In addition, I show this principle is also related to a condition that can be easily verified by numerical simulations in practice.

The second contribution of this paper is that it shows how to utilize the framework in (1) to conduct the counterfactual analysis under nonparametric distributional assumptions on the latent variables, even if the model is not point identified or has multiple counterfactual model predictions. The idea is to view the counterfactual outcomes of each individual as latent variables and treat the model predictions on the counterfactual outcomes as additional support restrictions. This way of thinking unites the model estimation and counterfactual analysis, and the identified set for the counterfactual result and other model parameters can be constructed jointly in one stage. This simple yet powerful idea is not specific to the identification approach considered in this paper. In fact, it can be easily extended to work with other identification methods such as Beresteanu, Molchanov and Molinari (2011), Schennach (2014) and Chesher and Rosen (2017) among others.

Finally, whenever the model is discrete, I show that the identification condition generated by the support-function-based approach can be greatly simplified and its corresponding inference problem can be transformed into an inference problem for linear systems with known coefficients. Following the recent developments in the related literature, the inference procedure can be implemented as a series of linear programming problems which greatly alleviates the computational complexity. See Fang, Santos, Shaikh and Torgovitsky (2020) among others in this literature.

This paper is mostly related to Ekeland, Galichon and Henry (2010) and Beresteanu, Molchanov and Molinari (2011), both of which study the identified set of models similar to those in (1). The sharp identification result in both of these two papers builds on some compactness assumptions of the model. The result in Ekeland, Galichon and Henry (2010) imposes the tightness and uniform integrability restriction on r, and the result in Beresteanu, Molchanov and Molinari (2011) assumes that $\Upsilon(z;\theta) := \{r(u,z;\theta) : (u,z) \in \Gamma(\theta)\}$ is an absolutely integrable random closed set for all θ. All these results exclude models whose $\Upsilon(z;\theta)$ is not compact. In empirical analysis, this non-compactness is not uncommon. For example, it arises naturally in models where agents make optimal discrete choices based on subjective expectations, or in regression models where the outcome is one-side censored. See examples in Section 2. Similar models have also been explored with an optimal transportation and capacity function approach in Galichon and Henry (2011). More recently, Chesher and Rosen (2017) discuss similar problems in the framework of generalized IV models and establish the equivalence between the use of random sets defined on the space of the observables and that on the space of the unobservables. They point out the advantage of the latter approach in deriving the sharp identification conditions. Both this paper and all the papers mentioned above derive testable restrictions which parameters in the identified set should satisfy, but this paper differs in how to establish the sharpness for these testable restrictions. All these papers in the literature focus on cases where the identified set is equal to the set of parameters satisfying the testable restrictions, but I study when these two sets are indistinguishable in finite samples. This unique approach enables me to derive much more general identification results than the existing ones.

While this paper and all the above literature try to characterize the identified set, Schennach (2014) characterizes the *moment closure* of the identified set using entropy-based approach under a dominating assumption and some other mild conditions. I will explore the relation between the identified set and its moment closure later in Section 4, but these two sets are different in general. By studying when the difference between the identified set and its moment closure is negligible, this paper complements the theoretical findings in Schennach (2014).

There is a growing literature on bounding counterfactual outcomes without imposing parametric distributions on the unobserved heterogeneity, especially in the context of discrete choice models. Manski (2007) shows that sharp bounds on counterfactual choice probabilities can be constructed by solving a linear programming problem given non-parametric constraints on agents' preferences revealed in the data. This idea is further developed in Tebaldi, Torgovitsky and Yang (2018) which also takes into account the endogeneity in prices. Chiong, Hsieh and Shum (2017) provide another non-sharp but computationally efficient way to bound counterfactual market shares based on the cyclic monotonicity implied by the optimality conditions in discrete choice models. More recently, Aguiar and Kashaev (2018) discusses a way to do counterfactual analysis in the context of the revealed preference axioms, which is similar to the general approach presented in this paper. Christensen and Connault (2019), which is concurrent with this paper, develops an alternative entropy-based approach which is able to conduct the sensitivity analysis of a specific parametric distribution and characterize the counterfactual outcome under nonparametric distribution assumptions.

This paper is also related to other papers which study the identification in structural models with nonparametric distributional assumptions on latent variables. Pakes (2010) and Pakes, Porter, Ho and Ishii (2015) have studied some important empirical models. They exploit the revealed preference conditions to construct moment inequalities, where some particular structures in payoff functions and information sets are imposed to cancel or integrate out the unobserved heterogeneity. The framework in this paper nests all the models they studied. In Example 1, I revisit one of their models and study the sharp identification results under various model restrictions.

As my identification conditions take the form of moment inequalities, this paper also relates to the literature on moment inequality inference. When the model is discrete, I show that the inference problem of my method can be transformed into the linear system inference problem studied in Fang, Santos, Shaikh and Torgovitsky (2020), which can be efficiently implemented as a sequence of linear programming problems. In small-scale discrete problems, the identification conditions can also be simplified into a finite number of moment inequalities. One can then apply inference methods in Chernozhukov, Hong and Tamer (2007), Andrews and Soares (2010) and recent Chernozhukov, Chetverikov and Kato (2018) to perform hypothesis testing and construct confidence region. In challenging cases where the identification conditions involve a continuum of moment inequalities, one could use inference procedures in Andrews and Shi (2017) and Chernozhukov, Lee and Rosen (2013). If the moment inequalities are conditional on other instruments, the inference can be conducted by following Andrews and Shi (2013) and Chernozhukov, Lee and Rosen (2013).

The rest of the paper is organized as follows. Section 2 describes the assumptions which I impose throughout this paper and introduces a running examples. Section 3 presents the support-function-based approach and shows that the set of parameters this approach characterizes is equal to the moment closure of the identified set. In Section 4, I study the relation between the identified set and its moment closure and explain why one should explicit write down all the support restrictions. Section 5 discusses how one can identify the counterfactual result in the same way as the structural parameters. Finally, I discuss the implementation of the inference for the discrete models in Section 6. Section 7 concludes the paper. The proofs of all the theorems are relegated to the appendix.

2 Assumptions and Motivating Examples

In this section, I am going to introduce the main assumptions and the running example of this paper. The following notations will be used throughout the paper: I use (Γ, r) to summerize the model restrictions in (1). The $\mathbb{R}$ will denote the real space. The $\|\cdot\|$ will stand for the Euclidean norm. I use the uppercase letters like U and Z for random variables, and their lowercase like u and z for the specific values of random variables. For any distribution F, the $\mathbb{P}_F$ and $\mathbb{E}_F$ refers to the probability and the expectation evaluated with respect to F.

Let $\mathcal{F}$ be a space of the probability distributions for Z. This $\mathcal{F}$ space can be the collection of all possible Borel distributions on $\mathcal{Z}$ or the collection of distributions that the researcher is willing to assume. I assume that the Θ and $\mathcal{F}$ satisfy the following regularity conditions.

Assumption 1. *Assume* $\dim(r) > 0$*. In addition, for any* $\theta \in \Theta$ *and any* $F \in \mathcal{F}$*, assume the following two conditions hold:*

(i) Set $\Gamma(\theta)$ *is a Borel set, and* $\Gamma(z;\theta)$ *defined as* $\Gamma(z;\theta) \coloneqq \{u \in \mathcal{U} : (u, z) \in \Gamma(\theta)\}$ *is a nonempty Borel set for every* z*. Moreover, the function* $r(u, z; \theta)$ *is Borel measurable in*

$\mathcal{U} \times \mathcal{Z}$.

(ii) There exists an Borel measurable function $g(\,\cdot\,;\theta, F)$ *such that* $\mathbb{E}_F g(Z;\theta,F) < \infty$ *and* $g(z;\theta,F) \geq \inf\{\|r(u,z;\theta)\| : u \in \Gamma(z;\theta)\}$ *for almost every* z.

The first condition in Assumption 1 is a basic measurability condition. The second condition ensures that there exists at least one distribution H for (U, Z) such that $\mathbb{P}_H\,(U,Z) \in \Gamma(\theta))$ and function r is integrable with respect to H. These conditions are very weak conditions to ensure the model in (1) is well-defined.

The first goal of this paper is to characterize the identified set of θ for the model in (1) given any distribution F in $\mathcal{F}$. Formally, for each θ in Θ and each $F \in \mathcal{F}$, define $\mathcal{H}(\theta, F)$ to be the set of all joint distributions H for (U, Z) which satisfy that $\mathbb{P}_H[(U,Z) \in \Gamma(\theta)] = 1$ and that H's marginal distribution for Z equals F. Then, the identified set for any given F can be defined as follows:

Definition 1 (identified set)**.** For any $F \in \mathcal{F}$, the *identified set*, denoted as $\Theta_I(F)$, is the set of all $\theta \in \Theta$ for which there exists some $H \in \mathcal{H}(\theta, F)$ such that $\mathbb{E}_H[r(U,Z;\theta)] = 0$, i.e.

$$\min_{H \in \mathcal{H}(\theta,F)} \|\mathbb{E}_H[r(U,Z;\theta)]\| = 0. \tag{2}$$

By definition, for each parameter θ in the identified set, there exists a distribution H which satisfies all assumptions in (1) and makes it observationally equivalent to the true parameter θ_0. This definition is in line with the definition used in most of the literature. See Roehrig (1988) and Ekeland, Galichon and Henry (2010) among many others. However, in addition to the identified set, it turns out that the moment closure defined in the following also plays an important role in the analysis.

Definition 2 (the moment closure of the identified set)**.** For any $F \in \mathcal{F}$, the *moment closure* of the identified set, denoted as $\Theta_I'(F)$, is the set of all $\theta \in \Theta$ for which the following equation holds:

$$\inf_{H \in \mathcal{H}(\theta,F)} \|\mathbb{E}_H[r(U,Z;\theta)]\| = 0. \tag{3}$$

This moment closure was the main analysis target in Schennach (2014). The main difference between the identified set and its moment closure is whether there exists a $H \in \mathcal{H}(\theta, F)$ which satisfies the moment restriction exactly. For an arbitrary $\theta \in \Theta_I'(F)$, the infimum in (3) may or may not be achieved by an $H \in \mathcal{H}(\theta, F)$. By definition, $\Theta_I(F) \subseteq \Theta_I'(F)$, but the reverse inclusion relation may not be true. Although the difference between $\Theta_I(F)$ and $\Theta_I'(F)$ seems to be small in definition, they can be very different in some cases. I will discuss the relation between $\Theta_I(F)$ and $\Theta_I'(F)$ in more details in Section 4.

In the rest of the paper, I sometimes abbreviate $\Theta_I(F)$ and $\Theta_I'(F)$ as Θ_I and Θ_I' respectively when there is no confusion on the underlying distribution F. I sometimes also write $\Theta_I(F)$ and $\Theta'_I(F)$ respectively as $\Theta_I(F;\Gamma,r)$ and $\Theta'_I(F;\Gamma,r)$ when I need to make their dependence on the model (Γ,r) explicit.

The framework in (1) covers a large class of structural models. I list two examples in the following.

Example 1 (binary choice model with subjective expectation)**.** Consider a binary choice model in which agents have limited information when making decisions. Let $Y_i \in \{0,1\}$ be agent i's choice. When $Y_i = 1$, the payoff π_i of player i is

$$\pi_i = X_i'\beta - \alpha$$

where X_i are some observable covariates and (α,β) are parameters to be estimated. When $Y_i = 0$, π_i is normalized to 0. Assume that, the agent i does not know the exact value of π_i when making decisions. Instead, the decision is based on agent i's subjective expectation $\mathbb{E}_s[\pi_i]$. Assume agent i chooses optimally, i.e.

$$Y_i = \begin{cases} 1 & \text{if } \mathbb{E}_s[\pi_i] > 0, \\ 0 & \text{if } \mathbb{E}_s[\pi_i] < 0. \end{cases} \tag{4}$$

Assume also that the agent's expectation is rational. That is, if we define the expectation error as $U_i \coloneqq \mathbb{E}_s[\pi_i] - \pi_i$, then $\mathbb{E}[U_i|Y_i] = 0$ almost surely.

This simple example has been studied in the literature. See, for example, Pakes (2010) and Pakes et al. (2015). Later in this paper, I will derive the sharp identification result for this model using the method in this paper. In Appendix A.2, I consider a more complicated version of this model where there also exists some payoff shocks that is not observabled in the data but is known to the agent when he/she chooses Y_i.

To see how this example fits into my framework, let $Z_i \coloneqq (Y_i, X_i)$ be the collection of observed variables and let $\theta \coloneqq (\alpha,\beta)$ be the vector of all parameters. Then, the optimality condition in (4) implies the following support restriction on (U_i, Z_i),

$$\mathbb{P}[(U_i,Z_i) \in \Gamma(\theta)] = 1, \text{ where } \Gamma(\theta) = \{(u_i,z_i) : (-1)^{y_i}[x_i'\beta - \alpha + U_i] \le 0\}. \tag{5}$$

Moreover, the rational expectation implies the following moment restriction:

$$\mathbb{E}r(U_i,Z_i;\theta) = 0, \text{ where } r(U_i,Z_i;\theta) = \begin{pmatrix} \mathbb{1}(Y_i = 1)U_i \\ \mathbb{1}(Y_i = 0)U_i \end{pmatrix}.$$

■

Example 2 (censored regression)**.** Consider the following regression model:

$$Y^* = X'\theta + \epsilon$$

where X is a vector of observable covariates, ϵ is the residual. Assume that we can only observe $Y = \max(Y^*, C)$ where C is an observable random variable. Assume that $\mathbb{E}[W\epsilon] = 0$ where W is a vector of instrumental variables which could overlap with X and/or C.

This model is related to but different from the interval censored regression model that have been widely studied in the literature, because the Y^* is only censored on one side. Analysis on the interval censored regression model can be found in Tamer (2010), Ponomareva and Tamer (2011) and Bontemps, Magnac and Maurin (2012). See also the discussions in Molinari (2020) and Chesher and Rosen (2020).

To see how this example fits into the framework, let $U \coloneqq (Y^*, \epsilon)$ collects all the unobservables, and let $Z \coloneqq (X, Y, C, W)$ collects all the observables. Then, the support restriction of this model is $\mathbb{P}[(U, Z) \in \Gamma(\theta)] = 1$ where

$$\Gamma(\theta) \coloneqq \left\{(u, z) : y^* = x'\theta + \epsilon \text{ and } y = \max(y^*, c)\right\}.$$

And, the moment restriction is $\mathbb{E}[r(U, Z; \theta)] = 0$ where $r(U, Z; \theta) = W\epsilon$. ■

3 Support-Function-Based Approach

In this section, I consider an identification approach for the model in (1) which turns out be similar to the identification strategy in Ekeland, Galichon and Henry (2010) and Beresteanu, Molchanov and Molinari (2011). However, the result which I am going to derive in this and the next sections builds on much weaker assumptions and covers a much wider range of models. In particular, it does not depend on the closedness and the absolute integrability assumptions as in these two papers.

For any F in $\mathcal{F}$, let θ be an arbitrary element in $\Theta_I(F)$. Then, there exists some H in $\mathcal{H}(\theta, F)$ such that $\mathbb{E}_H r(U, Z; \theta) = 0$. Define $\mathcal{S}$ to be the unit sphere in $\mathbb{R}^{\dim(r)}$, i.e. $\mathcal{S} \coloneqq \{\lambda \in \mathbb{R}^{\dim(r)} : \|\lambda\| = 1\}$. Then, $\mathbb{E}_H r(U, Z; \theta) = 0$ is equivalent to the following condition:

$$\forall \lambda \in \mathcal{S},\ \mathbb{E}_H[\lambda' r(U, Z; \theta)] = 0, \tag{6}$$

where λ' stands for the transpose of vector λ. Recall $\Gamma(z; \theta) = \{u : (u, z) \in \Gamma(\theta)\}$. Define function γ as $\gamma(\lambda, z; \theta) \coloneqq \sup_{u \in \Gamma(z;\theta)} \lambda' r(u, z; \theta)$, which is the support function of $\{r(u, z; \theta) : (u, z) \in \Gamma(\theta)\}$ given z. Then, $\mathbb{P}_H[(U, Z) \in \Gamma(\theta)] = 1$ implies that

$$\forall \lambda \in \mathcal{S}, \quad \mathbb{P}_H\left(\lambda' r(U, Z; \theta) \leq \gamma(\lambda, Z; \theta)\right) = 1. \tag{7}$$

Equation (6) and (7) then imply that:

$$\forall \lambda \in \mathcal{S},\ \mathbb{E}_H \gamma(\lambda, Z; \theta) \geq \mathbb{E}_H[\lambda' r(U, Z; \theta)] = 0. \tag{8}$$

Because $\gamma(\lambda, \cdot; \theta)$ is a function which only depends on z and the marginal distribution of H

on Z is equivalent to F, the condition in (8) is then equal to $\mathbb{E}_F\gamma(\lambda, Z; \theta) \geq 0$ for all $\lambda \in \mathcal{S}$, or, equivalently,

$$\inf_{\lambda \in S} \mathbb{E}_F\gamma(\lambda, Z; \theta) \geq 0. \tag{9}$$

The condition in (9) can be viewed as a group of moment equalities and inequalities which only depends on the observables and the parameters. The above derivation shows that (9) is a testable restriction which parameters in the identified set must satisfy. That is, $\Theta_I(F) \subseteq \{\theta \in \Theta : \theta \text{ satisfy Condition (9)}\}$. In the following theorem, I go one step beyond this result and characterize exactly what θ satisfies Condition (9).

Theorem 1. *Suppose Assumption 1 holds. Then, for any $F \in \mathcal{F}$,*

$$\Theta'_I(F) = \{\theta \in \Theta : \theta \textit{ satisfy Condition } (9)\}.$$

Theorem 1 shows that Condition (9) is an exact characterization of the moment closure of the identified set under very weak conditions. This result suggests that to study the property of the set of parameters which satisfies (9), one only needs to study the property of the moment closure Θ'_I. Moreover, note that the properties of Θ'_I are, in fact, intrinsic properties of model (Γ, r) instead of the properties of any specific identification strategy.

4 Relation Between Θ_I and Θ'_I

Given Theorem 1, the usefulness of Condition (9) depends on how close the identified set Θ_I and its moment closure Θ'_I are. As shown later, Θ'_I is generally different from Θ_I. In addition, the difference between Θ_I and Θ'_I can be quite small in some cases, and quite large in some other cases. Before I derive more theoretical results, I would like to present the following two heuristic examples which are intentionally constructed as simple as possible for illustration. However, the findings which I draw from these examples would apply to more complicated structural models.

Example 3. In this example, I will construct a simple model where $\Theta_I \neq \Theta'_I$ but the difference between these two is negligible. Consider a model that consists of random variable (U, X, Y), where U is unobserved and (X, Y) are observable. The space of Y is the binary set $\{0, 1\}$. Suppose the support of (U, X, Y) satisfies the following restriction almost surely:

$$U \in \begin{cases} [X, +\infty) & \text{if } Y = 1, \\ (-\infty, X] & \text{if } Y = 0. \end{cases}$$

Moreover, let the moment restriction be $\mathbb{E}[\Phi(U) - \theta] = 0$ where Φ is the cdf of the standard normal distribution. Let $\mathcal{F}$ be the set of all possible distributions of (X, Y) and let $\Theta = \mathbb{R}$.

In this example, one can show that, for any $F \in \mathcal{F}$,

$$\Theta_I(F) = \Big(\mathbb{E}_F[\mathbb{1}(Y=1)\Phi(X)],\ \mathbb{E}_F[\mathbb{1}(Y=0)\Phi(X) + \mathbb{1}(Y=1)]\Big).$$

Its lower and the upper bounds cannot be achieved exactly, but any θ close to the lower bound can be achieved by constructing a data generating process where $U = \mathbb{1}(Y=1)X + \mathbb{1}(Y=0)(X-q)$ and let q be a large enough number. Any θ close to the upper bound can be achieved in a similar way.

