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

Lixiong Li †

June 7, 2019

[click here for the latest version]

Abstract

Structural econometric models usually involve parametric distributional assumptions for unobserved heterogeneity. Although these assumptions are typically not informed by economic theory, and undermine the robustness of empirical results, they are generally thought to be necessary to simulate counterfactual predictions. In partially identified and incomplete structural models, counterfactual analysis is also hampered by the multiplicity of admissible structural parameter values and the multiplicity of counterfactual predictions for each structural parameter value. This paper shows how to construct identification conditions for both structural and counterfactual parameters in a large class of structural econometric models, including partially identified and incomplete ones, without imposing parametric distributional assumptions for unobserved variables. The identified set is characterized by moment inequalities, so that existing inferential methods can be applied, including subvector inference when only counterfactual parameters are of interest. The novelty and computational tractability of the methodology is illustrated on a class of discrete choice models and a class of entry models.

*I am grateful to Marc Henry, Joris Pinkse and Keisuke Hirano for their invaluable guidance and support throughout this project. I thank Michael Gechter, Paul Grieco, Patrik Guggenberger, Ismael Mourifié, Charles Murry, Mark Roberts, Karl Schurter and Neil Wallace for helpful discussions and suggestions. I would like to thank the seminar participants at IO Brown Bag and Econometrics Reading Group at Penn State, Econometrics Workshop at the University of Toronto and Cornell University for comments and suggestions.

†Department of Economics, The Pennsylvania State University. Email: lixiong.li@psu.edu

1 Introduction

In structural model estimation, unobserved heterogeneity is usually handled by imposing parametric distributional assumptions. For example, the distribution of unobserved heterogeneity is usually assumed to be Gumbel in discrete choice models, Pareto or Fréchet in structural trade models, and Normal in empirical entry games. Although these assumptions could lead to non-robust empirical results where parametric assumptions instead of data variation drive the identification, they are widely imposed in structural models. One main reason for this is that these parametric assumptions are generally thought to be necessary to simulate the counterfactual model predictions, which is usually the motivation to set up a structural model in the first place.

This paper develops a novel identification method to construct sharp bounds for both structural and counterfactual parameters under semiparametric assumptions on the unobserved heterogeneity. In contrast to the parametric assumptions, these semiparametric assumptions often stem from more robust and interpretable restrictions such as conditional mean/median independence, or quantile restrictions based on symmetric distributions or first-order stochastic dominance. Moreover, due to the generality of the method, one can use the method to construct bounds under a sequence of (possibly nested) assumptions. For example, one can start from some very robust conditions, which typically result in insufficiently informative bounds, and then impose more and more stringent assumptions until the result is informative enough for empirical analysis. Through this process, one can gain a good understanding of which assumption is essential to the final empirical results. This idea can be traced back to Manski (see for instance the discussion in Manski (1999) and Manski (2008)) but is only feasible when one can derive sharp identification conditions for each set of assumptions under evaluation, as I do in this paper.

More specifically, the semiparametric assumptions studied in this paper are moment conditions which involve parameters, observables, and unobservables. These moment conditions arise naturally in counterfactual analysis. As a simplistic example, consider a counterfactual exercise in a multinomial choice model where some characteristic of choices, say, the price, changes. Then, the counterfactual choice probability $\tilde{s}$ is an integration of each agent i's counterfactual choice decision $\widetilde{Y}_i$. In this example, the counterfactual parameter $\tilde{s}$ and one of the unobservables $\widetilde{Y}_i$ are involved in one moment equation, $\mathbb{E}[\widetilde{Y}_i - \tilde{s}] = 0$. This is different from the generalized method of moments (GMM) in which moment functions only depend on parameters and observables, not unobservables.

This paper has two contributions to the literature. First, it develops a new identification approach which can be applied to more general models and is sharp under weaker conditions than the existing literature. Second, it provides a new way to conduct the counterfactual analysis under semiparametric assumptions, even if the model is not point identified or has multiple counterfactual model predictions.

The identification approach transforms the structural model and these moment conditions into a set of moment inequalities which only depend on parameters and observables. The transformation builds on support functions which bound the moment functions given the model restrictions. It does not hinge on particular structures in which unobserved heterogeneity can be differenced out or integrated out, which makes it different from Pakes (2010) and Pakes et al. (2015). Moreover, I derive a set of sufficient conditions and a set of necessary conditions for the sharpness of the approach.

This paper overcomes several challenging technical difficulties. First of all, it covers the cases where moment functions are unbounded. The lack of boundedness violates regularity conditions of most duality and random set theoretic results that serve as the cornerstone of the existing identification results. All existing results on the sharp identification in similar econometric models impose either compactness conditions or integrable boundedness conditions, which rules out some interesting empirical applications. Ekeland et al. (2010) and Beresteanu et al. (2011) are examples of papers that make such assumptions. Secondly, the sharp identification conditions in this paper generally involve a continuum of moment conditions, which could make them hard to implement in practice. Similar challenges exist in Ekeland et al. (2010), Beresteanu et al. (2011) and Schennach (2014). Although there exist some inference procedures designed for infinitely many moment inequality models, their computational complexity can be overwhelming. Simplifying these moment conditions without losing (too much) information is a challenging task. There has been some related work in Galichon and Henry (2011), Chernozhukov et al. (2013), Chesher and Rosen (2017) and Luo and Wang (2017), which removes redundant moment inequalities among finitely many conditions. However, the problem here is even harder since the number of moment inequalities here could be infinite.

To solve the first issue, I develop a new regularization procedure. The idea is to first regularize the moment function by imposing constraints which, for example, bound the norm of the moment functions. The identification conditions for these regularized models can then be easily derived. Next, I weaken these regularization conditions by gradually lifting the bounds on the moment functions. During this process, the identified set of the regularized model will expand and eventually converge to a limiting set. It turns out that my identification approach is sharp if the set of parameters it characterizes is equal to this limiting set.

I also show that the first issue can be bypassed by relaxing the definition of the identified set to some enlarged version of it. This is in the spirit of Schennach (2014), but I only relax the moment restriction while keeping the other model restrictions intact. This relaxation turns out to be useful in that it removes the need for the boundedness conditions while preserving a good approximation for the identified set. Its similarities and differences with Schennach (2014) will be elaborated in Section 3.5.

To solve the second issue, I provide a new simplification algorithm. When the moment function only takes finitely many values, the algorithm finds a minimal set of moment inequalities which preserves sharpness, i.e., a minimal core determining class. The algorithm transforms the problem into a group of linear programming problems so that finding the minimal core determining class is equivalent to enumerating all vertices of a convex polyhedron. One can then apply any existing vertex enumeration algorithm to the core determining class problem. See, for example, Avis and Fukuda (1992) and, more recently, Avis and Jordan (2018). In cases where the moment function is not discrete, the algorithm can still be helpful. One can first discretize the moment function and find a minimal core determining class Λ for that discretized model. Such a set Λ will not be core-determining for the original model, but it should capture most of the information if the discretization is fine enough. This idea is formalized in Section 5.

1.1 Literature

This paper is not the first one to study the identification in structural models with semiparametric assumptions. 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 set 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 show their identification results are sharp under certain assumptions.

Our paper is mostly related to Ekeland, Galichon and Henry (2010), who propose a sharp identification approach which builds on the uniform integrability of the moment function and the tightness of the support of the unobserved heterogeneity. Those assumptions are very restrictive and rule out many models of interest, including the examples considered in this paper. As explained later in Section 3.4, part of the reason why they need these stringent assumptions is that their identification approach implicitly treats all model restrictions as moment conditions. In general cases, the bounds constructed in this paper are always weakly and sometimes strictly tighter than that in Ekeland, Galichon and Henry (2010). Similar identification ideas have been used in Beresteanu, Molchanov and Molinari (2011) for a different economic model, whose identification results also rely on certain compactness assumptions. Similar ideas have also been presented as an optimal transportation approach in Galichon and Henry (2011) in a parametric setting. Recently, Chesher and Rosen (2017) establish the equivalence between the identification conditions on the random set of the observables and those on the random set of the unobservables.

Another related paper is Schennach (2014), who proposes an entropy-based identification approach which is computation-friendly and works under very mild conditions. However, the identification approach in Schennach (2014) is not sharp under the assumptions stated therein. In fact, based on the approach in Schennach (2014), one could get two different identification results for the same model if the model is written in two different (yet equivalent) ways. As explained in Section 3.5, this is due to the fact that the theorem stated in Schennach (2014) is in terms of some enlarged version of the identified set. I clarify the relationship between the identified set and the enlarged identified set in Section 3.4.

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.

There is also a literature on performing counterfactual analysis in incomplete models. In the context of empirical games, computational methods have been developed for solving or bounding the multiple counterfactual equilibria. Methods based on lattice theory (Jia (2008), Uetake and Watanabe (2017)), mixed-integer linear programming (Reguant (2016)) and genetic algorithms (Aguirregabiria and Mira (2005)) have been proposed in the literature. In contrast to these contributions, this paper does not focus on computational methods for the counterfactuals. Instead, it proposes ways to perform counterfactual analysis without making parametric assumptions on the unobserved heterogeneity on which all the above contributions hinge.

As my identification conditions take the form of moment inequalities, this paper also relates to the literature on moment inequality inference. When my identification conditions are simplified into or approximated by a finite number of moment inequalities, one can apply methods in Chernozhukov, Hong and Tamer (2007), Andrews and Soares (2010) and Chernozhukov, Chetverikov and Kato (2018) to perform hypothesis testing and construct confidence region. 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). In general, cases where the identification conditions involve a continuum of moment inequalities, one can use inference procedures in Andrews and Shi (2017) and Chernozhukov, Lee and Rosen (2013).

The rest of the paper is organized as follows. Section 2 describes the class of models studied in this paper and provides several examples. Section 3 presents the identification approach developed in this paper. Section 4 discusses the counterfactual analysis. Section 5 provides a computational method. Section 6 illustrates the method with some simulations. Section 7 describes a further extension of the framework. Finally, Section 8 concludes the paper.

1.2 Notation

In the following, I use capital letters like U, Z to denote random variables, and use lower case letters like u, z to denote specific values of those random variables. Unless otherwise specified, the norm $\| \ \|$ stands for the Euclidean norm. For any positive integer k, I use S_k to denote the unit sphere in $\mathbb{R}^k$, i.e. $S_k \coloneqq \{\lambda \in \mathbb{R}^k : \|\lambda\| = 1\}$. I use $\mathbb{1}(\cdot)$ to denote the indicator function. Given a probability measure H, I use $\mathbb{E}_H$ and $\mathbb{P}_H$ to denote the expectation and probability taken with respect to H. If not explicitly specified, expectation $\mathbb{E}$ and probability $\mathbb{P}$ is take with respect to the true probability measure in the data generating process (DGP). Finally, I use Θ to stand for the parameter space.

2 Frameworks and Motivating Examples

This paper studies the identification for the following empirical model,

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

where U stands for the vector of unobservable variables, and Z is the vector of observables. Let $\mathcal{U}$ and $\mathcal{Z}$ be the space of U and Z respectively. $\Gamma(\theta)$ is a possibly θ-dependent set of (U, Z), which summarizes all structural model restrictions on (U, Z). The function r maps (U, Z) and θ to a Euclidean space $\mathbb{R}^{dr}$ where $dr < \infty$. The dimension of θ can be infinite. In the following, I denote the model restrictions in (1) by (Γ, r). I use the term *support restriction* to refer to the first part of the restriction in (1), and use *moment restriction* to refer to the second part.

For each $\theta \in \Theta$, set $\Gamma(\theta)$ and function $r(\cdot, \cdot; \theta)$ are known. The goal is to find all parameters θ which satisfy (1). Formally, let F_Z be the distribution of Z, which is identified from the data. I define the identified set as follows.

Definition 1 (identified set)**.** The identified set Θ_I is the set of all $\theta \in \Theta$ such that there exists a joint distribution H for (U, Z) which satisfies

(i) $\mathbb{P}_H[(U, Z) \in \Gamma(\theta)] = 1$,

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

(iii) H's marginal distribution for Z equals F_Z.

where the probability and expectation in (i) and (ii) are taken with respect to distribution H.

The first two conditions in the definition are the statement of the general framework in (1), and the last condition means the data generating process H is consistent with the data. In general, Θ_I may or may not be a singleton. For each parameter θ in the identified set Θ_I, there exists a distribution H which 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 others. I sometimes write Θ_I as $\Theta_I(\Gamma, r)$ when emphasizing the underlying model restrictions.

Framework in (1) covers a large class of structural models. In the following, I give several examples to illustrate how different kinds of structural models fit in the framework.

Example 1 (binary choice model with limited information)**.** 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, and π_i be her *ex post* payoff if she chooses $Y_i = 1$,

$$\pi_i = X_i\beta - \alpha + \epsilon_i.$$

Here, ϵ_i is agent i's payoff shock, X_i are covariates and (α, β) are parameters to be estimated. When making decisions, agent i knows ϵ_i, but she cannot observe X_i. Instead, she forms her expectation of X_i based on her information set $\mathcal{I}_i$. By the definition of $\mathcal{I}_i$, the payoff shock ϵ_i is contained in $\mathcal{I}_i$.[1] Let $\mathbb{E}_s[\pi_i|\mathcal{I}_i]$ be agent i's subjective expectation for π_i. Based on the principle of revealed preference, assume $\mathbb{E}_s[\pi_i|\mathcal{I}_i] \geq 0$ if she chooses $Y_i = 1$, and $\mathbb{E}_s[\pi_i|\mathcal{I}_i] \leq 0$ otherwise. Assume we have data on Y_i and realized X_i, but we don't know agent i's payoff shock ϵ_i or her information set $\mathcal{I}_i$. Instead, assume we observe a vector W_i of variables contained in $\mathcal{I}_i$. Define ν_i to be the expectation error of the agent, i.e. $\nu_i \coloneqq \mathbb{E}_s[X_i|\mathcal{I}_i] - X_i$.

Choice models with this information structure have been adopted to study various empirical applications. For example, Ho and Pakes (2014) uses a similar model to study hospital referral decisions, assuming there is no payoff shock, i.e. $\epsilon_i \equiv 0$. Dickstein and Morales (2018) adopt the same model to describe firms' exporting decision, under the extra assumption that ϵ_i follows a normal distribution and is independent of $\mathbb{E}_s[\pi_i|\mathcal{I}_i]$. The focus here is to obtain sharp bounds on (α, β) and to conduct counterfactual analysis under different sets of assumption so that we can later evaluate the identification power of different assumptions.

To see how this model fits in the general framework (1), let $U_i \coloneqq (\nu_i, \epsilon_i)$ be the collection of all unobserved variables, $Z_i \coloneqq (Y_i, X_i, W_i)$ be the collection of observed variables and let $\theta \coloneqq (\alpha, \beta)$ be the vector of all parameters. For the purpose of illustration, assume X_i is scalar. By the definition of ν_i, our revealed preference assumptions can be written as the following restriction,

$$\mathbb{P}[(U_i, Z_i) \in \Gamma(\theta)] = 1, \text{ where } \Gamma(\theta) = \{(u_i, z_i) : (-1)^{y_i}[(x_i + \nu_i)\beta - \alpha + \epsilon_i] \leq 0\}. \quad (2)$$

Equation (2) then serves as the support restriction in (1).

[1] Technically, the information set $\mathcal{I}_i$ is a σ-algebra. Hence, a more precise statement is that the σ-algebra generated by ϵ_i is a subset of $\mathcal{I}_i$.

As for the moment restrictions in (1), there are many possibilities. On the one hand, we could use the following weak assumptions,

$$\begin{aligned}\mathbb{E}[\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)|\mathcal{I}_i] &= 0,\\ \mathbb{E}[\mathbb{1}(\epsilon_i \geq 0) - \mathbb{1}(\epsilon_i \leq 0)|W_i] &= 0.\end{aligned}$$

which is equivalent to assuming that expectation error ν_i has zero median[2] conditional on $\mathcal{I}_i$ and ϵ_i has zero median conditional on W_i. Since $\mathcal{I}_i$ is not observable, we can instead work with the following implication

$$\begin{gathered}\mathbb{E}[\mathbb{1}(Y_i = 1, \epsilon_i \leq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big)|W_i] = 0,\\ \mathbb{E}[\mathbb{1}(Y_i = 1, \epsilon_i \geq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big)|W_i] = 0,\\ \mathbb{E}[\mathbb{1}(Y_i = 0, \epsilon_i \leq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big)|W_i] = 0,\\ \mathbb{E}[\mathbb{1}(Y_i = 0, \epsilon_i \geq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big)|W_i] = 0,\\ \mathbb{E}[\mathbb{1}(\epsilon_i \geq 0) - \mathbb{1}(\epsilon_i \leq 0)|W_i] = 0.\end{gathered} \tag{3}$$

Note that (3) is implied by the fact that ϵ_i belongs to agent i's information set $\mathcal{I}_i$ and Y_i is a function of $\mathcal{I}_i$.

On the other hand, we could also impose the following assumption on the conditional choice probability,

$$\mathbb{E}\left[\mathbb{1}(Y_i = 1) - \Phi\left(\frac{(X_i + \nu_i)\beta - \alpha}{\sigma}\right)\Big|W_i\right] = 0, \tag{4}$$

where Φ is the c.d.f. of the standard normal distribution. Condition (4) is consistent with the assumption that ϵ_i is distributed as $N(0, \sigma^2)$ and is independent of W_i and $\mathbb{E}_s[X_i|\mathcal{I}_i]$. One could then combine (4) and (3) to serve as the moment functions in (1).

Identification results in Section 3 will show how to derive moment inequality constraints which are equivalent to the above support and moment restrictions and, at the same time, only involve parameters and observables.

Consider now the counterfactual analysis. Following Dickstein and Morales (2018), for example, suppose each agent i is a firm deciding whether to export ($Y_i = 1$) or not ($Y_i = 0$), where X_i and α stand for its exporting profit and fixed cost respectively. I normalize X_i's slope coefficient to $\beta \equiv 1$. Consider a counterfactual in which the fixed cost has increased by 10%. Suppose one is interested in how many firms would continue to export after this counterfactual increase in the cost. In other words, suppose one want to know the following

[2] Alternatively, one could also assume agent i has rational expectation so that $\mathbb{E}[v_i|\mathcal{I}_i] = 0$. I will also discuss this setting in Section 3.

counterfactual parameter $\tilde{p}$:

$$\begin{aligned} \tilde{p} &= \mathbb{P}(\mathbb{E}_s[X_i|\mathcal{I}_i] - 1.1\alpha + \epsilon_i \geq 0) \\ &= \mathbb{P}(X_i + \nu_i - 1.1\alpha + \epsilon_i \geq 0), \end{aligned} \tag{5}$$

where the second equality follows from the definition $\nu_i := \mathbb{E}[X_i|\mathcal{I}_i] - X_i$. In general, this model is partially identified even when restrictions like (4) are imposed. In particular, the joint distribution of (ν_i, ϵ_i) is not point identified. Therefore, one cannot simply estimate α (and σ) and then solve for $\tilde{p}$ by simulation. However, one can view the counterfactual parameter $\tilde{p}$ as a model primitive together with structural parameters (α, β). In fact, equation (5) can be viewed as one of the moment restrictions in (1). Combining it with the support restrictions (2) and the moment restrictions (3) and/or (4), one can derive the identified set for $\tilde{p}$ and (α, β) jointly using identification methods developed in this paper. I provide more details in Section 4. ■

Example 2 (static entry game with complete information)**.** Consider a static entry game with complete information as in Bresnahan and Reiss (1991). Suppose there are I players. In each market m, player i could choose to enter the market ($Y_{i,m} = 1$) or not ($Y_{i,m} = 0$). When firm i chooses $Y_{i,m} = 0$, normalize its payoff at market m to 0; When firm i chooses $Y_{i,m} = 1$, assume its payoff equals $\pi_i(Y_{-i,m}, X_{i,m}; \theta) + U_{i,m}$ where[3]

$$\pi_i(Y_{-i,m}, X_{i,m}; \theta) = X'_{i,m}\alpha_i - \sum_{k \neq i} \Delta_k \cdot Y_{k,m},$$

and where $\theta = (\alpha_i, \Delta_i : i = 1, ..., I)$ is the parameter to be estimated, $\{X_{i,m}, i = 1, ..., I\}$ are firm-specific observable characteristics in market m, and $\{U_{i,m} : i = 1, ..., I\}$ are payoff shocks which are known to both firms but are unobserved in the data. A common interpretation for $U_{i,m}$ is that it captures the effect of all factors other than $X_{i,m}$ which also affects firm i's payoff. Assume $\Delta_i \geq 0$ for $i = 1, ..., I$ so that the pure-strategy Nash equilibrium always exists. Assume also that firms always play a pure-strategy Nash equilibrium. Define $U_m = (U_{i,m} : i = 1, ..., I)$, $Y_m = (Y_{i,m} : i = 1, ..., I)$ and $X_m = (X_{i,m} : i = 1, ..., I)$ as the vectors of all players' payoff shocks, entry decision and observable characteristics respectively.

Similar models have been studied in Berry (1992) and Ciliberto and Tamer (2009) to study market structures in Airline industries. More recently, Ciliberto, Murry and Tamer (2018) have considered a richer model in which firms not only make entry decisions but also set price after entry. In all these papers, the payoff shocks U_m are assumed to follow a multivariate Normal distribution and to be independent of X_m. However, as discussed in Section 6.2, the normality assumption seems to dominate the data variation and drives the

[3]Here, the linear form and the additive separability in the payoff function is assumed to keep the example simple for clear illustration. The method in paper applies to nonlinear and nonseparable payoff functions.

identification result. As an alternative, I propose a semiparametric assumption based on the median independence in the following.

First, let me show how the entry game fits in framework (1). Let $Z_m := (Y_m, X_m)$ collect all the observables. And, recall that U_m collects all the unobserved heterogeneity. As firms are assumed to play a pure-strategy Nash equilibrium, each firm's choice must be its best response to the other firm's choice so that we have the following support restriction $\mathbb{P}[(U_m, Z_m) \in \Gamma(\theta)] = 1$ where $\Gamma(\theta)$ is defined by

$$\Gamma(\theta) := \{(u_m, z_m) : \forall i,\ (-1)^{y_{i,m}}(\pi_i(y_{-i,m}, x_{i,m}; \theta) + u_{i,m}) \leq 0\}.$$

Now, instead of assuming that U_m follows a Normal distribution as in the literature, suppose U_m has zero median conditional on X_m.[4] Formally, assume $\mathbb{E}[r(U_m, Z_m; \theta)|X_m] = 0$ where $r = (r_i : i = 1, ..., I)$ and

$$r_i(u_m, z_m; \theta) = \mathbb{1}(u_{i,m} \geq 0) - \mathbb{1}(u_{i,m} \leq 0).$$

As shown later in Section 6.2, the identified set of (Γ, r) in this example can be characterized by a finite number of conditional moment inequalities. Moreover, the evaluation of these conditional moment inequalities does not require to solve the set of all pure-strategy Nash equilibria, which is a huge computational advantage over the existing literature. ■

Example 3 (Panel Multinomial Choice Model)**.** Consider a panel multinomial choice model, where agent i chooses from a choice set $\mathcal{J} = \{0, 1, ..., J\}$ at time periods $t = 1, ..., T$. T is a finite integer and $T \geq 2$. Agent i's indirect payoff from choosing option j at time t is U_{ijt} which is defined by

$$U_{ijt} = \pi(X'_{ijt}\theta, \nu_{ij}, \epsilon_{ijt})$$

where X_{ijt} is a vector of observable characteristics of option j specific to agent i at time t, ν_{ij} captures the fixed effect which agent i has on option j, and ϵ_{ijt} is agent i's transitory taste shock on option j at time t. ν_{ij} and ϵ_{ijt} can be of infinite dimension. Let U_i, Y_i, X_i, ν_i and ϵ_i be the collection of agent i's U_{ijt}, Y_{it}, X_{ijt}, ν_{ij} and ϵ_{ijt} over all choices and time periods respectively. Both ν_{ij} and ϵ_{ijt} are treated as unobserved heterogeneity. Assume that for each agent i and for an time t, $\{U_{ijt} : j \in \mathcal{J}\}$ has a unique maximizer,[5] and assume that agent i's observed choice Y_{it} at time t selects the best choice, $Y_{it} = \arg\max_{j \in \mathcal{J}} U_{ijt}$. The goal here is to identify parameter θ under mild condition on function π.

[4] Suppose, conditional on X_m, that U_m has zero mean instead of zero median. By applying results in Section 3.3, one can show that the identified set is the entire parameter space, which means that the zero mean assumption has no empirical content in this model.

[5] This is a standard assumption in the literature. Without this assumption, the identified set for θ equals the entire parameter space, as any choice probability and any value of θ can be rationalized by selecting π to be a constant function, i.e., $\pi \equiv 0$. This happens because function π is treated as a nonparametric function in this example.

This model has been studied in Pakes and Porter (2016) and Shi, Shum and Song (2018), both of which assume that $X_{ijt}'\theta$ and ϵ_{ijt} are additively separable in π. More recently, Gao and Li (2018) proposes a novel identification strategy which allows for nonseparable structures in π and infinite dimensional unobserved heterogeneity. This is a nontrivial contribution, as the fixed effect cannot be differenced out in a nonseparable payoff function. Nevertheless, in the following, I show their identification approach can be further improved using the method developed in this paper.

To see how this model fits in the framework. Let $Z_i = (Y_i, X_i)$ collect all the observables. Then, the revealed preference condition implies the following support restriction, $\mathbb{P}[(U_i, Z_i) \in \Gamma(\theta)] = 1$, where

$$\Gamma(\theta) = \Big\{(u_i, z_i) : \forall t = 1, ..., T,\ y_{it} = \arg\max_{j \in \mathcal{J}}\ u_{ijt}\Big\}. \tag{6}$$

To derive the moment restriction, following Gao and Li (2018), assume that

(i) Given ν_{ij} and ϵ_{ijt}, function π is weakly increasing in its first argument. That is, for any $\delta \geq \delta'$, $\pi(\delta, \nu_{ij}, \epsilon_{ijt}) \geq \pi(\delta', A_{ij}, \epsilon_{ijt})$.

(ii) $(Y_i, X_i, \nu_i, \epsilon_i)$ are i.i.d. across i.

(iii) The conditional distribution of $\epsilon_{it} = (\epsilon_{ijt} : j \in \mathcal{J})$ given (X_i, ν_i) is time invariant.

In Appendix H.3, I show that the above assumptions imply the following moment restrictions: For any two time periods s and t, define $\mathcal{A}_{ist}(\theta) = \{j \in \mathcal{J} : X_{ijs}'\theta \geq X_{ijt}'\theta\}$ and $\mathcal{B}_{ist}(\theta) = \{j \in \mathcal{J} : X_{ijs}'\theta \leq X_{ijt}'\theta\}$. For any two nonempty subsets $\mathcal{J}_1$ and $\mathcal{J}_2$ of $\mathcal{J}$, define $\rho_{ist}(\mathcal{J}_1, \mathcal{J}_2; \theta) = \mathbb{1}\big(\mathcal{J}_1 \subseteq \mathcal{A}_{ist}(\theta) \text{ and } \mathcal{J}_2 \subseteq \mathcal{B}_{ist}(\theta)\big)$. Then,

$$\mathbb{E}\Big[\rho_{ist}(\mathcal{J}_1, \mathcal{J}_2; \theta)\Big\{\mathbb{1}\Big(\max_{j \in \mathcal{J}_1} U_{ijs} \geq \max_{j \in \mathcal{J}_2} U_{ijs}\Big) - \mathbb{1}\Big(\max_{j \in \mathcal{J}_1} U_{ijt} \geq \max_{j \in \mathcal{J}_2} U_{ijt}\Big)\Big\}\Big| X_i\Big] \geq 0. \tag{7}$$

Moment restrictions as in (7) is an moment inequality restriction. Therefore, it does not directly fit into the framework in (1). In Section 7, however, I show that there is a straightforward extension of the framework in (1) to incorporate these moment inequality restrictions. In Section 7, I also work out the identification conditions derived from the support restriction in (6) and moment restrictions in (7). The set of identification constraints in Gao and Li (2018) are nested in these identification conditions. ■

3 Identification Results

3.1 Overview of Results and Basic Intuition

Before the formal statement of the general theorems, let's start with a simple special case to illustrate the main identification idea. Suppose the moment function $r(u, z; \theta)$ takes values in $\mathbb{R}$, i.e. $dr = 1$.

As a first step, we want to find the identification conditions implied by (1). Since the conditions in (1) involve unobservable variable U, they cannot be taken to the data directly. We need to transform them into conditions which only involves observable variable Z. To do so, let's define $\underline{r}$ and $\overline{r}$ to be the lower and upper bounds of r given the support restriction. That is,

$$\begin{aligned} \underline{r}(z;\theta) \coloneqq \inf_{u}\ & r(u,z;\theta) \\ \text{s.t. }\ & (u,z) \in \Gamma(\theta) \end{aligned} \qquad \text{and} \qquad \begin{aligned} \overline{r}(z;\theta) \coloneqq \sup_{u}\ & r(u,z;\theta) \\ \text{s.t. }\ & (u,z) \in \Gamma(\theta) \end{aligned} \tag{8}$$

The functions $\underline{r}(z;\theta)$ and $\overline{r}(z;\theta)$ are always well-defined, although they may take the values of $-\infty$ or $+\infty$, for example, when r and $\Gamma(\theta)$ are unbounded.

For any $\theta \in \Theta_I$, we have $\mathbb{P}[(U,Z) \in \Gamma(\theta)] = 1$, so that $\mathbb{P}[\underline{r}(Z;\theta) \leq r(U,Z;\theta) \leq \overline{r}(Z;\theta)] = 1$. Moreover, we know that $\mathbb{E}[r(U,Z;\theta)] = 0$. These lead to

$$\mathbb{E}[\underline{r}(Z;\theta)] \leq 0 \leq \mathbb{E}[\overline{r}(Z;\theta)]. \tag{9}$$

Note that the expectation in Condition (9) involves only the observable variable Z. As a result, Condition (9) can be taken to the data, and one can then compute confidence regions for θ using existing inference procedures for moment inequalities.

The transformation from (1) to (9) conveys the basic intuition of the identification strategy in this paper. By bounding the moment function with the support restriction, one can derive moment inequalities involving only observable variables.

Furthermore, there is no information loss in this transformation. Condition (9) is not only implied by (1), but is in fact equivalent to it. To provide the intuition for this sharpness result, I provide a heuristic proof in the following paragraph, and defer the formal proof to later sections.

Let θ^* be an arbitrary parameter for which Condition (9) holds. To show $\theta^* \in \Theta_I$, I'll construct a joint distribution H^* of (U,Z) so that

$$\mathbb{P}_{H^*}[(U,Z) \in \Gamma(\theta^*)] = 1 \quad \text{and} \quad \mathbb{E}_{H^*}[r(U,Z;\theta^*)] = 0, \tag{10}$$

where the expectation and probability are taken with respect to distribution H^*. Suppose the bounds, $\underline{r}$ and $\overline{r}$, can be achieved. That is, suppose there exist functions $\underline{u}(z;\theta^*)$ and $\overline{u}(z;\theta^*)$ such that

$$\begin{aligned} \underline{u}(z;\theta^*) &\in \operatorname*{arg\,min}_{u \in \Gamma(z;\theta^*)} r(u,z;\theta^*) \\ \overline{u}(z;\theta^*) &\in \operatorname*{arg\,max}_{u \in \Gamma(z;\theta^*)} r(u,z;\theta^*). \end{aligned} \tag{11}$$

where $\Gamma(z;\theta^*) \coloneqq \{u : (u,z) \in \Gamma(\theta^*)\}$.

Define $p^* := \mathbb{E}[\overline{r}(Z;\theta^*)]/(\mathbb{E}[\overline{r}(Z;\theta^*)]-\mathbb{E}[\underline{r}(Z;\theta^*)])$ and assume the denominator is nonzero. Since θ^* satisfies the moment inequalities in (9), we know p^* is between 0 and 1. Then, construct a discrete probability distribution $H^*_{U|Z}$ of U conditional on $Z = z$ as

$$U = \begin{cases} \underline{u}(z;\theta^*) & \text{with probability } p^* \\ \overline{u}(z;\theta^*) & \text{with probability } 1-p^*. \end{cases} \tag{12}$$

Let H^* be the joint distribution of (U, Z) constructed based on conditional distribution $H^*_{U|Z}$ of U and the marginal distribution of Z. Recall that the marginal distribution of Z is identified in the data, as Z is observable. Such an H^* satisfies Restriction (10). To see this, note that, by construction in (11) and (12),

$$\mathbb{P}_{H^*}[(U, Z) \in \Gamma(\theta^*)|Z = z] = 1 \text{ for almost every } z,$$

which further implies $\mathbb{P}_{H^*}[(U, Z) \in \Gamma(\theta^*)] = 1$. Moreover, we have

$$\mathbb{E}_{H^*}[r(U, Z;\theta^*)] = \mathbb{E}_{H^*}\left[\mathbb{E}_{H^*|Z}[r(U, Z;\theta^*)|Z]\right] = \mathbb{E}\left[p^*\underline{r}(Z;\theta^*) + (1-p^*)\overline{r}(Z;\theta^*)\right] = 0,$$

where the first equality follows from the law of iterated expectation and the second equality comes from the construction of $H^*_{U|Z}$, and the last equality is a result of the construction of p^*.

In words, by combining points which attain the bounds of the moment functions, we can construct a joint distribution of (U, Z) so that all model conditions are satisfied by the parameter θ^*. In the following sections, such idea will be formalized to prove the sharp identification results in general cases when moment functions are (integrably) bounded.

However, the moment functions r could be unbounded in some interesting applications. In those unbounded cases, our previous constructive argument fails, as bounds are infinite and boundary points like $\overline{u}(z;\theta)$ and $\underline{u}(z;\theta)$ no longer exists. Although the moment inequalities in (9) are still implied by the model restriction, they may or may not be sharp. I develop two approaches to study the identification in these cases. One involves regularization of the original model, which provides both sufficient conditions and necessary conditions for $\theta \in \Theta_I$. The other approach considers a weaker notion of the identified set. There two approaches are discussed in Section 3.3 and 3.4 respectively.

3.2 Basic Identification Results

In this section, I develop identification results in the general case. I show how to convert the conditions in Definition 1 into moment inequalities like those in (9), using the same intuition as in the previous heuristic derivation.

For any $\theta \in \Theta_I$ and one of its corresponding distributions H satisfying the conditions in

Definition 1, the following moment equality must hold for any vector λ in the unit sphere S_{dr}:

$$\mathbb{E}_H[\lambda' r(U, Z; \theta)] = 0. \tag{13}$$

Define $\Gamma(z;\theta)$ to be the projection of $\Gamma(\theta)$. That is,

$$\Gamma(z;\theta) \coloneqq \{u : (u, z) \in \Gamma(\theta)\}. \tag{14}$$

Then, $(U, Z) \in \Gamma(\theta)$ with probablity 1 implies that

$$\mathbb{P}_H \left(\lambda' r(U, Z; \theta) \leq \sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta) \right) = 1.$$

Therefore, we have

$$\mathbb{E}_H[\lambda' r(U, Z; \theta)] \leq \mathbb{E}_H \left[\sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta) \right] = \mathbb{E} \left[\sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta) \right]. \tag{15}$$

The $\mathbb{E}$ in the last term denotes the expectation with respect to the distribution of Z in the data. The last equality holds, because $\sup_{u \in \Gamma(z;\theta)}[\lambda' r(u, z; \theta)]$ is a function which only depends on z and Z is observed.

Combining (13) and (15), we conclude that, for any $\theta \in \Theta_I$,

$$\forall \lambda \in S_{dr}, \ \ \mathbb{E} \left[\sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta) \right] \geq 0. \tag{16}$$

Recall that S_{dr} denotes the unit sphere $\{\lambda \in \mathbb{R}^{dr} : \|\lambda\| = 1\}$ and $\Gamma(z;\theta)$ is defined in (14).

When $dr = 1$, the moment inequalities in (16) are equivalent to the two moment inequalities in (9). When $dr > 1$, the conditions in (16) typically involve a continuum of moment inequalities, which is similar to the cases in Ekeland, Galichon and Henry (2010), Beresteanu, Molchanov and Molinari (2011) and Schennach (2014). To do inference based on (16), one can adopt the procedures in Chernozhukov, Lee and Rosen (2013) or Andrews and Shi (2017). In some special cases, condition (16) can be further simplified to a finite number of moment inequalities. I discuss this issue in Section 5.

In the following, Theorem 1 formalizes the above heuristic derivation and shows the moment inequalities in (16) are valid identification conditions under very mild assumptions. Theorem 2 then provides sufficient conditions under which Condition (16) is a sharp characterization of the identified set. Results in Theorem 2 will be further generalized later in Theorem 3. The proof of Theorem 1 and 2 can be found in Appendix C.

Theorem 1. *Suppose that, for every $\theta \in \Theta$, the following conditions hold,*

C1 Set $\Gamma(\theta)$ is a Borel set and $\Gamma(z;\theta)$ is nonempty for almost every z. Moreover, the *function $r(u,z;\theta)$ is Borel measurable in $\mathcal{U}\times\mathcal{Z}$.*

C2 There exists an integrable function $g(\cdot;\theta)$ such that for almost every z,

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

Then, Condition (16) holds for any $\theta \in \Theta_I$.

Theorem 2. *Suppose, for every $\theta \in \Theta$, Condition C1 and the following conditions hold,*

C3 For almost every z, $\{r(u,z;\theta) : u \in \Gamma(z;\theta)\}$ is a closed set.

C4 There exists an integrable function $g(\cdot;\theta)$ such that for almost every z,

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

Then, $\theta \in \Theta_I$ if and only if θ satisfies Condition (16).

Remark 1. Condition C1 is a regularity condition on the measurability of $\Gamma(\theta)$ and $r(u,z;\theta)$. It is a very mild condition in most applications. Nevertheless, it could fail in cases when $r(u,z;\theta)$ stems from certain optimization problems. For example, let $\mathcal{X}$ be some unbounded and uncountable set and $r(u,z;\theta) = \sup_{x\in\mathcal{X}} \pi(u,x,z;\theta)$. Then, $r(u,z;\theta)$ will not necessarily be Borel measurable.

Condition C2 is a very mild regularity condition on the integrability of r. It ensures that the expectation in (16) is well defined, but it does not rule out the case where the expectation may be infinite for some λ.

Condition C3 holds in various cases. For example, it holds when $r(u,z;\theta)$ is continuous in u and $\Gamma(z;\theta)$ is a compact set. It also holds when $r(u,z;\theta)$ is a discrete function. However, it could be violated in some interesting cases as illustrated later in a concrete example.

Condition C4 is stronger than C2. It assumes that the moment function r restricted to Γ is bounded, and its bound is integrable. This is usually called the integrable boundedness condition, and is a standard assumption in random set theory. For example, see Chapter 2.1 in Molchanov (2005). This condition together with Condition C3 forms the regularity conditions under which $\underline{u}$ and $\overline{u}$ in (11) exist and are integrable in the constructive proof in the previous section. Condition C4 can be quite restrictive. For example, it can be violated whenever both the function r and the set $\Gamma(\theta)$ are unbounded. I will relax this condition in Theorem 3. □

In the following, I will use Example 1 as the leading example to illustrate when all conditions are satisfied and when Condition C3 and Condition C4 could fail.

Example 1 (continued)**.** Recall that, in Example 1, I have assumed the support restriction

$\mathbb{P}[(Z_i, U_i) \in \Gamma(\theta)] = 1$ where

$$\Gamma(\theta) = \{(z_i, u_i) : (-1)^{y_i}[(x_i + \nu_i)\beta - \alpha + \epsilon_i] \leq 0\}, \qquad \text{(2) revisited}$$

where y_i is agent i's choice, x_i stands for the covariates realized *ex post*, ν_i is agent i's expectation error and ϵ_i is her payoff shock. Let $z_i := (y_i, x_i)$ stand for all variables in the data, and let $u_i := (\nu_i, \epsilon_i)$ collect all latent variables. Recall also that I assume that the vector W_i is contained in agent i's information set.

In the rest of the discussion of the example, I condition on a value w_i of W_i and suppress it in the notation. The identified set discussed in the following should be viewed as the set of parameters which rationalize the model and the data conditional on $W_i = w_i$. The moment inequalities which I derive later should also be considered as moment inequalities conditional on $W_i = w_i$.

Case 1: Conditions C1-C4 are satisfied. Suppose we impose the following moment condition which stems from the zero median assumptions on ν_i and ϵ_i,

$$\begin{aligned}
\mathbb{E}[\mathbb{1}(Y_i = 1, \epsilon_i \leq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big)] &= 0,\\
\mathbb{E}[\mathbb{1}(Y_i = 1, \epsilon_i \geq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big)] &= 0,\\
\mathbb{E}[\mathbb{1}(Y_i = 0, \epsilon_i \leq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big)] &= 0,\\
\mathbb{E}[\mathbb{1}(Y_i = 0, \epsilon_i \geq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big)] &= 0,\\
\mathbb{E}[\mathbb{1}(\epsilon_i \geq 0) - \mathbb{1}(\epsilon_i \leq 0)] &= 0.
\end{aligned} \qquad \text{(3) revisited}$$

Since the moment functions in (3) can only take finitely many possible values, Condition C1-C4 are satisfied. Then, Theorem 2 implies that $\theta \in \Theta_I$ if and only if

$$\forall \lambda \in S_{dr}, \;\; \mathbb{E}\left[\sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta)\right] \geq 0, \tag{17}$$

where $dr = 5$ and

$$r(u, z; \theta) = \begin{pmatrix}
\mathbb{1}(y_i = 1, \epsilon_i \leq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big)\\
\mathbb{1}(y_i = 1, \epsilon_i \geq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big)\\
\mathbb{1}(y_i = 0, \epsilon_i \leq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big)\\
\mathbb{1}(y_i = 0, \epsilon_i \geq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big)\\
\mathbb{1}(\epsilon_i \geq 0) - \mathbb{1}(\epsilon_i \leq 0)
\end{pmatrix}. \tag{18}$$

In fact, when $\beta \neq 0$, one can simplify (17) into the following moment inequality conditions:

$$\begin{aligned}
\mathbb{E}\left[\mathbb{1}(Y_i = 1)\Big(3\mathbb{1}(X_i\beta - \alpha > 0) - \mathbb{1}(X_i\beta - \alpha < 0)\Big) + \mathbb{1}(Y_i = 0)\right] &\geq 0\\
\mathbb{E}\left[\mathbb{1}(Y_i = 0)\Big(3\mathbb{1}(X_i\beta - \alpha < 0) - \mathbb{1}(X_i\beta - \alpha > 0)\Big) + \mathbb{1}(Y_i = 1)\right] &\geq 0
\end{aligned}. \tag{19}$$

Intuitively, the moment conditions in (19) tend to be satisfied for parameters with which positive *ex post* mean utility, $\beta X_i - \alpha > 0$, often arises together with $Y_i = 1$, and tend to be violated otherwise. The moment conditions in (19) are also related to some classical identification conditions in the literature. In fact, if one assumes agent i has perfect expectations (i.e. $\nu_i \equiv 0$), the moment conditions in (17) reduce to the maximum score estimating equations in Manski (1975) and Manski (1988).

Case 2: Condition C4 is violated. In this case, assume there is no payoff shock and the expectation error has zero mean conditional on the agent's information set. That is, we assume $\epsilon_i \equiv 0$ and $\mathbb{E}[\nu_i|\mathcal{I}_i] = 0$. Under these assumptions, the support restriction can be simplified to

$$\mathbb{P}[(U,Z) \in \Gamma(\theta)] = 1 \text{ where } \Gamma(\theta) = \{(z_i, u_i) : (-1)^{y_i}[(x_i + \nu_i)\beta - \alpha] \leq 0\}. \tag{20}$$

Moreover, since agent i's decision Y_i is a function of her information set, $\mathbb{E}[\nu_i|\mathcal{I}_i] = 0$ implies that the following moment conditions hold for any covariate $W_i \in \mathcal{I}_i$:

$$\mathbb{E}[\mathbb{1}(Y_i = 0)\nu_i|W_i] = 0 \quad \text{and} \quad \mathbb{E}[\mathbb{1}(Y_i = 1)\nu_i|W_i] = 0. \tag{21}$$

Assume $\beta \neq 0$. Then, equation (21) is equivalent to

$$\mathbb{E}[r(U,Z)] = 0, \quad \text{where} \quad r(u,z) = \begin{pmatrix} \mathbb{1}(y_i = 1)\beta\nu_i \\ \mathbb{1}(y_i = 0)\beta\nu_i \end{pmatrix}, \tag{22}$$

where I suppress the conditioning on W_i in the notation as before.

One can show that moment inequality (17) with the above definition of r is now equivalent to

$$\mathbb{E}[\mathbb{1}(Y_i = 1)(X_i\beta - \alpha)] \geq 0 \quad \text{and} \quad \mathbb{E}[\mathbb{1}(Y_i = 0)(X_i\beta - \alpha)] \leq 0. \tag{23}$$

In fact, the two moment inequalities in (23) are the conditions one would get when applying approaches in Pakes (2010) and Pakes, Porter, Ho and Ishii (2015). However, Pakes (2010) and Pakes, Porter, Ho and Ishii (2015) do not provide any results on the sharpness of their approach. Whether or not (23) is a sharp characterization of the identified set is an open question in the literature. Later, I provide a proof of sharpness with my methodology in Section 3.3. In the next paragraph, I explain the difficulties involved.

Conditions C1-C3 are satisfied in this case, but Condition C4 is violated. To see these points, note that for some fixed $z := (x, y)$, the set $\{r(u,z) : (u,z) \in \Gamma(\theta)\}$ is equal to $\{0\} \times (-\infty, \alpha - x\beta]$ when $y = 0$, and is equal to $[\alpha - x\beta, +\infty) \times \{0\}$ when $y = 1$. Therefore, for any z, we have $\sup\{\|r(u,z;\theta)\| : u \in \Gamma(z;\theta)\} = +\infty$.

There is no easy way to restore Condition C4 without imposing extra restrictions. Therefore, one cannot apply Theorem 2 to establish the sharpness of (23). This motivates Theorem

3 in the next section which relaxes Condition C3 and C4.

Case 3: Condition C3 violated. In this case, I impose the parametric assumption on the conditional choice probability as in (4), which results in the following moment restriction $\mathbb{E}[r(U, Z; \theta)|W] = 0$ where

$$r(u_i, z_i; \theta) = \begin{pmatrix} \mathbb{1}(y_i = 1) - \Phi\left(\frac{(X_i+\nu_i)\beta - \alpha}{\sigma}\right) \\ \mathbb{1}(y_i = 1, \epsilon_i \le 0)\big(\mathbb{1}(\nu_i \ge 0) - \mathbb{1}(\nu_i \le 0)\big) \\ \mathbb{1}(y_i = 1, \epsilon_i \ge 0)\big(\mathbb{1}(\nu_i \ge 0) - \mathbb{1}(\nu_i \le 0)\big) \\ \mathbb{1}(y_i = 0, \epsilon_i \le 0)\big(\mathbb{1}(\nu_i \ge 0) - \mathbb{1}(\nu_i \le 0)\big) \\ \mathbb{1}(y_i = 0, \epsilon_i \ge 0)\big(\mathbb{1}(\nu_i \ge 0) - \mathbb{1}(\nu_i \le 0)\big) \\ \mathbb{1}(\epsilon_i \ge 0) - \mathbb{1}(\epsilon_i \le 0) \end{pmatrix}.$$

and Φ stands for the c.d.f. of the standard normal distribution. In this case, Condition C1, C2 and C4 are satisfied, as the function r is bounded. However, Condition C3 fails to hold. To see this, let $r_1(u, z) = \mathbb{1}(y_i = 1) - \Phi\left(\frac{(X_i+\nu_i)\beta-\alpha}{\sigma}\right)$. Then, it is easy to see that

$$\{r_1(u, z) : (u, z) \in \Gamma(\theta)\} = \begin{cases} (-1, 0) & \text{if } \ y = 0, \\ (0, 1) & \text{if } \ y = 1, \end{cases}$$

where the openness stems from the fact that the image of $\Phi(\cdot)$ on $\mathbb{R}$ is an open interval between 0 and 1. In other words, when $y_i = 1$, the result can always be rationalized by holding ϵ_i fixed and letting $\beta\nu_i \to +\infty$. When $\beta\nu_i \to +\infty$, $r_1(u_i, z_i; \theta) \to 0$. However, there does not exist any finite u which makes $r_1(u, z) = 0$. This loss of closedness cannot be ruled out without imposing more restrictions. ■

3.3 Identification Results with Regularization

Theorem 2 shows that the moment inequalities (16) are sharp under Conditions C1-C4. However, as illustrated in Example 1, Conditions C3 and C4 can be restrictive. To explore identification results without imposing Conditions C3 and C4, I first regularize the model so that all conditions are restored, then I show that the identification results in the regularized model have some implications on the original model. Based on such an approach, I can to derive a set of sufficient conditions and a set of necessary conditions under which the moment inequalities in (16) are sharp.

Let's start from the simple case where only Condition C4 is violated while Condition C1-C3 are satisfied. Define $\delta(z; \theta) := \inf\{\|r(u, z; \theta)\| : u \in \Gamma(z; \theta)\}$. For any $k > 0$, define the *regularized* support restriction $\Gamma_{k,\delta}$ as $\Gamma_{k,\delta}(\theta) := \Gamma(\theta) \cap \{(u, z) : \|r(u, z; \theta)\| \le \delta(z; \theta) + k\}$.

Compared to Γ, the regularized support restriction $\Gamma_{k,\delta}$ has an extra restriction which bounds the norm of the moment function. Also, $\Gamma_{k,\delta}$ is monotone in k as $\Gamma_{k,\delta}(\theta) \subseteq \Gamma_{k+1,\delta}(\theta)$

for any $k \geq 1$. When $k \to \infty$, one can show that $\Gamma_{k,\delta}(\theta)$ converges to $\Gamma(\theta)$ for each $\theta \in \Theta$ in the sense that $\Gamma(\theta) = \cup_{k>0}\Gamma_{k,\delta}(\theta)$.