The $\Theta'_I(F)$ in this example is

$$\Theta'_I(F) = \Big[\mathbb{E}_F[\mathbb{1}(Y=1)\Phi(X)],\ \mathbb{E}_F[\mathbb{1}(Y=0)\Phi(X) + \mathbb{1}(Y=1)]\Big],$$

which is almost the same as $\Theta_I(F)$ except that it is a closed interval. This $\Theta'_I(F)$ can be solved using Theorem 1. Although $\Theta_I(F) \neq \Theta'_I(F)$, their difference is so small that it is negligible in almost all empirical contexts. ■

Example 4. In this example, I will construct a simple model where the difference between Θ_I and Θ'_I is considerable. Let U be an unobserved random variable and (Z_1, Z_2) be two observable random variable. Consider a model (Γ, r) with

$$\Gamma(\theta) = \mathbb{R}^3 \quad \text{and} \quad r(u, z_1, z_2; \theta) = \begin{pmatrix} \mathbb{1}(z_1 \leq u \leq z_2) - 1 \\ u - \theta \end{pmatrix}.$$

Let $\Theta = \mathbb{R}$. And, let $\mathcal{F}$ be the distribution of (Z_1, Z_2) where $\mathbb{E}_F|Z_1| < \infty$, $\mathbb{E}_F|Z_2| < \infty$ and $\mathbb{P}_F(Z_1 \leq Z_2) = 1$.

In this example, the moment restriction $\mathbb{E}[r(U, Z_1, Z_2; \theta)] = 0$ implies that $\mathbb{P}(U \in [Z_1, Z_2]) = 1$. Therefore, for any $F \in \mathcal{F}$, $\Theta_I(F) = [\mathbb{E}_F Z_1, \mathbb{E}_F Z_2]$. At the same time, one can solve the support function $\gamma(\lambda, Z_1, Z_2; \theta)$ in this example as

$$\gamma(\lambda, z_1, z_2; \theta) = \begin{cases} \max(0, -\lambda_1) & \text{if } \lambda_2 = 0 \\ +\infty & \text{if } \lambda_2 \neq 0 \end{cases} \tag{10}$$

where $\lambda = (\lambda_1, \lambda_2)$ and λ_i is the multiplier corresponds to the ith dimension of $r(u, z_1, z_2; \theta)$. Equation (10) implies that $\mathbb{E}\gamma(\lambda, Z_1, Z_2; \theta) \geq 0$ for any $\lambda \in \mathcal{S}$ and any $\theta \in \mathbb{R}$. That is, Condition (9) is satisfied for any $\theta \in \mathbb{R}$. Therefore, Theorem 1 implies that $\Theta'_I(F) = \mathbb{R}$, which is much larger than $\Theta_I(F)$. ■

Given the difference between Θ_I and Θ'_I illustrated above, there are two ways to proceed: One way is to find sufficient conditions under which $\Theta_I = \Theta'_I$. This way is in line with most sharp identification results in the literature including those in Ekeland, Galichon and Henry (2010), Beresteanu, Molchanov and Molinari (2011) and Chesher and Rosen (2017). Another way is to study when the difference between Θ_I and Θ'_I is so small that it is negligible for the purpose of empirical analysis. In the following, I will present results following both approaches, but I consider the result in the second approach more empirically relevant. Moreover, although the analysis in the second approach is more challenging, it builds on much weaker assumptions and leads to more interpretable results.

4.1 when is Θ_I equal to Θ_I' ?

The following assumption is a sufficient condition for $\Theta_I = \Theta_I'$.

Assumption 2. *For any $\theta \in \Theta$ and any $F \in \mathcal{F}$, the following two conditions hold:*

(i) For almost every z, $\{r(u, z; \theta) : u \in \Gamma(z; \theta)\}$ is a closed set.

(ii) There exists an Borel measurable function $g(\cdot; \theta)$ such that $\mathbb{E}_F g(Z; \theta) < \infty$ and for almost every z,

$$g(z; \theta) \geq \sup\{\|r(u, z; \theta)\| : u \in \Gamma(z; \theta)\}.$$

Theorem 2. *Suppose Assumptions 1 and 2 hold. Then, $\Theta_I(F) = \Theta_I'(F)$ for all $F \in \mathcal{F}$.*

Assumption 2 is the same as the closedness and the absolute integrability condition imposed in Beresteanu, Molchanov and Molinari (2011), except that it does not involve the non-atomic restrictions therein. Because Assumption 2 implies that $\{r(u, z; \theta) : u \in \Gamma(z; \theta)\}$ should be a compact set almost surely, it essentially rules out models with noncompat $\{r(u, z; \theta) : u \in \Gamma(z; \theta)\}$. The limitation of Assumption 2 can also be illustrated by the examples in Section 2.

Example 1 (continued)**.** In Example 1, one can show that

$$\{r(u, z; \theta) : u \in \Gamma(z; \theta)\} = \begin{cases} \{(r_1, 0) : r_1 \geq -x_i'\beta + \alpha\} & \text{if } Y_i = 1 \\ \{(0, r_2) : r_2 \leq -x_i'\beta + \alpha\} & \text{if } Y_i = 0 \end{cases}$$

Because $\{r(u, z; \theta) : u \in \Gamma(z; \theta)\}$ is unbounded with probability 1, Assumption 2 does not hold in this example. However, the failure of Assumption 2 does not mean that the model in Example 1 does not have any empirical content. As shown in Appendix A.1, Condition (9) in this example can be simplified as

$$\begin{aligned} \mathbb{E}[\mathbb{1}(Y_i = 1)(X_i'\beta - \alpha)] &\geq 0, \\ \mathbb{E}[\mathbb{1}(Y_i = 0)(X_i'\beta - \alpha)] &\leq 0. \end{aligned} \tag{11}$$

Due to Theorem 1, we know Θ_I' is the set of (β, α) that satisfies the above two moment inequalities. The question is whether Θ_I is close or equal to Θ_I'. Because Assumption 2 fails to hold here, Theorem 2 cannot tell us whether these moment inequalities are also the sharp characterization of Θ_I. ■

Example 2 (continued)**.** In Example 2, one can show that

$$\{r(u,z;\theta) : u \in \Gamma(z;\theta)\} = \begin{cases} \{W(Y - X'\theta)\} & \text{if } Y > C \\ \{W\epsilon : \epsilon \leq C - X'\theta\} & \text{if } Y = C \end{cases}$$

To rule out the trivial case, assume $\mathbb{P}(Y = C) > 0$ and $\mathbb{P}(W = 0|Y = C) < 1$. Then, $\{r(u,z;\theta) : u \in \Gamma(z;\theta)\}$ is going to be unbounded with positive probability, which again leads to the failure of Assumption 2.

As before, let us take a look at Condition (9) in this example. Define $\mathcal{S}^+$ as $\mathcal{S}^+ := \{\lambda : \|\lambda\| = 1 \text{ and } \mathbb{P}(\lambda' W \geq 0|Y = C) = 1\}$. Then, Condition (9) in this example can be simplified as

$$\lambda' \mathbb{E} W(Y - X'\theta) \geq 0, \quad \forall \lambda \in \mathcal{S}^+. \tag{12}$$

Whether this condition is informative or not, depends on $\text{supp}(W|Y = C)$, i.e. the support of W conditional on the event that Y is censored. In the special case where $\text{supp}(W|Y = C) = \mathbb{R}^{\dim(W)}$, $\mathcal{S}^+$ is an empty set so that (12) does not impose any restrictions. Although we know from Theorem 1 that (12) characterizes the parameters in Θ_I', we still want to know whether (12) also characterizes the parameters in Θ_I. In particular, in the case where $\text{supp}(W|Y = C) = \mathbb{R}^{\dim(W)}$, we want to know whether the lack of informativeness is due to the fact that Condition (9) is not a sharp characterization for Θ_I in this example, or it is just that the assumptions that we imposed have little empirical content in the first place. ■

The above example illustrates the limitation of Assumption 2. On the other hand, if any of the two conditions in Assumption 2 is dropped, there are counterexamples where $\Theta_I \neq \Theta_I'$. Example 3 violates the first condition in Assumption 2, and Example 4 violates the second condition in Assumption 2. These counterexamples suggests that restrictions similar to those in Assumption 2 is inevitable if the goal is to ensure $\Theta_I' = \Theta_I$. Therefore, to be able to say something about all models, especially the models where $\{r(u,z;\theta) : u \in \Gamma(z;\theta)\}$ is noncompact, I will switch to the second approach in the following section.

4.2 when are Θ_I and Θ_I' indistinguishable?

In this section, I am going to study when Θ_I' is sufficiently close to Θ_I so that these two sets are not distinguishable in finite samples. As I will show later in Theorem 4, unlike the previous result which assumes compactness, Θ_I' and Θ_I are indistinguishable under very weak conditions. In fact, all that is needed is to ensure that the restrictions on the support of the latent variables have been explicitly included in Γ and none of the moment restrictions in (1) implicitly restricts the support of the latent variable. I will also present some interesting lemmas and easy verifiable conditions along the way.

To define the finite-sample indistinguishability formally, consider an i.i.d. sample $Z_1, ..., Z_n$ where n is the sample size. For any $\theta \in \Theta$, define $\mathcal{F}_\theta := \{F \in \mathcal{F} : \theta \in \Theta_I(F)\}$ and

$\mathcal{F}'_\theta := \{F \in \mathcal{F} : \theta \in \Theta'_I(F)\}$. Then, testing $\theta \in \Theta_I(F)$ is equivalent to testing $F \in \mathcal{F}_\theta$. Similarly, $\theta \in \Theta'_I(F)$ is equivalent to $F \in \mathcal{F}'_\theta$. Let ϕ_n be the critical function of a (randomized) test with $\phi_n(Z_1, ..., Z_n) \in [0, 1]$. For any $\theta \in \Theta$, the size of the test under the null hypothesis $H_0 : \theta \in \Theta_I(F)$ is $\sup_{F \in \mathcal{F}_\theta} \mathbb{E}_F \phi_n$. Because $\Theta_I(F) \subseteq \Theta'_I(F)$ for any F, we know $\mathcal{F}_\theta \subseteq \mathcal{F}'_\theta$ so that the following inequality is always true:

$$\sup_{F \in \mathcal{F}_\theta} \mathbb{E}_F \phi_n \leq \sup_{F \in \mathcal{F}'_\theta} \mathbb{E}_F \phi_n.$$

If the above weak inequality is in fact an equality, then the test ϕ_n would have no power against the alternative hypothesis $H_1 : \theta \in \Theta'_I(F) \backslash \Theta_I(F)$. In other words, the test ϕ_n cannot distinguish $\theta \in \Theta_I$ from $\theta \in \Theta'_I$ in finite samples.

Definition 3. For any $\theta \in \Theta$, I say that *it is impossible to distinguish $\theta \in \Theta_I$ from $\theta \in \Theta'_I$ in finite samples* if for any test ϕ_n, $\sup_{F \in \mathcal{F}_\theta} \mathbb{E}_F \phi_n = \sup_{F \in \mathcal{F}'_\theta} \mathbb{E}_F \phi_n$. If this is true for all θ, then I say that *it is impossible to distinguish Θ_I and Θ'_I in finite samples.*

For $\mathcal{F}^\dagger \in \{\mathcal{F}_\theta, \mathcal{F}'_\theta\}$, I use the following convention: if $\mathcal{F}^\dagger$ is empty, then $\sup_{F \in \mathcal{F}^\dagger} \mathbb{E}_F \phi_n = -\infty$. Therefore, in the trivial case where $\mathcal{F}'_\theta$ is empty and, hence, $\mathcal{F}_\theta$ is empty, it is impossible to distinguish $\theta \in \Theta_I$ from $\theta \in \Theta'_I$ in finite samples.

In the following, I am going to derive conditions under which $\theta \in \Theta_I$ and $\theta \in \Theta'_I$ are not distinguishable in finite samples. Note that, although the function r in model (Γ, r) can depend on U, it does not rule out special cases where some of its dimensions do not depend on U. Let θ be an arbitrary parameter in Θ. We can partition function $r(\cdot, \cdot; \theta)$ as

$$r(u, z; \theta) = \begin{pmatrix} r_1(z; \theta) \\ r_2(u, z; \theta) \end{pmatrix} \tag{13}$$

where function r_1 does not depend on the latent variables. Let $\dim(r_1)$ and $\dim(r_2)$ be the dimension of r_1 and r_2 respectively. If every dimension of $r(\cdot, \cdot; \theta)$ depends on U, then $\dim(r_1) = 0$. If none of the dimensions of $r(\cdot, \cdot; \theta)$ depend on U, then $\dim(r_2) = 0$. In most cases considered in this paper, $\dim(r_1) = 0$, but there does exist some interesting cases where $\dim(r_1) > 0$, one of which will be discussed later. Let $\mathcal{S}_2$ be the unit sphere in $\mathbb{R}^{\dim(r_2)}$, i.e. $\mathcal{S}_2 := \left\{\lambda \in \mathbb{R}^{\dim(r_2)} : \|\lambda\| = 1\right\}$. Then, the testable restriction (9) is equivalent to the following condition:

$$\mathbb{E}_F[r_1(Z; \theta)] = 0 \text{ and } \inf_{\lambda \in \mathcal{S}_2} \mathbb{E}_F \gamma_2(\lambda, Z; \theta) \geq 0, \tag{14}$$

where $\gamma_2(\lambda, z; \theta) := \sup_{u \in \Gamma(z;\theta)} \lambda' r_2(u, z; \theta)$. Condition (14) is similar to (9) except that the condition on r_1 is simplified into moment equalities. When $\dim(r_1) = 0$, (9) and (14) are identical. The following lemma plays a key role in my analysis and it is interesting in its own right.

Lemma 1. *Suppose Assumption 1 holds. Let F be an arbitrary element in $\mathcal{F}$ and let θ be an arbitrary parameter in Θ. Partition $r(u, z; \theta) = (r_1(z; \theta), r_2(u, z; \theta))$.*

(i) if $\mathbb{E}_F[r_1(Z; \theta)] = 0$ and $\inf_{\lambda \in \mathcal{S}_2} \mathbb{E}_F \gamma_2(\lambda, Z; \theta) > 0$, then $\theta \in \Theta_I(F)$. (Note that the $\inf_{\lambda \in \mathcal{S}_2} \mathbb{E}_F \gamma_2(\lambda, Z; \theta) > 0$ includes the case that $\inf_{\lambda \in \mathcal{S}_2} \mathbb{E}_F \gamma_2(\lambda, Z; \theta) = +\infty$.)

(ii) if $\theta \in \Theta'_I(F) \backslash \Theta_I(F)$, then

$$\mathbb{E}_F[r_1(Z; \theta)] = 0 \text{ and } \inf_{\lambda \in \mathcal{S}_2} \mathbb{E}_F \gamma_2(\lambda, Z; \theta) = 0,$$

or equivalently, $\inf_{\lambda \in \mathcal{S}} \mathbb{E}_F \gamma(\lambda, Z; \theta) = 0$.

Lemma 1 provide a sufficient condition for $\theta \in \Theta_I(F)$ without imposing assumptions as in Assumption 2. It also implies that Θ_I and Θ'_I can only differ in θs at which the testable restriction (9) is binding, i.e. $\inf_{\lambda \in \mathcal{S}} \mathbb{E} \gamma(\lambda, Z; \theta) = 0$. This result not only helps us to visualize the difference between Θ_I and Θ'_I, but also implies the following theorem.

Theorem 3. *Suppose Assumption 1 holds and $\mathcal{F}$ is a convex set. Let θ be an arbitrary parameter in Θ and partition $r(u, z; \theta) = (r_1(z; \theta), r_2(u, z; \theta))$. If there exists some $F^* \in \mathcal{F}$ such that*

$$\mathbb{E}_{F^*}[r_1(Z; \theta)] = 0 \text{ and } \inf_{\lambda \in \mathcal{S}_2} \mathbb{E}_{F^*} \gamma_2(\lambda, Z; \theta) > 0, \tag{15}$$

then it is impossible to distinguish $\theta \in \Theta_I$ from $\theta \in \Theta'_I$ in finite samples. Note that the $\inf_{\lambda \in \mathcal{S}_2} \mathbb{E}_{F^*} \gamma_2(\lambda, Z; \theta) > 0$ in (15) includes the case that $\inf_{\lambda \in \mathcal{S}_2} \mathbb{E}_{F^*} \gamma_2(\lambda, Z; \theta) = +\infty$.

To see the intuition of this result, consider the simple case where $\dim(r_1) = 0$. In this case, (15) is the same as

$$\inf_{\lambda \in \mathcal{S}} \mathbb{E}_{F^*} \gamma(\lambda, Z; \theta) > 0. \tag{16}$$

For any distribution $F^\dagger$ in $\mathcal{F}'_\theta$, and any $k \geq 1$, define $F_k = (1 - \frac{1}{k})F^\dagger + \frac{1}{k}F^*$. Because $\inf_{\lambda \in \mathcal{S}} \mathbb{E}_{F^\dagger} \gamma(\lambda, Z; \theta) \geq 0$, and because $\inf_{\lambda \in \mathcal{S}} \mathbb{E}_F \gamma(\lambda, Z; \theta)$ is concave function of F, it must be true that $\inf_{\lambda \in \mathcal{S}} \mathbb{E}_{F_k} \gamma(\lambda, Z; \theta) > 0$ for any k. Lemma 1 then implies $F_k \in \mathcal{F}_\theta$ for any k, so that any test ϕ_n must satisfy $\mathbb{E}_{F_k} \phi_n \leq \sup_{F \in \mathcal{F}_\theta} \mathbb{E}_F \phi_n$ for all k. Let $k \to \infty$, we conclude that $\mathbb{E}_{F^\dagger} \phi_n \leq \sup_{F \in \mathcal{F}_\theta} \mathbb{E}_F \phi_n$. Since $F^\dagger$ is an arbitrary element in $\mathcal{F}'_\theta$, this implies that any test cannot distinguish $\theta \in \Theta_I(F)$ and $\theta \in \Theta'_I(F)$ in finite samples.

To apply Theorem 3, one only needs to find one $F^* \in \mathcal{F}$ to satisfy (15) and such F^* need not be the true distribution which generates the data. In practice, the simplest way to verify (15) is to construct or simulate a data generating process for (U, Z) in which θ is the true parameter and verify whether the resulting distribution for Z satisfies (15). In general models, one needs to check (15) for each $\theta \in \Theta$ in order to ensure indistinguishability for each $\theta \in \Theta$. This is the best result one can possibly get without imposing more structure on how (Γ, r) depends on θ. In practice, (Γ, r) usually has more structure. For example, r_1 and γ_2 may depend on θ and Z only through some index $W(Z; \theta)$. This structure can be utilized to simplify computation greatly. To illustrate how Theorem 3 can be operationalized in practice, let me revisit Example 1 and 2.