When there is a violation of Condition C4, studying the regularized model $(\Gamma_{k,\delta},\, r)$ can help us understand the original model $(\Gamma,\, r)$. First, if θ is in the identified set of model $(\Gamma_{k,\delta},\, r)$, then θ is also in the identified set of the original model. Second, even when Condition C4 is violated in the original model, it always holds in the regularized model $(\Gamma_{k,\delta},\, r)$. As a result, Theorem 2 implies that θ is in the identified set of model $(\Gamma_{k,\delta},\, r)$ if and only if the following moment inequalities are satisfied,

$$\forall \lambda \in S_{dr},\ \ \mathbb{E}\left[\sup_{u \in \Gamma_{k,\delta}(Z;\theta)} \lambda' r(u, Z; \theta)\right] \geq 0, \tag{24}$$

where $\Gamma_{k,\delta}(z;\theta) \coloneqq \{u : (u,z) \in \Gamma_{k,\delta}(\theta)\}$ is the projection of $\Gamma_{k,\delta}$. Therefore, (24) serves as a sufficient condition for $\theta \in \Theta_I$.

As for a necessary condition for $\theta \in \Theta_I$, given the fact that $\Gamma_{k,\delta}(\theta)$ converges to $\Gamma(\theta)$, it is natural to conjecture that, for any $\theta \in \Theta_I$, the moment inequality (24) is satisfied in the limit when $k \to \infty$. This conjecture is confirmed later in Theorem 3.

In general cases where both Condition C3 and C4 could be violated, we define the regularized model in the following way.

Definition 2. Given model (Γ, r), we say (Γ', r) is a regularized model of (Γ, r) if, for any $\theta \in \Theta$,

(i) $\Gamma'(\theta) \subseteq \Gamma(\theta)$

(ii) Condition C1-C4 hold for model (Γ', r).

Furthermore, let $\{(\Gamma_k, r) : k \geq 1\}$ be a sequence of regularized models of (Γ, r). We say that (Γ_k, r) converges to (Γ, r) if, for any $\theta \in \Theta$,

(i) for any $k \geq 1$, $\Gamma_k(\theta) \subseteq \Gamma_{k+1}(\theta)$.

(ii) $\Gamma(\theta) = \cup_{k \geq 1}\Gamma_k(\theta)$.

The following theorem now formalizes and extends our previous discussions on $(\Gamma_{k,\delta}, r)$ to more general cases. Its proof can be found in Appendix D.

Theorem 3. *Suppose all θ in Θ satisfy Conditions C1-C2 for model (Γ, r). Let $\{(\Gamma_k, r) : k \geq 1\}$ be a sequence of regularized models of (Γ, r).*

(i) If (Γ_k, r) converges to (Γ, r), then for any $\theta \in \Theta_I$, we have

$$\lim_{k\to\infty} \inf_{\lambda \in S_{dr}} \mathbb{E}\left[\sup_{u \in \Gamma_k(Z;\theta)} \lambda' r(u, Z; \theta)\right] \geq 0. \tag{25}$$

(ii) $\theta \in \Theta_I$ *if there exists some* $k \geq 1$ *such that*

$$\inf_{\lambda \in S_{dr}} \mathbb{E}\left[\sup_{u \in \Gamma_k(Z;\theta)} \lambda' r(u, Z; \theta)\right] \geq 0. \tag{26}$$

Although Theorem 3 only applies to the regularized models, it actually implies both Theorem 1 and 2. To see this, note that for any $k \geq 0$, we have $\Gamma_k \subseteq \Gamma$ so that for any $\lambda \in S_{dr}$,

$$\mathbb{E}\left[\sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta)\right] \geq \mathbb{E}\left[\sup_{u \in \Gamma_k(Z;\theta)} \lambda' r(u, Z; \theta)\right].$$

As a result, (25) implies (16) so that Result (i) in Theorem 3 implies Theorem 1. Furthermore, if we set $\Gamma_k = \Gamma$, Result (ii) in Theorem 3 trivially implies Theorem 2.

Moreover, Theorem 3 also implies a set of necessary conditions and a set of sufficient conditions under which Condition (16) is a sharp characterization of the identified set.

Corollary 1. *Suppose all* θ *in* Θ *satisfy Condition C1-C2 for model* (Γ, r)*. Let* $\{(\Gamma_k, r) : k \geq 1\}$ *be a sequence of regularized models which converges to* (Γ, r)*. Let* $\widetilde{\Theta}$ *be the set of parameters which satisfy Condition* (16),

$$\forall \lambda \in S_{dr}, \;\; \mathbb{E}\left[\sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta)\right] \geq 0. \qquad \text{(16) revisited}$$

(i) $\Theta_I = \widetilde{\Theta}$ *implies that for any* $\varepsilon > 0$ *and any* θ *satisfying* (16), *there exists some* $k > 0$ *such that*

$$\forall \lambda \in S_{dr}, \;\; \mathbb{E}\left[\sup_{u \in \Gamma_k(Z;\theta)} \lambda' r(u, Z; \theta)\right] \geq -\varepsilon. \tag{27}$$

(ii) $\Theta_I = \widetilde{\Theta}$ *if for each* θ *satisfying* (16), *there exists some* $k \geq 0$ *such that*

$$\forall \lambda \in S_{dr}, \;\; \mathbb{E}\left[\sup_{u \in \Gamma_k(Z;\theta)} \lambda' r(u, Z; \theta)\right] \geq 0. \tag{28}$$

In the following, I show how Theorem 3 and Corollary 1 can be applied to a concrete example.

Example 1 (continued)**.** Let's now revisit Case 2 in Example 1 where Condition C4 is violated. Recall that, in this case, we have Γ defined in (20) and r defined in (22). Consider Γ_k defined by

$$\Gamma_k(\theta) = \Gamma(\theta) \cap \{(u, z) : \|r(u, z)\| \leq |x\beta - \alpha| + k\} \tag{29}$$

Then, one can verify that Conditions C1-C4 hold for each (Γ_k, r).

For any $k \geq 1$, one can show that the moment inequalities (28) in this case can be simplified to the following restrictions:

$$\begin{aligned}
&\mathbb{E}[\mathbb{1}(Y_i = 1)(X_i\beta - \alpha)] \geq 0 \\
&\mathbb{E}[\mathbb{1}(Y_i = 0)(X_i\beta - \alpha)] \leq 0 \\
&\mathbb{E}[\mathbb{1}(Y_i = 1)(X_i\beta - \alpha)] \leq k \\
&\mathbb{E}[\mathbb{1}(Y_i = 0)(X_i\beta - \alpha)] \geq -k.
\end{aligned} \tag{30}$$

Compared to the moment inequalities (23) derived from the original model, we have two extra inequalities in (30) which involve k. Based on Corollary 1, a sufficient condition for (23) to be a sharp characterization is that, for each parameter (α, β), (23) implies (30) for a large enough k.

Now, it's easy to see that, for any $k \geq \mathbb{E}[|X_i\beta - \alpha|]$, the last two inequalities in (30) always hold and, hence, are redundant. As a result, (23) and (30) are equivalent for any $k \geq \mathbb{E}[|X_i\beta - \alpha|]$. This proves that (23) is a sharp characterization of the identified set in this case.

Similarly, by applying Corollary 1, one can show Condition (16) also sharply characterize the identified set in Case 3 of Example 1. ■

3.4 Enlarged Identified Set

There are two ways to understand the identification property of Condition (16) derived in this paper. One way is to study conditions under which Condition (16) is a sharp characterization of the identified set, as I did in previous sections. Another way is to study what parameter value actually satisfy Condition (16) and what properties they have without imposing any restrictive assumptions, which is the topic of this section.

Let's start from the following weaker notion of identified set, which is the same as the identified set in Definition 1 except that the moment restrictions are slightly relaxed.

Definition 3 (ε-enlarged identified set)**.** For any $\varepsilon > 0$, define Θ_I^ε as the set of all $\theta \in \Theta$ such that there exists a joint distribution H of (U, Z) which satisfies

(i) $\mathbb{P}_H(U \in \Gamma(Z; \theta)) = 1$

(ii) $\|\mathbb{E}_H r(U, Z; \theta)\| \leq \varepsilon$

(iii) H's marginal distribution for Z equals F_Z.

In addition, define $\Theta_I' := \bigcap_{\varepsilon > 0} \Theta_I^\varepsilon$.

By the definition of Θ_I^ε, for any $\varepsilon_1 < \varepsilon_2$, we have $\Theta_I^{\varepsilon_1} \subseteq \Theta_I^{\varepsilon_2}$. Therefore, Θ_I' could be viewed as the limit of Θ_I^ε. It turns out that Θ_I' is actually equal to the set of all parameters which satisfy Condition (16), as shown in the following theorem. Its proof is in Appendix C.

Theorem 4. *Suppose Conditions C1 and C2 hold for every $\theta \in \Theta$. Then, $\theta \in \Theta_I'$ if and only if θ satisfies (16).*

Note that Theorem 4 should not be interpreted as a sharpness result. Although based on Theorem 2 and Corollary 1, we already know that Θ_I' can equal Θ_I under some regularity conditions, Θ_I' could be very different from Θ_I in some other cases. For some model (Γ, r), it is possible to transform it into another equivalent model $(\widetilde{\Gamma}, \tilde{r})$ such that $\Theta_I(\Gamma, r) = \Theta_I(\widetilde{\Gamma}, \tilde{r})$ but $\Theta_I'(\Gamma, r) \neq \Theta_I'(\widetilde{\Gamma}, \tilde{r})$. I first illustrate this point using our leading example, and then discuss its implications.

Example 1 (continued)**.** In what follows, I revisit Case 2 in Example 1. I'm going to show three different representations of this model. Θ_I does not vary with the representation, but Θ_I' does.

Representation 1 In previous discussions, we have been using the following definition of Γ and r in Case 2:

$$\Gamma_1(\theta) = \{(u, z) : (-1)^y[(x+v)\beta - \alpha] \leq 0\} \text{ and } r_1(u, z) = \begin{pmatrix} \mathbb{1}(y=1)\beta\nu \\ \mathbb{1}(y=0)\beta\nu \end{pmatrix}.$$

As shown in the previous section, Condition (16) is a sharp characterization of the identified set in this case. Therefore, Theorem 4 implies $\Theta_I(\Gamma_1, r_1) = \Theta_I'(\Gamma_1, r_1)$. (Note that, to emphasize the underlying model, I write $\Theta_I(\Gamma_1, r_1)$ for Θ_I and $\Theta_I'(\Gamma_1, r_1)$ for Θ_I'.)

Representation 2 Alternatively, one might view the support restriction as one of the moment restrictions, which results in the following choice for (Γ_2, r_2):

$$\Gamma_2(\theta) = \mathbb{R}^2 \times \{0,1\} \text{ and } r_2(u, z) = \begin{pmatrix} \mathbb{1}\Big((-1)^y[(x+v)\beta - \alpha] \leq 0\Big) - 1 \\ \mathbb{1}(y=1)\beta\nu \\ \mathbb{1}(y=0)\beta\nu \end{pmatrix}.$$

where $\mathbb{R}^2 \times \{0,1\}$ is the entire space of (u, z) or (ν, x, y). In this representation, the support restriction $\mathbb{P}[(U, Z) \in \Gamma_2(\theta)] = 1$ always holds and does not have any identifying power. All restrictions are now summarized in the moment condition $\mathbb{E}[r_2(U, Z)] = 0$. As (Γ_1, r_1) and (Γ_2, r_2) are two representations of the same model, we have $\Theta_I(\Gamma_1, r_1) = \Theta_I(\Gamma_2, r_2)$. However, by solving Condition (16) for this model and applying Theorem 4, one can show that $\Theta_I'(\Gamma_2, r_2)$ equals the entire parameter space, i.e. $\Theta_I'(\Gamma_2, r_2) = \Theta$.

Representation 3 In this representation, we perform a change of variable. Recall the agent's payoff $\pi = (x+\nu)\beta - \alpha$. Therefore, we have $\beta\nu = \pi - (x\beta - \alpha)$. Furthermore, by the fact that $(-1)^y\pi \leq 0$ holds with probablity 1, we know $\beta\nu = \mathbb{1}(y=1)|\pi| - \mathbb{1}(y=0)|\pi| - (x\beta - \alpha)$. This motivates the following representation,

$$\Gamma_3(\theta) = \mathbb{R}^2 \times \{0,1\}, \text{ and } r_3(\pi, z) = \begin{pmatrix} \mathbb{1}(y=1)(|\pi| - (\beta x - \alpha)) \\ \mathbb{1}(y=0)(-|\pi| - (\beta x - \alpha)) \end{pmatrix}.$$

where π is now the only unobserved random variable in the model. As in the second representation, the support restriction $\mathbb{P}[(\pi, Z) \in \Gamma_3] = 1$ trivially holds and all possible identifying power comes from the moment condition $\mathbb{E}[r_3(\pi, Z)] = 0$. Yet, in this case, by applying Corollary 1 and Theorem 4, one can show that $\Theta_I'(\Gamma_3, r_3) = \Theta_I(\Gamma_3, r_3)$.

In the above, we have three equivalent representations of the same model. In the first and the third representation, we have $\Theta_I = \Theta_I'$, whereas Θ_I and Θ_I' differ a lot in the second representation. These results are consistent with our previous findings in Corollary 1. By constructing a sequence of regularized models as in (29), one can show that the sufficient conditions in Corollary 1(ii) for $\Theta_I = \Theta_I'$ are satisfied in Representations 1 and 3, whereas the necessary conditions in Corollary 1(i) for $\Theta_I = \Theta_I'$ are violated in Representation 2. ■

Therefore, Theorem 4 should not be interpreted as a sharpness result. As illustrated in the previous example, Θ_I' may be different for different representations of the same model. Technically, this is due to the fact that the set of all probability measures of U and Z is in general not compact in the topology of weak convergence. To see this, let $\{\varepsilon_n\}$ be any positive sequence converging to zero. For any $\theta \in \Theta_I'$, there exists a sequence of probability measures $\{H_n\}$ satisfying $\|\mathbb{E}_{H_n} r(U, Z; \theta)\| \leq \varepsilon_n$ and the other two conditions in Definition 3. However, without imposing more regularity conditions, $\{H_n\}$ need not be uniformly tight. Therefore, even if $\mathbb{E}_{H_n}[r(U, Z; \theta)]$ converges to zero, there still might not be a distribution H satisfying $\mathbb{E}_H[r(U, Z; \theta)] = 0$.

However, Theorem 4 does shed some light on the identification property of Condition (16). It suggests that the support restriction should not be treated as one of the moment restrictions. To make this point concrete, for any model (Γ, r), define model $(\widetilde{\Gamma}, \tilde{r})$ as

$$\widetilde{\Gamma}(\theta) = \mathcal{U} \times \mathcal{Z} \quad \text{and} \quad \tilde{r}(u, z; \theta) = \begin{pmatrix} \mathbb{1}\Big((u, z) \in \Gamma(\theta)\Big) - 1 \\ r(u, z; \theta) \end{pmatrix} \tag{31}$$

where $\mathcal{U}$ and $\mathcal{Z}$ are the space of random vectors U and Z respectively. By the definition of Θ_I', one can show that $\Theta_I'(\Gamma, r) \subseteq \Theta_I'(\widetilde{\Gamma}, \tilde{r})$. Then, Theorem 4 implies that Condition (16) always generates weakly tighter bounds on θ in model (Γ, r) than that in model $(\widetilde{\Gamma}, \tilde{r})$. When Conditions C1-C4 or sufficient conditions in Corollary 1(ii) hold for model $(\widetilde{\Gamma}, \tilde{r})$, we have $\Theta_I(\Gamma, r) = \Theta_I(\widetilde{\Gamma}, \tilde{r}) = \Theta_I'(\Gamma, r) = \Theta_I'(\widetilde{\Gamma}, \tilde{r})$. However, those conditions can be quite restrictive. For example, when both r and $\{u : (u, z) \in \Gamma(\theta)\}$ are unbounded for some fixed z, Conditions C1-C4 for model $(\widetilde{\Gamma}, \tilde{r})$ require the support of U to be compact. Moreover, when the necessary condition in Corollary 1(i) is violated for model $(\widetilde{\Gamma}, \tilde{r})$, we could still have the sufficient condition in Corollary 1(ii) holding for model (Γ, r), in which case $\Theta_I(\Gamma, r) = \Theta_I(\widetilde{\Gamma}, \tilde{r}) = \Theta_I'(\Gamma, r) \subsetneqq \Theta_I'(\widetilde{\Gamma}, \tilde{r})$. In the latter case, the difference between $\Theta_I'(\Gamma, r)$ and $\Theta_I'(\widetilde{\Gamma}, \tilde{r})$ can be considerable, as illustrated in the previous example.

3.5 Relation with the Entropy Based Approach

An entropy based approach has been proposed in Schennach (2014). To compare my approach with Schennach (2014), it's helpful to restate her theorem in my notation. In order to apply the method in Schennach (2014), one has to choose a dominating measure μ_θ on (U, Z) for each θ, where μ_θ could be θ-dependent. Define $\mu \coloneqq \{\mu_\theta : \theta \in \Theta\}$. Suppose the user-specified μ satisfies the following assumption.

Assumption S. *Suppose for any $\theta \in \Theta$,*

(i) μ_θ's marginal distribution for Z equals F_Z. (Recall F_Z stands for the distribution of Z identified in the data.)

(ii) For any $\lambda \in \mathbb{R}^{dr}$,

$$\int \log\left(\mathbb{E}_{\mu_\theta}[\exp(\lambda' r(U, z; \theta))|Z = z]\right) \mathrm{d}F_Z(z)$$

is finite and differentiable with respect to λ.

(iii) $\|r\|$ is an integrable function with respect to μ_θ.

(iv) $\mathbb{P}_{\mu_\theta}[(U, Z) \in \Gamma(\theta)] = 1$.

Assumption S is a restatement of Definition 2.2 in Schennach (2014) when there is no effective support restriction, i.e. $\Gamma(\theta) = \mathcal{U} \times \mathcal{Z}$. (Recall that $\mathcal{U}$ and $\mathcal{Z}$ are the spaces of U and Z respectively.) If a probability distribution H is absolutely continuous with respect to μ_θ, we write $H \ll \mu_\theta$. In the following, I define the set $\Theta'_{I,\mu}$ as the enlarged identified set with respect to the dominating measure μ.

Definition 4 (ϵ-enlarged μ-identified set)**.** Given any dominating measure μ_θ which satisfies Assumption S and any $\epsilon > 0$, define $\Theta^\epsilon_{I,\mu}$ as the set of all $\theta \in \Theta$ such that there exists a probability measure H of (U, Z) which satisfies

(i) $\mathbb{P}_H[(U, Z) \in \Gamma(\theta)) = 1$,

(ii) $\|\mathbb{E}_H r(U, Z; \theta)\| \leq \epsilon$,

(iii) H's marginal distribution for Z equals F_Z,

(iv) $H \ll \mu_\theta$.

Moreover, define $\Theta'_{I,\mu} \coloneqq \bigcap_{\epsilon>0} \Theta^\epsilon_{I,\mu}$.

Compared to the enlarged identified set Θ'_I defined in Definition 3, the dominated enlarged identified set $\Theta'_{I,\mu}$ defined in Definition 4 requires H to be dominated by μ_θ, i.e. $H \ll \mu_\theta$ in Condition (iv). This condition is stronger than that the support of H is included in the support of μ_θ. For example, if one assumes μ_θ to be the standard normal distribution, then $H \ll \mu_\theta$ implies that H cannot have any mass point.

A generalized version of the identification results in Schennach (2014) is stated in the following theorem.

Theorem 5 (Schennach (2014))**.** *Suppose Condition C1-C2 holds for each $\theta \in \Theta$. Then, for any dominating measure μ which satisfies Assumption S, $\theta \in \Theta'_{I,\mu}$ if and only if*

$$\inf_{\lambda \in \mathbb{R}^{d_r}} \left\| \mathbb{E}_{H_{\lambda,\theta}} r(U, Z; \theta) \right\| = 0, \tag{32}$$

where the Radon-Nikodym derivative of $H_{\lambda,\theta}$ with respect to μ_θ is defined by

$$h_{\lambda,\theta}(u, z) = \frac{\exp(\lambda' r(u, z; \theta))}{\mathbb{E}_{\mu_\theta}[\exp(\lambda' r(U, z; \theta))|Z = z]}. \tag{33}$$

Remark 2. When μ_θ is the same for all θ and there is no effective support restriction i.e. $\Gamma(\theta) = \mathcal{U} \times \mathcal{Z}$, Theorem 5 is the same as Theorem 2.1 in Schennach (2014) and equation (32) is equivalent to Schennach (2014)'s identification condition (i.e. equation (6) therein). Although not explicitly stated in Theorem 2.1 in Schennach (2014), Condition (iv) in Definition 4 is in fact necessary for Schennach (2014)'s identification result. Counterexamples can be constructed when Condition (iv) is removed. See Appendix E.1 for a concrete example. When the support restriction is effective, i.e. $\Gamma(\theta) \subsetneq \mathcal{U} \times \mathcal{Z}$, Theorem 5 formalizes the discussion in Section 4.1 in Schennach (2014). I provide a proof for Theorem 5 in Appendix E for completeness. □

In theory, Theorem 5 differs from my results in two aspects. First, since Θ'_I and $\Theta'_{I,\mu}$ are different in general, Theorem 4 and 5 imply that my approach and the approach in Schennach (2014) characterize two different sets of parameters. Θ'_I only depends on the structures defined in (Γ, r), whereas $\Theta'_{I,\mu}$ also depends on the choice of dominating measure due to its extra dominance restriction. This dominance restriction could be helpful, for example, if applied researchers have prior knowledge that the distribution of U has no mass point. But, as discussed later, it could also cause pitfalls when implementing the entropy based method.

Second, $\Theta'_{I,\mu}$ is also an enlarged identified set. Conditions in the definition of $\Theta'_{I,\mu}$ do not require distribution H and parameter θ to satisfy moment restrictions exactly. One can define a notion of identified set analogous to Θ_I when using the entropy based approach.

Definition 5 (μ-identified set)**.** Given any dominating measure μ_θ which satisfies Assumption S, define $\Theta_{I,\mu}$ as the set of all $\theta \in \Theta$ such that there exists a probability measure H of (U, Z) which satisfies

(i) $\mathbb{P}_H[(U, Z) \in \Gamma(\theta)) = 1$,

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

(iii) H's marginal distribution for Z equals F_Z,

(iv) $H \ll \mu_\theta$.

Just as Θ_I and Θ_I' could be very different as discussed in Section 3.4, the enlarged identified set $\Theta_{I,\mu}'$ could be much larger than the identified set $\Theta_{I,\mu}$. Theorem 2.1 in Schennach (2014) and Theorem 5 in this paper did not establish the equivalence between that $\theta \in \Theta_{I,\mu}$ and that θ satisfies (32). Also, Schennach (2014) did not distinguish the difference between support restrictions and moment restriction. But as pointed out in Section 3.4, it's generally a good idea to treat these two restrictions differently. This could also be true for entropy based approach, but, to the best of my knowledge, there is no such theoretical analysis for entropy based approach in the literature as I've done in the previous sections. Nor does there exist any formal results showing sufficient or necessary conditions for $\Theta_{I,\mu}' = \Theta_{I,\mu}$.

Let me now discuss the difference between the two approaches in terms of numerical computation. When implementing the entropy based approach, one often approximates the dominating measure μ_θ by some discrete distribution $\hat{\mu}_\theta$ whose support for U conditional on Z is discrete and finite. This is usually done by sampling from μ_θ's conditional distribution for U given Z. One can then easily calculate the expectation in (32) with μ_θ replaced by $\hat{\mu}_\theta$. In this way, one attempts to approximate $\Theta_{I,\mu}'$ by $\Theta_{I,\hat{\mu}}'$.

The same approximation idea could be used to implement my method. Let $\text{supp}(\hat{\mu}_\theta, z)$ be $\hat{\mu}_\theta$'s support for U conditional on $Z = z$. One can then approximate (16) by the following moment inequality,

$$\forall \lambda \in S_{dr}, \;\; \mathbb{E}\left[\max\{\lambda' r(u, Z; \theta) : u \in \Gamma(Z;\theta) \cap \text{supp}(\hat{\mu}_\theta, Z)\}\right] \geq 0. \tag{34}$$

The max inside the expectation can then be easily calculated, since $\text{supp}(\hat{\mu}_\theta, z)$ has a finite number of elements. To see why approximation (16) makes sense, note that $\Theta_I'(\Gamma \cap \text{supp}(\hat{\mu}), r) = \Theta_{I,\hat{\mu}}'$ whenever the $\text{supp}(\hat{\mu}_\theta, Z)$ is discrete. Theorem 4 then implies that $\theta \in \Theta_{I,\hat{\mu}}'$ if and only if θ satisfies (34).