Example 1 (continued)**.** In Example 1, $\dim(r_1) = 0$ so that (15) can be simplified to (16). Recall $Z = (Y, X)$ is the observables. For any $\theta = (\beta, \alpha)$ and any $\lambda \in \mathcal{S}$, the $\gamma(\lambda, Z; \theta)$, by its definition, is the result of the following optimization problem:

$$\begin{aligned}\gamma(\lambda, Z; \theta) \quad &= \quad \sup_u \lambda_1 \mathbb{1}(Y=1)u + \lambda_2 \mathbb{1}(Y=0)u \\ &\quad\quad \text{s.t. } u \in \begin{cases} [-X'\beta + \alpha, +\infty) & \text{if } Y = 1 \\ (-\infty, -X'\beta + \alpha] & \text{if } Y = 0 \end{cases}\end{aligned}$$

Therefore, $\gamma(\lambda, Z; \theta)$ depends on Z and θ only through $X'\beta - \alpha$ and Y. For any θ with $\beta \neq 0$, consider a data generating process (DGP) H_θ for (U, Z) as follows: $U \equiv 0$, $X'\beta - \alpha \sim N(0, 1)$ and $Y = \mathbb{1}(X'\beta - \alpha + U \geq 0)$. This DGP satisfies all model restrictions for θ. Let F_θ be the marginal distribution of H_θ for $Z = (Y, X)$. In addition, by construction, the resulting distribution G of $(X'\beta - \alpha, Y)$ is the same for all θ with $\beta \neq 0$. Let $W = X'\beta - \alpha$ and, with a slight abuse of the notation, write $\gamma(\lambda, Z; \theta) = \gamma(\lambda, W, Y)$. Then, for all $\lambda \in \mathcal{S}$, $\mathbb{E}_{F_\theta}\gamma(\lambda, Z; \theta) = \mathbb{E}_G \gamma(\lambda, W, Y)$. Numerical simulation shows that $\inf_{\lambda \in \mathcal{S}} \mathbb{E}_{F_\theta}\gamma(\lambda, Z; \theta) = \inf_{\lambda \in \mathcal{S}} \mathbb{E}_G \gamma(\lambda, W, Y) \approx 0.399 > 0$. Therefore, Theorem 3 implies that for any θ with $\beta \neq 0$, one cannot distinguish $\theta \in \Theta$ from $\theta \in \Theta'$ in finite samples. ■

Example 2 (continued)**.** In Example 2, $\dim(r_1) = 0$ so that we only need to check (16). Recall that $Y^* = X'\theta + \epsilon$ but we observe $Y = \max(Y^*, C)$. The $Z = (Y, C, W, X)$ collects all the observables and $U = (Y^*, \epsilon)$ is the latent variables. For any $\theta \in \Theta$ and any $\lambda \in \mathcal{S}$, the $\gamma(\lambda, Z; \theta)$ in this example is equal to

$$\gamma(\lambda, Z; \theta) = \begin{cases} \lambda' W(Y - X'\theta) & \text{if } Y > C \text{ or } \lambda' W \geq 0 \\ +\infty & \text{if otherwise} \end{cases}$$

For any $\theta \in \Theta$, one can always construct a data generating process (DGP) H for (U, Z) such that (i) $Y^* = X'\theta + \epsilon$, $Y = \max(Y^*, C)$ and $\mathbb{E}_H W\epsilon = 0$, (ii) $\mathbb{P}_H(Y^* \leq C) > 0$ and $\mathbb{P}_H(Y^* = C | Y^* \leq C) = 0$, and (iii) $\text{supp}(W | Y^* \leq C)$ is of dimension $\dim(W)$, i.e. it cannot be included within a hyperplane in $\mathbb{R}^{\dim(W)}$. Let F be the marginal distribution of H for Z. Then, for any $\lambda \in \mathcal{S}$, one can show that

$$\mathbb{E}_F \gamma(\lambda, Z; \theta) = \begin{cases} \mathbb{E}_H \mathbb{1}(Y^* \leq C)\lambda' W(C - Y^*) & \text{if } \mathbb{P}_H(\lambda' W \geq 0 | Y^* \leq C) = 1 \\ +\infty & \text{if otherwise} \end{cases}$$

Because $\text{supp}(W | Y^* \leq C)$ is of dimension $\dim(W)$, we know $\mathbb{P}_H(\lambda' W = 0 | Y^* \leq C) < 1$ for any $\lambda \in \mathcal{S}$. Because $\mathbb{P}_H(Y^* = C) = 0$, this implies that $\mathbb{E}_H \mathbb{1}(Y^* \leq C)\lambda' W(C - Y^*) > 0$ for any λ satisfying $\mathbb{P}_F(\lambda' W \geq 0 | Y = C) = 1$. Therefore, $\mathbb{E}_F \gamma(\lambda, Z; \theta) > 0$ for any $\lambda \in \mathcal{S}$.

The restrictions (i)-(iii) in the above paragraph are very weak conditions. As long as the $\mathcal{F}$ that the researcher assumes is compatible with these restrictions on the DGP, Theorem 3 implies that, for any $\theta \in \Theta$ one cannot distinguish $\theta \in \Theta$ from $\theta \in \Theta'$ in finite samples. ■

4.3 Irreducibility

Because (15) is only a sufficient condition for the indistinguishability between $\theta \in \Theta_I$ and $\theta \in \Theta_I'$, one would like to know how restrictive it is. Moreover, because (15) is a purely technical condition, one would like to know the interpretation and what to do when this condition fails to hold. In the rest of this section, I am going to explore the answer to these questions, which leads to a general principle that all restrictions on the support of the latent variables should be explicitly stated in $\Gamma(\theta)$ and the support restrictions should be treated differently from the moment restrictions. To motivate the analysis, let me revisit Example 4.

Example 4 (continued). I have already shown that Θ_I and Θ_I' are very different in Example 4. In the following, I point out two more observations on this example. First of all, given the $\gamma(\lambda, Z; \theta)$ solved in (10), one can show that $\inf_{\lambda \in \mathcal{S}} \mathbb{E}_F \gamma(\lambda, Z; \theta) = 0$ for any $\theta \in \Theta$ and any $F \in \mathcal{F}$. Since $\dim(r_1) = 0$ in this example, it implies that inequality (15) fails for all $\theta \in \Theta$ and for all $F \in \mathcal{F}$. This observation confirms the result in Theorem 3.

Secondly, recall that the moment restriction in this example is

$$\begin{aligned} \mathbb{E}[\mathbb{1}(Z_1 \le U \le Z_2) - 1] &= 0, \\ \mathbb{E}[U - \theta] &= 0. \end{aligned}$$

Note that the first moment restriction can be *reduced* to the support restriction, $\mathbb{P}(Z_1 \le U \le Z_2) = 1$. In other words, I can define a reduced model with moment restriction $\mathbb{E}[U - \theta] = 0$ and support restriction $\mathbb{P}(Z_1 \le U \le Z_2) = 1$. This reduced model, denoted as $(\tilde{\Gamma}, \tilde{r})$, is equivalent to the original model (Γ, r) in the sense that $\Theta_I(F; \Gamma, r) = \Theta_I(F; \tilde{\Gamma}, \tilde{r})$ for all $F \in \mathcal{F}$.

It turns out that these two observations are connected. As I will show later in a general setting, the reducibility noted in the second observation is the exact reason why (15) fails for all $F \in \mathcal{F}$. ■

In Example 4, after some of the moment restrictions are reduced to a support restriction, the reduced model is equivalent to the original model. Such reducibility is going to play an essential role in the following analysis. To state the general result, I am going to define such reducibility formally in the following.

Definition 4. Let (Γ, r) be an arbitrary model and let θ be an arbitrary element in Θ. Partition $r(u, z; \theta) = (r_1(z; \theta), r_2(u, z; \theta))$. We say model $(\tilde{\Gamma}, \tilde{r})$ is a *reduced model* of (Γ, r) at θ if $\dim(r_2) > 0$ and there exists $\dim(r_2)$ number of linearly independent vectors $\lambda_1, ..., \lambda_{\dim(r_2)} \in$

$\mathbb{R}^{\dim(r_2)}$ such that

$$\tilde{\Gamma}(\theta) = \Big\{(u,z) \in \Gamma(\theta) : \lambda_1' r_2(u,z;\theta) = \gamma_2(\lambda_1, z;\theta)\Big\}$$

where $\gamma_2(\lambda_1, z;\theta) \coloneqq \sup_{u\in\Gamma(z;\theta)} \lambda_1' r_2(u,z;\theta)$, and

$$\tilde{r}(u,z;\theta) = \begin{pmatrix} r_1(z;\theta) \\ \gamma_2(\lambda_1, z;\theta) \\ \lambda_2' r_2(u,z;\theta) \\ \lambda_3' r_2(u,z;\theta) \\ \vdots \\ \lambda_{\dim(r_2)}' r_2(u,z;\theta) \end{pmatrix}.$$

When model (Γ, r) reduces to $(\tilde{\Gamma}, \tilde{r})$, both the support and moment restrictions are modified. The reduced model includes an extra support restriction that the value of $\lambda_1' r_2(u,z;\theta)$ should equal $\gamma_2(\lambda_1, z;\theta)$ almost surely, and one of its moment restriction (after some possible rotation) is reduced to a condition which only involves Z and θ but not U. As a result, if we partition $\tilde{r}(u,z;\theta) = (\tilde{r}_1(z;\theta), \tilde{r}_2(u,z;\theta))$, $\dim(\tilde{r}_1) > \dim(r_1)$. One can show that $\Theta_I(F;\tilde{\Gamma},\tilde{r}) \subseteq \Theta_I(F;\Gamma,r)$, but the reverse inclusion is not true in general. In fact, it is possible that $\Theta_I(F;\tilde{\Gamma},\tilde{r}) = \emptyset$ but $\Theta_I(F;\Gamma,r) \neq \emptyset$. However, in some special cases as in Example 4, the original and the reduced model could be equivalent in the sense that they always have the same identified set. When there is such equivalence, I say the model is reducible.

Definition 5. Model (Γ, r) is *reducible* at θ if there exist some reduced model $(\tilde{\Gamma}, \tilde{r})$ at θ such that, for any $F \in \mathcal{F}$, $\theta \in \Theta_I(F;\Gamma,r)$ if and only if $\theta \in \Theta_I(F;\tilde{\Gamma},\tilde{r})$. Model (Γ, r) is *reducible* if it is reducible at some $\theta \in \Theta$. Reversely, the model is *irreducible* at θ if it is not reducible at θ and it is *irreducible* if it is not reducible.

By definition, if model (Γ, r) is reducible at θ, then some of its moment restrictions (possibly after some rotation) evaluated at θ can be equivalently transformed and separated into a support restriction and a moment restriction which only involves Z. In other words, a model is reducible if and only if some restriction on the support of the latent variables is implicitly stated in its moment restrictions. Reversely, a model is irreducible if and only if all of its support restrictions have been explicitly specified and none of the moment restrictions can be further reduced. When a reducible model is transformed into one of its equivalent reduced models, it will not change its identified set, but it could shrink the moment closure of the identified set. In fact, the following theorem shows that, if a model is, or has been transformed to, an irreducible one, the difference between the identified set and its moment closure cannot be distinguished in finite samples.

Theorem 4. *Suppose Assumption 1 hold for model* (Γ, r) *and* $\mathcal{F}$ *is a convex set. The following two results hold for any* $\theta \in \Theta$*,* *(i) if model* (Γ, r) *is irreducible at* θ*, then it is impossible to distinguish* $\theta \in \Theta_I$ *from* $\theta \in \Theta_I'$ *in finite samples.*

(ii) if (15) *fails to hold for all* $F \in \mathcal{F}$*, then one of the following statements must be true:*

(a) $\mathcal{F}_\theta'$ *is empty or, equivalently,* $\theta \notin \Theta_I'(F)$ *for all* $F \in \mathcal{F}$*,*

(b) *model* (Γ, r) *is reducible.*

Theorem 4 shows that Θ_I and Θ_I' are indistinguishable in finite samples if the model is irreducible or, to put differently, if all the support restrictions have been explicitly stated and none of the moment restrictions implicitly restrict the support of the latent variable. Note that whether or not a model is reducible is not an intrinsic property of the model. Instead, it depends on the way in which the model is written down. If model (Γ, r) is reducible at θ, then one could always reformulate it into an equivalent reduced model $(\tilde{\Gamma}, \tilde{r})$. If the resulted $(\tilde{\Gamma}, \tilde{r})$ is still reducible, then one can reduce $(\tilde{\Gamma}, \tilde{r})$ even further. After at most $\dim(r_2)$ rounds of reduction, any model can be rewritten as an equivalent irreducible model. Therefore, Theorem 4 essentially implies that Θ_I' of any model in the framework of (1), after proper reformulation and reduction, is indistinguishable from Θ_I in finite samples. Along with the result in Theorem 1, this result establishes the sharp identification results for all models.

Theorem 4 also enumerates all the cases in which the inequality (15) could fail to hold at θ for all $F \in \mathcal{F}$. In case (a), because $\mathcal{F}_\theta' = \emptyset$ also implies that $\mathcal{F}_\theta = \emptyset$, both $\theta \in \Theta_I(F)$ and $\theta \in \Theta_I'(F)$ will always be rejected in this trivial case. In fact, if $\mathcal{F}_\theta'$ is empty, one should not include this θ in the parameter space in the first place, because this θ can never be the true parameter. One can detect this case by trying to construct a DGP with θ being the true parameter. Case (a) cannot happen if such construction is possible. If one can rule out case (a), then (15) holds for some $F \in \mathcal{F}$ if and only if model (Γ, r) is irreducible at θ. This provides a way to verify the irreducibility using numerical simulations instead of analysis skills.

5 Counterfactual Analysis

In this section, I show how the identification approach discussed in the previous sections can be used to conduct counterfactual analysis. In the following, I call the parameters of interest in the counterfactual analysis *counterfactual parameters*, and call the other parameters *structural parameters*. Before discussing the general results, let me illustrate the basic idea with the running example.

Example 1 (continued)**.** Recall that, the support restriction in Example 1 is, $\mathbb{P}[(Z_i, U_i) \in \Gamma(\theta)] = 1$ where

$$\Gamma(\theta) = \{(z_i, u_i) : (-1)^{y_i}[x_i'\beta - \alpha + u_i] \leq 0\}, \qquad (5) \text{ revisited}$$

where y_i is agent i's choice, x_i stands for the covariates, u_i is the expectation error. Here, u_i

is the only unobserved variables, $z_i = (y_i, x_i)$ stands for the observables, and $\theta \coloneqq (\alpha, \beta)$ is the parameter.

Let us now consider a counterfactual setting in which parameter α changes to a hypothetical value $\tilde{\alpha}$, for example, because of a hypothetical change to the fixed cost of agent's choice. $\tilde{\alpha}$ could be a fixed value or a value related to α such as $\tilde{\alpha} = \frac{1}{2}\alpha$ or $\tilde{\alpha} = \alpha - 1$. Let $\widetilde{Y}_i$ be agent i's choice in this counterfactual. Assume $\mathbb{E}[U_i|\widetilde{Y}_i] = 0$ almost surely. Suppose that we are interested in the counterfactual choice probability $\tilde{p}$ defined as $\tilde{p} = \mathbb{P}(\widetilde{Y}_i = 1)$. Given the fact that agent i's counterfactual choice $\widetilde{Y}_i$ is not observed, how to find the identified set for the counterfactual parameter $\tilde{p}$?

It turns out that even if $\widetilde{Y}_i$ is not observed, it must satisfy the following restriction almost surely

$$\widetilde{Y}_i \in \begin{cases} \{1\} & \text{if } X_i'\beta - \tilde{\alpha} + U_i > 0, \\ \{0\} & \text{if } X_i'\beta - \tilde{\alpha} + U_i < 0, \\ \{0, 1\} & \text{if } X_i'\beta - \tilde{\alpha} + U_i = 0. \end{cases}$$

In other words, if I let $\tilde{u}_i \coloneqq (u_i, \tilde{y}_i)$ be the collection of all unobserved variables including the counterfactual choice, then I can write a new support restriction which restricts the latent variable U_i as well as $\widetilde{Y}_i$ as follows:

$$\mathbb{P}[(\widetilde{U}_i, Z_i) \in \widetilde{\Gamma}(\theta)] = 1 \text{ where } \widetilde{\Gamma}(\theta) = \{(\tilde{u}_i, z_i) : (-1)^{y_i}[x_i'\beta - \alpha + u_i] \leq 0,\ (-1)^{\tilde{y}_i}[x_i'\beta - \tilde{\alpha} + u_i] \leq 0\}.$$

Next, let us find out the moment restrictions which involves $\widetilde{Y}_i$. Analogous to $\mathbb{E}[U_i|Y_i] = 0$, I assume $\mathbb{E}[U_i|\widetilde{Y}_i] = 0$, which implies that $\mathbb{E}[\mathbb{1}(\widetilde{Y}_i = 1)U_i] = 0$ and $\mathbb{E}[\mathbb{1}(\widetilde{Y}_i = 0)U_i] = 0$. In addition, we can treat $\tilde{p}$ as an extra model parameter and view $\tilde{p} = \mathbb{E}[\mathbb{1}(\widetilde{Y} = 1)]$ as an additional moment restrictions. Append these restrictions to the original moment restrictions, and we get the following updated moment restrictions:

$$\mathbb{E}[\tilde{r}(\tilde{U}_i, Z_i; \tilde{\theta})] = 0, \text{ where } \tilde{r}(\tilde{U}_i, Z_i; \tilde{\theta}) = \begin{pmatrix} \mathbb{1}(Y_i = 1)U_i \\ \mathbb{1}(Y_i = 0)U_i \\ \mathbb{1}(\widetilde{Y}_i = 1)U_i \\ \mathbb{1}(\widetilde{Y}_i = 0)U_i \\ \mathbb{1}(\widetilde{Y}_i = 1) - \tilde{p} \end{pmatrix}$$

and $\tilde{\theta} \coloneqq (\alpha, \beta, \tilde{p})$ is the collection of all parameters including the counterfactual parameter $\tilde{p}$.

We now have a new model $(\widetilde{\Gamma}, \tilde{r})$ which incorporates all the restrictions on the counterfactual choice $\widetilde{Y}_i$. All the identification results discussed in the preceding sections can then be applied to $(\widetilde{\Gamma}, \tilde{r})$, in the same way as it applies to (Γ, r). Then, Theorem 1 implies that Condition (9), with (Γ, r) replaced by $(\widetilde{\Gamma}, \tilde{r})$, characterizes the Θ_I' for $\tilde{\theta}$. When Theorem 2 or 3 apply, it also characterizes the Θ_I for $\tilde{\theta}$.

To illustrate the connection between the Condition (9) for $(\widetilde{\Gamma}, \tilde{r})$ and that for (Γ, r) more vividly, note that, when X_i is continuously distributed and $\tilde{\alpha} < \alpha$,[1] the Condition (9) for $(\widetilde{\Gamma}, \tilde{r})$ in this example can be simplified as the following condition, as shown in Appendix A.1:

$$\begin{aligned}
\mathbb{E}[\mathbb{1}(Y_i = 1)(X_i'\beta - \alpha)] &\geq 0, \\
\mathbb{E}[\mathbb{1}(Y_i = 0)(X_i'\beta - \alpha)] &\leq 0, \\
\mathbb{E}[\mathbb{1}(Y_i = 1)] &\leq \tilde{p}, \\
\mathbb{E}[\mathbb{1}(Y_i = 1)] + \mathbb{E}[\mathbb{1}(Y_i = 0,\ X_i'\beta - \tilde{\alpha} \geq \mu)] &\geq \tilde{p}.
\end{aligned} \tag{17}$$

where

$$\mu = \begin{cases} -\infty & \text{if } \mathbb{E}[\mathbb{1}(Y_i = 0)(X'\beta - \tilde{\alpha})] \geq 0, \\ \mu \text{ that solves } \mathbb{E}[\mathbb{1}(Y_i = 0, X_i'\beta - \tilde{\alpha} \geq \mu)(X'\beta - \tilde{\alpha})] = 0 & \text{if } \mathbb{E}[\mathbb{1}(Y_i = 0)(X'\beta - \tilde{\alpha})] < 0. \end{cases}$$

Note that the first two conditions (17) are used to characterize the parameters in the original model. Because the counterfactual analysis does not impose extra identification restrictions for (α, β), it is not surprising that the first two conditions (17) coincide with the Condition (9) for the original model shown in (11). The last two conditions in (17) are the new conditions induced from the restrictions that are imposed to the counterfactual choice $\tilde{Y}_i$. They provide lower and upper bounds for the counterfactual parameter $\tilde{p}$. When $\mathbb{E}[\mathbb{1}(Y_i = 0)(X'\beta - \tilde{\alpha})] \geq 0$, the upper bound for $\tilde{p}$ is 1. When $\mathbb{E}[\mathbb{1}(Y_i = 0)(X'\beta - \tilde{\alpha})] < 0$, the upper bound for $\tilde{p}$ is nontrivial.