Though easy to implement in practice, this approximation idea does have some pitfalls. For any two dominance measures μ_θ and η_θ, if $\mu_\theta \ll \eta_\theta$ and $\eta_\theta \ll \mu_\theta$, then $\Theta_{I,\mu}' = \Theta_{I,\eta}'$ in theory. In practice, however, the support of distributions $\hat{\mu}_\theta$ and $\hat{\eta}_\theta$, which approximates μ_θ and η_θ respectively, can be very different. For example, if the μ_θ's and η_θ's conditional distribution for U given Z are $N(0,1)$ and $N(10,1)$, and $\hat{\mu}_\theta$ and $\hat{\eta}_\theta$ are generated by sampling k random points from μ_θ and η_θ respectively. Then, it's likely that the support of $\hat{\mu}_\theta$ and $\hat{\eta}_\theta$ can be contained in two disjoint intervals even when k is large. As a result, $\Theta_{I,\hat{\mu}}'$ and $\Theta_{I,\hat{\eta}}'$ could look very different. In this way, this approximation idea makes the computation results sensitive to the choice of dominating measures, even for those which yield equivalent results in theory. In practice, this can lead to nonrobust and misleading results for empirical analysis. In addition, there is no formal results in the literature which establish that $\Theta_{I,\hat{\mu}}'$ would converges to $\Theta_{I,\mu}'$ when $\hat{\mu}(\theta)$ converges to μ_θ. In Section 5, I provide an alternative method which is more reliable and sometimes more efficient.

4 Counterfactual Analysis

In this section, I show how the identification approach developed in the previous section 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's first look at the following example.

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

$$\Gamma(\theta) = \{(z_i, u_i) : (-1)^{y_i}[(x_i + \nu_i)\beta - \alpha + \epsilon_i] \leq 0\}, \qquad \text{(2) revisited}$$

where y_i is agent i's choice, x_i stands for the covariates realized *ex post*, ν_i is agent i's expectation error and ϵ_i is her payoff shock. Here, $u_i = (\nu_i, \epsilon_i)$ stands for all unobservable variables, $z_i = (y_i, x_i)$ collects all observables and $\theta \coloneqq (\alpha, \beta)$ stands for all parameters. In addition, suppose we impose the moment restrictions in (3).

Let's now consider a counterfactual setting in which parameter α changes to a hypothetical value $\tilde{\alpha}$. Let $\widetilde{Y}_i$ be agent i's choice in this counterfactual. Suppose we are interested in the counterfactual choice probability $\tilde{p}$ defined as $\tilde{p} = \mathbb{E}[\mathbb{1}(\widetilde{Y}_i = 1)]$. Given the fact that agent i's counterfactual choice $\widetilde{Y}_i$ is not observed, how can we 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

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

In other words, if we let $\tilde{u}_i \coloneqq (\nu_i, \epsilon_i, \tilde{y}_i)$ be the collection of all unobserved variables including the counterfactual choice, the new support restriction can be written as

$$\widetilde{\Gamma}(\theta) = \{(z_i, \tilde{u}_i) : (-1)^{y_i}[(x_i + \nu_i)\beta - \alpha + \epsilon_i] \leq 0 \text{ and } (-1)^{\tilde{y}_i}[(x_i + \nu_i)\beta - \tilde{\alpha} + \epsilon_i] \leq 0\}.$$

Moreover, we can treat $\tilde{p}$ as one of the model primitives and view $\tilde{p} = \mathbb{E}[\mathbb{1}(\widetilde{Y} = 1)]$ as one of the moment restrictions. That is, let $\tilde{\theta} \coloneqq (\alpha, \beta, \tilde{p})$ be the collection of all parameters including the counterfactual parameter $\tilde{p}$. The new set of moment conditions can then be written as $\mathbb{E}[\tilde{r}(\widetilde{U}, Z; \tilde{\theta})] = 0$, where

$$\tilde{r}(\tilde{u}, z; \tilde{\theta}) = \begin{pmatrix} \mathbb{1}(\tilde{y}_i = 1) - \tilde{p} \\ \mathbb{1}(y_i = 1, \epsilon_i \leq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big) \\ \mathbb{1}(y_i = 1, \epsilon_i \geq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big) \\ \mathbb{1}(y_i = 0, \epsilon_i \leq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big) \\ \mathbb{1}(y_i = 0, \epsilon_i \geq 0)\big(\mathbb{1}(\nu_i \geq 0) - \mathbb{1}(\nu_i \leq 0)\big) \\ \mathbb{1}(\epsilon_i \geq 0) - \mathbb{1}(\epsilon_i \leq 0) \end{pmatrix}. \tag{35}$$

The moment function $\tilde{r}$ here is the result of combining $\mathbb{E}[\mathbb{1}(\widetilde{Y} = 1) - \tilde{p}] = 0$ with the existing moment conditions in (18).

Theorem 2 then implies that the identified set for $\tilde{\theta}$ is characterized by Condition (16) with (Γ, r) replaced by $(\widetilde{\Gamma}, \tilde{r})$. In fact, when $\beta \neq 0$, one can show that, if X_i is a continuous random variable and $\tilde{\alpha} \geq \alpha$,[6] Condition (16) of model $(\widetilde{\Gamma}, \tilde{r})$ can be simplified to the following two sets of moment inequalities,

$$\mathbb{E}[\Lambda_1 Q] \geq 0 \text{ and } \mathbb{E}[\Lambda_2 Q + \gamma \tilde{p}] \geq 0, \tag{36}$$

where Λ_1 is a 2×5 matrix defined by

$$\Lambda_1 = \begin{pmatrix} -1 & 3 & 1 & 1 & 1 \\ 1 & 1 & 3 & 3 & -1 \end{pmatrix},$$

Λ_2 is a 4×5 matrix and γ is a 4×1 vector, each of which is defined by

$$\Lambda_2 = \begin{pmatrix} -1 & 3 & -1 & 3 & 3 \\ 1 & 1 & -1 & 3 & 3 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 7 & 3 & 3 \end{pmatrix}, \quad \gamma = \begin{pmatrix} 4 \\ 4 \\ -1 \\ -4 \end{pmatrix},$$

and Q is a 5×1 vector of indicators defined by

$$Q = \begin{pmatrix} \mathbb{1}(Y_i = 0, X_i\beta - \alpha > 0) \\ \mathbb{1}(Y_i = 0, X_i\beta - \alpha < 0) \\ \mathbb{1}(Y_i = 1, X_i\beta - \alpha > 0, X_i\beta - \tilde{\alpha} > 0) \\ \mathbb{1}(Y_i = 1, X_i\beta - \alpha > 0, X_i\beta - \tilde{\alpha} < 0) \\ \mathbb{1}(Y_i = 1, X_i\beta - \alpha < 0). \end{pmatrix}.$$

Note that the first part of (36), $\mathbb{E}[\Lambda_1 Q] \geq 0$, is actually equivalent to the moment inequalties (19) we get in the original model without considering the counterfactual. The second part of

[6] Analogous moment inequalities also exists when $\tilde{\alpha} \leq \alpha$.

the inequality in (36), $\mathbb{E}[\Lambda_1 Q + \gamma \tilde{p}] \geq 0$, provides lower and upper bounds for the counterfactual parameter $\tilde{p}$. With (36) 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,

$$\mathbb{E}[g(\widetilde{Y}_i, U_i, Z_i; \theta, \tilde{p})] = 0. \tag{37}$$

This moment condition usually holds by the definition of $\tilde{p}$ itself 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. We can now define a new support restriction $\mathbb{P}[(\widetilde{U}_i, Z_i) \in \widetilde{\Gamma}(\theta)] = 1$ based on the original restrictions as well as the restrictions on the counterfactuals, i.e.

$$\widetilde{\Gamma}(\theta) := \{(\tilde{u}_i, z_i) : (u_i, z_i) \in \Gamma(\theta) \text{ and } \tilde{y}_i \in \mathcal{C}(u_i, z_i; \theta)\}. \tag{38}$$

Finally, let $\tilde{\theta} := (\theta, \tilde{p})$ be the collection of both structural and counterfactual parameters. Then, we can 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 (37) and defining

$$\tilde{r}(\tilde{u}_i, z_i; \tilde{\theta}) = \begin{pmatrix} g(\tilde{y}_i, u_i, z_i; \tilde{\theta}) \\ r(u_i, z_i; \theta) \end{pmatrix}. \tag{39}$$

One can then view $\tilde{\theta}$ as a model primitive and apply the identification method in Section 3 to $(\widetilde{\Gamma}, \tilde{r})$. Depending on the goal of the empirical analysis, Condition (16) 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 model predictions with the simulated sample to recover the counterfactual parameters. Such 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 unobervables, I directly utilize the restrictions on the unobservables in the original data. Heuristically, if we were able to observe U_i in the data, counterfactual analysis would be straightforward and there would be no need to simulate the unobservables. In practice, we don't observe U_i in the data, but what we actually observe puts 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.

5 Core Determining Class

In general cases, condition (16) consists of a continuum of moment inequalities. Although one could adopt the inference procedures in Chernozhukov, Lee and Rosen (2013) and Andrews and Shi (2017) to do inference based on (16), these test procedures could be hard to implement when the dimension of r is relative large. In this section, I propose one way to simplify (16) into finite dimensional moment inequalities with little or no loss of identification power.

Let's first define the notion of core determining class, which is first introduced in Galichon and Henry (2011).

Definition 6 (core determining class)**.** We call a subset Λ in $\mathbb{R}^{dr}$ a core determining class for model (Γ, r), if for any distribution F of Z and any $\theta \in \Theta$, the following two conditions,

$$\forall \lambda \in S_{dr}, \ \ \mathbb{E}_F\left[\sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta)\right] \geq 0 \tag{40}$$

and

$$\forall \lambda \in \Lambda, \ \ \mathbb{E}_F\left[\sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta)\right] \geq 0. \tag{41}$$

are equivalent. We call a subset Λ in $\mathbb{R}^{dr}$ a *minimal* core determining class if none of its proper subsets are core determining classes.

Here, (40) is the same as (16) except the expectation is now taken with respect to an arbitrary distribution F instead of the true distribution of Z.[7] By definition, any set which

[7] One can also define a F_Z-specific core determining class by restricting the F in (40) to be F_Z. Finding a minimal F_Z-specific core determining class would require a data-dependent procedure similar to the general moment selection

contains the unit sphere S_{d_r} is a core-determining class, but this is of little use. Ideally, one would like to find a minimal core determining class.

However, in general, finding a minimal core determining class can be hard. There are some related results in Galichon and Henry (2011), Chesher, Rosen and Smolinski (2013), Chesher and Rosen (2017) and Luo and Wang (2017), but the problem here is more challenging, as the identification condition in (16) involves a continuum of moment inequalities. Therefore, I take a different approach from the literature. Instead of finding a minimal core determining class for a general model, I first approximate a general model (Γ, r) by a discretized model $(\Gamma, r^\dagger)$ in which the moment function $r^\dagger$ is discrete and only takes a finite number of possible values. Then, I develop a method to find a minimal core determining class $\Lambda^\dagger$ for this discretized model $(\Gamma, r^\dagger)$. Finally, I show that one does not lose too much information by using $\Lambda^\dagger$ in the identification conditions of the original model (Γ, r). This idea is formalized in the following section.

5.1 Approximation

Let $\widetilde{\Lambda}^\dagger$ be a core determining class for model $(\Gamma, r^\dagger)$. Define $\Lambda^\dagger := \{\lambda / \|\lambda\| : \lambda \in \widetilde{\Lambda}^\dagger \text{ and } \lambda \neq 0\}$. After normalization, $\Lambda^\dagger$ is a subset of S_{d_r} and is still core determining for model $(\Gamma, r^\dagger)$. Suppose a model (Γ, r) can be approximated well enough by $(\Gamma, r^\dagger)$. Then, the following proposition shows that we lose little by using $\Lambda^\dagger$ in model (Γ, r)'s identification conditions.

Proposition 1. *Suppose Conditions C1-C2 holds for all $\theta \in \Theta$ in model (Γ, r) and $(\Gamma, r^\dagger)$. Suppose there exists some $\epsilon > 0$ such that*

$$\forall \theta \in \Theta,\ \forall (u, z) \in \Gamma(\theta), \quad \left\| r(u, z; \theta) - r^\dagger(u, z; \theta) \right\| \leq \epsilon.$$

Let $\Lambda^\dagger$ be some core determining class for model $(\Gamma, r^\dagger)$ with $\Lambda^\dagger \subseteq S_{d_r}$. Define $\widetilde{\Theta}^\dagger$ to be the set of θ which satisfies

$$\forall \lambda \in \Lambda^\dagger, \quad \mathbb{E}\left[\sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta) \right] \geq 0. \tag{42}$$

Then, $\Theta_I \subseteq \widetilde{\Theta}^\dagger \subseteq \Theta_I^\epsilon$, where Θ_I^ϵ is defined in Definition 3.

When Conditions C1-C4 or conditions in Corollary 1(ii) hold, Theorem 4 implies that Θ_I^ϵ converges to Θ_I as ϵ converges to 0. As a result, Proposition 1 implies a trade off between the power of (42) and the computational complexity. the difference between $\widetilde{\Theta}^\dagger$ and Θ_I can be made arbitrary small, by selecting $r^\dagger$ to be a good enough approximation for r. At the same time, as the approximation becomes finer and finer, it generally become more and more computationally demanding to find a minimal core determining class for model $(\Gamma, r^\dagger)$.

mechanism in Andrews and Soares (2010), which is beyond the scope of this paper.

Given the result in Proposition 1, we only need to find minimal core determining classes for a special class of models, which can approximate a general model and, at the same time, whose minimal core determining class is easy to find. This is the topic of the next section.

5.2 Minimal Core Determining Classes in Discrete Models

In this section, I assume the moment function r can be written in the following form,

$$r(u, z; \theta) = m(u, z; \theta) + \psi(\theta) \tag{43}$$

where $m(u, z; \theta)$ is a discrete function and only takes a finite number of possible values, and $\psi(\theta)$ is a bounded function which only depends on θ. If we define $\mathcal{M} \coloneqq \{m(u, z; \theta) : \theta \in \Theta, (u, z) \in \Gamma(\theta)\}$ as the image of $m(u, z; \theta)$, then the cardinality $|\mathcal{M}|$ of $\mathcal{M}$ should be finite. In previous example, moment functions in (18) and (35) take this form. Models whose moment function r takes the form of (43) can approximate any general model whose moment function is bounded within the support of (U, Z) and the parameter space Θ.

Let's now start to solve a minimal core determining class. For each z and θ, define $\mathcal{M}(z; \theta) \coloneqq \{m(u, z; \theta) : u \in \Gamma(z; \theta)\}$. By definition, we have

$$\sup\{\lambda' r(u, z; \theta) : u \in \Gamma(z; \theta)\} = \max\{\lambda' t : t \in \mathcal{M}(z; \theta)\} + \lambda' \psi(\theta). \tag{44}$$

Since $\mathcal{M}(z; \theta)$ is a nonempty subset of $\mathcal{M}$, $\mathcal{M}(z; \theta)$ can take at most $2^{|\mathcal{M}|} - 1$ different values. Enumerate all the possible values of $\mathcal{M}(z; \theta)$ as $\{\mathcal{M}_1, ..., \mathcal{M}_K\}$, so that, for any z and θ, we have $\mathcal{M}(z; \theta) = \mathcal{M}_k$ for some $k \in \{1, ..., K\}$. For any distribution F of Z, define $p_{k,F}(\theta) \coloneqq \mathbb{P}_F[\mathcal{M}(Z; \theta) = \mathcal{M}_k]$. Then, (44) implies that (40) is equivalent to the following condition:

$$\forall \lambda \in S_{dr}, \quad \sum_{k=1}^{K} p_{k,F}(\theta) \max\{\lambda' t : t \in \mathcal{M}_k\} + \lambda' \psi(\theta) \geq 0. \tag{45}$$

Since $\max\{\lambda' t : t \in \mathcal{M}_k\}$ is positively homogeneous with respect to λ, i.e.

$$\forall \alpha > 0, \max\{(\alpha\lambda)' t : t \in \mathcal{M}_k\} = \alpha \max\{\lambda' t : t \in \mathcal{M}_k\},$$

we know that (45) is also equivalent to

$$\forall \lambda \in [-1, 1]^{dr}, \quad \sum_{k=1}^{K} p_{k,F}(\theta) \max\{\lambda' t : t \in \mathcal{M}_k\} + \lambda' \psi(\theta) \geq 0.$$

Define $v_k \coloneqq \max\{\lambda' t : t \in \mathcal{M}_k\}$, $v = (v_1, ..., v_K)$ and $p_F(\theta) = (p_{1,F}(\theta), ..., p_{K,F}(\theta))$. Then, the above condition can be rewritten as the following inequality which involves a linear programming problem:

$$\begin{aligned} 0 \quad \leq \quad & \inf \ p_F(\theta)'v + \lambda'\psi(\theta) \\ & \text{s.t. } v_k \geq \lambda' t, \quad \forall k = 1, ..., K, \ \forall t \in \mathcal{M}_k, \\ & \quad \lambda \in [-1,1]^{dr}. \end{aligned} \tag{46}$$

Up to now, I have established the equivalence between (40) and (46). To construct a core determining class, let $\mathcal{P}$ stand for the following polyhedron,

$$\mathcal{P} \coloneqq \{(v,\lambda) : \lambda \in [-1,1]^{dr}, v_k \geq \lambda' t, \forall k = 1, ..., K, \forall t \in \mathcal{M}_k\}. \tag{47}$$

Note that $\mathcal{P}$ is constructed based on the knowledge of $\mathcal{M}$, and it does not depend on $p_F(\theta)$ or $\psi(\theta)$. Let $\mathcal{V}$ be the set of all extreme points of $\mathcal{P}$. Since $p_{k,F}(\theta) \geq 0$ for any k and F, we know the infimum in (46) must be finite. By the properties of linear programming, this implies that the infimum in (46) can always be achieved by points within $\mathcal{V}$. Therefore, (40) is equivalent to the following finite collection of moment inequalities,

$$\forall (v,\lambda) \in \mathcal{V}, \quad \mathbb{E}_F\left[\sum_{k=1}^{K} \mathbb{1}(\mathcal{M}(Z;\theta) = \mathcal{M}_k)v_k + \lambda'\psi(\theta)\right] \geq 0. \tag{48}$$

Define $\mathcal{V}_\lambda$ to be the projection of $\mathcal{V}$ onto the space of λ. Then, $\mathcal{V}_\lambda$ is a finite core determining class for model (Γ, r).

Let $\Delta_K \coloneqq \{p \in \mathbb{R}^K : \sum_k p_k = 1 \text{ and } p_k \geq 0, \ \forall k = 1, ..., K\}$ be the K-dimensional simplex. Let $\Psi \coloneqq \{\psi(\theta) : \theta \in \Theta\}$ be the image of ψ. Then, by the definition of $p_F(\theta)$ and $\gamma(\mathcal{V})$, (48) is satisfied if and only if

$$\big(p_F(\theta), \psi(\theta)\big) \in \big\{(p,t) \in \Delta_K \times \Psi : p'v + t'\lambda \geq 0, \forall (v,\lambda) \in \mathcal{V}\big\}.$$

Define $\gamma(\mathcal{V}) \coloneqq \{(p,t) \in \Delta_K \times \Psi : p'v + t'\lambda \geq 0, \forall (v,\lambda) \in \mathcal{V}\}$. If $\mathcal{V}$ is not a minimal core determining class, one can find some redundant point (v',λ') in $\mathcal{V}$ so that $\gamma(\mathcal{V})$ remains the same after (v',λ') is removed, i.e., $\gamma(\mathcal{V}) = \gamma(\mathcal{V} \setminus \{(v',\lambda')\})$. One can keep removing these redundant points until we find a minimal subset $\mathcal{V}^*$ such that $\gamma(\mathcal{V}^*) = \gamma(\mathcal{V})$ and, for any proper subset $\mathcal{V}'$ of $\mathcal{V}^*$, $\gamma(\mathcal{V}') \neq \gamma(\mathcal{V})$. Then, $\mathcal{V}^*$'s projection onto the space of λ is a minimal core determining class.

The following proposition summarizes the the above derivation.

Proposition 2 (core determining class). *Suppose Conditions C1-C2 holds for any θ in model (Γ, r) and r is in the form of (43). For any subset $\mathcal{V}'$ of $\mathcal{V}$, define $\mathcal{V}'_\lambda$ as the projection of $\mathcal{V}'$ onto the space of λ. Then,*

(i) if a subset $\mathcal{V}'$ of $\mathcal{V}$ satisfies $\gamma(\mathcal{V}') = \gamma(\mathcal{V})$, then $\mathcal{V}'_\lambda$ is a core determining class.

(ii) if a subset $\mathcal{V}^*$ of $\mathcal{V}$ is a minimal subset which satisfies $\gamma(\mathcal{V}^*) = \gamma(\mathcal{V})$, then $\mathcal{V}^*_\lambda$ is a minimal core determining class.

(iii) if a subset $\mathcal{V}'$ of $\mathcal{V}$ satisfies $\gamma(\mathcal{V}') = \gamma(\mathcal{V})$, then Condition (16) in Section 3.2 is equivalent to

$$\forall (v, \lambda) \in \mathcal{V}', \quad \mathbb{E}\left[\sum_{k=1}^{K} \mathbb{1}(\mathcal{M}(Z;\theta) = \mathcal{M}_k) v_k + \lambda' \psi(\theta)\right] \geq 0.$$

In practice, constructing a minimal core determining class based on Proposition 2 involves two major steps: (i) find the set $\mathcal{V}$ of all extreme points in polyhedron $\mathcal{P}$; and (ii) construct $\mathcal{V}^*$ from $\mathcal{V}$ by removing all redundant points.

In computational geometry, Step (i) is often called the vertex enumeration problem. Since polyhedron $\mathcal{P}$ is unbounded, this problem is NP-hard as shown in Khachiyan, Boros, Borys, Gurvich and Elbassioni (2009). Moreover, in the worst case, K could be as large as $2^{|\mathcal{M}|} - 1$, which makes the problem even harder. However, in some applications, K can be much smaller than $2^{|\mathcal{M}|} - 1$. For example, for r defined in (35), we have $|\mathcal{M}| = 27$ and $K = 5$. In addition, there exists several efficient implementations of the vertex enumeration algorithms. For example, see `Parma Polyhedra Library` in Bagnara, Hill and Zaffanella (2008) for a single thread implementation, and `mplrs` in Avis and Jordan (2018) for a multi-thread implementation.

When Ψ is a polyhedron as in (18) and (35), Step (ii) can be implemented by the algorithm in Clarkson (1994). In the worst case, it needs to solve $|\mathcal{V}|$ linear programming problems, where $|\mathcal{V}|$ is the cardinality of $\mathcal{V}$. In practice, it often takes less time to complete Step (ii) compared to Step (i). A good implementation can be found in `Parma Polyhedra Library`. When Ψ is not a polyhedron, finding a minimal core determining class is not easy. What we can do in this case is to remove all redundant points in $\mathcal{V}$ so that there is no redundant linear constrain on (p, t) in the following,

$$\begin{aligned} \forall (v, \lambda) \in \mathcal{V}, \qquad & p'v + t'\lambda \geq 0, \\ \forall k = 1, ..., K, \qquad & p_k \geq 0, \\ \text{and} \qquad & \sum_k p_k = 1. \end{aligned}$$

Let $\mathcal{V}^\dagger$ be the resulting minimal subset of $\mathcal{V}$. Then, $\mathcal{V}^\dagger_\lambda$ is always core determining, though it may not be minimal.

6 Applications

6.1 Exporting Decision with Limited Information

My first empirical example uses the setting in Dickstein and Morales (2018) (hereafter, "DM"), which examines the export decisions of Chilean firms with weak assumptions on firms' information set. The goal here is two-fold. First of all, I study the sharp identified set under different sets of assumptions. By comparing those results, we can see which key assumptions have to be made to get informative empirical results. Secondly, I illustrate how counterfactual analysis can be conducted using the method developed in previous sections.

The empirical model studied in DM is very similar to Example 1. In the benchmark model, DM assume that the profit of exporter i exporting to country j at period t is

$$\pi_{ijt} = \beta X_{ijt} - \alpha_0 - \alpha_1 dist_{ij} + \epsilon_{ijt},$$

where X_{ijt} is firm i's exporting revenue, βX_{ijt} stands for firm i's exporting revenue net of production costs, $dist_{ij}$ stands for the geographic distance between firm i's home country and country j, and ϵ_{ijt} is the unobserved heterogeneity. DM normalize $\beta = 0.2$ and treat (α_0, α_1) as parameters to be estimated. Firm i observes ϵ_{ijt} and $dist_{ij}$, but it does not observe X_{ijt} when making exporting decisions. Instead, firm i forms a subjective expectation $\mathbb{E}_s[X_{ijt}|\mathcal{I}_{it}]$ based on its information set $\mathcal{I}_{it}$ at time t. Assume $\mathbb{E}_s[\pi_{ijt}|\mathcal{I}_{it}] \geq 0$ if firm i exports, and $\mathbb{E}_s[\pi_{ijt}|\mathcal{I}_{it}] \leq 0$ if it doesn't.