The inference is not the main focus here, but I want to point out that (17) can be rewritten equivalently as the following group of moment inequalities:

$$\begin{aligned}
\mathbb{E}[\mathbb{1}(Y_i = 1)(X_i'\beta - \alpha)] &\geq 0, \\
\mathbb{E}[\mathbb{1}(Y_i = 0)(X_i'\beta - \alpha)] &\leq 0, \\
\mathbb{E}[\mathbb{1}(Y_i = 1)] &\leq \tilde{p}, \\
\mathbb{E}[\mathbb{1}(Y_i = 1)] + \mathbb{E}[\mathbb{1}(Y_i = 0)\max(1 + \mu(X'\beta - \tilde{\alpha}), 0)] &\geq \tilde{p},\ \forall \mu \geq 0
\end{aligned} \tag{18}$$

where μ is a scalar and ranges over all non-negative values. With (18) in hand, one can then construct a confidence region for the structural parameter (α, β) and the counterfactual parameter $\tilde{p}$ jointly. Or, if one only cares about the counterfactual parameter $\tilde{p}$, one can conduct subvector inference directly on $\tilde{p}$ and treat (α, β) as nuisance parameters. ■

In general, counterfactual analysis can be conducted in the following way. Let $\widetilde{Y}_i$ denote the counterfactual model prediction. Suppose the counterfactual parameter $\tilde{p}$ satisfies the following moment conditions for some known function g,

$$\mathbb{E}[g(\widetilde{Y}_i, U_i, Z_i; \theta, \tilde{p})] = 0. \tag{19}$$

This moment condition usually comes from the definition of $\tilde{p}$ and some extra restrictions on

[1] One can derive similar conditions in the case of $\tilde{\alpha} > \alpha$.

$\widetilde{Y}_i$ and U_i as in the above example. In general cases, $\tilde{p}$ could be a vector, and function g could also be a vector function.

Given the unobservable and observed characteristics (u_i, z_i), define $\mathcal{C}(u_i, z_i; \theta)$ to be the set of all counterfactual behaviors which are consistent with the model assumptions. Then, the model restrictions on the counterfactual behaviors can be written as

$$\mathbb{P}[\widetilde{Y}_i \in \mathcal{C}(U_i, Z_i; \theta)] = 1.$$

Define $\widetilde{U}_i := (U_i, \widetilde{Y}_i)$ to be the collection of all unobservables including the counterfactual model prediction. Define a new support restriction $\mathbb{P}[(\widetilde{U}_i, Z_i) \in \widetilde{\Gamma}(\theta)] = 1$ based on the original support restrictions as well as the restrictions on the counterfactuals, i.e.

$$\widetilde{\Gamma}(\theta) := \{(\tilde{u}_i, z_i) : (u_i, z_i) \in \Gamma(\theta) \text{ and } \tilde{y}_i \in \mathcal{C}(u_i, z_i; \theta)\}. \tag{20}$$

Finally, let $\tilde{\theta} := (\theta, \tilde{p})$ be the collection of both structural and counterfactual parameters. Then, construct the new moment restriction $\mathbb{E}[\tilde{r}(\widetilde{U}_i, Z_i; \tilde{\theta})] = 0$ by combining the original moment restriction $\mathbb{E}[r(U_i, Z_i; \theta)] = 0$ with (19) and defining

$$\tilde{r}(\tilde{u}_i, z_i; \tilde{\theta}) = \begin{pmatrix} g(\tilde{y}_i, u_i, z_i; \tilde{\theta}) \\ r(u_i, z_i; \theta) \end{pmatrix}. \tag{21}$$

One can then view $\tilde{\theta}$ as a model primitive and apply the method in Section 3 to $(\widetilde{\Gamma}, \tilde{r})$. Depending on the goal of the empirical analysis, Condition (9) can be used to find the sharp identified set for θ and $\tilde{p}$ jointly or the projected identified set only for $\tilde{p}$.

In contrast to the above procedure, the traditional simulation-based counterfactual analysis is usually conducted as follows: One first sets up an empirical model in which the distribution of all random variables can be point identified. Then, the structural parameters are estimated. Finally, one simulates the unobservables with the estimated distribution and explicitly solves for the exact value or the bound of the model predictions with the simulated sample to recover the counterfactual parameters. Such an approach only works if the distribution of unobservables is point identified, but the point identification of the distribution often hinges on stringent restrictions like parametric assumptions on the distribution of unobservables, or large support assumptions for the covariates.

The approach developed in this section works under very mild conditions. Instead of simulating the unobservables, I directly utilize the restrictions on the unobservables in the original data. Heuristically, if it is possible to observe U_i in the data, the counterfactual analysis would be straightforward and there would be no need to simulate the unobservables. In practice, it is impossible to observe or calculate U_i in a possibly partially identified model, but the observed variables and the model do have restrictions on U_i, which further restricts the possible values of the counterfactuals. By exploiting these restrictions, one can then derive bounds on the counterfactual parameters. This is the basic intuition behind the construction of $(\widetilde{\Gamma}, \tilde{r})$ and also the major distinction between my approach and the traditional simulation-based approach.

6 Inference Method for Discrete Models

In this section, I discuss efficient ways to conduct inference when model (Γ, r) is discrete. The general form of the identification condition in (9) involves an infinite number of moment inequalities, whose inference problem is typically computationally challenging. However, as I will show later, when the model is discrete there exist ways to alleviate the computational complexity.

Definition 6. Model (Γ, r) is said to be *discrete* in the support if the set $\{r(u, z; \theta) : (u, z) \in \Gamma(\theta)\}$ is finite for any $\theta \in \Theta$.

A model (Γ, r) can be discrete if function $r(\cdot, \cdot; \theta)$ is discrete or if the support of (Z, U) is discrete. When a model (Γ, r) is discrete, Theorem 2 implies that $\Theta_I = \Theta_I'$. In the remaining of this section, I am going to show that, when (Γ, r) is discrete, testing $H_0 : \theta \in \Theta_I$ is equivalent to the inference for a linear system with known coefficients studied in Fang, Santos, Shaikh and Torgovitsky (2020).

For each $z \in \mathcal{Z}$, recall that $\Upsilon(z; \theta) \coloneqq \{r(u, z; \theta) : (u, z) \in \Gamma(\theta)\}$ is the set of all possible values of r at given z and θ. Define $\Upsilon(\theta) \coloneqq \{\Upsilon(z; \theta) : z \in \mathcal{Z}\}$. For a discrete model, both $\Upsilon(z; \theta)$ and $\Upsilon(\theta)$ are finite sets. We can then enumerate $\Upsilon(\theta)$ as $\Upsilon(\theta) = \{\Upsilon_1(\theta), ..., \Upsilon_K(\theta)\}$, where K can also depend on θ though I leave such dependence implicit to avoid heavy notations. Then, for any $z \in \mathcal{Z}$ with $\Upsilon(z; \theta) = \Upsilon_k(\theta)$, we know that

$$\gamma(\lambda, z; \theta) = \max\{\lambda' r : r \in \Upsilon_k(\theta)\}.$$

Therefore, for any $F \in \mathcal{F}$, $\inf_{\lambda \in S} \mathbb{E}_F \gamma(\lambda, Z; \theta) \geq 0$ is equivalent to

$$\inf_{\lambda \in \mathbb{R}^{\dim(r)}} \sum_{k=1}^{K} p_{F,k}(\theta) \max\{\lambda' r : r \in \Upsilon_k(\theta)\} \geq 0,$$

where $p_{F,k}(\theta) = \mathbb{P}_F(\Upsilon(Z; \theta) = \Upsilon_k)$. After introducing a vector of auxiliary variables $t \in \mathbb{R}^K$ and letting $t_k = \max\{\lambda' r : r \in \Upsilon_k(\theta)\}$, I can rewrite the identification condition as the following inequality:

$$\begin{aligned} 0 \leq \inf_{\lambda \in \mathbb{R}^{\dim(r)}, t \in \mathbb{R}^K} \quad & \sum_{k=1}^{k} p_{F,k}(\theta) t_k \\ \text{s.t.} \quad & t_k \geq \lambda' r, \quad \forall k = 1, ..., K, \forall r \in \Upsilon_k(\theta). \end{aligned} \tag{22}$$

Note that the right-hand side in the inequality (22) is a linear programming problem. To make the notation more concise, define $p_F(\theta) = (p_{F,1}(\theta), ..., p_{F,K}(\theta))'$. Enumerate $\Upsilon_k(\theta) =$

$\{r_{k,1}(\theta), ..., r_{k,m_k}(\theta)\}$ with m_k being the number of elements in $\Upsilon_k(\theta)$. Here, I leave the dependence of m_k on θ implicit to simplify the notations. Let $M = \sum_{k=1}^{K} m_k$. Define a $(\dim(r) + K) \times M$ matrix $A(\theta)$ as follows:

$$A(\theta) = \begin{pmatrix} -r_{1,1}(\theta) & \cdots & -r_{1,m_1}(\theta) & -r_{2,1}(\theta) & \cdots & -r_{2,m_2}(\theta) & \cdots & -r_{K,1}(\theta) & \cdots & -r_{K,m_K}(\theta) \\ 1 & \cdots & 1 & & & & & & & \\ & & & 1 & \cdots & 1 & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & 1 & \cdots & 1 \end{pmatrix}.$$

Then, inequality (22) can be written in the matrix notation:

$$\begin{aligned} 0 \leq \inf_{\lambda \in \mathbb{R}^{\dim(r)}, t \in \mathbb{R}^K} \quad & p_F(\theta)' t \\ \text{s.t.} \quad & A(\theta)' \begin{pmatrix} \lambda \\ t \end{pmatrix} \geq 0. \end{aligned} \tag{23}$$

After invoking the Farkas' lemma, inequality (23) holds if and only if the linear system in the following proposition has a nonnegative solution.

Proposition 1. *Suppose (Γ, r) is discrete. Then, $\theta \in \Theta_I(F)$ if and only if there exists some $\nu \in \mathbb{R}^M$ such that $\nu \geq 0$ and*

$$A(\theta)\nu = \begin{pmatrix} 0_{dim(r)} \\ p_F(\theta) \end{pmatrix}$$

where $0_{\dim(r)}$ is a zero vector of dimension $\dim(r)$ and $p_F(\theta) = (\mathbb{P}_F(\Upsilon(Z;\theta) = \Upsilon_k) : k = 1, ..., K)'$.

Note that $A(\theta)$ is a known and nonstochastic matrix function of θ. In finite samples. one only need to estimate vector $p_F(\theta)$. Therefore, the inferece problem for $\theta \in \Theta_I$ in a discrete model, after being transformed as in Proposition 1, fits in the framework of Fang, Santos, Shaikh and Torgovitsky (2020). One can then utilize the inference procedure therein to conduct the test for θ. Since their inference procedures can be implemented as a series of linear programming problems, the transformed hypothesis as in Proposition 1 is suitable for large scale problems.

For small scale problems, one can also enumerate all the vertices of the following polyhedron

$$\left\{ \begin{pmatrix} \lambda \\ t \end{pmatrix} : A(\theta)' \begin{pmatrix} \lambda \\ t \end{pmatrix} \geq 0 \right\}$$

using efficient software packages in computational geometry such as the package `lrs` by David Avis among others. Suppose the above polyhedron has N vertices which are enumerated as $(\lambda_1(\theta), t_1(\theta))$, ..., $(\lambda_N(\theta), t_N(\theta))$. Here, N can also depend on the value of θ, though I leave this dependence implicit to simplify notations. For any linear programming problem whose optimal value is finite, its optimal value can always be achieved at some vertex of the its feasible region. Therefore, inequality (23) holds if and only if the following moment inequalities holds:

$$\forall i = 1, ..., N, \; p_F'(\theta)t_i(\theta) \geq 0, \tag{24}$$

which further implies that $\theta \in \Theta_I(F)$ if and only if (24) holds for θ.

Again, note that $t_i(\theta)$ is nonstochastic and known given the value of θ. Therefore, one can use inference for moment inequalities to test inequality (24). See Andrews and Soares (2010) and Romano, Shaikh and Wolf (2014) among many others.

7 Conclusion

The sharp identification analysis in the partial identification literature typically focuses on methods and conditions under which the set of parameters satisfying the testable restrictions is equal to the identified set exactly. What makes the analysis in this paper different is that I switch to a slightly different analysis goal and study when these two sets are indistinguishable in finite samples. This leads to less restrictive regularity conditions and a wider class of applicable models. I applied this idea to models in framework (1) in this paper, but the same idea could be carried over to identification analysis for other models in the future as well.

Another distinctive feature of this paper is the unified treatment for both structural and counterfactual parameters. The reason why this is possible even for a partially identified incomplete model is that I allow the moment restrictions to depend on the latent variables. Indeed, this unified treatment is always possible whenever the identification strategy is general enough to incorporate the latent variables in the model restrictions. Hence, this idea can also be applied to methods in Beresteanu, Molchanov and Molinari (2011), Chesher and Rosen (2017) and Schennach (2014) among others.

This paper also illustrates the computation advantage of a discrete model. For a non-discrete model not covered in Section 6, a natural extension is to think about discretizing it or approximating it with another discrete models. It is possible to establish consistent results that the difference between the identified set of the continuous model and the discretized model would converge to zero as the approximation error goes to zero. However, what is more challenging is to bound the difference between these two sets for a given value of the approximation error instead of that in the limit. This approximation idea is also related to the sensitivity analysis where the model in hand could be slightly misspecified. Exploring these ideas would be an interesting topic for future work.

Appendices

A Additional Results on Examples

A.1 Results related to Example 1

Simplification of Condition (9): Note that, for each $\lambda \in \mathcal{S}$,

$$\mathbb{E}_F \gamma(\lambda, Z; \theta) = \mathbb{E}\left[\mathbb{1}(Y=1) \sup_{u \in [-X'\beta+\alpha, +\infty)} \lambda_1 u + \mathbb{1}(Y=0) \sup_{u \in (-\infty, -X'\beta+\alpha]} \lambda_2 u\right].$$

Therefore, we only need to focus on $\lambda \in \mathcal{S}' \coloneqq \{\lambda \in \mathcal{S} : \lambda_1 \leq 0, \lambda_2 \geq 0\}$ at which $\mathbb{E}_F \gamma(\lambda, Z; \theta) < +\infty$. For any $\lambda \in \mathcal{S}'$,

$$\mathbb{E}_F \gamma(\lambda, Z; \theta) = \lambda_1 \mathbb{E}\mathbb{1}(Y=1)(-X'\beta+\alpha) + \lambda_2 \mathbb{E}\mathbb{1}(Y=0)(-X'\beta+\alpha)$$

Note that $\mathbb{E}_F \gamma(\lambda, Z; \theta) \geq 0$ for all $\lambda \in \mathcal{S}'$ if and only if

$$\begin{aligned} \mathbb{E}_F[\mathbb{1}(Y=1)(X'\beta-\alpha)] &\geq 0, \\ \mathbb{E}_F[\mathbb{1}(Y=0)(X'\beta-\alpha)] &\leq 0. \end{aligned}$$

Therefore, Condition (9) holds if and only if the above moment inequality holds.

Simplification of Condition (9) in the counterfactual analysis: In the following, I only study the case where $\tilde{\alpha} < \alpha$. The analysis for $\tilde{\alpha} > \alpha$ is very similar. By definition, the γ function for (Γ', r') is

$$\begin{aligned} \gamma(\lambda, Z; \theta) \;=\; & \sup_{u, \widetilde{Y}} \lambda_1 \mathbb{1}(Y=1)u + \lambda_2 \mathbb{1}(Y=0)u + \lambda_3 \mathbb{1}(\widetilde{Y}=1)u + \lambda_4 \mathbb{1}(\widetilde{Y}=0)u + \lambda_5(\mathbb{1}(\widetilde{Y}=1) - \tilde{p}) \\ & \text{s.t. } u \in \begin{cases} [-X'\beta+\alpha, +\infty) & \text{if } Y = \widetilde{Y} = 1 \\ [-X'\beta+\tilde{\alpha}, -X'\beta+\alpha] & \text{if } Y = 0, \widetilde{Y} = 1 \\ (-\infty, -X'\beta+\tilde{\alpha}] & \text{if } Y = \widetilde{Y} = 0 \end{cases} \end{aligned}$$

Therefore,

$$\gamma(\lambda, Z; \theta) = \begin{cases} (\lambda_1+\lambda_3)(-X'\beta+\alpha) + \lambda_5(1-\tilde{p}) & \text{if } Y=1,\ \lambda_1+\lambda_3 \leq 0 \\ \max\big\{(\lambda_2+\lambda_3)(-X'\beta+\tilde{\alpha}) + \lambda_5(1-\tilde{p}), & \\ \quad (\lambda_2+\lambda_3)(-X'\beta+\alpha) + \lambda_5(1-\tilde{p}), & \text{if } Y=0, \lambda_2+\lambda_4 \geq 0 \\ \quad (\lambda_2+\lambda_4)(-X'\beta+\tilde{\alpha}) - \lambda_5\tilde{p}\big\} & \\ +\infty & \text{if otherwise} \end{cases}$$

As a result, we only need to focus on $\lambda \in \mathcal{S}'' \coloneqq \{\lambda \in \mathcal{S} : \lambda_1 + \lambda_3 \leq 0,\ \lambda_2 + \lambda_4 \geq 0\}$. In the following, I am going to show that Condition (9) is equivalent to (17) which are further equivalent to (18). I divide the proof into three parts.

Part 1: I claim that $\mathbb{E}\gamma(\lambda, Z; \theta) \geq 0$ for any $\lambda \in \mathcal{S}''$ with $\lambda_5 = 0$ if and only if

$$\begin{aligned} &\mathbb{E}\mathbb{1}(Y=1)(X'\beta - \alpha) \geq 0, \\ &\mathbb{E}\mathbb{1}(Y=0)(X'\beta - \alpha) \leq 0. \end{aligned} \tag{25}$$

To see why it is so, note that, for any $\lambda \in \mathcal{S}''$ with $\lambda_5 = 0, \mathbb{E}\gamma(\lambda, Z; \theta) \geq 0$ is equivalent to

$$\begin{aligned} &\mathbb{E}\mathbb{1}(Y=1)(\lambda_1+\lambda_3)(-X'\beta+\alpha) + \\ &\mathbb{E}\mathbb{1}(Y=0)\max\left\{(\lambda_2+\lambda_3)(-X'\beta+\tilde{\alpha}), (\lambda_2+\lambda_3)(-X'\beta+\alpha), (\lambda_2+\lambda_4)(-X'\beta+\tilde{\alpha})\right\} \geq 0. \end{aligned}$$

Therefore, $\mathbb{E}\gamma(\lambda, Z; \theta) \geq 0$ for any $\lambda \in \mathcal{S}''$ with $\lambda_5 = 0$ if and only if $\mathbb{E}\mathbb{1}(Y=1)(X'\beta - \alpha) \geq 0$ and the following inequality holds for any $\lambda \in \mathcal{S}''$ with $\lambda_5 = 0$,

$$\mathbb{E}\mathbb{1}(Y=0)\max\left\{(\lambda_2+\lambda_3)(-X'\beta+\tilde{\alpha}), (\lambda_2+\lambda_3)(-X'\beta+\alpha), (\lambda_2+\lambda_4)(-X'\beta+\tilde{\alpha})\right\} \geq 0. \tag{26}$$

Moreover, (26) holds for all $\lambda \in \mathcal{S}''$ with $\lambda_5 = 0$ if and only if $\mathbb{E}[\mathbb{1}(Y=0)(X'\beta - \alpha) \leq 0$. This is because (26) implies $\mathbb{E}[\mathbb{1}(Y=0)(X'\beta - \alpha) \leq 0$ when $\lambda_2 + \lambda_3 = \lambda_2 + \lambda_4 = 1$, and because $\mathbb{E}[\mathbb{1}(Y=0)(X'\beta - \alpha) \leq 0$ implies (26) for all $\lambda_2 + \lambda_3 \geq 0$, and because (26) always hold when $\lambda_2 + \lambda_3 \leq 0$.

Part 2: I claim that, when (25) hold, $\mathbb{E}\gamma(\lambda, Z; \theta) \geq 0$ for all $\lambda \in \mathcal{S}''$ with $\lambda_5 < 0$ if and only if $\tilde{p} \geq \mathbb{E}\mathbb{1}(Y=1)$. To see this, let $a = -(\lambda_1+\lambda_3)/\lambda_5$, $b = -(\lambda_2+\lambda_3)/\lambda_5$ and $c = -(\lambda_2+\lambda_4)/\lambda_5$. Then, $\mathbb{E}\gamma(\lambda, Z; \theta) \geq 0$ for all $\lambda \in \mathcal{S}''$ with $\lambda_5 < 0$ is equivalent to the following condition for all $a \leq 0$, $b \in \mathbb{R}$ and $c \geq 0$,

$$\begin{aligned} &a\mathbb{E}\mathbb{1}(Y=1)[(-X'\beta+\alpha)] \\ &\qquad + \mathbb{E}\mathbb{1}(Y=0)\max\left\{b(-X'\beta+\tilde{\alpha}), b(-X'\beta+\alpha), c(-X'\beta+\tilde{\alpha})+1\right\} \geq 1-\tilde{p} \end{aligned} \tag{27}$$

Since (25) hold, we know that (27) holds for all $a \leq 0$, $b \in \mathbb{R}$ and $c \geq 0$ is equivalent to that the following condition holds for all $b \in \mathbb{R}$ and $c \geq 0$,

$$\mathbb{E}\mathbb{1}(Y=0)\max\left\{b(-X'\beta+\tilde{\alpha}), b(-X'\beta+\alpha), c(-X'\beta+\tilde{\alpha})+1\right\} \geq 1-\tilde{p}. \tag{28}$$

When $b = c = 0$, (28) implies that $\tilde{p} \geq \mathbb{E}\mathbb{1}(Y=1)$. Moreover, for any $b \in \mathbb{R}$ and $c \geq 0$,

$$\begin{aligned} &\mathbb{E}\mathbb{1}(Y=0)\max\left\{b(-X'\beta+\tilde{\alpha}), b(-X'\beta+\alpha), c(-X'\beta+\tilde{\alpha})+1\right\} \\ \geq\ &c\mathbb{E}\mathbb{1}(Y=0)(-X'\beta+\tilde{\alpha}) + \mathbb{E}\mathbb{1}(Y=1) \\ \geq\ &c\mathbb{E}\mathbb{1}(Y=0)(-X'\beta+\alpha) + \mathbb{E}\mathbb{1}(Y=1) \\ \geq\ &\mathbb{E}\mathbb{1}(Y=1), \end{aligned}$$

where the second last inequality follows from $\tilde{\alpha} < \alpha$ and the last inequality follows from (25).
Part 3: I claim that, when (25) hold, $\mathbb{E}\gamma(\lambda, Z; \theta) \geq 0$ for all $\lambda \in \mathcal{S}''$ with $\lambda_5 > 0$ if and only if the last inequality in (17) hold. To see this, let $a = (\lambda_1 + \lambda_3)/\lambda_5$, $b = (\lambda_2 + \lambda_3)/\lambda_5$ and $c = (\lambda_2 + \lambda_4)/\lambda_5$. Then, $\mathbb{E}\gamma(\lambda, Z; \theta) \geq 0$ for all $\lambda \in \mathcal{S}''$ with $\lambda_5 > 0$ is equivalent to the following condition for all $a \leq 0$, $b \in \mathbb{R}$ and $c \geq 0$,

$$
\begin{aligned}
&a\mathbb{E}\mathbb{1}(Y=1)[(-X'\beta + \alpha)] + \mathbb{E}\mathbb{1}(Y=1) \\
&\qquad + \mathbb{E}\mathbb{1}(Y=0)\max\left\{b(-X'\beta + \tilde{\alpha}) + 1, b(-X'\beta + \alpha) + 1, c(-X'\beta + \tilde{\alpha})\right\} \geq \tilde{p} \quad (29)
\end{aligned}
$$

Since (25) hold, that (29) holds for all $a \leq 0$, $b \in \mathbb{R}$ and $c \geq 0$ is equivalent to that the following condition hold for all $b \in \mathbb{R}$ and $c \geq 0$,

$$\mathbb{E}\mathbb{1}(Y=1) + \mathbb{E}\mathbb{1}(Y=0)\max\left\{b(-X'\beta + \tilde{\alpha}) + 1, b(-X'\beta + \alpha) + 1, c(-X'\beta + \tilde{\alpha})\right\} \geq \tilde{p} \quad (30)$$

When $b \geq 0$, (30) is equivalent to

$$\mathbb{E}\mathbb{1}(Y=1) + \mathbb{E}\mathbb{1}(Y=0)\max\left\{b(-X'\beta + \alpha) + 1, c(-X'\beta + \tilde{\alpha})\right\} \geq \tilde{p}. \quad (31)$$

Note that when $b = c = 0$, (31) implies $1 \geq \tilde{p}$. Reversely, for any $b \geq 0$ and $c \geq 0$,

$$
\begin{aligned}
& \mathbb{E}\mathbb{1}(Y=1) + \mathbb{E}\mathbb{1}(Y=0)\max\left\{b(-X'\beta + \alpha) + 1, c(-X'\beta + \tilde{\alpha})\right\} \\
\geq\ & \mathbb{E}[\mathbb{1}(Y=1)] + \mathbb{E}[\mathbb{1}(Y=0)(b(-X'\beta + \alpha) + 1)] \\
\geq\ & \mathbb{E}[\mathbb{1}(Y=1)] + \mathbb{E}[\mathbb{1}(Y=0)] \\
=\ & 1
\end{aligned}
$$

where the last inequality follows from (25). Hence, that (30) holds for all $b \geq 0$ and $c \geq 0$ is equivalent to $1 \geq \tilde{p}$.

When $b \leq 0$, (30) is equivalent to

$$\mathbb{E}\mathbb{1}(Y=1) + \mathbb{E}\mathbb{1}(Y=0)\max\left\{b(-X'\beta + \tilde{\alpha}) + 1, c(-X'\beta + \tilde{\alpha})\right\} \geq \tilde{p}.$$

Because $\tilde{\alpha} < \alpha$ and (25) hold, the above condition holds for all $b \leq 0$ and $c \geq 0$ if and only if the following condition holds:

$$\inf_{\rho \geq 0} \mathbb{E}[\mathbb{1}(Y=0)\max\left\{\rho(X'\beta - \tilde{\alpha}) + 1, 0)\right\} \geq \tilde{p} - \mathbb{E}[\mathbb{1}(Y=1)] \quad (32)$$

Note that the left-hand side of (32) is a convex minimization problem. Its Karush–Kuhn–Tucker (KKT) conditions for the optimal ρ^* can be written as

$$
\begin{aligned}
\rho^* \leq 0,\ \mathbb{E}\mathbb{1}(Y=0, \rho^*(X'\beta - \tilde{\alpha}) + 1 \geq 0)(X'\beta - \tilde{\alpha}) \geq 0 \\
\rho^*\mathbb{E}\mathbb{1}(Y=0, \rho^*(X'\beta - \tilde{\alpha}) + 1 \geq 0)(X'\beta - \tilde{\alpha}) = 0
\end{aligned}
$$

If $\mathbb{E}\mathbb{1}(Y=0)(X'\beta-\tilde{\alpha})\geq 0$, then $\rho^*=0$ satisfies the KKT conditions. If $\mathbb{E}\mathbb{1}(Y=0)(X'\beta-\tilde{\alpha})<0$, then $\rho^*>0$ and it solves

$$\mathbb{E}\mathbb{1}\left(Y=0,(X'\beta-\tilde{\alpha})\geq-\frac{1}{\rho^*}\right)(X'\beta-\tilde{\alpha})=0$$

which coincide with the last condition in (17).

A.2 additional example

Example 5 (binary choice model with both known and unknown errors to the agents)**.** Consider the same model as in Example 1 except that the agents now might know some payoff shocks when making the decision. As before, let $Y_i\in\{0,1\}$ be agent i's choice. When $Y_i=1$, the payoff π_i of player i is

$$\pi_i=X_i'\beta-\alpha+\epsilon_i,$$

where ϵ_i is the payoff shocks that are known to the agent i but unobservable to the researchers. As in Example 1, assume agent i chooses optimally based on his subjective expectation,

$$Y_i=\begin{cases}1 & \text{if } \mathbb{E}_s[\pi_i]>0,\\ 0 & \text{if } \mathbb{E}_s[\pi_i]<0.\end{cases} \tag{33}$$

Assume the expectation is rational so that the expectation error $\nu_i=\mathbb{E}_s[\pi_i]-\pi_i$ satisfy $\mathbb{E}[\nu_i|Y_i]=0$ almost surely. Assume also that ϵ_i has median zero, i.e. $\mathbb{P}(\epsilon_i\leq 0)=\mathbb{P}(\epsilon_i\geq 0)$.

To fit this model into the framework, let $Z_i:=(Y_i,X_i)$, $U_i:=(\epsilon_i,\nu_i)$ and $\theta:=(\alpha,\beta)$. Then, the support restriction is

$$\mathbb{P}[(U_i,Z_i)\in\Gamma(\theta)]=1,\ \text{ where } \Gamma(\theta)=\{(u_i,z_i):(-1)^{y_i}[x_i'\beta-\alpha+\nu_i+\epsilon_i]\leq 0\}. \tag{34}$$

Moreover, moment restriction is:

$$\mathbb{E}r(U_i,Z_i;\theta)=0,\ \text{ where } r(U_i,Z_i;\theta)=\begin{pmatrix}\mathbb{1}(Y_i=1)\nu_i\\ \mathbb{1}(Y_i=0)\nu_i\\ \mathbb{1}(\epsilon_i\leq 0)-\mathbb{1}(\epsilon_i\geq 0)\end{pmatrix}.$$

■

B Basic Concepts of Random Set Theory

This section collects some basic concepts and results of random sets and measurable functions used in the paper. Throughout the paper, the random set is defined on a finite-dimensional Euclidean space. I follow the notation in Molchanov (2005) whenever possible.

Definition B.1 (Random Set)**.** Let $(\Omega, \mathscr{S}, P)$ be a probability space. A correspondence $Y : \Omega \rightrightarrows \mathbb{R}^d$ is said to be a *random closed set* if (i) $Y(\omega)$ is closed almost surely; (ii) for each compact set K in $\mathbb{R}^d$, $\{\omega \in \Omega : Y(\omega) \cap K \neq \emptyset\} \in \mathscr{S}$.

Fix a complete probability space $(\Omega, \mathscr{S}, P)$. Let $L^1(\Omega; \mathbb{R}^d)$ denote the set of all integrable functions $f : \Omega \mapsto \mathbb{R}^d$. The following introduces the expectation concept of random set theory.

Definition B.2 (integrable selections)**.** If Y is a random closed set, then $S^1(Y)$ denotes the family of all integrable selections of Y. That is,

$$S^1(Y) \coloneqq \{f \in L^1(\Omega; \mathbb{R}^d) : f(\omega) \in Y(\omega) \text{ almost surely}\}$$

Definition B.3 (integration of random set)**.** Let Y be a random closed set. Its *Aumann integral* $\mathbb{E}_I Y$ is defined as the set of all expectations of integrable selections,

$$\mathbb{E}_I Y \coloneqq \{\mathbb{E} f : f \in S^1(Y)\}$$

Its *selection expectation* $\mathbb{E}Y$ is defined as the closure of $\mathbb{E}_I Y$,

$$\mathbb{E}Y \coloneqq \text{cl}\{\mathbb{E} f : f \in S^1(Y)\}$$

Finally, the following introduces a boundedness concept on random sets.

Definition B.4 (integrable random set)**.** A random closed set Y is called *integrable* if $S^1(Y) \neq \emptyset$. A random closed set Y is called *integrably bounded* if $\|Y\| \coloneqq \sup\{\|t\| : t \in Y\}$ has finite expectation, i.e. $\|Y\| \in L^1(\Omega; \mathbb{R})$.

The following lemma summarizes the results I used to prove the theorems in the paper.

Lemma B.1. *Let Y be a random closed set, whose realization is a subset of $\mathbb{R}^d$.*

(i) $S^1(Y) \neq \emptyset$ *if and only if* $\inf\{\|t\| : t \in Y\}$ *is integrable.*

(ii) *If Y is integrably bounded, $\mathbb{E}_I Y$ is a compact set and $\mathbb{E}Y = \mathbb{E}_I Y$.*

(iii) *If a function $\zeta : \mathbb{R}^d \mapsto \mathbb{R} \cup \{\pm\infty\}$ is upper or lower semicontinuous , then* $\inf\{\zeta(t) : t \in Y\}$ *is a random variable. Moreover, if $S^1(Y) \neq \emptyset$ and $\mathbb{E}\zeta(f)$ is defined for all $f \in S^1(Y)$ and $\mathbb{E}\zeta(f) < \infty$ for at least one $f \in S^1(Y)$, then*

$$\inf_{f \in S^1(Y)} \mathbb{E}\zeta(f) = \mathbb{E} \inf_{t \in Y} \zeta(t)$$

(iv) *If $S^1(Y) \neq \emptyset$, then $\mathbb{E}\overline{\text{co}}(Y) = \overline{\text{co}}\mathbb{E}Y$ where $\overline{\text{co}}$ stands for the closure of the convex hull.*

Proof. For results (i), (iii) and (iv), see Molchanov (2005), Theorem 1.7 (p.149), Theorem 1.10 (p. 150) and Theorem 1.17 (p. 154) respectively.

For result (ii), Theorem 1.24 on page 158 in Molchanov (2005) implies $\mathbb{E}_I Y$ is a closed set. Moreover, since $\|v\| \leq \mathbb{E}\|Y\|$, $\forall v \in \mathbb{E}_I Y$, $\mathbb{E}_I Y$ is bounded. Since $\mathbb{E}_I Y \subseteq \mathbb{R}^d$, $\mathbb{E}_I Y$ is compact. □

C Selection Theorem

This section collects some concepts and results on measurable selection which will be cited later in the proof.

Definition C.1 (universally measurable set)**.** Let S be a Polish space and let $\mathscr{B}_S$ be its Borel sigma algebra. A subset S' of S is a *universally measurable* set if for any complete probability space $(S, \mathscr{F}, F)$ with $\mathscr{B}_S \subseteq \mathscr{F}$, $S' \in \mathscr{F}$.

Definition C.2 (universally measurable function)**.** Let S be a Polish space and let $\mathscr{B}_S$ be its Borel sigma algebra, and T be some topological space. A function $f : S \mapsto T$ is *universally measurable* if for any Borel set B of T, $\{s \in S : f(s) \in B\}$ is universally measurable.

By definition, if a function is uniformly measurable, then it's also measurable in the completion of any Borel probability space. Moreover, any Borel set in a Polish space is universally measurable. Given $D \subseteq S \times T$, define $\text{proj}_S(D) \coloneqq \{s \in S : \exists t \in T, (s,t) \in D\}$ and $D_s \coloneqq \{t \in T : (t,s) \in D\}$. The following lemma is a simplified version of Proposition 7.50(b) in Bertsekas and Shreve (1978).

Lemma C.1 (measurable selection)**.** *Let S and T be Polish spaces, let $D \subseteq S \times T$ be a Borel set, and let $f : D \to \mathbb{R}$ be a Borel measurable function. Define $f^* : proj_S(D) \to \mathbb{R} \cup \{-\infty\}$ by*

$$f^*(s) = \inf_{t \in D_s} f(s,t).$$

Suppose $f^*(s) > -\infty$ for any $s \in proj_S(D)$. Then, the set

$$I \coloneqq \{s \in proj_S(D) : \exists t_s \in D_s, f(s,t_s) = f^*(s)\}$$

is universally measurable. And, for every $\epsilon > 0$, there exists a universally measurable function $\phi : proj_S(D) \mapsto T$ such that (i) $Gr(\phi) \subseteq D$; (ii) for all $s \in proj_S(D)$, $f(s,\phi(s)) \leq f^*(s) + \epsilon,\ \forall s \in S$ and, (iii) for all $s \in I$, $f(s,\phi(s)) = f^*(s)$.

Proof. Since

- every Borel set is an analytic set,
- every Polish space is a Borel space as defined in Definition 7.7 in Bertsekas and Shreve (1978) (page 118),
- every Borel measurable function is lower semianalytic function as defined in Definition 7.21 in Bertsekas and Shreve (1978) (page 177),

the result follows from Proposition 7.50(b) on page 184 in Bertsekas and Shreve (1978). □

D Proof of Theorem 1, Theorem 2 and Lemma 1

The proof builds on the result of random sets and measurable selections listed in Appendix B and C. To state the proof, I also need the following extra notation:

Notation For any $F \in \mathcal{F}$, let $\widetilde{\Theta}(F)$ be the set of all θ which satisfies (9). For any set A in an Euclidean space, I use int$A$ to denote its interior , cl$A$ to denote its closure, co$A$ to denote its convex hull and $\overline{\text{co}}A$ to denote the closure of its convex hull. Given any topological space X, let $\mathscr{B}_{\mathcal{X}}$ denote all Borel sets on $\mathcal{X}$, and $\mathscr{P}_{\mathcal{X}}$ denote the set of all probability measures on measurable space $(\mathcal{X}, \mathscr{B}_{\mathcal{X}})$. Recall that $\mathcal{U}$ and $\mathcal{Z}$ denote the space of U and Z respectively. For any $F \in \mathcal{F}$, let the probability space $(\mathcal{Z}, \mathscr{Z}, F)$ be the completion of $(\mathcal{Z}, \mathscr{B}_{\mathcal{Z}}, F)$. Moreover, recall $\Gamma(z;\theta) := \{u \in \mathcal{U} : (u, z) \in \Gamma(\theta)\}$. Define $\Upsilon(z;\theta)$ as the image of $\Gamma(z;\theta)$ by r, i.e.

$$\Upsilon(z;\theta) := \{r(u, z;\theta) : u \in \Gamma(z;\theta)\}.$$

Then, (9) can be rewritten as

$$\forall \lambda \in \mathcal{S},\ \mathbb{E}_F\left[\sup_{t \in \Upsilon(Z;\theta)} \lambda' t\right] \geq 0.$$

In the following, I first prove Lemma D.1 which establishes some useful properties for $\Upsilon(z;\theta)$ as a random set. Then, I prove Theorem 1 first, and then Theorem 2 and finally Lemma 1.

D.1 Property of $\Upsilon(z;\theta)$

In the following, I use Assumption 1(i) and 1(ii) to denote the first and the second condition in Assumption 1 respectively. Similarly, I use Assumption 2 (i) and Assumption 2 (ii) to denote the first and the second condition in Assumption 2. The following lemma provides some basic results needed for the proof of all theorems.

Lemma D.1. *Let F be an arbitrary element in $\mathcal{F}$.*

(i) Suppose Assumption 1(i) holds. Then, for each $\theta \in \Theta$, $\text{cl}\Upsilon(\cdot;\theta)$ is a random closed set in probability space $(\mathcal{Z}, \mathscr{Z}, F)$.

(ii) Suppose Assumption 1 hold. Then, for each $\theta \in \Theta$, $\text{cl}\Upsilon(\cdot;\theta)$ is an integrable random closed set in probability space $(\mathcal{Z}, \mathscr{Z}, F)$.

(iii) Suppose Assumption 1(i) and Assumption 2(ii) hold. Then, for each $\theta \in \Theta$, random closed set $\text{cl}\Upsilon(\cdot;\theta)$ is integrably bounded in probability space $(\mathcal{Z}, \mathscr{Z}, F)$.

Proof of Lemma D.1. (i) I first show $\text{cl}\Upsilon(\cdot;\theta)$ is a random closed set under Assumption 1(i).

Let $D = \{t_1, t_2, ...\}$ be a countable set dense in $\mathbb{R}^{\dim(r)}$. For each $t_i \in D$, consider the following optimization problem ,

$$\inf_{u \in \Gamma(z;\theta)} \|t_i - r(u, z; \theta)\|$$

Given that $\|t_i - r(u, z; \theta)\|$ is a Borel measurable function of (u, z), that $\Gamma(\theta)$ is a Borel set, and that $\Gamma(z; \theta)$ is nonempty almost surely, Lemma C.1 implies that, for any $n \in \mathbb{N}$, there exists a universally measurable function $f_{i,n} : \mathcal{Z} \mapsto \mathcal{U}$ such that for any $z \in Z$, $f_{i,n}(z) \in \Gamma(z; \theta)$ and

$$\|t_i - r(f_{i,n}(z), z; \theta)\| \leq \frac{1}{n} + \inf_{u \in \Gamma(z;\theta)} \|t_i - r(u, z; \theta)\| .$$

See Definition C.2 for the definition of a universal measurable function. Since $(\mathcal{Z}, \mathscr{Z}, F)$ is the completion of the Borel probability space $(\mathcal{Z}, \mathscr{B}_Z, F)$, by the definition of universally measurable functions, $f_{i,n}(z)$ is also $\mathscr{Z}$-measurable.