Following DM, I assume we do not observe $\mathbb{E}_s[\pi_{ijt}|\mathcal{I}_{it}]$ or $\mathcal{I}_{it}$. Instead, we observe some instrument W_{ijt} within $\mathcal{I}_{ijt}$. In DM, W_{ijt} includes firm i's exporting revenue at time $t-1$, the aggregate exports from firm i's home country to country j at time $t-1$ and the distance $dist_{ij}$. As DM constructs X_{ijt} from firm j's sales revenue in its home country, X_{ijt} is observable to us for all firms.

As in Example 1, let Y_{ijt} be the exporting decision, $Z_{ijt} \coloneqq (Y_{ijt}, X_{ijt}, dist_{ij}, W_{ijt})$ be the collection of the observables, $\nu_{ijt} \coloneqq \mathbb{E}_s[X_{ijt}|\mathcal{I}_{it}] - X_{ijt}$ be firm i's expectation error, and $U_{ijt} = (\nu_{ijt}, \epsilon_{ijt})$ be the collection of the unobservables. Whenever there is no confusion, I omit i, j, t in the subscript for ease of notation.

In this example, the support restriction can be written as $\mathbb{P}[(Z,U) \in \Gamma(\theta)] = 1$, where

$$\Gamma(\theta) = \{(z,u) : (-1)^y[(x+\nu)\beta - \alpha_0 - \alpha_1 dist + \epsilon] \leq 0\}. \tag{49}$$

I consider the moment restrictions under the following two sets of assumptions.

AS1: Assume (i) ν has zero median conditional on ϵ, $\mathbb{E}[X|\mathcal{I}]$ and W, (ii) ϵ has zero median conditional on W, and (iii) assume the sign of ϵ is independent of $\mathbb{E}[X|\mathcal{I}]$ conditional on W. Let $\mathcal{A}_0 = (-\infty, 0]$ and $\mathcal{A}_1 = [0, \infty)$. Since $\mathbb{E}[X|\mathcal{I}] = X + \nu$ by definition, we can write **AS1**

as $\mathbb{E}[r_1(U,Z;\theta)|W]=0$ where $r_1=(r_{1,k,k'}:k,k'\in\{0,1\})$ and

$$r_{1,k,k'}(u,z;\theta)=\begin{pmatrix}[\mathbb{1}(\epsilon\in\mathcal{A}_k)-0.5]\cdot\mathbb{1}(x+\nu\in\mathcal{A}_{k'})\\ y\cdot\mathbb{1}(\epsilon\in\mathcal{A}_k\text{ and }x+\nu\in\mathcal{A}_{k'})\cdot(\mathbb{1}(\nu\geq 0)-\mathbb{1}(\nu\leq 0))\\ (1-y)\cdot\mathbb{1}(\epsilon\in\mathcal{A}_k\text{ and }x+\nu\in\mathcal{A}_{k'})\cdot(\mathbb{1}(\nu\geq 0)-\mathbb{1}(\nu\leq 0))\end{pmatrix}. \tag{50}$$

The first function in (50) stems from condition (ii) and (iii) in **AS1**, and the last two functions are due to condition (i) in **AS1**.

AS2: Assume (i) ν has zero median conditional on ϵ, $\mathbb{E}[X|\mathcal{I}]$ and W, (ii) ϵ has a Normal distribution $N(0,\sigma^2)$, (iii) ϵ is independent of $\mathbb{E}[X|\mathcal{I}]$ and W. Let K be some integer and $\{\mathcal{B}_1,...,\mathcal{B}_K\}$ be K intervals which partition $\mathbb{R}$. Then, **AS2** implies $\mathbb{E}[r_2(U,Z;\theta)|W]=0$ where $r_2=(r_{2,k,k'}:k,k'\in\{1,\cdots,K\})$ and

$$r_{2,k,k'}(u,z;\theta)=\begin{pmatrix}[\mathbb{1}(\epsilon\in\mathcal{B}_k)-p_k]\cdot\mathbb{1}(x+\nu\in\mathcal{B}_{k'})\\ y\cdot\mathbb{1}(\epsilon\in\mathcal{B}_k\text{ and }x+\nu\in\mathcal{B}_{k'})\cdot(\mathbb{1}(\nu\geq 0)-\mathbb{1}(\nu\leq 0))\\ (1-y)\cdot\mathbb{1}(\epsilon\in\mathcal{B}_k\text{ and }x+\nu\in\mathcal{B}_{k'})\cdot(\mathbb{1}(\nu\geq 0)-\mathbb{1}(\nu\leq 0))\\ \mathbb{1}(x+\nu\in\mathcal{B}_{k'})(y-\Phi[\sigma^{-1}(\beta(x+\nu)-\alpha_0-\alpha_1 dist)])\end{pmatrix} \tag{51}$$

where p_k is equal to $\mathbb{P}_{N(0,\sigma^2)}(\epsilon\in\mathcal{B}_k)$. The last function in (51) is due to the fact that the conditional choice probability $\mathbb{E}[y|W,\mathbb{E}(X|\mathcal{I})]$ equals $\Phi[\sigma^{-1}(\beta(x+\nu)-\alpha_0-\alpha_1 dist)]$ under **AS2**. In the results reported below, I set $K=8$.

The identification conditions in DM build on the same assumptions as in **AS2** except that they assume ν has zero mean instead of zero median. I write their assumptions in the following for comparison.

DM: Assume (i) ν has zero mean conditional on ϵ, $\mathbb{E}[X|\mathcal{I}]$ and W (ii) ϵ is has a Normal distribution $N(0,\sigma^2)$, (iii) ϵ is independent of $\mathbb{E}[X|\mathcal{I}]$ and W. Under the assumptions, DM derive the following moment inequalities,

$$\mathbb{E}[m(Z;\theta)]\geq 0,\ \text{where } m(z;\theta)=\begin{pmatrix}y(1-\Phi(\Delta))/\Phi(\Delta)-(1-y)\\ (1-y)\Phi(\Delta)/(1-\Phi(\Delta))-y\\ -(1-y)\Delta+y\phi(\Delta)/\Phi(\Delta)\\ y\Delta+(1-y)\phi(\Delta)/\Phi(\Delta)\end{pmatrix}, \tag{52}$$

and where $\Delta=\sigma^{-1}(\beta x-\alpha_0-\alpha_1 dist)$, and ϕ is the p.d.f. of the standard normal distribution.

Since I don't have access to the original data in DM, I simulate a sample based on the data statistics and estimation results reported in DM. The simulation is designed as follows: (a) In the simulation, I set the value of (α_0,α_1) and σ to the middle point of their reported confidence interval; (b) I assume there are two periods, period 0 and 1. (c) Fix $t=1$, the distribution of $X_{ij,t-1}$ is the Fréchet distribution whose c.d.f. $F(x)=\exp(-Tx^{-\gamma})$. I set $\gamma=4$ and calibrate the value of T so that the resulting export probability matches that of the

Chilean chemical industry in Year 2000, reported in Table 1 of DM; (d) Firm i's expectation $\mathbb{E}_s[X_{ijt}|\mathcal{I}_{it}]$ equals $X_{ij,t-1}$; (e) ν follows the Normal distribution $N(0, \sigma_\nu^2)$ where $\sigma_\nu = 0.5\sigma$. (f) The sample size is the same as than in DM. See Appendix H.1 for more details.

Table 1 summarizes the confidence interval for the structural parameter $(\alpha_0, \alpha_1, \sigma)$. The results under Assumption **AS1** and **AS2** are based on the support restriction with $\Gamma(\theta)$ defined in (49) and moment restrictions with r defined in (50) and (51) respectively. The result based on DM's moment inequality (52) is reported in the last column. I construct the confidence interval for each parameter by projecting the confidence region as in Andrews and Soares (2010). See Appendix G for a way of computing the identified set using support vector machines.

Table 1: Confidence Interval for Structural Parameters

Parameters	True Value	AS 1	AS 2	DM
α_0	72.0	$[0, 150.1]$	$[0, 129.3]$	$[0, 389.0]$
α_1	168.4	$[55.6, 324.9]$	$[91.6, 323.2]$	$[0, 1071.7]$
σ	100.5	$-$	$[64.6, 173.9]$	$[0, 127.6]$

Next, I consider a counterfactual exercise in which the fixed exporting cost increases by \$10,000. In the following, I show how to construct an augmented model $(\widetilde{\Gamma}, \tilde{r})$ to identify the counterfactual export probability.

Let $\widetilde{Y}_{ijt}$ be the counterfactual export decision, and $\widetilde{U}_{ijt} = (\nu_{ijt}, \epsilon_{ijt}, \widetilde{Y}_{ijt})$ be the collection of all unobservables including the counterfactuals. Then, the support restriction can now be written as $\mathbb{P}[(Z, \widetilde{U}) \in \widetilde{\Gamma}(\theta)] = 1$, where

$$\begin{aligned}\widetilde{\Gamma}(\theta) = \{(z, \tilde{u}) : & (-1)^y[(x+\nu)\beta - \alpha_0 - \alpha_1 dist + \epsilon] \leq 0 \\ & (-1)^{\tilde{y}}[(x+\nu)\beta - \tilde{\alpha}_0 - \alpha_1 dist + \epsilon] \leq 0 \quad \}.\end{aligned} \tag{53}$$

In this exercise, when estimating the counterfactual export probability, I only consider exporters with export distance ($dist_{ij}$) between 40% and 50% quantiles. For any interval $\mathcal{C}$ of $X_{ij,t-1}$, let $\tilde{p}(\mathcal{C})$ the be counterfactual export probability for firms whose $X_{ij,t-1}$ is within $\mathcal{C}$. To ease the notation, I use X_{-1} to denote $X_{ij,t-1}$. By the definition of $\tilde{p}(\mathcal{C})$, we have

$$\mathbb{E}[(\tilde{Y} - \tilde{p}(\mathcal{C}))\mathbb{1}(X_{-1} \in \mathcal{C})] = 0 \tag{54}$$

When imposing assumptions in **AS1**, I can now append (54) to the moment restrictions defined in (50) as the new set of moment restrictions. When imposing assumptions in **AS2**, I append (54) and the following

$$\mathbb{E}[\mathbb{1}(X+\nu \in \mathcal{B}_{k'})(\widetilde{Y} - \Phi[\sigma^{-1}(\beta(X+\nu) - \tilde{\alpha}_0 - \alpha_1 dist)])|W] = 0$$

to the moment restrictions defined in (51). Finally, when imposing assumptions in **AS3**, DM derive bounds for the counterfactual export probability in Theorem 3.

In Table 2, I report the confidence interval for $\tilde{p}(\mathcal{C})$ where $\mathcal{C}$ is intervals of $20\% - 30\%$ quantiles, $50\% - 60\%$ quantiles and $70\% - 80\%$ quantiles of export revenues at $t-1$.

This exercise has the following two implications: (i) By the results under **AS1** and **AS2** in Table 1 and 2, I find that the normality assumption on ϵ in **AS2** plays a limited role in the estimation of the structural parameters, but is essential to provide an informative lower bound for the counterfactual export probability; (ii) The results generated by the moment inequalities in DM are not very informative. One reason for this is that their identification conditions are not sharp. Another reason is that function $m(z;\theta)$ in (52) involves the ratio between $1-\Phi(\Delta)$ and $\Phi(\Delta)$ so that the variance of $m(Z;\theta)$ can be very large.[8]

Table 2: Confidence Interval for Counterfactual Export Probability

quantiles of revenue $X_{ij,t-1}$	True Value	AS 1	AS 2	DM
$20\% - 30\%$ quantile	0.23	$[0.08, 0.26]$	$[0.14, 0.26]$	$[0, 1]$
$50\% - 60\%$ quantile	0.32	$[0, 0.37]$	$[0.16, 0.35]$	$[0, 1]$
$70\% - 80\%$ quantile	0.41	$[0, 0.47]$	$[0.29, 0.45]$	$[0, 1]$

6.2 Entry Game with Complete Information

In this section, I revisit Example 2 to study the identification conditions under semiparametric assumptions and to compare the identified set under the different types of assumptions.

Example 2 (continued). Recall that, in Example 2, I've assumed the following support restriction $\mathbb{P}[(U_m, Z_m) \in \Gamma(\theta)] = 1$ where

$$\Gamma(\theta) := \left\{(u_m, z_m) : \forall i = 1, ..., I,\ (-1)^{y_{i,m}}(\pi_i(y_{-i,m}, x_{i,m}; \theta) + u_{i,m}) \leq 0\right\},$$

and where $y_{i,m}$ is firm i's entry decision in market m, $x_{i,m}$ is firm i's observed characteristics

[8] In fact, one can show the variance of $m(Z;\theta)$ is infinite if the distribution of export revenues is Pareto or any other fat-tailed distribution. In the simulation setting, I let the export revenue follow the Fréchet distribution under which the variance of $m(Z;\theta)$ is finite in theory but is still very large in practice, especially with small values of σ.

and $u_{i,m}$ is the unobserved heterogeneity of firm i. Recall also that

$$\pi_i(y_{-i,m}, x_{i,m}; \theta) = x'_{i,m}\alpha_i - \sum_{k \neq i} \Delta_k \cdot y_{k,m}.$$

is the mean utility of player i when choosing $y_{i,m} = 1$. In addition, I impose the following zero median assumptions on U_m, that for each player i,

$$\mathbb{E}[\mathbb{1}(U_{i,m} \geq 0) - \mathbb{1}(U_{i,m} \leq 0) | X_m] = 0. \tag{55}$$

Since the moment functions in (55) consist of indicator functions, Conditions C1-C4 are satisfied. Conditional on a value x_m of X_m, Theorem 2 then implies that θ is in the identified set if and only if

$$\forall \lambda \in S_{dr}, \ \mathbb{E}\left[\sup_{u \in \Gamma(Z;\theta)} \lambda' r(u, Z; \theta) \ \Big| \ X_m = x_m\right] \geq 0,$$

where dr equals the number of players and $r = (r_i : i = 1, ..., I)$ with $r_i(u_m, z_m; \theta) = \mathbb{1}(u_{i,m} \geq 0) - \mathbb{1}(u_{i,m} \leq 0)$. One can further simplify the above conditions to the following moment inequalities that for each player $i = 1, ..., I$ and for almost every X_m,

$$\begin{array}{l} \mathbb{E}[\mathbb{1}(Y_{i,m} = 0, \ \pi_i(Y_{-i,m}, X_{i,m}; \theta) > 0) - 0.5 \mid X_m] \leq 0 \\ \mathbb{E}[\mathbb{1}(Y_{i,m} = 1, \ \pi_i(Y_{-i,m}, X_{i,m}; \theta) < 0) - 0.5 \mid X_m] \leq 0 \end{array}. \tag{56}$$

One nice property of (56) is that its evaluation does not require solving the set of all Nash equilibria, whose computational complexity increases exponentially in the number of players. This is in contrast to the identification conditions under Normality assumptions. For example, under the assumption that the conditional distribution of U_m given X_m is $N(0, \Sigma)$, Ciliberto and Tamer (2009) derived that, for any $y \in \{0, 1\}^I$ and almost every X_m,

$$\begin{aligned} &\int \mathbb{1}(\{y\} = NE(X_m, u_m)) \ \mathrm{d}\Phi_\Sigma(u_m) \\ &\qquad \leq \mathbb{P}(Y_m = y | X_m) \leq \int \mathbb{1}(y \in NE(X_m, u_m)) \ \mathrm{d}\Phi_\Sigma(u_m), \end{aligned} \tag{57}$$

where $NE(x_m, u_m)$ stands for the set of Nash equilibria given (x_m, u_m) and Φ_Σ is the probability measure of $N(0, \Sigma)$.

Let Θ_I^{ZC} be the set of parameters which satisfies (56), and let Θ_I^N be the set of parameters which satisfies (57). To illustrate the difference between Θ_I^{ZC} and Θ_I^N and to see the how these assumptions and the data variation jointly shapes the identified set, I design a simulation experiment in which the data generating process (DGP) is the following:

(i) There are two players, i.e. $I = 2$.

(ii) For each $i = 1, 2$, $U_{i,m}$ follows a t-distribution with degree of freedom d.

(iii) For each $i = 1, 2$, $X_{i,m}$ is scalar and the support of $X_{i,m}$ is K evenly spaced points within interval $[0, 5]$. In total, X_m has K^2 support points.

(iv) $U_{1,m}$, $U_{2,m}$, $X_{1,m}$ and $X_{2,m}$ are mutually independent.

(v) When there is more than one pure-strategy Nash equilibrium, each equilibrium occurs with equal probability.

(vi) For each $i = 1, 2$, I normalize $\alpha_i = 1$ and let $\Delta_i = 1$.

This DGP can be indexed by (d, K), where d changes the underlying distribution for $U_{i,m}$ and K controls how rich the data variation is. I compute the identified set for Θ_I^{ZC} and Θ_I^N for different DGPs. When computing both identified sets, I normalize $\alpha_i = 1$. I also treat Σ as a nuisance parameters when computing Θ_I^N. Table 3 reports the identified set for (Δ_1, Δ_2) in Θ_I^{ZC} and Θ_I^N. [9]

Table 3: Identified Set under Different DGPs

DGP Settings		Θ_I^{ZC}	Θ_I^N
$d = 1$	$k = 2$	$[0.00, 5.00]^2$	$\emptyset$
$d = 1$	$k = 20$	$[0.79, 1.32]^2$	$\emptyset$
$d = 1$	$k = 100$	$[0.91, 1.16]^2$	$\emptyset$
$d = 10$	$k = 2$	$[0.00, 5.00]^2$	$\emptyset$
$d = 10$	$k = 20$	$[0.79, 1.05]^2$	$\emptyset$
$d = 10$	$k = 100$	$[0.96, 1.01]^2$	$\emptyset$
$d = \infty$	$k = 2$	$[0.00, 5.00]^2$	$\{(1, 1)\}$
$d = \infty$	$k = 20$	$[0.79, 1.05]^2$	$\{(1, 1)\}$
$d = \infty$	$k = 100$	$[0.96, 1.01]^2$	$\{(1, 1)\}$

First, when $d < \infty$, $U_{i,m}$ follows a t distribution instead of a Normal distribution. That is, moment inequalities in (57) are misspecified. Therefore, it is not surprising that Θ_I^N is the empty set when $d < \infty$. However, Θ_I^N becomes a singleton as long as $d = \infty$, even if there is very little data variation. Note that, when $k = 2$, $X_{i,m}$ has the minimum data variation: it can only take two possible values. This suggests that the identification result hinges on whether or not the Normality assumption is correct. When the Normality assumption is correct, the model is point identified even if there is little data variation. When the Normality assumption

[9] It is nontrivial to check whether Θ_I^N is indeed empty or singleton. In Appendix H.2, I develop an algorithm to handle this issue.

is misspecifed, no parameter can rationalize the data. The extent of data variation seems to play little role in this example.

Next, when $k = 2$, Θ_I^{ZC} is not very informative. It should not be surprising as there is little data variation when $k = 2$. However, as k increases, the support of $X_{i,m}$ becomes richer and Θ_I^{ZC} gradually becomes more informative.

In practice, when one works with finite sample size, the situation may be less startling. The confidence set of Θ_I^N is less likely to be nonempty even if it is misspecified. Also, the empirical result under the zero median assumption may or may not be informative. However, given the result in Table 3, it is interesting to study how the empirical results change over various sets of semiparametric/parametric assumptions and see the tradeoff between robustness and informativeness in the real data. I leave this for future work. ■

7 Extension to Moment Inequality Restrictions

As illustrated in Example 3, the moment restrictions in some interesting applications sometimes take the form of inequalities instead of equalities. In this section, I extend the framework in (1) to the following,

$$\mathbb{P}[(U, Z) \in \Gamma(\theta)] = 0, \quad \mathbb{E}[r_1(U, Z; \theta)] = 0, \text{ and } \mathbb{E}[r_2(U, Z; \theta)] \geq 0. \tag{58}$$

Let dr_1 and dr_2 be the dimension of r_1 and r_2 respectively.

One way to find the identification conditions for models satisfying (58) is to introduce a slackness variable V with $V \in \mathbb{R}_+^{dr_2}$. Let $\widetilde{U} = (U, V)$ and construct the moment function $\tilde{r}$ as

$$\tilde{r}(\tilde{u}, z; \theta) = \begin{pmatrix} r_1(u, z; \theta) \\ r_2(u, z; \theta) - v \end{pmatrix}. \tag{59}$$

Moreover, construct the support restriction as $\mathbb{P}[(\widetilde{U}, Z) \in \widetilde{\Gamma}(\theta)] = 1$ with $\widetilde{\Gamma}$ defined as

$$\widetilde{\Gamma}(\theta) = \{(\tilde{u}, z) : (u, z) \in \Gamma(\theta) \text{ and } v \geq 0\}. \tag{60}$$

Then, the model in (58) is equivalent to $\mathbb{P}[(\widetilde{U}, Z) \in \widetilde{\Gamma}(\theta)] = 1$ and $\mathbb{E}[\tilde{r}(\widetilde{U}, Z; \theta)] = 0$. Define $S_{dr_1,dr_2} = \{(\lambda_1, \lambda_2) \in \mathbb{R}^{dr_1} \times \mathbb{R}_+^{dr_2} : \|(\lambda_1, \lambda_2)\| = 1\}$. To derive its identification conditions, note that Condition (16) for $(\widetilde{\Gamma}, \tilde{r})$ can be simplified to

$$\forall (\lambda_1, \lambda_2) \in S_{dr_1,dr_2}, \ \mathbb{E}\left[\sup_{u \in \Gamma(Z;\theta)} \lambda_1' r_1(u, Z; \theta) + \lambda_2' r_2(u, Z; \theta)\right] \geq 0. \tag{61}$$

Let $r = (r_1, r_2)$. Condition (61) is very similar to Condition (16) for (Γ, r) with moment equality constraints, except that the Lagrange multiplier λ_2 for r_2 only takes nonnegative values in (61).

Due to the presence of the slackness variables, Condition C4 does not hold for $(\widetilde{\Gamma}, \tilde{r})$ even if the norm of both r_1 and r_2 are bounded. Hence, Theorem 2 cannot be applied directly. In the following, however, I establish the sharpness of (61) as a corollary of Theorem 3, the proof of which can be found in Appendix F.

Corollary 2. *Given the model in* (58)*, define its identified set to be the identified set of* $(\widetilde{\Gamma}, \tilde{r})$ *in Definition 1. Let* $r = (r_1, r_2)$*. Suppose all* $\theta \in \Theta$ *satisfies Conditions C1 and C2 applied to* (Γ, r)*. Then, Condition* (61) *holds for any* θ *in the identified set. Suppose, in addition, all* $\theta \in \Theta$ *satisfies Conditions C3 and C4 applied to* (Γ, r)*. Then,* θ *is in the identified set if and only if* θ *satisfies Condition* (61)*.*

Corollary 2 can be viewed as a counterpart of Theorem 1 and 2 for models with moment inequality restrictions. Under the conditions in Theorems 3 and 4, one could also establish similar sharp identification results for the model with moment inequality restrictions. Let us now revisit Example 3.

Example 3 (continued)**.** Recall that, the support restriction in Example 3 can be written as $\mathbb{P}[(U_i, Z_i) \in \Gamma(\theta)] = 1$, where

$$\Gamma(\theta) = \Big\{(u_i, z_i) : \forall t = 1, ..., T,\ y_{it} = \arg\max_{j \in \mathcal{J}}\ u_{ijt}\Big\}.$$

In addition, recall that under the same assumptions in Gao and Li (2018), the following moment inequality restrictions hold: for any two time periods s and t, and any two nonempty subsets $\mathcal{J}_1$ and $\mathcal{J}_2$ of choice set $\mathcal{J}$,

$$\mathbb{E}\left[\rho_{ist}(\mathcal{J}_1, \mathcal{J}_2; \theta)\left\{\mathbb{1}\left(\max_{j\in\mathcal{J}_1} U_{ijs} \geq \max_{j\in\mathcal{J}_2} U_{ijs}\right) - \mathbb{1}\left(\max_{j\in\mathcal{J}_1} U_{ijt} \geq \max_{j\in\mathcal{J}_2} U_{ijt}\right)\right\}\middle| X_i\right] \geq 0, \quad (62)$$

where

$$\begin{aligned} \mathcal{A}_{ist}(\theta) &= \{j \in \mathcal{J} : X_{ijs}'\theta \geq X_{ijt}'\theta\} \\ \mathcal{B}_{ist}(\theta) &= \{j \in \mathcal{J} : X_{ijs}'\theta \leq X_{ijt}'\theta\} \\ \rho_{ist}(\mathcal{J}_1, \mathcal{J}_2; \theta) &= \mathbb{1}\big(\mathcal{J}_1 \subseteq \mathcal{A}_{ist}(\theta) \text{ and } \mathcal{J}_2 \subseteq \mathcal{B}_{ist}(\theta)\big). \end{aligned}$$

One can check that Conditions C1-C4 are satisfied in this example. Let function r collect all the moment functions in (62). Let $S_{dr}^{+} = \{\lambda \in \mathbb{R}_{+}^{dr} : \|\lambda\| = 1\}$. Conditional on one value x_i of X_i, Corollary 2 implies that the identified set of model (Γ, r) in this example is characterized by the following set of moment inequalities,

$$\forall \lambda \in S_{dr}^{+},\ \mathbb{E}\left[\sup_{u_i \in \Gamma(Z_i;\theta)} \lambda' r(u_i, Z_i; \theta)\ \middle|\ X_i = x_i\right] \geq 0. \quad (63)$$

When $T = 2$, I show in Appendix H.4 that the above condition can be further simplified into the following moment inequalities,

$$\forall s, t = 1, ..., T, \quad \mathbb{E}[\mathbb{1}(Y_{is} \in \mathcal{A}_{ist}(\theta)) - \mathbb{1}(Y_{it} \in \mathcal{A}_{ist}(\theta)) \mid X_i] \geq 0. \tag{64}$$

In other words, the moment inequality in (64) is a sharp identification condition. When $T > 2$, the conditions in (64) are still valid but the sharp identification conditions could be more complicated.