Fix an arbitrary z. Since, by construction, $f_{i,n}(z) \in \Gamma(z; \theta)$, we know $\mathrm{cl}\{r\big(f_{i,n}(z), z\big) : i, n \in \mathbb{N}\} \subseteq \mathrm{cl}\Upsilon(z; \theta)$. On the other hand, for any $t \in \mathrm{cl}\Upsilon(z; \theta)$ and any $\epsilon > 0$, there must exists some $t_i \in D$ such that $\|t - t_i\| \leq \epsilon/3$, and there must exists some $n \in \mathbb{N}$ such that $\|t_i - r(f_{i,n}(z), z; \theta)\| \leq 2\epsilon/3$. Hence, for any $t \in \mathrm{cl}\Upsilon(z; \theta)$ and any $\epsilon > 0$, there exists some $\tilde{t} \in \{r\big(f_{i,n}(z), z\big) : i, n \in \mathbb{N}\}$ such that $\big\|t - \tilde{t}\big\| \leq \epsilon$. Hence, $\mathrm{cl}\Upsilon(z; \theta) = \mathrm{cl}\{r\big(f_{i,n}(z), z\big) : i, n \in \mathbb{N}\}$. By Theorem 2.3 on page 26 of Molchanov (2005), $\mathrm{cl}\Upsilon(z; \theta)$ is a random closed set in $(\mathcal{Z}, \mathscr{Z}, F)$.

(*ii*) Suppose, in addition, Assumption 1(ii) holds. The fact that $\mathrm{cl}\Upsilon(z; \theta)$ is a random closed set implies $z \mapsto \inf\{\|t\| : t \in \mathrm{cl}\Upsilon(z; \theta)\}$ is measurable in $(\mathcal{Z}, \mathscr{Z})$ (See result (iii) in Lemma B.1). Moreover, note that

$$\inf\{\|t\| : t \in \Upsilon(z; \theta)\} = \inf\{\|t\| : t \in \mathrm{cl}\Upsilon(z; \theta)\}.$$

Assumption 1(ii) then implies $z \mapsto \inf\{\|t\| : t \in \mathrm{cl}\Upsilon(z; \theta)\}$ is an integrable function. By Definition B.4 and Lemma B.1(i), $\mathrm{cl}\Upsilon(\cdot; \theta)$ is integrable.

(*iii*) Finally, given the first result in this lemma, Assumption 2(ii) directly implies $\mathrm{cl}\Upsilon(\cdot; \theta)$ is integrably bounded by definition. □

D.2 Proof of Theorem 1

Let me state the following two lemmas first, the proof of which will be presented after I prove Theorem 1.

Lemma D.2. *Suppose set A is a nonempty closed convex set in $\mathbb{R}^d$. Then $0 \in A$ if and only if*

$$\inf_{\lambda \in \mathbb{R}^d} \sup\{\lambda' t : t \in A\} \geq 0. \tag{35}$$

Note that (35) *includes the case that* $\inf_{\lambda\in\mathbb{R}^d}\sup\{\lambda' t : t\in A\}\geq 0$ *which could happen when* $A=\mathbb{R}^d$.

Lemma D.3. *Suppose Assumption 1 hold. Then, for any* $F\in\mathcal{F}$, $0\in\overline{\mathrm{co}}\mathbb{E}_F\mathrm{cl}\Upsilon(Z;\theta)$ *implies* $\theta\in\Theta_I'(F)$.

Proof of Theorem 1. Fix an arbitrary element F in $\mathcal{F}$. In the following proof, I will abbreviate $\mathbb{E}_F$ as $\mathbb{E}$, $\Theta_I'(F)$ as Θ_I', and $\widetilde{\Theta}(F)$ as $\widetilde{\Theta}$. Recall that $\mathbb{E}_I$ stands for the Aumann integral.

First of all, I'm going to show $\widetilde{\Theta}(F)\subseteq\Theta_I'(F)$. Lemma D.1 implies that $\mathrm{cl}\Upsilon(\cdot;\theta)$ is an integrable random closed set in $(\mathcal{Z},\mathscr{Z},F)$. Suppose, for the purpose of contradiction, there exists $\theta\in\widetilde{\Theta}$ such that $\theta\notin\Theta_I'$. Then, by Lemma D.3, $0\notin\overline{\mathrm{co}}\mathbb{E}\mathrm{cl}\Upsilon(Z;\theta)$. Lemma D.2 then implies that the following inequality holds:

$$\inf_{\lambda\in\mathbb{R}^{\dim(r)}}\sup\{\lambda' t : t\in\overline{\mathrm{co}}\mathbb{E}\mathrm{cl}\Upsilon(Z;\theta)\}<0$$

By Lemma B.1(iv), and the fact that $\overline{\mathrm{co}}\Upsilon(Z;\theta)\subseteq\overline{\mathrm{co}}\mathrm{cl}\Upsilon(Z;\theta)$, and that the Aumann integral $\mathbb{E}_I\overline{\mathrm{co}}\Upsilon(Z;\theta)\subseteq\mathbb{E}\overline{\mathrm{co}}\Upsilon(Z;\theta)$, we know

$$\inf_{\lambda\in\mathbb{R}^{\dim(r)}}\sup\{\lambda' t : t\in\mathbb{E}_I\overline{\mathrm{co}}\Upsilon(Z;\theta)\}<0 \tag{36}$$

Choose any $\tilde\lambda$ such that $\sup\{\tilde\lambda' t : t\in\mathbb{E}_I\overline{\mathrm{co}}\Upsilon(Z;\theta)\}<0$. Note that

$$\sup\{\tilde\lambda' t : t\in\mathbb{E}_I\overline{\mathrm{co}}\Upsilon(Z;\theta)\}=-\inf_{f\in S^1(\overline{\mathrm{co}}\Upsilon(Z;\theta))}\mathbb{E}[-\tilde\lambda' f] \tag{37}$$

where S^1 is defined in Definition B.2. Apply Lemma B.1(iii) with $\zeta(t)=-\lambda' t$ to get

$$\begin{aligned}&-\inf_{f\in S^1(\overline{\mathrm{co}}\Upsilon(Z;\theta))}\mathbb{E}[-\tilde\lambda' f]\\ =\;&-\mathbb{E}\inf\{-\tilde\lambda' t : t\in\overline{\mathrm{co}}\Upsilon(Z;\theta)\}\\ =\;&\mathbb{E}\sup\{\tilde\lambda' t : t\in\overline{\mathrm{co}}\Upsilon(Z;\theta)\}.\end{aligned} \tag{38}$$

Equation (37) and (38) imply

$$\mathbb{E}\sup\{\tilde\lambda' t : t\in\overline{\mathrm{co}}\Upsilon(Z;\theta)\}=\sup\{\tilde\lambda' t : t\in\mathbb{E}_I\overline{\mathrm{co}}\Upsilon(Z;\theta)\}<0. \tag{39}$$

In addition, since $\Upsilon(z;\theta)\subseteq\mathbb{R}^{\dim(r)}$,

$$\sup\{\tilde\lambda' t : t\in\overline{\mathrm{co}}\Upsilon(z;\theta)\}=\sup\{\tilde\lambda' t : t\in\Upsilon(z;\theta)\}, \tag{40}$$

equation (39) and (40) imply

$$\inf_{\lambda\in\mathbb{R}^{\dim(r)}}\mathbb{E}\sup\{\lambda' t : t\in\Upsilon(Z;\theta)\}<0.$$

This contradicts $\theta\in\widetilde{\Theta}$. This proves $\widetilde{\Theta}\subseteq\Theta_I'$.

To show $\Theta_I' \subseteq \widetilde{\Theta}$. Fix any $\theta \in \Theta_I'$ and any $\epsilon > 0$, there exists a distribution H of (U, Z) such that (i) $\|\mathbb{E}r(U, Z; \theta)\| \leq \epsilon$; (ii) $\mathbb{P}_H(U \in \Gamma(Z; \theta)) = 1$; (iii) the marginal distribution of H on Z equals to F. For any $\lambda \in \mathcal{S}$,

$$
\begin{aligned}
-\epsilon &\leq \mathbb{E}_H(\lambda' r(U, Z; \theta)) \\
&\leq \mathbb{E}_H \left\{ \sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta) \right\} \\
&= \mathbb{E} \left\{ \sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta) \right\}
\end{aligned}
$$

where the first inequality comes from Cauchy-Schwarz inequality, the second inequality comes form $\mathbb{P}_H(U \in \Gamma(Z; \theta)) = 1$, and the last equality follows from the fact that $\sup\{\lambda' r(u, z; \theta) : u \in \Gamma(z; \theta)\}$ only depends on z. Hence,

$$
-\epsilon \leq \inf_{\lambda \in \mathcal{S}} \mathbb{E} \left[\sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta) \right].
$$

Since these holds with any $\epsilon > 0$,

$$
0 \leq \inf_{\lambda \in \mathcal{S}} \mathbb{E} \left[\sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta) \right],
$$

which implies $\theta \in \widetilde{\Theta}$. □

Proof of Lemma D.2. This is a classic result of the support function. See, for example, Theorem 2.2.2 in Hiriart-Urruty and Lemaréchal (2001) for its proof. □

Proof of Lemma D.3. Fix an arbitrary $F \in \mathcal{F}$. In the following proof, I will abbreviate $\mathbb{E}_F$ as $\mathbb{E}$, and $\Theta_I'(F)$ as Θ_I'. Under Assumption 1, $\mathrm{cl}\Upsilon(Z; \theta)$ is an integrable random closed set in $(\mathcal{Z}, \mathscr{Z}, F)$. Suppose $0 \in \overline{\mathrm{co}}\mathbb{E}\mathrm{cl}\Upsilon(Z; \theta)$ is true, I want to prove that $\theta \in \Theta_I'$.

Fix an arbitrary $\epsilon > 0$. By the fact that $\overline{\mathrm{co}}A = \overline{\mathrm{co}}\mathrm{cl}A$ for any subset A in finite dimensional Euclidean space, and that $\mathbb{E}\mathrm{cl}\Upsilon(Z; \theta) = \mathrm{cl}(\mathbb{E}_I \mathrm{cl}\Upsilon(Z; \theta))$ by Definition B.4, $0 \in \overline{\mathrm{co}}\mathbb{E}\mathrm{cl}\Upsilon(Z; \theta)$ must imply $0 \in \overline{\mathrm{co}}\mathbb{E}_I \mathrm{cl}\Upsilon(Z; \theta)$. Hence, there exists some $v \in \mathrm{co}\mathbb{E}_I \mathrm{cl}\Upsilon(Z; \theta)$ such that $\|v\| \leq \epsilon$. By Carathéodory's theorem, there must exists $p_0, p_1, ..., p_{\dim(r)} \in [0, 1]$ and $v_0, ..., v_{\dim(r)} \in \mathbb{E}_I \mathrm{cl}\Upsilon(Z; \theta)$ such that $\sum_{j=0}^{\dim(r)} p_j = 1$ and $v = \sum_{j=0}^{\dim(r)} p_j v_j$.For each $j = 0, ..., \dim(r)$, there exists $f_j \in S^1(\mathrm{cl}\Upsilon(Z; \theta))$ such that $v_j = \mathbb{E} f_j(Z)$. Hence,

$$
\left\| \sum_{j=0}^{\dim(r)} p_j \mathbb{E} f_j(Z) \right\| \leq \epsilon.
$$

By the definition of $S^1(\mathrm{cl}\Upsilon(Z; \theta))$, each f_j is measure and integrable in $(\mathcal{Z}, \mathscr{Z}, F)$.

Let T be a random variable independent with Z, which is supported on $\{0, 1, ..., \dim(r)\}$

and is distributed as the following,

$$\mathbb{P}(T=j)=p_j,\ \forall j\in\{0,1,...,\dim(r)\}.$$

Construct random variable $R\in\mathbb{R}^{\dim(r)}$ from T and Z as

$$R=\sum_{j=0}^{\dim(r)}\mathbb{1}\{T=j\}f_j(Z).$$

Let H' denote the joint distribution of (Z,R) in measurable space $(\mathcal{Z}\times\mathbb{R}^{\dim(r)},\mathscr{B}_{\mathcal{Z}\times\mathbb{R}^{\dim(r)}})$. By construction, H''s marginal distribution for Z equals F, and

$$\mathbb{P}_{H'}(R\in \mathrm{cl}\Upsilon(Z;\theta))=1.$$

Also,

$$\|\mathbb{E}_{H'}R\|=\left\|\int\mathbb{E}_{H'}[R|Z=z]\mathrm{d}F_Z\right\|=\left\|\mathbb{E}\sum_{j=0}^{\dim(r)}p_jf_j(Z)\right\|=\left\|\sum_{j=0}^{\dim(r)}p_j\mathbb{E}f_j(Z)\right\|\leq\epsilon.$$

Now consider H' as in the completion of probability space $(\mathcal{Z}\times\mathbb{R}^{\dim(r)},\mathscr{B}_{\mathcal{Z}\times\mathbb{R}^{\dim(r)}},H')$. Since $\mathbb{P}_{H'}(R\in \mathrm{cl}\Upsilon(Z;\theta))=1$, the definition of $\Upsilon(Z;\theta)$ implies

$$\mathbb{P}_{H'}\Big(\inf_{u\in\Gamma(Z;\theta)}\|r(u,Z;\theta)-R\|=0\Big)=1$$

Since $\{(z,u):u\in\Gamma(z;\theta)\}\times\mathbb{R}^{\dim(r)}$ is a Borel set, and that $(u,z,t)\mapsto\|r(u,z;\theta)-t\|$ is a Borel measurable function in $\mathcal{U}\times\mathcal{Z}\times\mathbb{R}^{\dim(r)}$, Lemma C.1 in Appendix C implies that there exists a universally measurable function $g:\mathcal{Z}\times\mathbb{R}^{\dim(r)}\mapsto\mathcal{U}$, such that for any $t\in\mathbb{R}^{\dim(r)}$ and any $z\in\mathcal{Z}$, $g(z,t)\in\Gamma(z;\theta)$ and

$$\left\|r\big(g(z,t),z\big)-t\right\|\leq\epsilon+\inf_{u\in\Gamma(z;\theta)}\|r(u,z;\theta)-t\|\,.$$

Construct random variable $U=g(Z,R)$. Let H be the joint distribution of (U,Z) in the measurable space $(\mathcal{U}\times\mathcal{Z},\mathscr{B}_{\mathcal{U}\times\mathcal{Z}})$. Then, $\mathbb{P}_H(U\in\Gamma(Z;\theta))=1$ and

$$\mathbb{P}_H(\|r(U,Z;\theta)-R\|\leq\epsilon)=1,$$

so that

$$\|\mathbb{E}_Hr(U,Z;\theta)\|\leq\epsilon+\|\mathbb{E}_HR\|\leq 2\epsilon$$

This completes the proof that $\theta\in\Theta_I'$. □

D.3 Proof of Theorem 2

Before the main proof, I need an extra lemma, the proof of which is presented after the proof of Theorem 2.

Lemma D.4. *Suppose Assumption 1 and 2 hold. Then, for any $F \in \mathcal{F}$, $0 \in \overline{\text{co}}\mathbb{E}_F \text{cl}\Upsilon(Z;\theta)$ implies $\theta \in \Theta_I(F)$.*

Proof of Theorem 2. Fix an arbitrary F in $\mathcal{F}$. Because I have shown in Theorem 1 that $\Theta_I'(F) = \widetilde{\Theta}(F)$, and because $\Theta_I(F) \subseteq \Theta_I'(F)$, I only need to prove $\widetilde{\Theta}(F) \subseteq \Theta_I(F)$. To show $\widetilde{\Theta}(F) \subseteq \Theta_I(F)$, suppose, for the purpose of contradiction, there exists some $\theta \in \widetilde{\Theta}(F)$ such that $\theta \notin \Theta_I(F)$. Then, by Lemma D.4, $0 \notin \overline{\text{co}}\mathbb{E}_F \text{cl}\Upsilon(Z;\theta)$. Yet, as shown in the proof of Theorem 1, this contradicts the fact that $\theta \in \widetilde{\Theta}(F)$. □

Proof of Lemma D.4. Fix an arbitrary $F \in \mathcal{F}$. In the following proof, I will abbreviate $\mathbb{E}_F$ as $\mathbb{E}$, and $\Theta_I(F)$ as Θ_I. Recall also that $\mathbb{E}_I$ stands for the Aumann integral.

The proof of this lemma is similar to that of Lemma D.3. One only needs to notice that under Assumption 1 and 2, $0 \in \overline{\text{co}}\mathbb{E}\text{cl}\Upsilon(Z;\theta)$ not only implies $0 \in \overline{\text{co}}\mathbb{E}_I \text{cl}\Upsilon(Z;\theta)$ but also implies $0 \in \text{co}\mathbb{E}_I \Upsilon(Z;\theta)$. For clarity, I provide the entire proof.

Suppose $0 \in \overline{\text{co}}\mathbb{E}\text{cl}\Upsilon(Z;\theta)$, I want to show $\theta \in \Theta_I$. First of all, note that $0 \in \overline{\text{co}}\mathbb{E}\text{cl}\Upsilon(Z;\theta)$ is equivalent to $0 \in \overline{\text{co}}\mathbb{E}\Upsilon(Z;\theta)$ under Assumption 2(i). Moreover, Assumption 2(ii) together with Lemma D.1 also implies $\Upsilon(Z;\theta)$ is an integrably bounded random closed set. By Lemma B.1(ii), $\mathbb{E}\Upsilon(Z;\theta)$ is a compact set and $\mathbb{E}\Upsilon(Z;\theta) = \mathbb{E}_I\Upsilon(Z;\theta)$. Since $\mathbb{E}\Upsilon(Z;\theta) \subseteq \mathbb{R}^{\dim(r)}$, Carathéodory's theorem implies $\text{co}\mathbb{E}\Upsilon(Z;\theta)$ is also compact. Hence, $0 \in \overline{\text{co}}\mathbb{E}\text{cl}\Upsilon(Z;\theta)$ implies $0 \in \text{co}\mathbb{E}_I\Upsilon(Z;\theta)$.

Given $0 \in \text{co}\mathbb{E}_I\Upsilon(Z;\theta)$, Carathéodory's theorem also implies that there must exists $p_0, p_1, ..., p_{\dim(r)} \in [0,1]$ and $v_0, ..., v_{\dim(r)} \in \mathbb{E}_I\Upsilon(Z;\theta)$ such that $\sum_{j=0}^{\dim(r)} p_j = 1$ and $\sum_{j=0}^{\dim(r)} p_j v_j = 0$.

For each $j = 0, ..., \dim(r)$, there exists $f_j \in S^1(\Upsilon(Z;\theta))$ such that $v_j = \mathbb{E}f_j(Z)$. Hence,

$$\sum_{j=0}^{\dim(r)} p_j \mathbb{E} f_j(Z) = 0.$$

Recall that $(\mathcal{Z}, \mathscr{Z}, F)$ denotes the completion of Borel probability space $(\mathcal{Z}, \mathscr{B}_{\mathcal{Z}}, F)$. By the definition of $S^1(\Upsilon(Z;\theta))$, each f_j is measure and integrable in $(\mathcal{Z}, \mathscr{Z}, F)$.

The remainder of the proof is similar to that in Lemma D.3. Let T be a random variable independent of Z, which is supported on $\{0, 1, ..., \dim(r)\}$ and is distributed as the following,

$$\mathbb{P}(T = j) = p_j, \ \forall j \in \{0, 1, ..., \dim(r)\}.$$

Construct random variable $R \in \mathbb{R}^{\dim(r)}$ from T and Z as

$$R = \sum_{j=0}^{\dim(r)} \mathbb{1}\{T = j\} f_j(Z)$$

Let H' denote the joint distribution of (Z, R) in measurable space $(\mathcal{Z} \times \mathbb{R}^{\dim(r)}, \mathscr{B}_{\mathcal{Z}\times\mathbb{R}^{\dim(r)}})$. By construction, H''s marginal distribution for Z equals F_Z, and

$$\mathbb{P}_{H'}(R \in \Upsilon(Z;\theta)) = 1,$$

and

$$\mathbb{E}_{H'} R = \int \mathbb{E}_{H'}[R|Z = z] \mathrm{d}F_Z(z) = \mathbb{E} \sum_{j=0}^{\dim(r)} p_j f_j(Z) = \sum_{j=0}^{\dim(r)} p_j \mathbb{E} f_j(Z) = 0.$$

Now consider H' as in the completion of probability space $(\mathcal{Z} \times \mathbb{R}^{\dim(r)}, \mathscr{B}_{\mathcal{Z}\times\mathbb{R}^{\dim(r)}}, H')$. Since $\mathbb{P}_{H'}(R \in \Upsilon(Z;\theta)) = 1$, the definition of $\Upsilon(Z;\theta)$ implies

$$\mathbb{P}_H\Big(\min_{u\in\Gamma(Z;\theta)} \|r(u, Z;\theta) - R\| = 0\Big) = 1.$$

Since $\{(z,u) : u \in \Gamma(z;\theta)\} \times \mathbb{R}^{\dim(r)}$ is a Borel set, and $(u, z, t) \mapsto \|r(u, z;\theta) - t\|$ is a Borel measurable function in $\mathcal{U} \times \mathcal{Z} \times \mathbb{R}^{\dim(r)}$, Lemma C.1 in Appendix C implies that there exists a universally measurable function $g : \mathcal{Z} \times \mathbb{R}^{\dim(r)} \mapsto \mathcal{U}$, such that, for any $z \in \mathcal{Z}$ and $t \in \mathbb{R}^{\dim(r)}$, $g(z,t) \in \Gamma(z;\theta)$. In addition, for any $z \in \mathcal{Z}$ and $t \in \mathbb{R}^{\dim(r)}$ which satisfies

$$\inf_{u\in\Gamma(Z;\theta)} \|r(u, z;\theta) - t\| = \min_{u\in\Gamma(z;\theta)} \|r(u, z;\theta) - t\|,$$

it must be true that

$$\|r(g(z,t), z) - t\| = \min_{u\in\Gamma(z;\theta)} \|r(u, z;\theta) - t\|.$$

Construct random variable $U = g(Z, R)$. Let H be the joint distribution of (U, Z) in the measurable space $(\mathcal{U} \times \mathcal{Z}, \mathscr{B}_{\mathcal{U}\times\mathcal{Z}})$. Then, $\mathbb{P}_H(U \in \Gamma(Z;\theta)) = 1$ and

$$\mathbb{P}_H(r(U, Z;\theta) = R) = 1,$$

so that

$$\mathbb{E}_H r(U, Z;\theta) = \mathbb{E}_H R = 0$$

This completes the proof that $\theta \in \Theta_I$. □

D.4 Proof of Lemma 1

The proof of this result builds on the following lemmas, whose proofs will be presented later.

Lemma D.5. *Suppose set A is a nonempty closed convex set in $\mathbb{R}^d$. Let $\text{int}A$ denote the interior of A. Then, $x \in \text{int}A$ if and only if*

$$\inf_{\lambda\in\mathbb{R}^d:\|\lambda\|=1} \sup\{\lambda'(t-x) : t \in A\} > 0. \tag{41}$$

Note that (41) includes the case that $\inf_{\lambda\in\mathbb{R}^d:\|\lambda\|=1} \sup\{\lambda'(t-x) : t \in A\} = +\infty$ which could happen when $A = \mathbb{R}^d$.

Lemma D.6. *Suppose set A is a nonempty set in $\mathbb{R}^d$. Suppose $x \in \text{int}(\text{co}A)$, there there exists some $\epsilon > 0$, a positive integer $K > 0$ and $a_1, ..., a_K \in A$, such that, $x \in \text{int}(\text{co}\{a'_1, ..., a'_K\})$ for any $a'_1, ..., a'_K$ with $\|a_i - a'_i\| < \epsilon$ for $i = 1, ..., K$.*

Lemma D.7. *Suppose Assumption 1 hold. Then, for any $F \in \mathcal{F}$, $0 \in \text{int}(\overline{\text{co}}\mathbb{E}_F\text{cl}\Upsilon(Z;\theta))$ implies $\theta \in \Theta_I(F)$.*

Proof of Lemma 1. Fix an arbitrary $F \in \mathcal{F}$. In the following of the proof, I will abbreviate $\mathbb{E}_F$ as $\mathbb{E}$, and $\Theta_I(F)$ as Θ_I. Recall also that $\mathbb{E}_I$ stands for the Aumann integral. Lemma D.1 implies that $\text{cl}\Upsilon(\cdot;\theta)$ is an integrable random closed set. The proof will be conducted in three steps.

Step 1: Lemma 1(i) holds when $\dim(r_1) = 0$.

Suppose $\dim(r_1) = 0$. I need to prove that $\inf_{\lambda\in\mathcal{S}} \mathbb{E}_F\gamma(\lambda, Z;\theta) > 0$ implies $\theta \in \Theta_I$ in this step. Suppose $\inf_{\lambda\in\mathcal{S}} \mathbb{E}_F\gamma(\lambda, Z;\theta) > 0$. I'm going to prove $\theta \in \Theta_I$ by contradiction.

Suppose, for the purpose of contradiction, that $\theta \notin \Theta_I$. Then, Lemma D.5 and D.7 implies that

$$\inf_{\lambda\in\mathcal{S}} \sup\{\lambda't : t \in \overline{\text{co}}\mathbb{E}\text{cl}\Upsilon(Z;\theta)\} \leq 0.$$

By Lemma B.1(iv), and the fact that $\overline{\text{co}}\Upsilon(Z;\theta) \subseteq \overline{\text{co}}\text{cl}\Upsilon(Z;\theta)$, and that $\mathbb{E}_I\overline{\text{co}}\Upsilon(Z;\theta) \subseteq \mathbb{E}\overline{\text{co}}\Upsilon(Z;\theta)$, we know

$$\inf_{\lambda\in\mathcal{S}} \sup\{\lambda't : t \in \mathbb{E}_I\overline{\text{co}}\Upsilon(Z;\theta)\} \leq 0 \tag{42}$$

Since $\sup\{\lambda't : t \in \mathbb{E}_I\overline{\text{co}}\Upsilon(Z;\theta)\}$ is a lower semi-continuous function of λ and $\mathcal{S}$ is compact, there exists some $\tilde{\lambda}$ such that $\sup\{\tilde{\lambda}'t : t \in \mathbb{E}_I\overline{\text{co}}\Upsilon(Z;\theta)\} \leq 0$. Note that

$$\sup\{\tilde{\lambda}'t : t \in \mathbb{E}_I\overline{\text{co}}\Upsilon(Z;\theta)\} = -\inf_{f\in S^1(\overline{\text{co}}\Upsilon(Z;\theta))} \mathbb{E}[-\tilde{\lambda}'f] \tag{43}$$

where S^1 is defined in Definition B.2. Apply Lemma B.1(iii) with $\zeta(t) = -\lambda't$ to get

$$\begin{aligned} & -\inf_{f\in S^1(\overline{\text{co}}\Upsilon(Z;\theta))} \mathbb{E}[-\tilde{\lambda}'f] \\ = \; & -\mathbb{E}\inf\{-\tilde{\lambda}'t : t \in \overline{\text{co}}\Upsilon(Z;\theta)\} \\ = \; & \mathbb{E}\sup\{\tilde{\lambda}'t : t \in \overline{\text{co}}\Upsilon(Z;\theta)\}. \end{aligned} \tag{44}$$

Equation (43) and (44) imply

$$\mathbb{E}\sup\{\tilde{\lambda}'t : t \in \overline{\text{co}}\Upsilon(Z;\theta)\} = \sup\{\tilde{\lambda}'t : t \in \mathbb{E}_I\overline{\text{co}}\Upsilon(Z;\theta)\} \leq 0. \tag{45}$$

In addition, since $\Upsilon(z;\theta)$ is a subset of the Euclidean space,

$$\sup\{\tilde{\lambda}'t : t \in \overline{\text{co}}\Upsilon(z;\theta)\} = \sup\{\tilde{\lambda}'t : t \in \Upsilon(z;\theta)\}. \tag{46}$$

Equation (45) and (46) then imply

$$\inf_{\lambda\in\mathbb{R}^{\dim(r)}} \mathbb{E}\sup\{\lambda't : t \in \Upsilon(Z;\theta)\} \leq 0.$$

This contradicts the fact that θ satisfy $\inf_{\lambda\in\mathcal{S}} \mathbb{E}_F\gamma(\lambda, Z;\theta) > 0$.

Step 2: Lemma 1(i) holds when $\dim(r_1) > 0$.

Recall that $\mathcal{H}(\theta, F)$ is defined as the set of all joint distributions H for (U, Z) which satisfy that $\mathbb{P}_H[(U, Z) \in \Gamma(\theta)] = 1$ and that H's marginal distribution for Z equals F.

- When $\dim(r_2) = 0$, for any $H \in \mathcal{H}(\theta, F)$, we have $\mathbb{E}_H[r(U, Z;\theta)] = \mathbb{E}_F[r_1(Z;\theta)]$. Therefore, (2) is equivalent to $\mathbb{E}_F[r_1(Z;\theta)] = 0$. Hence, $\mathbb{E}_F r_1(Z;\theta) = 0$ implies $\theta \in \Theta_I(F;\Gamma, r_1) = \Theta_I(F;\Gamma, r)$ by Definition 1.
- When $\dim(r_2) > 0$, note that (2) is equivalent to the following condition:

  $$\mathbb{E}_F[r_1(Z;\theta)] = 0 \text{ and } \inf_{H\in\mathcal{H}(\theta,F)} \|\mathbb{E}_H[r_2(U, Z;\theta)]\| = 0.$$

  which implies that $\Theta_I(F;\Gamma, r) = \Theta_I(F;\Gamma, r_1) \cap \Theta_I(F;\Gamma, r_2)$ by Definition 1. Following the same proof in the previous paragraph, we know that $\mathbb{E}_F[r_1(Z;\theta)] = 0$ implies $\theta \in \Theta_I(F;\Gamma, r_1)$. Following the same proof in Step 1, we know that $\inf_{\lambda\in\mathcal{S}_2} \mathbb{E}_F\gamma_2(\lambda, Z;\theta) > 0$ implies $\theta \in \Theta_I(F;\Gamma, r_2)$. As a result, $\mathbb{E}_F[r_1(Z;\theta)] = 0$ and $\inf_{\lambda\in\mathcal{S}_2} \mathbb{E}_F\gamma_2(\lambda, Z;\theta) > 0$ implies $\theta \in \Theta_I(F;\Gamma, r)$.

Step 1 and 2 completes the proof for Lemma 1(i).

Step 3: Lemma 1(ii) holds.

Suppose $\theta \in \Theta'_I(F)\backslash\Theta_I(F)$. Because $\theta \in \Theta'_I(F)$, we know

$$\mathbb{E}_F[r_1(Z;\theta)] = 0 \text{ and } \inf_{\lambda\in\mathcal{S}_2} \mathbb{E}_F\gamma_2(\lambda, Z;\theta) \geq 0.$$

Moreover, $\theta \notin \Theta_I(F)$ implies that there is no

$$\mathbb{E}_F[r_1(Z;\theta)] = 0 \text{ and } \inf_{\lambda\in\mathcal{S}_2} \mathbb{E}_F\gamma_2(\lambda, Z;\theta) > 0.$$

Hence, we must have $\mathbb{E}_F[r_1(Z;\theta)] = 0$ and $\inf_{\lambda\in\mathcal{S}_2} \mathbb{E}_F\gamma_2(\lambda, Z;\theta) = 0$. □

Proof of Lemma D.5. By Theorem 2.2.3 (on page 138) in Hiriart-Urruty and Lemaréchal (2001), we know that $x \in \text{int}A$ if and only if for any λ with $\|\lambda\| = 1$, $\sup\{\lambda'(t-x) : t \in A\} > 0$.

Therefore, I only need to show that $\sup\{\lambda'(t-x) : t \in A\} > 0$ for any λ with $\|\lambda\| = 1$ if and only if

$$\inf_{\lambda \in \mathbb{R}^d : \|\lambda\|=1} \sup\{\lambda'(t-x) : t \in A\} > 0.$$

The "if" part of this claim follows from the definition of inf. To show the "only if" part of this claim, note that $\sup\{\lambda'(t-x) : t \in A\}$ is a lower semi-continuous function of λ and that $\{\lambda \in \mathbb{R}^d : \|\lambda\| = 1\}$ is a compact set. Note also that there must be $\inf_{\lambda \in \mathbb{R}^d : \|\lambda\|=1} \sup\{\lambda'(t-x) : t \in A\} >= 0$. Therefore, if $\inf_{\lambda \in \mathbb{R}^d : \|\lambda\|=1} \sup\{\lambda'(t-x) : t \in A\} < +\infty$, this infimum is achieved by some λ with $\|\lambda\| = 1$ so that $\inf_{\lambda \in \mathbb{R}^d : \|\lambda\|=1} \sup\{\lambda'(t-x) : t \in A\} > 0$. If $\inf_{\lambda \in \mathbb{R}^d : \|\lambda\|=1} \sup\{\lambda'(t-x) : t \in A\} = +\infty$, then we automatically have $\inf_{\lambda \in \mathbb{R}^d : \|\lambda\|=1} \sup\{\lambda'(t-x) : t \in A\} > 0$. □

Proof of Lemma D.6. By Gustin (1947), there exists some $a_1, ..., a_K \in A$ such that x is in the interior of $\text{co}\{a_1, ..., a_K\}$ and $K \leq 2d$. By Lemma D.5, we know that $\lambda'(a_i - x) > 0$ for all λ with $\|\lambda\| = 1$ and for all $i = 1, ..., K$. Therefore, there exists some $\epsilon > 0$ such that for any $i = 1, ..., K$ and for any a_i' with $\|a_i' - a_i\| < \epsilon$, we have $\lambda'(a_i' - x) > 0$. By Lemma D.5, this is equivalent to that $x \in \text{int}(\text{co}\{a_1', ..., a_K'\})$ for any $a_1', ..., a_K'$ with $\|a_i - a_i'\| < \epsilon$ for $i = 1, ..., K$. □

Proof of Lemma D.7. Fix an arbitrary $F \in \mathcal{F}$. In the following, I will abbreviate $\mathbb{E}_F$ as $\mathbb{E}$, and $\Theta_I(F)$ as Θ_I. Recall that $\mathbb{E}_I$ stands for the Aumann integral. Recall also that the probability space $(\mathcal{Z}, \mathscr{X}, F)$ denotes the completion of Borel probability space $(\mathcal{Z}, \mathscr{B}_Z, F)$.

Under Assumption 1, $\text{cl}\Upsilon(Z;\theta)$ is an integrable random closed set in $(\mathcal{Z}, \mathscr{X}, F_Z)$. Suppose $0 \in \text{int}(\overline{\text{co}}\mathbb{E}\text{cl}\Upsilon(Z;\theta))$ is true, I want to prove that $\theta \in \Theta_I$.

Because $\overline{\text{co}}A = \overline{\text{co}}\text{cl}A$ for any subset A in an Euclidean space, and because $\mathbb{E}\text{cl}\Upsilon(Z;\theta) = \text{cl}(\mathbb{E}_I\text{cl}\Upsilon(Z;\theta))$ by Definition B.4, $0 \in \text{int}(\overline{\text{co}}\mathbb{E}\text{cl}\Upsilon(Z;\theta))$ imply $0 \in \text{int}(\overline{\text{co}}\mathbb{E}_I\text{cl}\Upsilon(Z;\theta))$. Furthermore, because Proposition 2.1.8 in Hiriart-Urruty and Lemaréchal (2001) implies that $\text{int}(\overline{\text{co}}A) = \text{int}(\text{co}A)$ for any subset A in an Euclidean space, $0 \in \text{int}(\overline{\text{co}}\mathbb{E}_I\text{cl}\Upsilon(Z;\theta))$ implies that $0 \in \text{int}(\text{co}\mathbb{E}_I\text{cl}\Upsilon(Z;\theta))$. By Lemma D.6, we know there exists some $\epsilon > 0$, some positive integer K and some $v_1, ..., v_K \in \mathbb{E}_I\text{cl}\Upsilon(Z;\theta)$ such that $0 \in \text{int}(\text{co}\{\tilde{v}_1, ..., \tilde{v}_K\})$ for any $(\tilde{v}_1, ..., \tilde{v}_K)$ with $\|\tilde{v}_i - v_i\| < \epsilon$ for any $i = 1, ..., K$.

For any $k = 1, ..., K$. Because $v_k \in \mathbb{E}_I\text{cl}\Upsilon(Z;\theta)$, there exists $f_k \in S^1(\text{cl}\Upsilon(Z;\theta))$ such that $v_k = \mathbb{E}f_k(Z)$. Because every measurable function in $(\mathcal{Z}, \mathscr{X}, F_Z)$ can be well approximated by a Borel measurable function, there exists some Borel function $\tilde{f}_k$ such that $\mathbb{P}(f_k(Z) = \tilde{f}_k(Z)) = 1$. Therefore, we know $\mathbb{E}\tilde{f}_k(Z) = v_k$ and

$$\mathbb{P}\left(\inf_{u \in \Gamma(Z;\theta)} \left\| r(u, Z;\theta) - \tilde{f}_k(Z) \right\| = 0\right) = 1.$$

Since $\{(z, u) : u \in \Gamma(z;\theta)\} \times \mathbb{R}^{\dim(r)}$ is a Borel set, and that $(u, z) \mapsto \left\| r(u, z;\theta) - \tilde{f}_k(z) \right\|$ is a Borel measurable function, Lemma C.1 in Appendix C implies that there exists a universally measurable function $g : \mathcal{Z} \mapsto \mathcal{U}$, such that for almost every $z \in \mathcal{Z}$, $g_k(z) \in \Gamma(z;\theta)$ and

$$\left\| r\big(g_k(z), z\big) - \tilde{f}_k(z) \right\| \leq \epsilon + \inf_{u \in \Gamma(z;\theta)} \left\| r(u, z; \theta) - \tilde{f}_k(z) \right\|.$$

By the construction of g_k, $\|v_k - \mathbb{E}r(g_k(Z), Z)\| < \epsilon$.

As a result, I have shown that there exists function $g_1, ..., g_K$ in $(\mathcal{Z}, \mathscr{Z}, F_Z)$ such that $\mathbb{P}(g_k(Z) \in \Gamma(Z;\theta)) = 1$ for each $k = 1, ..., K$ and $0 \in \text{co}\{\mathbb{E}r(g_1(Z), Z), ..., \mathbb{E}r(g_K(Z), Z)\}$. This implies that there exists a joint distribution H for (U, Z) such that (i) H's marginal distribution for Z is F_Z, (ii) $\mathbb{P}_H (U \in \Gamma(Z;\theta)) = 1$ and (iii) $\mathbb{E}_H r(U, Z; \theta) = 0$. Hence, $\theta \in \Theta_I$. □

E Proof of Theorem 3

Define set $\mathcal{F}_\theta^*$ as

$$\mathcal{F}_\theta^* := \left\{ F \in \mathcal{F} : \mathbb{E}_F[r_1(Z;\theta)] = 0 \text{ and } \inf_{\lambda \in \mathcal{S}_2} \mathbb{E}_F \gamma_2(\lambda, Z; \theta) > 0. \right\}$$

Since (15) holds, we know $\mathcal{F}_\theta^*$ is nonempty. By Lemma 1, $\mathcal{F}_\theta^* \subseteq \mathcal{F}_\theta \subseteq \mathcal{F}_\theta'$. Hence, both $\sup_{F \in \mathcal{F}_\theta} \mathbb{E}_F \phi_n$ and $\sup_{F \in \mathcal{F}_\theta'} \mathbb{E}_F \phi_n$ are well defined and finite. $\mathcal{F}_\theta^* \subseteq \mathcal{F}_\theta \subseteq \mathcal{F}_\theta'$ also implies that $\sup_{F \in \mathcal{F}_\theta^*} \mathbb{E}_F \phi_n \leq \sup_{F \in \mathcal{F}_\theta} \mathbb{E}_F \phi_n \leq \sup_{F \in \mathcal{F}_\theta'} \mathbb{E}_F \phi_n$. Therefore, to show the desired result, we only need to show that for any $F \in \mathcal{F}_\theta'$, $\mathbb{E}_F \phi_n \leq \sup_{F \in \mathcal{F}_\theta^*} \mathbb{E}_F \phi_n$.