To compare with Gao and Li (2018), given any X_i, define $\gamma_{ist}(j;\theta)$ for any $j \in \mathcal{J}$ and any two time periods s, t by, $\gamma_{ist}(j;\theta) = \mathbb{1}(X'_{ijs}\theta \geq X'_{ijt}\theta \text{ and } \forall k \neq j,\ X'_{iks}\theta \leq X'_{ikt}\theta)$. The identification condition in Gao and Li (2018) can be written as the following moment inequalities,

$$\forall j \in \mathcal{J},\ \forall s, t = 1, ..., T, \quad \mathbb{E}[\gamma_{ist}(j;\theta)\,(\mathbb{1}(Y_{is} = j) - \mathbb{1}(Y_{it} = j))\,|X_i] \geq 0,$$

which is nested in (64). In fact, when $X'_{ijt}\theta$ changes across time for all choices, the identification condition in Gao and Li (2018) is the same as (64) except that they only pick up the cases when $\mathcal{A}_{ist}(\theta)$ is singleton, which does not exhaust all the information in the data variation. ■

8 Conclusion

In this paper, I developed a new identification approach for structural models with semiparametric assumptions on the unobserved heterogeneity. It characterizes the identified set for both structural and counterfactual parameters using a set of moment inequalities. I also derive the sufficient and necessary conditions for the sharpness of the procedure. In addition, the results on the enlarged identified set can be helpful to better understand existing results in the literature.

The generality of the framework makes it possible to apply the method to various structural models. In this paper, I worked out the analytic identification conditions for three examples, which may be of independent interest. In the future, for instance, identification conditions in Example 1 could be extended to other structural models with similar information structures. It is also worth revisiting Example 2 to investigate what kinds of semiparametric assumptions achieve the balance between the informativeness and the robustness in practice. The identification constraints derived in Example 3 could also be applied to empirical analysis. Also, the treatment on the monotonicity conditions in Example 3 could be extended to other settings.

The main factor which limits the scope of the empirical applications is the computational issue. The computational procedures in Section 5 can be further improved in the future.

For example, it is unclear what is the best way to construct the discretized approximation. Also, there might be ways to construct a minimal core determining class which is specific to the joint distribution in the data. Although raising some challenging statistical issues, such data dependent core determining classes could decrease the computational complexity even further. Another possibility is to explore minimal core determining classes when (Γ, r) has some special structure, as did in Galichon and Henry (2011) for models with monotonicity structures.

Appendices

A 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 A.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 A.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 A.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 A.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 contains all the results I used to prove our theorems in the paper.

Lemma A.1. *Let Y be a closed random 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 semicontinous , 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 $\mathbb{E}\overline{\mathrm{co}}\ (Y) = \overline{\mathrm{co}}\ \mathbb{E}Y$ where $\overline{\mathrm{co}}$ 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. □

B Selection Theorem

I also need a measurable selection theorem presented later in Lemma B.2.

Definition B.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 B.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. Some basic relation between Borel sets and universally measurable sets are listed in the following lemma.

Lemma B.1.

(i) In a Polish space, every Borel set is universally measurable.

(ii) For a function $f : S \mapsto T$ between Polish spaces, the following statements are equivalent.

- *f is Borel measurable*
- *$\mathrm{Gr}f$ is a Borel subset of $S \times T$, where $\mathrm{Gr}f := \{(s, f(s)) : s \in S\}$ is the graph of f.*

Proof. As for (i), see Corollary 12.27 and Theorem 12.41 in Aliprantis and Border (2007).

As for (ii), see Theorem 12.28 in Aliprantis and Border (2007) □

Given $D \subseteq S \times T$, define $\text{proj}_S(D) := \{s \in S : \exists t \in T, (s,t) \in D\}$ and $D_s := \{t \in T : (t,s) \in D\}$. The following lemma is a simplified version of Proposition 7.50(b) in Bertsekas and Shreve (1978).

Lemma B.2 (measurable selection). *Let S and T be Polish spaces, let $D \subseteq S \times T$ be a Borel set, and le $f : D \to \mathbb{R}$ 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

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

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). □

C Proof Theorems 1, 2 and 4

The proof builds on the theory of random sets. Readers who are not familiar with those concepts are invited to consult Appendix A.

We need the following extra notation: 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. Recall also that F_Z denote the distribution of Z identified in the data. Let the probability space $(Z, \mathscr{Z}, F_Z)$ be the completion of $(Z, \mathscr{B}_{\mathcal{Z}}, F_Z)$. 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, (16) can be rewritten as

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

Finally, let $\widetilde{\Theta}$ be the set of all θ which satisfies (16).

In the following, I first prove Lemma C.1 which establishes some useful properties for $\Upsilon(z;\theta)$ as a random set. Then, I prove Theorem 4 first, and then Theorem 1 and finally Theorem 2.

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

Lemma C.1. *(i) Suppose Condition C1 holds. Then,* cl $\Upsilon(\cdot;\theta)$ *is a random closed set.*

(ii) Suppose Conditions C1 and C2 hold. Then, cl $\Upsilon(\cdot;\theta)$ *is an integrable random closed set.*

(iii) Suppose Conditions C1 and C4 hold. Then, random closed set cl $\Upsilon(\cdot;\theta)$ *is integrably bounded.*

Proof of Lemma C.1. (i) We first show cl $\Upsilon(\cdot;\theta)$ is a random closed set under Condition C1.

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

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

We know 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 B.2 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 B.2 for the definition of a universal measurable function. Since $(\mathcal{Z}, \mathscr{Z}, F_Z)$ is the completion of the Borel probability space $(\mathcal{Z}, \mathscr{B}_Z, F_Z)$, 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 cl $\{r\big(f_{i,n}(z), z\big) : i, n \in \mathbb{N}\} \subseteq$ cl $\Upsilon(z;\theta)$. On the other hand, for any $t \in$ 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 $t \in$ 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, cl $\Upsilon(z;\theta) =$ cl $\{r\big(f_{i,n}(z), z\big) : i, n \in \mathbb{N}\}$. By Theorem 2.3 on page 26 of Molchanov (2005), cl $\Upsilon(z;\theta)$ is a random closed set in $(\mathcal{Z}, \mathscr{Z}, F_Z)$.

(ii) Suppose, in addition, Condition C2 holds. The fact that cl $\Upsilon(z;\theta)$ is a random closed set implies $z \mapsto \inf\{\|t\| : t \in \text{cl } \Upsilon(z;\theta)\}$ is measurable in $(Z, \mathscr{Z})$ (See result (iii) in Lemma

A.1). Moreover, note that

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

Condition C2 then implies $z \mapsto \inf\{\|t\| : t \in \text{cl } \Upsilon(z;\theta)\}$ is an integrable function. By Definition A.4 and Lemma A.1(i), cl $\Upsilon(\cdot;\theta)$ is integrable.

(*iii*) Finally, Condition C4 directly implies cl $\Upsilon(\cdot;\theta)$ is integrably bounded by definition. □

C.2 Proof of Theorem 4

I first state the following two lemmas, the proof of which will be presented after I prove Theorem 4.

Lemma C.2. *If set A is a closed convex set in $\mathbb{R}^d$, then $0 \in A$ if and only if*

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

Lemma C.3. *Suppose Conditions C1 and C2 hold. Then, $0 \in \overline{\text{co}}\ \mathbb{E}\text{cl } \Upsilon(Z;\theta)$ implies $\theta \in \Theta_I'$.*

Proof of Theorem 4. First of all, Lemma C.1 implies that cl $\Upsilon(\cdot;\theta)$ is an integrable random closed set.

Let's now show $\widetilde{\Theta} \subseteq \Theta_I'$. Suppose there exists $\theta \in \widetilde{\Theta}$ such that $\theta \notin \Theta_I'$. Then, by Lemma C.3, $0 \notin \overline{\text{co}}\ \mathbb{E}\text{cl } \Upsilon(Z;\theta)$. Lemma C.2 then implies that the following inequality

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

holds almost surely.

By Lemma A.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

$$\inf_{\lambda \in \mathbb{R}^{dr}} \sup\{\lambda' t : t \in \mathbb{E}_I\overline{\text{co}}\ \Upsilon(Z;\theta)\} < 0 \tag{65}$$

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

$$\sup\{\tilde{\lambda}' t : t \in \mathbb{E}_I\overline{\text{co}}\ \Upsilon(Z;\theta)\} = -\inf_{f \in S^1(\overline{\text{co}}\ \Upsilon(Z;\theta))} \mathbb{E}[-\tilde{\lambda}' f] \tag{66}$$

where S^1 is defined in Definition A.2. Applying Lemma A.1(iii) with $\zeta(t) = -\lambda' t$, we know

$$\begin{aligned} & -\inf_{f \in S^1(\overline{\text{co}}\ \Upsilon(Z;\theta))} \mathbb{E}[-\tilde{\lambda}' f] \\ = & -\mathbb{E}\inf\{-\tilde{\lambda}' t : t \in \overline{\text{co}}\ \Upsilon(Z;\theta)\} \\ = & \ \mathbb{E}\sup\{\tilde{\lambda}' t : t \in \overline{\text{co}}\ \Upsilon(Z;\theta)\}. \end{aligned} \tag{67}$$

Combining equation (66) and (67), we know

$$\mathbb{E}\sup\{\tilde{\lambda}'t : t \in \overline{\text{co}}\ \Upsilon(Z;\theta)\} = \sup\{\tilde{\lambda}'t : t \in \mathbb{E}_I\overline{\text{co}}\ \Upsilon(Z;\theta)\} < 0. \tag{68}$$

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

$$\sup\{\tilde{\lambda}'t : t \in \overline{\text{co}}\ \Upsilon(z;\theta)\} = \sup\{\tilde{\lambda}'t : t \in \Upsilon(z;\theta)\}. \tag{69}$$

Combine equation (68) and (69), we conclude

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

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) $P_H(U \in \Gamma(Z;\theta)) = 1$; (iii) the marginal distribution of H on Z equals to F_Z. For any $\lambda \in \mathbb{R}^{dr}$ with $\|\lambda\| = 1$, we have

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

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,

$$-\epsilon \leq \inf\left\{\mathbb{E}_{F_Z}\left[\sup_{u\in\Gamma(Z;\theta)} \lambda' r(u,Z;\theta)\right] : \lambda \in \mathbb{R}^{dr}, \|\lambda\| = 1\right\}.$$

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

$$0 \leq \inf\left\{\mathbb{E}_{F_Z}\left[\sup_{u\in\Gamma(Z;\theta)} \lambda' r(u,Z;\theta)\right] : \lambda \in \mathbb{R}^{dr}, \|\lambda\| = 1\right\}.$$

Since $\eta(0,Z;\theta) \equiv 0$, we conclude that $\theta \in \widetilde{\Theta}$. □

Proof of Lemma C.2. If $0 \in A$, we know $\sup\{\lambda't : t \in A\} \geq 0$ for any λ. Hence, $0 \in A$ implies

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

Suppose $0 \notin A$, the strict hyperplane separation theorem implies there exists some $\lambda \neq 0$

and $c \in \mathbb{R}$, such that

$$\lambda \cdot 0 > c > \lambda' t, \ \forall t \in A.$$

Therefore, $0 \notin A$ implies

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

This completes the proof. □

Proof of Lemma C.3. Let probability space $(\mathcal{Z}, \mathscr{Z}, F_Z)$ denote the completion of Borel probability space $(\mathcal{Z}, \mathscr{B}_Z, F_Z)$. Under Condition C1 and C2, cl $\Upsilon(Z;\theta)$ is an integrable random closed set in $(\mathcal{Z}, \mathscr{Z}, F_Z)$. Suppose $0 \in \overline{\text{co}}\ \mathbb{E}\text{cl}\ \Upsilon(Z;\theta)$ is true, we want to prove that $\theta \in \Theta_I'$. (In the rest of proof, we write $\mathbb{E}$ for $\mathbb{E}_{F_Z}$ whenever there is no possible confusion.)

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

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

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

Let T be a random variable independent with Z, which is supported on $\{0, 1, ..., dr\}$ and is distributed as the following,

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

Construct random variable $R \in \mathbb{R}^{dr}$ from T and Z as

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

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

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

Also,

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

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

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

Since $\{(z,u) : u \in \Gamma(z)\} \times \mathbb{R}^{dr}$ is a Borel set, and that $(u,z,t) \mapsto \|r(u,z;\theta) - t\|$ is a Borel measurable function in $\mathcal{U} \times \mathcal{Z} \times \mathbb{R}^{dr}$, Lemma B.2 in Appendix B implies that there exists a universally measurable function $g : \mathcal{Z} \times \mathbb{R}^{dr} \mapsto \mathcal{U}$, such that for any $t \in \mathbb{R}^{dr}$ and any $z \in \mathcal{Z}$, $g(z,t) \in \Gamma(z;\theta)$ and

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

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

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

so that

$$\|\mathbb{E}_H r(U,Z;\theta)\| \le \epsilon + \|\mathbb{E}_H R\| \le 2\epsilon$$

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

C.3 Proof of Theorem 1

By the definition of Θ_I and Θ_I', we know $\Theta_I \subseteq \Theta_I'$. Theorem 4 then implies Theorem 1.

C.4 Proof of Theorem 2

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

Lemma C.4. *Suppose Conditions C1-C4 hold. Then,* $0 \in \overline{\text{co}}\ \mathbb{E}\text{cl } \Upsilon(Z;\theta)$ *implies* $\theta \in \Theta_I$.

Proof of Theorem 2. Since we've already proven $\Theta_I \subseteq \widetilde{\Theta}$ in Theorem 1, we only need to prove $\widetilde{\Theta} \subseteq \Theta_I$. To show $\widetilde{\Theta} \subseteq \Theta_I$, suppose, for the purpose of contradiction, there exists some $\theta \in \widetilde{\Theta}$ such that $\theta \notin \Theta_I$. Then, by Lemma C.4, $0 \notin \overline{\text{co}}\ \mathbb{E}\text{cl } \Upsilon(Z;\theta)$. Yet, as shown in the proof of Theorem 4, this contradicts the fact that $\theta \in \widetilde{\Theta}$. □

Proof of Lemma C.4. The proof of this lemma is similar to that of Lemma C.3. One only needs to notice that under Conditions C1-C4, $0 \in \overline{\text{co}}\ \mathbb{E}\text{cl } \Upsilon(Z;\theta)$ 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 $0 \in \overline{\text{co}}\ \mathbb{E}\text{cl}\ \Upsilon(Z;\theta)$, we 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 Condition C3. Moreover, Condition C4 together with Lemma C.1 also implies $\Upsilon(Z;\theta)$ is an integrably bounded random closed set. By Lemma A.1(ii), we know that $\mathbb{E}\Upsilon(Z;\theta)$ is a compact set and $\mathbb{E}\Upsilon(Z;\theta) = \mathbb{E}_I\Upsilon(Z;\theta)$. Since $\mathbb{E}\Upsilon(Z;\theta) \subseteq \mathbb{R}^{dr}$, 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 $p_0, p_1, ..., p_{dr} \in [0,1]$ and $v_0, ..., v_{dr} \in \mathbb{E}_I\Upsilon(Z;\theta)$ such that $\sum_{j=0}^{dr} p_j = 1$ and $\sum_{j=0}^{dr} p_j v_j = 0$.

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

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

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

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

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

Construct random variable $R \in \mathbb{R}^{dr}$ from T and Z as

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

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

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

and

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

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

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

Since $\{(z, u) : u \in \Gamma(z)\} \times \mathbb{R}^{dr}$ is a Borel set, and $(u, z, t) \mapsto \|r(u, z;\theta) - t\|$ is a Borel measurable function in $\mathcal{U} \times \mathcal{Z} \times \mathbb{R}^{dr}$, Lemma B.2 in Appendix B implies that there exists a universally measurable function $g : \mathcal{Z} \times \mathbb{R}^{dr} \mapsto \mathcal{U}$, such that, for any $z \in \mathcal{Z}$ and $t \in \mathbb{R}^{dr}$, $g(z,t) \in \Gamma(z;\theta)$. In addition, for any $z \in \mathcal{Z}$ and $t \in \mathbb{R}^{dr}$ which satisfies

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

we have

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

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 $\theta \in \Theta_I$. $\square$

D Proof of Theorem 3

Proof. I first prove Result(i). Fix any $\theta \in \Theta_I$ and one of its corresponding distributions H in Definition 1. For each $k > 0$, define $\Upsilon_k(z;\theta) \coloneqq \{r(u,z;\theta) : (u,z) \in \Gamma_k(\theta)\}$. As (Γ_k, r) is a regularized model, we know, by Lemma C.1, $\Upsilon_k(Z;\theta)$ is an integrably bounded random closed set. Therefore, there exists some integrable function $t_1(z)$ such that $t_1(z) \in \Upsilon_1(z;\theta)$.

For each $k \geq 1$, construct $g_k : \mathcal{U} \times \mathcal{Z} \mapsto \mathbb{R}^{dr}$ as

$$g_k(u,z) = \begin{cases} r(u,z;\theta) & \text{if } (u,z) \in \Gamma_k(z;\theta) \\ t_1(z) & \text{if } (u,z) \notin \Gamma_k(z;\theta). \end{cases}$$

By construction, for each $\lambda \in S_{dr}$, we have $\sup_{u \in \Gamma_k(z;\theta)} \lambda' r(u,z;\theta) = \sup_{u \in \Gamma(z;\theta)} \lambda' g_k(u,z;\theta)$.

Since (Γ_k, r) converges to (Γ, r), we know g_k converges to r everywhere. Moreover, for each k,

$$\mathbb{E}_H[\|g_k(U,Z) - r(U,Z;\theta)\|] \leq \mathbb{E}[2\|r(U,Z;\theta)\| + \|t(Z)\|] < +\infty.$$

Therefore, by the dominated convergence theorem, $\tau_k \coloneqq \mathbb{E}\|g_k(U,Z) - r(U,Z)\|$ converges to 0 as $k \to \infty$.

For any $\lambda \in S_{dr}$, $\mathbb{E}_H[\lambda' r(U,Z;\theta)] = 0$ so that $\mathbb{E}[\lambda'(g_k(U,Z) - r(U,Z))] = \mathbb{E}[\lambda' g_k(U,Z)]$.

Then, we have

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

where the first inequality comes from Cauchy–Schwarz inequality and $\|\lambda\| = 1$, and the last equality follows from the construction of g_k.

Since τ_k does not depend on λ, we have

$$
-\tau_k \leq \inf_{\lambda\in S_{dr}} \mathbb{E}\left[\sup_{u\in\Gamma_k(Z;\theta)} \lambda' r(U,Z;\theta)\right].
$$

Let $k \to \infty$ and we've proved Result (i).

Finally, Result (ii) follows trivially from Theorem 2 and the fact that $\Theta_I(\Gamma_k, r) \subseteq \Theta_I(\Gamma, r)$.

□

E Results Related to the Entropy Based Approach

In the following, I need some extra notation. Let $\widetilde{\Theta}_\mu$ be the set of all θ which satisfies (32). For any probability measure G, let $\mathcal{L}_1(G)$ denotes the set of all Borel measurable functions which are integrable with respect to G.

E.1 $\Theta'_{I,\mu}$ is μ-dependent

In the following, I provide a simple example which illustrates Condition (iv) in Definition 4 is in fact necessary for Schennach (2014)'s identification result. This example also illustrates that $\Theta'_{I,\mu}$ could depend on user specified dominating measure μ.

Suppose both U and Z are scalars in $\mathbb{R}$. Let $\Gamma(z;\theta) := [z-1, z+1]$, $r(u,z;\theta) := \mathbb{1}(z = u) - \theta$ and F_Z equals standard normal distribution. It's easy to see $\Theta_I = \Theta'_I = [0,1]$.

In this example, one can construct the following dominating measure μ. For each θ, μ_θ's marginal distribution for Z equals F_Z. Conditional on each z, μ_θ's conditional distribution for U equals the uniform distribution on interval $[z-1, z+1]$. By construction, μ_θ remains the same for all θ. With such μ, one can show $\Theta'_{I,\mu} = \widetilde{\Theta}_\mu = \{0\}$. Recall $\widetilde{\Theta}_\mu$ is the set of θ which satisfies (32).

Alternatively, we can construct another dominating measure μ' as follows. For each θ, the marginal distribution of μ'_θ for Z equals F_Z. Moreover, conditional on each z, let the conditional distribution of μ'_θ for U equal the Dirac measure δ_z defined as

$$\forall A \in \mathscr{B}_{\mathcal{U}},\ \delta_z(A) = \mathbb{1}(z \in A).$$

where $\mathscr{B}_{\mathcal{U}}$ is the set of all Borel sets in $\mathcal{U}$. By construction, μ'_θ remains the same for all θ. With such μ', one can show $\Theta'_{I,\mu'} = \widetilde{\Theta}_{\mu'} = \{1\}$.

Finally, if we construct dominating measure μ'' as $\mu''_\theta := 0.5\mu_\theta + 0.5\mu'_\theta$, then $\Theta'_{I,\mu''} = \widetilde{\Theta}_{\mu''} = [0,1]$.

E.2 Proof of Theorem 5

To prove Theorem 5, I need to introduce another definition. Let H and G be two probability measures. The relative entropy $D(H\|G)$ between probability H and G, also known as Kullback-Leibler divergence, is defined as

$$D(H\|G) = \begin{cases} \int h \log(h)\ \mathrm{d}G & \text{if } H \ll G \\ +\infty & \text{if } H \not\ll G \end{cases}$$

where h is H's Radon–Nikodym derivative with respect to G.

Definition 7. Given any dominating measure μ_θ which satisfies Assumption S and any $\epsilon > 0$, define $\Theta^\epsilon_{KL,\mu}$ as the set of all $\theta \in \Theta$ such that there exists a probability measure H of (U, Z) which satisfies

(i) $\mathbb{P}_H[(U, Z) \in \Gamma(\theta)) = 1$,

(ii) $\|\mathbb{E}_H r(U, Z; \theta)\| \leq \epsilon$,

(iii) H's marginal distribution for Z equals F_Z,

(iv) $D(H\|\mu_\theta) < \infty$.

Moreover, define $\Theta'_{KL,\mu} := \bigcap_{\epsilon>0} \Theta^\epsilon_{KL,\mu}$.

Condition (iv) in Definition 7 might look slightly stronger than Condition (iv) in Definition 4, but $\Theta'_{KL,\mu}$ and $\Theta'_{I,\mu}$ are in fact the same according to the following lemma.

Lemma E.1. *Suppose Condition C1-C2 holds for each $\theta \in \Theta$. Then, for any dominating measure μ which satisfies Assumption S, $\Theta'_{KL,\mu} = \Theta'_{I,\mu}$.*

Proof. Since $D(H\|\mu_\theta) < \infty$ implies $H \ll \mu_\theta$, we know $\Theta'_{KL,\mu} \subseteq \Theta'_{I,\mu}$. Therefore, we only need to show that $\Theta'_{I,\mu} \subseteq \Theta'_{KL,\mu}$.