For each $F \in \mathcal{F}$, define Define $\psi(F) = \inf_{\lambda \in \mathcal{S}_2} \mathbb{E}_F \gamma_2(\lambda, Z; \theta)$. For any $F_1, F_2 \in \mathcal{F}$, let $F_\delta = \delta F_1 + (1-\delta) F_2$ for any $\delta \in [0, 1]$. Then,

$$\begin{aligned}
\psi(F_\delta) &= \inf_{\lambda \in \mathcal{S}_2} \Big(\delta \mathbb{E}_{F_1} \gamma_2(\lambda, Z; \theta) + (1-\delta) \mathbb{E}_{F_2} \gamma_2(\lambda, Z; \theta) \Big) \\
&\geq \inf_{\lambda \in \mathcal{S}_2} \delta \mathbb{E}_{F_1} \gamma_2(\lambda, Z; \theta) + \inf_{\lambda \in \mathcal{S}_2} (1-\delta) \mathbb{E}_{F_2} \gamma_2(\lambda, Z; \theta) \\
&= \delta \psi(F_1) + (1-\delta) \psi(F_2)
\end{aligned}$$

Therefore, ψ is a concave function.

Now, fix an arbitrary $F \in \mathcal{F}_\theta'$. For any $F^* \in \mathcal{F}^*$ and any $k \geq 1$, define $F_k := (1 - \frac{1}{k}) F + \frac{1}{k} F^*$. Since $F \in \mathcal{F}_\theta'$, $\mathbb{E}_F[r_1(Z;\theta)] = 0$ and $\inf_{\lambda \in \mathcal{S}_2} \mathbb{E}_F \gamma_2(\lambda, Z; \theta) \geq 0$. Therefore, the concavity of ψ implies that $F_k \in \mathcal{F}_\theta^*$ for all $k \geq 1$. Since $\mathbb{E}_F \phi_n = \lim_{k \to \infty} \mathbb{E}_{F_k} \phi_n$, we know

$$\mathbb{E}_F \phi_n \leq \sup_{k \geq 1} \mathbb{E}_{F_k} \phi_n \leq \sup_{F' \in \mathcal{F}^*} \mathbb{E}_{F'} \phi_n.$$

This completes the proof.

F Proof of Theorem 4

I first prove the second result, then prove the first result.

The second result in the theorem follows immediately from the following lemma, which will be proved at the end of this section.

Lemma F.1. *Suppose Assumption 1 hold and $\mathcal{F}$ is convex. Let θ be an arbitrary parameter in Θ and partition $r(u,z;\theta) = (r_1(z;\theta), r_2(u,z;\theta)$. Suppose $\mathcal{F}'_\theta$ is nonempty, and that the inequality (15) fails to hold for all $F \in \mathcal{F}$. Then,*

(i) there exists some $\tilde{\lambda} \in \mathcal{S}_2$ such that $\mathbb{E}_F \gamma_2(\tilde{\lambda}, Z; \theta) = 0$ for all $F \in \mathcal{F}'_\theta$.

(ii) model (Γ, r) is reducible at θ. In particular, for any $\lambda_2, ..., \lambda_{\dim(r_2)}$ such that $\tilde{\lambda}, \lambda_2, ..., \lambda_{\dim(r_2)}$ are linearly independent, define reduced model $(\tilde{\Gamma}, \tilde{r})$ as

$$\begin{aligned} \tilde{\Gamma}(\theta) &= \left\{(u,z) \in \Gamma(\theta) : u \in \arg\max_{u \in \Gamma(z;\theta)} \tilde{\lambda}' r_2(u,z;\theta)\right\}, \\ \tilde{r}(u,z;\theta) &= \begin{pmatrix} r_1(Z;\theta) \\ \gamma_2(\tilde{\lambda}, z;\theta) \\ \lambda_2' r(u,z;\theta) \\ \lambda_3' r(u,z;\theta) \\ \vdots \\ \lambda'_{\dim(r)} r(u,z;\theta) \end{pmatrix}. \end{aligned}$$

Then, for any $F \in \mathcal{F}$, $\theta \in \Theta_I(F; \Gamma, r)$ if and only if $\theta \in \Theta_I(F; \tilde{\Gamma}, \tilde{\gamma})$.

To prove the first result in the theorem, suppose the model (Γ, r) is irreducible at θ. Consider the following cases:

- When $\mathcal{F}'_\theta$ is empty, $\mathcal{F}_\theta$ is also empty so that both $\theta \in \Theta_I(F)$ and $\theta \in \Theta'_I(F)$ are false for any $F \in \mathcal{F}$, which implies that $\theta \in \Theta'_I$ and $\theta \in \Theta_I$ cannot be distinguished in finite samples.
- When $\mathcal{F}'_\theta$ is nonempty, the second result of this theorem implies that there exists some $F \in \mathcal{F}$ which satisfies (15). Theorem 3 then implies that $\theta \in \Theta'_I$ and $\theta \in \Theta_I$ cannot be distinguished in finite samples.

Since $\theta \in \Theta'_I$ and $\theta \in \Theta_I$ are indistinguishable in both cases, the proof is now complete.

Proof of Lemma F.1. Fix θ to be an arbitrary parameter with which (15) does not hold for all $F \in \mathcal{F}$. The proof will be divided into two parts: Part 1 deals with the first part of the result and Part 2 deals with the second part of the result.

Part 1 First of all, the fact that $\mathcal{F}'_\theta$ is nonempty and (15) fails to hold for all $F \in \mathcal{F}$ implies that $\dim(r_2) > 0$. For any $F \in \mathcal{F}'_\theta$, Theorem 1 implies that $\inf_{\lambda \in \mathcal{S}} \mathbb{E}_F \gamma(\lambda, Z; \theta) \geq 0$, which is equivalent to

$$\mathbb{E}_F[r_1(Z;\theta)] = 0 \text{ and } \inf_{\lambda\in\mathcal{S}_2} \mathbb{E}_F\gamma_2(\lambda, Z;\theta) \geq 0.$$

Because (15) fails to hold for all $F \in \mathcal{F}$, we know that for any $F \in \mathcal{F}'_\theta$, $\inf_{\lambda\in\mathcal{S}_2} \mathbb{E}_F\gamma_2(\lambda, Z;\theta) = 0$. Since $\gamma_2(\lambda, Z;\theta)$ is lower semi-continuous in λ and $\mathcal{S}_2$ is a compact set, we know that for each $F \in \mathcal{F}'_\theta$, there exists some $\lambda \in \mathcal{S}_2$ such that $\mathbb{E}_F\gamma_2(\lambda, Z;\theta) = 0$. For each $F \in \mathcal{F}'_\theta$, define $\Lambda(F) \coloneqq \{\lambda \in \mathcal{S}_2 : \mathbb{E}_F\gamma_2(\lambda, Z;\theta) = 0\}$. Then, for each $F \in \mathcal{F}'_\theta$, $\Lambda(F)$ is nonempty. To show the first result of Lemma F.1, I only need to show that there exists some $F^* \in \mathcal{F}'_\theta$ such that $\cap_{F\in\mathcal{F}'_\theta}\Lambda(F) = \Lambda(F^*)$. When $\mathcal{F}'_\theta$ only contains one element F^*, it's trivially true that $\cap_{F\in\mathcal{F}'_\theta}\Lambda(F) = \Lambda(F^*)$. So, I suppose $\mathcal{F}'_\theta$ contains at least two elements in the remaining of the proof in this part.

Note that $\mathcal{F}'_\theta$ is a convex set because $\mathcal{F}$ is convex and $\mathcal{F}'_\theta = \{F \in \mathcal{F} : \mathbb{E}_F\gamma(\lambda, Z;\theta) \geq 0, \forall\lambda \in \mathcal{S}\}$. The relative interior $\text{ri}\mathcal{F}'_\theta$ defined as $\text{ri}\mathcal{F}'_\theta \coloneqq \{F \in \mathcal{F}'_\theta : \forall F' \in \mathcal{F}'_\theta,\ \exists\delta > 1 \text{ such that } \delta F + (1-\delta)F' \in \mathcal{F}'_\theta\}$ should contain at least two elements because $\mathcal{F}'_\theta$ contains at least two elements.

To proceed, I claim that for any $F_1, F_2 \in \mathcal{F}'_\theta$ and any $\delta \in (0,1)$, $\Lambda(F_1)\cap\Lambda(F_2) = \Lambda(F_\delta)$ where $F_\delta \coloneqq \delta F_1 + (1-\delta)F_2$. To see why this is true, note that for any $\lambda \in \Lambda(F_1)\cap\Lambda(F_2)$, $\mathbb{E}_{F_\delta}\gamma_2(\lambda, Z;\theta) = \delta\mathbb{E}_{F_1}\gamma_2(\lambda, Z;\theta) + (1-\delta)\mathbb{E}_{F_2}\gamma_2(\lambda, Z;\theta) = 0$. Hence, $\Lambda(F_\delta) \supseteq \Lambda(F_1)\cap\Lambda(F_2)$. Now, for any $\lambda \in \mathcal{S}_2\backslash(\Lambda(F_1)\cap\Lambda(F_2))$, we know the following is true:

- $\mathbb{E}_{F_1}\gamma_2(\lambda, Z;\theta) \geq 0$, because $F_1 \in \mathcal{F}'_\theta$;
- $\mathbb{E}_{F_2}\gamma_2(\lambda, Z;\theta) \geq 0$, because $F_1 \in \mathcal{F}'_\theta$;
- either $\mathbb{E}_{F_1}\gamma_2(\lambda, Z;\theta) > 0$ or $\mathbb{E}_{F_2}\gamma_2(\lambda, Z;\theta) > 0$, because $\lambda \notin \Lambda(F_1)\cap\Lambda(F_2)$.

Therefore, $\mathbb{E}_{F_\delta}\gamma_2(\lambda, Z;\theta) = \delta\mathbb{E}_{F_1}\gamma_2(\lambda, Z;\theta) + (1-\delta)\mathbb{E}_{F_2}\gamma_2(\lambda, Z;\theta) > 0$. Hence, for any $\lambda \in \mathcal{S}_2\backslash(\Lambda(F_1)\cap\Lambda(F_2))$, $\lambda \notin \Lambda(F_\delta)$. Hence, $\Lambda(F_\delta) \subseteq \Lambda(F_1)\cap\Lambda(F_2)$. Combine both results, I conclude that $\Lambda(F_1)\cap\Lambda(F_2) = \Lambda(F_\delta)$ for any $\delta \in (0,1)$.

Next, I claim that for any two F_1, F_2 in $\text{ri}\mathcal{F}'_\theta$, $\Lambda(F_1) = \Lambda(F_2)$. To see why this is true, note that by the definition of $\text{ri}\mathcal{F}'_\theta$, there must exists F_3 and F_4 in $\mathcal{F}'_\theta$ and $\delta_1, \delta_2 \in (0,1)$ such that $F_1 = \delta_1 F_3 + (1-\delta_1)F_4$ and $F_2 = \delta_2 F_3 + (1-\delta_2)F_4$. By the preceding result, we know $\Lambda(F_1) = \Lambda(F_2) = \Lambda(F_3)\cap\Lambda(F_4)$.

Finally, let F^* be an arbitrary element in $\text{ri}\mathcal{F}'_\theta$. I claim that $\cap_{F\in\mathcal{F}'_\theta}\Lambda(F) = \Lambda(F^*)$. To see why this is true, note that for any $F \in \mathcal{F}'_\theta$, there must exists $F' \in \mathcal{F}'_\theta$ with $F' \neq F$ because $\mathcal{F}'_\theta$ is assumed to have at least two elements. Because $\frac{1}{2}F + \frac{1}{2}F' \in \text{ri}\mathcal{F}_\theta$, the claims which I proved in the above paragraphs implies that $\Lambda(F^*) = \Lambda(F)\cap\Lambda(F')$. As a result, $\Lambda(F^*) \subseteq \Lambda(F)$ for all $F \in \mathcal{F}_\theta$. Hence, $\cap_{F\in\mathcal{F}'_\theta}\Lambda(F) = \Lambda(F^*)$.

Part 2 I am going to prove the second part of the result in two steps.

Step 1: $\forall F \in \mathcal{F}$, $\theta \in \Theta_I(F;\tilde{\Gamma},\tilde{r})$ implies that $\theta \in \Theta_I(F;\Gamma,r)$. To prove this result, suppose $\theta \in \Theta_I(F;\tilde{\Gamma},\tilde{r})$ for some $F \in \mathcal{F}$. Because of the definition of $\Theta_I(F;\tilde{\Gamma},\tilde{r})$, there exists some joint distribution H of (U,Z) such that (i) $\mathbb{P}_H((U,Z) \in \tilde{\Gamma}(\theta)) = 1$; (ii) $\mathbb{E}_H r_1(Z;\theta) = 0$,

$\mathbb{E}_H\gamma_2(\tilde{\lambda}, Z; \theta) = 0$ and $\mathbb{E}_H\lambda_i' r_2(U, Z; \theta) = 0$ for $i = 2, ..., \dim(r_2)$; (iii) the marginal distribution of H for Z is F.

Because of the construction of $\tilde{\Gamma}$ in this lemma, and because $\mathbb{P}_H((U, Z) \in \tilde{\Gamma}(\theta)) = 1$, we know that $\mathbb{P}_H(\lambda_1' r(U, Z; \theta) = \gamma_2(\tilde{\lambda}, Z; \theta)) = 1$. Therefore, in addition to $\mathbb{E}_H\lambda_i' r_2(U, Z; \theta) = 0$ for each $i = 2, ..., \dim(r_2)$, we also have $\mathbb{E}_H\tilde{\lambda}' r_2(U, Z; \theta) = 0$. Because $\tilde{\lambda}$, λ_2, ..., $\lambda_{\dim(r_2)}$ are linearly independent, this implies that $\mathbb{E}_H r_2(U, Z; \theta) = 0 \in \mathbb{R}^{\dim(r_2)}$. Moreover, since $\tilde{\Gamma}(\theta) \subseteq \Gamma(\theta)$, $\mathbb{P}_H((U, Z) \in \Gamma(\theta)) = 1$. As a result, $\theta \in \Theta_I(F; \Gamma, r)$.

Step 2: $\forall F \in \mathcal{F}$, $\theta \in \Theta_I(F; \Gamma, r)$ implies that $\theta \in \Theta_I(F; \tilde{\Gamma}, \tilde{r})$. To prove this result, suppose $\theta \in \Theta_I(F; \Gamma, r)$ for some $F \in \mathcal{F}_\theta$. By the definition of $\Theta_I(F; \Gamma, r)$, there exists some joint distribution H of (U, Z) such that (i) $\mathbb{P}_H((U, Z) \in \Gamma(\theta)) = 1$; (ii) $\mathbb{E}_H r_1(Z; \theta) = 0$ and $\mathbb{E}_H r_2(U, Z; \theta) = 0$; and (iii) the marginal distribution of H for Z is F. Note that since $F \in \mathcal{F}_\theta \subseteq \mathcal{F}_\theta'$, we know $\mathbb{E}_F\gamma_2(\tilde{\lambda}, Z; \theta) = 0$ by the construction of $\tilde{\lambda}$.

Define $\phi(u, z; \theta) = \gamma_2(\tilde{\lambda}, z; \theta) - \tilde{\lambda}' r_2(u, z; \theta)$. To show $\theta \in \Theta_I(F; \tilde{\Gamma}, \tilde{r})$, I only need to verify that $\mathbb{P}_H(\phi(U, Z; \theta) = 0) = 1$. By the construction of H, there is $\mathbb{E}_H\tilde{\lambda}' r_2(U, Z; \theta) = 0$. Moreover, by the construction of $\tilde{\lambda}$, there is $\mathbb{E}_F\gamma_2(\tilde{\lambda}, Z; \theta) = 0$ which implies that $\mathbb{E}_H\gamma_2(\tilde{\lambda}, Z; \theta) = 0$. Therefore, we have $\mathbb{E}_H\phi(U, Z; \theta) = 0$. Recall $\gamma_2(\tilde{\lambda}, z; \theta) := \sup_{u \in \Gamma(z;\theta)} \tilde{\lambda}' r_2(u, z; \theta)$. Because $\mathbb{P}_H((U, Z) \in \Gamma(\theta)) = 1$, there is $\mathbb{P}_H(\phi(U, Z; \theta) \geq 0) = 1$. Combine this result with $\mathbb{E}_H\phi(U, Z; \theta) = 0$, it must be true that $\mathbb{P}_H(\phi(U, Z; \theta) = 0) = 1$. This proves the desired result. □

References

Aguiar, V. and N. Kashaev, "Stochastic Revealed Preferences with Measurement Error," *Working Paper*, 2018.

Andrews, D. and G. Soares, "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 2010, *78* (1), 119–157.

_ and X. Shi, "Inference Based on Conditional Moment Inequalities," *Econometrica*, 2013, *81* (2), 609–666.

_ and _ , "Inference Based on Many Conditional Moment Inequalities," *Journal of Econometrics*, 2017, *196* (2), 275–287.

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

Bertsekas, D. P. and S. E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, 1978.

Bontemps, C., T. Magnac, and E. Maurin, "Set Identified Linear Models," *Econometrica*, 2012, *80* (3), 1129–1155.

Chernozhukov, V., D. Chetverikov, and K. Kato, "Inference on Causal and Structural Parameters Using Many Moment Inequalities," *Working Paper*, 2018.

_ , H. Hong, and E. Tamer, "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 2007, *75* (5), 1243–1284.

_ , S. Lee, and A. M. Rosen, "Intersection Bounds: Estimation and Inference," *Econometrica*, 2013, *81* (2), 667 – 737.

Chesher, A. and A. M. Rosen, "Generalized Instrumental Variable Models," *Econometrica*, 2017, *85* (3), 959–989.

_ and _ , "Instrumental Variable Models for Censored Outcomes," 2020.

Chiong, K., Y.-W. Hsieh, and M. Shum, "Counterfactual Estimation in Semiparametric Discrete-choice Models," *Working Paper*, 2017.

Christensen, T. and B. Connault, "Counterfactual Sensitivity and Robustness," *Working Paper*, April 2019.

Ekeland, I., A. Galichon, and M. Henry, "Optimal Transportation and the Falsifiability of Incompletely Specified Economic Models," *Economic Theory*, 2010, *42* (2), 355–374.

Fang, Z., A. Santos, A. Shaikh, and A. Torgovitsky, "Inference for Large-Scale Linear Systems with Known Coefficients," 2020.

Galichon, A. and M. Henry, "Set Identification in Models with Multiple Equilibria," *The Review of Economic Studies*, 2011, *78* (4), 1264–1298.

Gustin, W., "On the interior of the convex hull of a euclidean set," *Bulletin of the American Mathematical Society*, 1947, *53* (4), 299–301.

Hiriart-Urruty, J.-B. and C. Lemaréchal, *Fundamentals of convex analysis*, Springer, 2001.

Manski, C. F., "Partial Identification of Counterfactual Choice Probabilities," *International Economic Review*, 2007, *48* (4), 1393–1410.

Molchanov, I., *Theory of Random Sets*, Springer, 2005.

Molinari, F., "Microeconometrics with partial identification," in "Handbook of Econometrics," Vol. 7, Elsevier, 2020, pp. 355–486.

Pakes, A., "Alternative Models for Moment Inequalities," *Econometrica*, 2010, *78* (6), 1783–1822.

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

Ponomareva, M. and E. Tamer, "Misspecification in moment inequality models: back to moment equalities?," *The Econometrics Journal*, 2011, *14* (2), 186–203.

Roehrig, C. S., "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 1988, *56* (2), 433–447.

Romano, J., A. Shaikh, and M. Wolf, "A Practical Two-Step Method for Testing Moment Inequalities," *Econometrica*, 2014, *82* (5), 1979–2002.

Schennach, S. M., "Entropic Latent Variable Integration Via Simulation," *Econometrica*, 2014, *82* (1), 345–385.

Tamer, E., "Partial Identification in Econometrics," *Annual Review of Economics*, September 2010, *2* (1), 167–195.

Tebaldi, P., A. Torgovitsky, and H. Yang, "Nonparametric Estimates of Demand in the California Health Insurance Exchange," *Working Paper*, 2018.