For any $\epsilon > 0$ and any $\theta \in \Theta^\epsilon_{I,\mu}$, let H be a probability measure which satisfies Conditions (i)-(iv) in Definition 4. Let h be H's Radon–Nikodym derivative with respect to G. For any

$k > 0$, define $\tilde{h}_k(u,z)$ as

$$\tilde{h}_k(u,z) := \begin{cases} h(u,z) & \text{if } h(u,z) \le k \\ 0 & \text{if } h(u,z) > k \end{cases}$$

and $h_k(u,z)$ as $h_k(u,z) := \tilde{h}_k(u,z) + (\mathbb{E}_{\mu_\theta}[h(U,Z)|Z=z] - \mathbb{E}_{\mu_\theta}[\tilde{h}_k(U,Z)|Z=z])$. Let H_k be the measure whose density with respect to μ_θ is h_k. Then, H_k satisfies Condition (i), (iii) and (iv) by construction. By the monotone convergence theorem, $\mathbb{E}_{\mu_\theta}[\tilde{h}_k(U,Z)|Z=z] \to \mathbb{E}_{\mu_\theta}[h(U,Z)|Z=z]$ almost surely as $k \to \infty$, which implies that $h_k \to h$ almost surely. Moreover, since $h_k \le h+1$, the dominated convergence theorem implies that as $k \to \infty$,

$$\|\mathbb{E}_H r(U,Z;\theta) - \mathbb{E}_{H_k} r(U,Z;\theta)\| \to 0.$$

Therefore, there exists a large enough K such that for any $k \ge K$,

$$\|\mathbb{E}_{H_k} r(U,Z;\theta)\| \le 2\epsilon.$$

This implies that for any $\epsilon > 0$, $\Theta^{\epsilon}_{I,\mu} \subseteq \Theta^{2\epsilon}_{KL,\mu}$. Hence, $\Theta'_{I,\mu} \subseteq \Theta'_{KL,\mu}$. □

Proof of Theorem 5. Recall $\widetilde{\Theta}_\mu := \{\theta \in \Theta : (32) \text{ is satisifed}\}$. By Lemma E.1, we only need to prove that $\Theta'_{KL,\mu} = \widetilde{\Theta}_\mu$.

First of all, we show $\widetilde{\Theta}_\mu \subseteq \Theta'_{KL,\mu}$. Fix an arbitrary θ in $\widetilde{\Theta}_\mu$. For any $\epsilon > 0$, there exists some $\lambda \in \mathbb{R}^{dr}$ such that

$$\left\|\mathbb{E}_{H_{\lambda,\theta}} r(U,Z;\theta)\right\| \le \epsilon$$

which implies $H_{\lambda,\theta}$ satisfies Condition (ii) in Definition 7. Moreover, by construction, $H_{\lambda,\theta}$ satisfies Condition (i) and (iii) in Definition 7. Finally, note

$$\begin{aligned}
& D(H_{\lambda,\theta}\|\mu_\theta) \\
= & \int \log(h_{\lambda,\theta})\, \mathrm{d}H_{\lambda,\theta} \\
= & \int \lambda' r\, \mathrm{d}H_{\lambda,\theta} - \int \log\left(\mathbb{E}_{\mu_\theta}[\exp(\lambda' r(U,z;\theta))|Z=z]\right) \mathrm{d}H_{\lambda,\theta} \\
= & \lambda' \mathbb{E}_{H_{\lambda,\theta}}[r(U,Z;\theta)] - \int \log\left(\mathbb{E}_{\mu_\theta}[\exp(\lambda' r(U,z;\theta))|Z=z]\right) \mathrm{d}F_Z
\end{aligned}$$

where $h_{\lambda,\theta}$ is the density of $H_{\lambda,\theta}$ with respect to μ_θ, and the last equality follows from $H_{\lambda,\theta}$ satisfying Condition (iii) in Definition 7. Since dominating measure μ satisfies Assumption S,

$$\int \log\left(\mathbb{E}_{\mu_\theta}[\exp(\lambda' r(U,z;\theta))|Z=z]\right) \mathrm{d}F_Z$$

is finite. This result together with the fact that $\left\|\mathbb{E}_{H_{\lambda,\theta}}[r(U,Z;\theta)]\right\| \le \epsilon$, we know $D(H_{\lambda,\theta}\|\mu_\theta)$ must be finite. Hence, Condition (iv) in Definition 7 is also satisfied. Therefore, $\theta \in \Theta'_{KL,\mu}$.

Since θ is an arbitrary element in $\widetilde{\Theta}_\mu$, we conclude $\widetilde{\Theta}_\mu \subseteq \Theta'_{KL,\mu}$.

Next, we show $\Theta'_{KL,\mu} \subseteq \widetilde{\Theta}_\mu$. Recall that $\mathcal{L}_1(\mu_\theta)$ stands for the set of all measurable functions defined on $\mathcal{U} \times \mathcal{Z}$ which is integrable with respect to μ_θ. Define $\mathcal{M}$ to be the set of all functions $h \in \mathcal{L}_1(\mu_\theta)$ such that (i) $h(u,z) \geq 0$ for μ_θ almost every (u,z); (ii) for any $q \in \mathcal{L}_1(F_Z)$, $\mathbb{E}_{\mu_\theta}[q(Z)h(U,Z) - q(Z)] = 0$. By construction, for any probability distribution H with $H \ll \mu_\theta$, H satisfies Condition (i) and (iii) in Definition 7 if and only if H's density with respect to μ_θ belongs to $\mathcal{M}$.

Recall function r maps $\mathcal{U} \times \mathcal{Z}$ to $\mathbb{R}^{dr}$, and we can write r as

$$r(u,z) = (r_1(u,z), ..., r_{dr}(u,z)).$$

For each $i = 1, ..., dr$, let R_i be the nonnegative finite measure whose density with respect to μ_θ equals $|r_i|$. Let $\mathcal{L} \coloneqq \mathcal{L}_1(\mu_\theta) \times \mathcal{L}_1(R_1) \times \cdots \times \mathcal{L}_1(R_{dr})$ and define mapping $A : \mathcal{L} \mapsto \mathbb{R}^{dr}$ where, for any $(h, \tilde{h}_1, ..., \tilde{h}_{dr}) \in \mathcal{L}$,

$$A(h, \tilde{h}_1, ..., \tilde{h}_{dr}) \coloneqq \left(\int r_1 \tilde{h}_1 \, \mathrm{d}\mu_\theta, \cdots, \int r_{dr} \tilde{h}_{dr} \, \mathrm{d}\mu_\theta \right) \in \mathbb{R}^{dr}. \tag{70}$$

Note, for each i, $\int r_i \tilde{h}_i \, \mathrm{d}\mu_\theta$ is well defined and finite for any $\tilde{h}_i \in \mathcal{L}_1(R_i)$. It's easy to see A is linear and continuous. By construction, for any probability H absolutely continuous with respect to μ_θ, H satisfies Condition (ii) in Definition 7 if and only if its density $h \in \mathcal{L}_1(\mu_\theta) \cap \mathcal{L}_1(R_1) \cap \cdots \cap \mathcal{L}_1(R_{dr})$ and $\|A(h, h, ..., h)\| \leq \epsilon$.

Define function $f : \mathcal{L} \mapsto \mathbb{R} \cup \{+\infty\}$ as

$$f(h, \tilde{h}_1, ..., \tilde{h}_{dr}) \coloneqq \begin{cases} \mathbb{E}_{\mu_\theta} h \log(h) & \text{if } h \in \mathcal{M} \text{ and } \forall i = 1, ..., dr, h = \tilde{h}_i, \text{ a.s} \\ +\infty & \text{if otherwise} \end{cases} \tag{71}$$

And, for any $\epsilon > 0$, define $g_\epsilon : \mathbb{R}^{dr} \mapsto \mathbb{R} \cup \{+\infty\}$ as

$$g_\epsilon(\lambda) \coloneqq \begin{cases} 0 & \text{if } \|\lambda\| \leq \epsilon \\ +\infty & \text{if } \|\lambda\| > \epsilon \end{cases} \tag{72}$$

By the construction of f and g_ϵ, for any $\theta \in \Theta'_{KL,\mu}$ and any $\epsilon > 0$,

$$\inf \left\{ f(h, \tilde{h}_1, ..., \tilde{h}_{dr}) + g_\epsilon\big(A(h, \tilde{h}_1, ..., \tilde{h}_{dr})\big) : (h, \tilde{h}_1, ..., \tilde{h}_{dr}) \in \mathcal{L} \right\} < +\infty. \tag{73}$$

We are going to study the Fenchel duality of the infimum in (73). For any $(h, \tilde{h}_1, ..., \tilde{h}_{dr}) \in \mathcal{L}$, define its norm as $\|h\|_1 + \sum_{i=1}^{dr} \|\tilde{h}_i\|_1$ where $\|h\|_1$ is h's L_1 norm in $\mathcal{L}_1(\mu_\theta)$ and $\|\tilde{h}_i\|_1$ is $\tilde{h}_i$'s L_1 norm in $\mathcal{L}_1(R_i)$. Under this norm, $\mathcal{L}$ is a Banach space. Also, it's easy to see that g_ϵ is continuous on $\{\lambda \in \mathbb{R}^{dr} : \|\lambda\| < \epsilon\}$. Moreover, for any $\theta \in \Theta'_{KL,\mu}$, we know $A(\text{ dom } f) \cap \{b \in \mathbb{R}^{dr} : \|b\| < \epsilon\}$ is nonempty. By Fenchel's duality theorem, (see, for example, Theorem 4.4.3 in Borwein and Zhu (2005)), we know the infimum in (73) equals

$$\sup\Big\{-f^*(A^*\lambda)-g^*(-\lambda):\lambda\in\mathbb{R}^{dr}\Big\} \tag{74}$$

where A^* stands for the adjoint of A, and f^* and g^* are the convex conjugate of f and g respectively. Moreover, Fenchel's duality theorem also implies the supremum in (74) can be achieved by some $\lambda^*\in\mathbb{R}^{dr}$ when $\theta\in\Theta'_{KL,\mu}$.

By Lemma E.2 stated and proved below,

$$\sup\Big\{-f^*(A^*\lambda)-g_\epsilon^*(-\lambda):\lambda\in\mathbb{R}^{dr}\Big\}=-\inf_{\lambda\in\mathbb{R}^{dr}}\phi_\epsilon(\lambda)$$

where

$$\phi_\epsilon(\lambda):=\int\log\Big(\mathbb{E}_{\mu_\theta}[\exp(\lambda' r(U,z;\theta))|Z=z]\Big)\,\mathrm{d}F_Z+\epsilon\,\|\lambda\|\,.$$

Lemma E.2 also implies function ϕ_ϵ is a convex function. In addition, λ^* achieves the infimum $\inf\phi_\epsilon(\lambda)$ when $\theta\in\Theta'_{KL,\mu}$.

Fix any $\theta\in\Theta'_{KL,\mu}$ Since μ satisfies Assumption S, the domain of ϕ_ϵ equals $\mathbb{R}^{dr}$ so that its subgradient exists at any $\lambda\in\mathbb{R}^{dr}$. Let $\partial\phi(\lambda^*)$ be the subgradient of ϕ at λ^*. By Condition (ii) in Assumption S, we know

$$\partial\phi(\lambda^*)=\mathbb{E}_{H_{\lambda^*,\theta}}r+\epsilon\cdot\partial(\|\lambda^*\|)$$

where $\partial(\|\lambda^*\|)$ is the subgradient of $\|\lambda\|$ at λ^*. Since λ^* minimizes $\phi(\lambda)$, we know

$$0\in\partial\phi_\epsilon(\lambda^*)$$

which is equivalent to

$$-\mathbb{E}_{H_{\lambda^*,\theta}}r\in\epsilon\cdot\partial(\|\lambda^*\|).$$

Since for any $\gamma\in\partial(\|\lambda^*\|)$, we have $\|\gamma\|\le 1$. Therefore,

$$\Big\|\mathbb{E}_{H_{\lambda^*,\theta}}r\Big\|\le\epsilon.$$

Since ϵ can be any positive number, this implies, $\theta\in\widetilde{\Theta}_\mu$. Since the above result holds for any $\theta\in\Theta'_{KL,\mu}$, we conclude $\Theta'_{KL,\mu}\subseteq\widetilde{\Theta}_\mu$. □

Lemma E.2. *Assume all conditions in Theorem 5 hold. Let A, f and g_ϵ be defined as in equation (70), (71) and (72) respectively. Let A^*, f^* and g_ϵ^* be the adjoint of A, convex conjugate of f and g_ϵ respectively. Then, for any $\lambda\in\mathbb{R}^{dr}$,*

$$\begin{aligned} f^*(A^*\lambda) &= \int\log\big(\mathbb{E}_{\mu_\theta}[\exp(\lambda' r(U,z;\theta))|Z=z]\big)\,\mathrm{d}F_Z\\ g_\epsilon^*(-\lambda) &= \epsilon\,\|\lambda\| \end{aligned}$$

Proof. By definition of $f^*(A^*\lambda)$, we know,

$$\begin{aligned} f^*(A^*\lambda) &= \sup\left\{\int\Big(\sum_{i=1}^{dr}\lambda_i r_i \tilde{h}_i\Big)\ \mathrm{d}\mu_\theta - f(h,\tilde{h}_1,...,\tilde{h}_{dr}) : (h,\tilde{h}_1,...,\tilde{h}_{dr}) \in \mathcal{L}\right\} \\ &= \sup\left\{\int\Big(\sum_{i=1}^{dr}\lambda_i r_i h\Big)\ \mathrm{d}\mu_\theta - \int h\log(h)\ \mathrm{d}\mu_\theta : h \in \mathcal{M}\cap\mathcal{L}_1(R_1)\cap\cdots\cap\mathcal{L}_1(R_{dr})\right\} \end{aligned}$$

where the second equality comes from the definition of f. Let $h_{\lambda,\theta}$ be as defined in (33). Then, by the fact that μ satisfies Assumption S and by the construction of $h_{\lambda,\theta}$, we know $h_{\lambda,\theta} \in \mathcal{M}\cap\mathcal{L}_1(R_1)\cap\cdots\cap\mathcal{L}_1(R_{dr})$, so that

$$\begin{aligned} f^*(A^*\lambda) &\geq \int\Big(\sum_{i=1}^{dr}\lambda_i r_i h_{\lambda,\theta}\Big)\ \mathrm{d}\mu_\theta - \int h_{\lambda,\theta}\log(h_{\lambda,\theta})\ \mathrm{d}\mu_\theta \\ &= \int \log\big(\mathbb{E}_{\mu_\theta}[\exp(\lambda' r(U,z;\theta))|Z=z]\big)\ \mathrm{d}F_Z \end{aligned} \tag{75}$$

where the last equality can be derived after substituting $h_{\lambda,\theta}$ with the formula in (33). We're going to show the reverse of the above inequality also holds.

Define $\widetilde{\mathcal{L}} \coloneqq \mathcal{L}_1(\mu)\cap\mathcal{L}_1(R_1)\cap\cdots\cap\mathcal{L}_1(R_{dr})$. Note that

$$\begin{aligned} & f^*(A^*\lambda) \\ = \;& \sup\left\{\int\Big(\sum_{i=1}^{dr}\lambda_i r_i h\Big)\ \mathrm{d}\mu_\theta - \int h\log(h)\ \mathrm{d}\mu_\theta : h \in \mathcal{M}\cap\mathcal{L}_1(R_1)\cap\cdots\cap\mathcal{L}_1(R_{dr})\right\} \\ = \;& \sup_{h\in\widetilde{\mathcal{L}},\ h\geq 0} \int_{\{(u,z):u\in\Gamma(z;\theta)\}} \Big(\sum_{i=1}^{dr}\lambda_i r_i - \log(h)\Big)h\ \mathrm{d}\mu_\theta \\ & \text{s.t. } \forall\varphi\in\mathcal{L}_1(F_Z), \int \mathbb{1}(u\in\Gamma(z;\theta))\varphi h\ \mathrm{d}\mu_\theta = \int\varphi\ \mathrm{d}F_Z \\ = \;& \sup_{h\in\widetilde{\mathcal{L}},\ h\geq 0}\ \inf_{\varphi\in\mathcal{L}_1(F_Z)} \int_{\{(u,z):u\in\Gamma(z;\theta)\}} \Big(\sum_{i=1}^{dr}\lambda_i r_i - \log(h) + \varphi\Big)h\ \mathrm{d}\mu_\theta - \int\varphi\ \mathrm{d}\mu_\theta \\ \leq \;& \inf_{\varphi\in\mathcal{L}_1(F_Z)}\ \sup_{h\in\widetilde{\mathcal{L}},\ h\geq 0} \int_{\{(u,z):u\in\Gamma(z;\theta)\}} \Big(\sum_{i=1}^{dr}\lambda_i r_i - \log(h) + \varphi\Big)h\ \mathrm{d}\mu_\theta - \int\varphi\ \mathrm{d}\mu_\theta \\ \leq \;& \inf_{\varphi\in\mathcal{L}_1(F_Z)} \int_{\{(u,z):u\in\Gamma(z;\theta)\}} \left[\sup_{h\in\mathbb{R},\ h\geq 0}\Big(\sum_{i=1}^{dr}\lambda_i r_i - \log(h) + \varphi\Big)h\right]\ \mathrm{d}\mu_\theta - \int\varphi\ \mathrm{d}\mu_\theta \end{aligned}$$

Also, note that

$$\sup_{h\in\mathbb{R},\ h\geq 0}\Big(\sum_{i=1}^{dr}\lambda_i r_i(u,z;\theta) - \log(h) + \varphi(z)\Big)h = \exp(\lambda' r(u,z;\theta) + \varphi(z) - 1)$$

Hence,

$$
\begin{aligned}
f^*(A^*\lambda) &\leq \inf_{\varphi\in\mathcal{L}_1(F_Z)} \int \left[\exp(\lambda' r(u,z;\theta)+\varphi(z)-1)-\varphi(z)\right] \mathrm{d}\mu_\theta \\
&= \inf_{\varphi\in\mathcal{L}_1(F_Z)} \int \left[\exp(\varphi(z)-1)\cdot\mathbb{E}_{\mu_\theta}\left[\exp(\lambda' r(U,z;\theta))|Z=z\right]-\varphi(z)\right] \mathrm{d}F_Z
\end{aligned}
$$

Let $\widetilde{\varphi}(z) \coloneqq 1-\log(\mathbb{E}_{\mu_\theta}\left[\exp(\lambda' r(U,z;\theta))|Z=z\right])$. Then, it's easy to show for F_Z almost all z,

$$
\widetilde{\varphi}(z) = \arg\min_{\varphi\in\mathbb{R}} \left[\exp(\varphi-1)\cdot\mathbb{E}_{\mu_\theta}\left[\exp(\lambda' r(U,z;\theta))|Z=z\right]-\varphi\right]. \tag{76}
$$

Also, since μ satisfies Assumption S, we know $\widetilde{\varphi}\in\mathcal{L}_1(F_Z)$. Hence,

$$
\begin{aligned}
&\int \left[\exp(\widetilde{\varphi}(z)-1)\cdot\mathbb{E}_{\mu_\theta}\left[\exp(\lambda' r(U,z;\theta))|Z=z\right]-\widetilde{\varphi}(z)\right] \mathrm{d}F_Z \\
\geq\ & \inf_{\varphi\in\mathcal{L}_1(F_Z)} \int \left[\exp(\varphi(z)-1)\cdot\mathbb{E}_{\mu_\theta}\left[\exp(\lambda' r(U,z;\theta))|Z=z\right]-\varphi(z)\right] \mathrm{d}F_Z \\
\geq\ & \int \inf_{\varphi\in\mathbb{R}} \left[\exp(\varphi-1)\cdot\mathbb{E}_{\mu_\theta}\left[\exp(\lambda' r(U,z;\theta))|Z=z\right]-\varphi\right] \mathrm{d}F_Z \\
\geq\ & \int \left[\exp(\widetilde{\varphi}(z)-1)\cdot\mathbb{E}_{\mu_\theta}\left[\exp(\lambda' r(U,z;\theta))|Z=z\right]-\widetilde{\varphi}(z)\right] \mathrm{d}F_Z
\end{aligned}
$$

where the first inequality comes from the fact that $\widetilde{\varphi}\in\mathcal{L}_1(F_Z)$ and the last inequality comes from (76). Therefore,

$$
\begin{aligned}
f^*(A^*\lambda) &\leq \int \left[\exp(\widetilde{\varphi}(z)-1)\cdot\mathbb{E}_{\mu_\theta}\left[\exp(\lambda' r(U,z;\theta))|Z=z\right]-\widetilde{\varphi}(z)\right] \mathrm{d}F_Z \\
&= \int \log\left(\mathbb{E}_{\mu_\theta}[\exp(\lambda' r(U,z;\theta))|Z=z]\right) \mathrm{d}F_Z
\end{aligned}
$$

Combine the above result with (75), we conclude

$$
f^*(A^*\lambda) = \int \log\left(\mathbb{E}_{\mu_\theta}[\exp(\lambda' r(U,z;\theta))|Z=z]\right) \mathrm{d}F_Z.
$$

Finally, it's easy to see that

$$
\begin{aligned}
g_\epsilon^*(-\lambda) &= \sup_{\gamma\in\mathbb{R}^{d_r}} -\gamma'\lambda - g_\epsilon(\gamma) \\
&= \sup_{\gamma\in\mathbb{R}^{d_r},\|\gamma\|\leq\epsilon} -\gamma'\lambda \\
&= \epsilon\,\|\lambda\|
\end{aligned}
$$

This completes the proof. $\square$

F Proof of Corollary 2

Let $r = (r_1, r_2)$, and let $\widetilde{\Gamma}$ be defined as in (60) and $\tilde{r}$ be defined as in (59). Corollary 2 is an immediate result of the following three lemmas.

Lemma F.1. *Suppose all $\theta \in \Theta$ satisfies Conditions C1 and C2 for (Γ, r). Then, all $\theta \in \Theta$ satisfies Conditions C1 and C2 for model $(\widetilde{\Gamma}, \tilde{r})$.*

Lemma F.2. *Suppose all $\theta \in \Theta$ satisfies Conditions C1 and C2 for (Γ, r). Then, (61) is equivalent to*

$$\forall \lambda \in S_{dr_1+dr_2}, \ \mathbb{E}\left[\sup_{\tilde{u}\in\widetilde{\Gamma}(Z;\theta)} \lambda' \tilde{r}(\tilde{u}, Z; \theta)\right] \geq 0. \tag{77}$$

Lemma F.3. *Suppose all $\theta \in \Theta$ satisfies Conditions C1-C4 for (Γ, r). Then, $\theta \in \Theta_I(\widetilde{\Gamma}, \tilde{r})$ if and only if (77) holds.*

Proof of Corollary 2. I first prove the first part of the theorem. Suppose all $\theta \in \Theta$ satisfies Conditions C1 and C2 for (Γ, r). Then, by Lemma F.1, F.2 and Theorem 1, any θ in the identified set must satisfies (61).

Next, I prove the second part of the theorem. Suppose all $\theta \in \Theta$ satisfies Conditions C1-C4 for model (Γ, r). Then, by Lemma F.2, F.3 and Theorem 2, we know that θ is in the identified set if and only if θ satisfy (61). $\square$

Proof of Lemma F.1. The proof of Lemma F.1 is straightforward, as long as one notes that

$$\|\tilde{r}(\tilde{u}, z; \theta)\| \leq \|r(u, z; \theta)\| + \|v\|$$

so that

$$\inf\{\|\tilde{r}(\tilde{u}, z; \theta)\| : \tilde{u} \in \widetilde{\Gamma}(z; \theta)\} \leq \inf\{\|r(u, z; \theta)\| : u \in \Gamma(z; \theta)\}.$$

$\square$

Proof of Lemma F.2. First, $V \geq 0$ implies that for any $\lambda \in S_{dr_1, dr_2}$,

$$\mathbb{E}\left[\sup_{u\in\Gamma(Z;\theta)} \lambda_1' r_1(u, Z; \theta) + \lambda_2' r_2(u, Z; \theta)\right] = \mathbb{E}\left[\sup_{\tilde{u}\in\widetilde{\Gamma}(Z;\theta)} \lambda' \tilde{r}(\tilde{u}, Z; \theta)\right].$$

Second, note that for any $\lambda \in S_{dr_1+dr_2} \backslash S_{dr_1, dr_2}$ and almost every Z,

$$\sup_{\tilde{u}\in\widetilde{\Gamma}(Z;\theta)} \lambda' \tilde{r}(\tilde{u}, Z; \theta) = +\infty.$$

Hence, for any $\lambda \in S_{dr_1+dr_2} \backslash S_{dr_1,dr_2}$, we have

$$\mathbb{E}\left[\sup_{\tilde{u} \in \widetilde{\Gamma}(Z;\theta)} \lambda' \tilde{r}(\tilde{u}, Z; \theta)\right] = +\infty.$$

This implies that (61) are equivalent to (77). $\square$

Proof of Lemma F.3. I prove this lemma by applying Corollary 1. Suppose all $\theta \in \Theta$ satisfies Conditions C1-C4 for (Γ, r). Define $\widetilde{\Theta}(\widetilde{\Gamma}, \tilde{r})$ to be the set of θ which satisfy (77). Given Lemma F.1 and Theorem 1, I only need to show $\widetilde{\Theta}(\widetilde{\Gamma}, \tilde{r}) \subseteq \Theta_I(\widetilde{\Gamma}, \tilde{r})$.

Fix any $\theta \in \widetilde{\Theta}(\widetilde{\Gamma}, \tilde{r})$. For any $k \geq 0$, define $\widetilde{\Gamma}_k(\theta) = \widetilde{\Gamma}(\theta) \cap \{(\tilde{u}, v) : \|v\| \leq k\}$. Since Conditions C1-C4 hold for (Γ, r), $(\widetilde{\Gamma}_k, \tilde{r})$ is a sequence of regularized models. Note that, for any $\lambda \in S_{dr_1,dr_2}$, we have

$$\mathbb{E}\left[\sup_{\tilde{u} \in \widetilde{\Gamma}(Z;\theta)} \lambda' \tilde{r}(\tilde{u}, Z; \theta)\right] = \mathbb{E}\left[\sup_{\tilde{u} \in \widetilde{\Gamma}_k(Z;\theta)} \lambda' \tilde{r}(\tilde{u}, Z; \theta)\right].$$

For any λ_2, define λ_2^+ and λ_2^- by $\lambda_{2,j}^+ = \max(0, \lambda_{2,j})$ and $\lambda_{2,j}^- = \max(0, -\lambda_{2,j})$, for any $j = 1, ..., dr_2$. Then, for any $\lambda = (\lambda_1, \lambda_2) \in S_{dr_1+dr_2} \backslash S_{dr_1,dr_2}$, we have

$$\begin{aligned}
& \mathbb{E}\left[\sup_{\tilde{u} \in \widetilde{\Gamma}_k(Z;\theta)} \lambda' \tilde{r}(\tilde{u}, Z; \theta)\right] \\
= \; & \mathbb{E}\left[\sup_{u \in \Gamma(Z;\theta)} \lambda_1' r_1(u, Z; \theta) + \lambda_2' r_2(u, Z; \theta)\right] + \left\|\lambda_2^-\right\| k \\
\geq \; & \mathbb{E}\left[\sup_{u \in \Gamma(Z;\theta)} \lambda_1' r_1(u, Z; \theta) + \lambda_2^{+\prime} r_2(u, Z; \theta)\right] - \left\|\lambda_2^-\right\| \cdot \mathbb{E}\left[\sup_{u \in \Gamma(Z;\theta)} \|r_2(u, Z; \theta)\|\right] + \left\|\lambda_2^-\right\| k.
\end{aligned}$$

Note also that, by Condition C4, there exists some integrable function $g(z;\theta)$ such that $g(z;\theta) \geq \sup\{\|r_2(u, z; \theta)\| : u \in \Gamma(z;\theta)\}$. Therefore, for any $\lambda = (\lambda_1, \lambda_2) \in S_{dr_1+dr_2} \backslash S_{dr_1,dr_2}$, we have

$$\mathbb{E}\left[\sup_{\tilde{u} \in \widetilde{\Gamma}_k(Z;\theta)} \lambda' \tilde{r}(\tilde{u}, Z; \theta)\right] \geq \mathbb{E}\left[\sup_{u \in \Gamma(Z;\theta)} [\lambda_1' r_1(u, Z; \theta) + \lambda_2^{+\prime} r_2(u, Z; \theta)]\right] + \left\|\lambda_2^-\right\| (k - \mathbb{E}[g(Z;\theta)]).$$

Since $\theta \in \widetilde{\Theta}(\widetilde{\Gamma}, \tilde{r})$, we know

$$\mathbb{E}\left[\sup_{u \in \Gamma(Z;\theta)} [\lambda_1' r_1(u, Z; \theta) + \lambda_2^{+\prime} r_2(u, Z; \theta)]\right] \geq 0.$$

Therefore, for any $k \geq \mathbb{E}[g(Z;\theta)]$ and any $\lambda \in S_{dr_1+dr_2} \backslash S_{dr_1,dr_2}$, we have

$$\mathbb{E}\left[\sup_{\tilde{u}\in\widetilde{\Gamma}_k(Z;\theta)} \lambda'\tilde{r}(\tilde{u}, Z;\theta)\right] \geq 0.$$

By Corollary 1, this implies $\theta \in \Theta_I(\widetilde{\Gamma}, \tilde{r})$. This completes the proof. □

G Support Vector Machine and Boundaries of Confidence Set

In this section, I describe the way I used to find the boundaries of the confidence region of parameters, when the confidence region is constructed by inverting a test. Although I was not aware of it at the time, the method I used is similar to that proposed by Bar and Molinari (2018).

Let $t : \Theta \in \{0, 1\}$ be a test function. For any $\theta \in \Theta$, $t(\theta) = 0$ if this θ is rejected by the test, and $t(\theta) = 1$ otherwise. Let γ be some known vector which has the same dimension as θ. Consider the following optimization problem,

$$\begin{aligned} \kappa(\gamma) = \quad & \sup \quad \gamma'\theta \\ & \text{s.t.} \quad t(\theta) = 1. \end{aligned}$$

For example, when $\gamma = (1, 0, \cdots, 0)$, $\kappa(\gamma)$ solves the upper bound of the confidence interval of the first element of θ.

To compute $\kappa(\gamma)$, I use the following algorithm. Let q be some some prespecified value between $(0, 1)$ and $\overline{M}$ be some large integer.

Step 1 Find some parameter θ_1 such that $t(\theta_1) = 1$, and find some parameter θ_2 such that $t(\theta_2) = 0$.

Step 2 Randomly draw $M - 2$ parameters, θ_3, ..., θ_M, inside Θ.

Step 3 Let $\Xi_M = \{(\theta_m, t(\theta_m)) : m = 1, 2, ..., M\}$. Take Ξ_M as input and train a support vector machine whose prediction for θ is $\tilde{t}_M(\theta)$.

Step 4 With probability q, find $\tilde{\kappa}_M = \sup\{\gamma'\theta : \tilde{t}_M(\theta) = 1\}$ and let $\tilde{\theta}_{M+1}$ be the parameter which achieves $\tilde{\kappa}_M$, i.e. $\tilde{\kappa}_M = \gamma'\tilde{\theta}_{M+1}$ and $\tilde{t}_M(\tilde{\theta}_{M+1}) = 1$. With probability $1 - q$, draw $\tilde{\theta}_{M+1}$ randomly from the parameter space Θ.

Step 5 Compute $t(\tilde{\theta}_{M+1})$.

Step 6 Let $\Xi_{M+1} = \Xi_M \cup \{(\tilde{\theta}_M, t(\tilde{\theta}_M))\}$. Repeat Steps 3, 4 and 5 with Ξ_M replaced by Ξ_{M+1} and $M = M + 1$. Stop when $M > \overline{M}$.

Step 7 Then, $\kappa(\gamma)$ can be approximated by $\max\{\gamma'\theta : (t(\theta), \theta) \in \Xi_{\overline{M}}$ and $t(\theta) = 1\}$.

H Details in Examples 1-3

H.1 Example 1: Simulation Design

The result reported in Table 1 is based on a simulated sample constructed as follows:

(a) DM report the confidence interval for parameters (α_0, α_1) and σ in the Table 2 of their paper. Based on the odds-based and revealed-preference based moment inequalities, they report that the confidence intervals for α_0, α_1 and σ in the chemical industry are $[62.8, 81.1]$, $[142.5, 194.2]$ and $[85.1, 115.9]$ respectively. In the simulation, I set α_0, α_1 and σ to be the middle point of their reported confidence intervals.

(b) The distribution of ν is the normal distribution $N(0, \sigma_\nu^2)$, where $\sigma_\nu = 0.5\sigma$.

(c) The distribution of $\mathbb{E}_s[X|\mathcal{I}]$ is the Fréchet distribution whose c.d.f. $F(x) = \exp(-Tx^{-\gamma})$. I set $\gamma = 4$ and calibrate the value of T so that the resulting export probability matches that of the Chilean chemical industry in the Year 2000, as reported in Table 1 of DM.

(d) By the definition of ν, I set $X = \mathbb{E}_s[X|\mathcal{I}] - \nu$.

(e) The distribution of $dist$ is drawn from the uniform distribution between $\underline{d}$ and $\overline{d}$. The $\underline{d}$ is set to be the geographic distance between Chile and Brazil, and the $\overline{d}$ is set to be the geographic distance between Chile and Japan. Based on the CEPII database, I set $\underline{d} = 1,128km$ and $\overline{d} = 17,247km$.

(f) $\mathbb{E}_s[X|\mathcal{I}]$, ν, ϵ and $dist$ are mutually independent.

(g) For any $\tau \in [0, 1]$, define $q_1(\tau)$ and $q_2(\tau)$ to be the τ-quantiles of $dist$ and $\mathbb{E}_s[X|\mathcal{I}]$ respectively. Let $M = 10$. For any $m = 0, 1, ..., M-1$, define $\delta_{1,m} = \mathbb{1}(dist \in [q_1(m/M), q_1((m+1)/M)]$ and $\delta_{2,m} = \mathbb{1}(\mathbb{E}_s[X|\mathcal{I}] \in [q_2(m/M), q_2((m+1)/M)])$. I construct the instrument $W = (\delta_{1,m} \cdot \delta_{2,m'} : m, m' = 0, 1, ..., M-1)$. The total number of instruments is M^2.

(h) The sample size is set to 58,520, which is the same as that in DM.

H.2 Example 2: Solving Θ_I^N in Table 3

In this section, I suppress the m subscript to ease the notation. Θ_I^N is defined to be the set of all $(\Delta_1, \Delta_2) \in \mathbb{R}^2$ such that there exits a covariance matrix Σ which satisfies the following (in)equalities for each $x \in \mathcal{X}$,

$$\begin{aligned} p(0,0,x) &= \mathbb{P}_\Sigma(x_i + U_i \leq 0, \forall i) && (78) \\ p(1,1,x) &= \mathbb{P}_\Sigma(x_i - \Delta_i + U_i \geq 0, \forall i) && (79) \\ p(1,0,x) &\leq \mathbb{P}_\Sigma(x_1 - \Delta_1 + U_1 \geq 0, X_2 + U_2 \leq 0) && (80) \\ p(0,1,x) &\leq \mathbb{P}_\Sigma(x_2 - \Delta_2 + U_2 \geq 0, X_1 + U_1 \leq 0) && (81) \end{aligned}$$

where $p(y_1, y_2, x)$ denotes $\mathbb{P}(Y = (y_1, y_2)|X)$ in DGP, and $\mathbb{P}_\Sigma$ denotes the probability with respect to normal distribution $N(0, \Sigma)$ of U.

Generally speaking, finding all solution of nonlinear (in)equalities (78)-(81) is nontrivial in practice. Hence, special effort is needed to ensure the numerical robustness of the result of Θ_I^N in Table 3.

Let σ_i be the standard error of U_i, and ρ be the correlation between U_1 and U_2. Then,

$$\Sigma \coloneqq \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{bmatrix}$$

I state the following lemma, whose proof is straightforward.

Lemma H.1.

(i) *When $x = (0, 0)$, $\mathbb{P}_\Sigma(x_i + U_i \leq 0, \forall i)$ only depends on ρ and is strictly increasing in ρ.*

(ii) *When $x = (x_1, 0)$ with $x_1 \neq 0$, $\mathbb{P}_\Sigma(x_i + U_i \leq 0, \forall i)$ only depends on ρ and σ_1 and is strictly increasing in σ_1 given each ρ.*

(iii) *When $x = (0, x_2)$ with $x_2 \neq 0$, $\mathbb{P}_\Sigma(x_i + U_i \leq 0, \forall i)$ only depends on ρ and σ_2 and is strictly increasing in σ_2 given each ρ.*

(iv) *Given x, Σ and Δ_1 (or Δ_2), $\mathbb{P}_\Sigma(x_i - \Delta_i + U_i \geq 0, \forall i)$ is strictly decreasing in Δ_2 (or Δ_1). Moreover, if both (Δ_1, Δ_2) and (Δ_1', Δ_2') solve equation (79) given x and Σ, then $\Delta_1 < \Delta_1'$ if and only if $\Delta_2 > \Delta_2'$.*

Based on the first three results in Lemma H.1, we know, for each k and m, if equation (78) admits covariance matrix Σ as a solution, such solution is unique and can be found easily.

Suppose such covariance matrix Σ is found. For our purpose, we need an algorithm whose result $\widetilde{\Theta}_I^N$ has the following property:

(1) $\widetilde{\Theta}_I^N$ is empty if and only if Θ_I^N is empty.

(2) $\widetilde{\Theta}_I^N$ is singleton if and only if Θ_I^N is singleton. In this case, $\widetilde{\Theta}_I^N = \Theta_I^N$.

Given Σ, by the last result in Lemma H.1, equation (79) defines an implicit function $\Delta_1(\Delta_2, x, \Sigma)$ and $\Delta_2(\Delta_1, x, \Sigma)$ for each x as long as equation (79) admits a solution. Moreover, such function $\Delta_1(\Delta_2, x, \Sigma)$ (or, $\Delta_2(\Delta_1, x, \Sigma)$) is strictly decreasing in Δ_2 (or, Δ_1). This suggests the following iterative algorithm.

Algorithm H.1. *Suppose we have initial bounds $\{(\underline{\Delta}_i^{(0)}, \overline{\Delta}_i^{(0)}) : i = 1, 2\}$ such that*

$$\Theta_I^N \subseteq \times_{i=1}^2 [\underline{\Delta}_i^{(0)}, \overline{\Delta}_i^{(0)}] \coloneqq \{\Delta : \underline{\Delta}_i^{(0)} \leq \Delta_i \leq \overline{\Delta}_i^{(0)}, \forall i\}.$$

Return $\widetilde{\Theta}_I^N = \emptyset$, if equations (78) doesn't have a solution. If there is a solution, denote it Σ and conduct the following iterative steps.

In Step j, update bounds $\{(\underline{\Delta}_i^{(j)}, \overline{\Delta}_i^{(j)}) : i = 1, 2\}$ as follows,

$$\begin{aligned}
\overline{\Delta}_1^{(j)} &= \min\left\{\overline{\Delta}_1^{(j-1)}, \inf_{x\in\mathcal{X}} \Delta_1(\underline{\Delta}_2^{(j-1)}, x, \Sigma)\right\} \\
\overline{\Delta}_2^{(j)} &= \min\left\{\overline{\Delta}_2^{(j-1)}, \inf_{x\in\mathcal{X}} \Delta_2(\underline{\Delta}_1^{(j-1)}, x, \Sigma)\right\} \\
\underline{\Delta}_1^{(j)} &= \max\left\{\underline{\Delta}_1^{(j-1)}, \sup_{x\in\mathcal{X}} \Delta_1(\overline{\Delta}_2^{(j)}, x, \Sigma)\right\} \\
\underline{\Delta}_2^{(j)} &= \max\left\{\underline{\Delta}_2^{(j-1)}, \sup_{x\in\mathcal{X}} \Delta_2(\overline{\Delta}_1^{(j)}, x, \Sigma)\right\}.
\end{aligned}$$

If $\underline{\Delta}_i^{(j)} > \overline{\Delta}_i^{(j)}$ for some i, then return $\widetilde{\Theta}_I^N = \emptyset$; If $\underline{\Delta}_i^{(j)} = \underline{\Delta}_i^{(j-1)}$ and $\overline{\Delta}_i^{(j)} = \overline{\Delta}_i^{(j-1)}$ for all i, then return $\widetilde{\Theta}_I^N = \times_{i=1}^2[\underline{\Delta}_i^{(j)}, \overline{\Delta}_i^{(j)}]$; Otherwise, repeat step with $j + 1$.

When calculating results in Table 3, I first solve Θ_I^N using grid search. Then, use the bounds in Θ_I^N as the initial bounds for Algorithm H.1. By Lemma H.1, the $\widetilde{\Theta}_I^N$ generated by Algorithm H.1 satisfies Conditions (1) and (2).

H.3 Example 3: Derivation of Moment Restriction (7)

Given all the assumptions in Example 3, and for any two subsets $\mathcal{J}_1$ and $\mathcal{J}_2$ of $\mathcal{J}$, moment inequalities in (7) hold trivially if $\mathcal{J}_1$ is not included in $\mathcal{A}_{ist}(\theta)$ or $\mathcal{J}_2$ is not included in $\mathcal{B}_{ist}(\theta)$. Now, suppose $\mathcal{J}_1 \subseteq \mathcal{A}_{ist}(\theta)$ and $\mathcal{J}_2 \subseteq \mathcal{B}_{ist}(\theta)$. Then, we have

$$\begin{aligned}
& \mathbb{E}\left[\mathbb{1}\left(\max_{j\in\mathcal{J}_1} U_{ijs} \geq \max_{j\in\mathcal{J}_2} U_{ijs}\right)\Big|\nu_i, X_i\right] \\
= \; & \mathbb{E}\left[\mathbb{1}\left(\max_{j\in\mathcal{J}_1} \pi(X'_{ijs}\theta, \nu_{ij}, \epsilon_{ijs}) \geq \max_{j\in\mathcal{J}_2} \pi(X'_{ijs}\theta, \nu_{ij}, \epsilon_{ijs})\right)\Big|\nu_i, X_i\right] \\
= \; & \mathbb{E}\left[\mathbb{1}\left(\max_{j\in\mathcal{J}_1} \pi(X'_{ijs}\theta, \nu_{ij}, \epsilon_{ijt}) \geq \max_{j\in\mathcal{J}_2} \pi(X'_{ijs}\theta, \nu_{ij}, \epsilon_{ijt})\right)\Big|\nu_i, X_i\right] \\
\geq \; & \mathbb{E}\left[\mathbb{1}\left(\max_{j\in\mathcal{J}_1} \pi(X'_{ijt}\theta, \nu_{ij}, \epsilon_{ijt}) \geq \max_{j\in\mathcal{J}_2} \pi(X'_{ijt}\theta, \nu_{ij}, \epsilon_{ijt})\right)\Big|\nu_i, X_i\right] \\
= \; & \mathbb{E}\left[\mathbb{1}\left(\max_{j\in\mathcal{J}_1} U_{ijt} \geq \max_{j\in\mathcal{J}_2} U_{ijt}\right)\Big|\nu_i, X_i\right]
\end{aligned}$$

where the first equality follows from the definition of U_{ijs}, and the second equality follows from the assumption that the distribution of ϵ_{it} conditional on (A_i, X_i) does not depend on t. And, the next inequality comes from the fact that $\mathcal{J}_1 \subseteq \mathcal{A}_{ist}(\theta)$ and $\mathcal{J}_2 \subseteq \mathcal{B}_{ist}(\theta)$, the definition of $\mathcal{A}_{ist}(\theta)$ and $\mathcal{B}_{ist}(\theta)$ and the assumption that π is weakly increasing in its first argument. Finally, the last equality follows form the definition of U_{ijt}. The law of iterated expectation then implies (7).

H.4 Example 3: Establishing the Equivalence between (63) and (64) when $T = 2$

First, I show (64) implies (63) when $T = 2$. Fix any $\theta \in \Theta$. Since $T = 2$, it is without loss of generality to let $s = 1$ and $t = 2$. Conditional on a value x_i of X_i, let $\mathcal{A} = \mathcal{A}_{ist}(\theta)$ and $\mathcal{B} = \mathcal{B}_{ist}(\theta)$. For any $z_i = (y_i, x_i)$, define $u'_{ijs}(z_i)$ by

$$u'_{ijs}(z_i) = \begin{cases} 1 & \text{if } j = y_{is}, \\ 0 & \text{if } j \neq y_{is} \text{ and } j \in \mathcal{A}, \\ -1 & \text{otherwise.} \end{cases}$$

Define u'_{ijt} by

$$u'_{ijt}(z_i) = \begin{cases} 1 & \text{if } j = y_{it}, \\ 0 & \text{if } j \neq y_{it} \text{ and } j \in \mathcal{B}, \\ -1 & \text{otherwise.} \end{cases}$$

Let $u'_i(z_i) = (u'_{ijs}(z_i), u'_{ijt}(z_i) : j \in \mathcal{J})$. Then, by construction, $(u'_i(z_i), z_i) \in \Gamma(\theta)$, or equivalently, $u'_i(z_i) \in \Gamma(z_i; \theta)$. Hence, for any z_i and any $\lambda \in S^+_{dr}$,

$$\sup_{u_i \in \Gamma(z_i;\theta)} \lambda' r(u_i, z_i; \theta) \geq \lambda' r(u'_i(z_i), z_i; \theta).$$

Therefore,

$$\forall \lambda \in S^+_{dr}, \quad \mathbb{E}[\lambda' r(u'_i(Z_i), Z_i; \theta) | X_i = x_i] \geq 0 \tag{82}$$

implies (63). As a result, I only need to show (64) implies (82). To prove this, let $\lambda_{st}(\mathcal{J}_1, \mathcal{J}_2)$ be the corresponding Lagrange multiplier for moment restriction in (62) for any nonempty subset $\mathcal{J}_1$ of $\mathcal{A}$ and any nonempty subset $\mathcal{J}_2$ of $\mathcal{B}$. Then,

$$\begin{aligned} & \lambda' r(u'_i(z_i), z_i; \theta) \\ = & \sum_{\mathcal{J}_1, \mathcal{J}_2} \lambda_{st}(\mathcal{J}_1, \mathcal{J}_2) \left[\mathbb{1}\left(\max_{j \in \mathcal{J}_1} u'_{ijs}(z_i) \geq \max_{j \in \mathcal{J}_2} u'_{ijs}(z_i) \right) - \mathbb{1}\left(\max_{j \in \mathcal{J}_1} u'_{ijt}(z_i) \geq \max_{j \in \mathcal{J}_2} u'_{ijt}(z_i) \right) \right], \end{aligned}$$

where $\mathcal{J}_1$ and $\mathcal{J}_2$ in the summation ranges over all the nonempty set of $\mathcal{A}$ and $\mathcal{B}$ respectively. By the construction of $u'_i(z_i)$, we have

$$\begin{aligned} \mathbb{1}\left(\max_{j \in \mathcal{J}_1} u'_{ijs}(z_i) \geq \max_{j \in \mathcal{J}_2} u'_{ijs}(z_i) \right) &= 1 - \mathbb{1}(y_{is} \in \mathcal{J}_2 \backslash \mathcal{J}_1) \\ \mathbb{1}\left(\max_{j \in \mathcal{J}_1} u'_{ijt}(z_i) \geq \max_{j \in \mathcal{J}_2} u'_{ijt}(z_i) \right) &= \mathbb{1}(y_{it} \in \mathcal{J}_1) \end{aligned}$$

Therefore, (82) is equivalent to that for any nonempty subset $\mathcal{J}_1$ of $\mathcal{A}$ and any nonempty subset $\mathcal{J}_2$ of $\mathcal{B}$,

$$\mathbb{E}[1 - \mathbb{1}(Y_{is} \in \mathcal{J}_2 \backslash \mathcal{J}_1) - \mathbb{1}(Y_{it} \in \mathcal{J}_1)|X_i = x_i] \geq 0. \tag{83}$$

By the fact that $\mathcal{J}_1$ is a subset of $\mathcal{A}$ and $\mathcal{J}_2$ is a subset of $\mathcal{B}$, we have $1 - \mathbb{1}(Y_{is} \in \mathcal{J}_2 \backslash \mathcal{J}_1) \geq \mathbb{1}(Y_{is} \in \mathcal{A})$ and $\mathbb{1}(y_{it} \in \mathcal{J}_1) \leq \mathbb{1}(Y_{it} \in \mathcal{A})$. Therefore, (83) is implied by

$$\mathbb{E}[\mathbb{1}(Y_{is} \in \mathcal{A}) - \mathbb{1}(Y_{it} \in \mathcal{A})|X_i = x_i] \geq 0$$

which proves that (64) implies (82), hence, (63).

Next, I show that (63) implies (64). By Theorem 2, (63) is equivalent to (62). Let $\mathcal{J}_1 = \mathcal{A}_{ist}$ and $\mathcal{J}_2 = \mathcal{B}_{ist}$. Then, (62) implies

$$\mathbb{E}\left[\mathbb{1}\left(\max_{j \in \mathcal{A}_{ist}} U_{ijs} \geq \max_{j \in \mathcal{B}_{ist}} U_{ijs}\right) - \mathbb{1}\left(\max_{j \in \mathcal{A}_{ist}} U_{ijt} \geq \max_{j \in \mathcal{B}_{ist}} U_{ijt}\right) \Big| X_i = x_i\right] \geq 0.$$

One can then verify that

$$\begin{aligned}
\mathbb{1}\left(\max_{j \in \mathcal{A}_{ist}} U_{ijs} \geq \max_{j \in \mathcal{B}_{ist}} U_{ijs}\right) &= \mathbb{1}(Y_{is} \in \mathcal{A}_{ist}) \\
\mathbb{1}\left(\max_{j \in \mathcal{A}_{ist}} U_{ijt} \geq \max_{j \in \mathcal{B}_{ist}} U_{ijt}\right) &= \mathbb{1}(Y_{it} \in \mathcal{A}_{ist}).
\end{aligned}$$

Hence, (63) implies (64).

References

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

Aguirregabiria, V. and P. Mira, "A Hybrid Genetic Algorithm for the Maximum Likelihood Estimation of Models with Multiple Equilibria: A First Report," *New Mathematics and Natural Computation*, 2005, *1* (2), 295–303.

Aliprantis, C. and K. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer, 2007.

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.

Avis, D. and C. Jordan, "mplrs: A Scalable Parallel Vertex/Facet Enumeration Code," *Mathematical Programming Computation*, 2018, *10* (2), 267–302.

_ **and K. Fukuda**, "A Pivoting Algorithm for Convex Hulls and Vertex Enumeration of Arrangements and Polyhedra," *Discrete & Computational Geometry*, 1992, *8* (3), 295–313.

Bagnara, R., P. Hill, and E. Zaffanella, "The Parma Polyhedra Library: Toward a Complete Set of Numerical Abstractions for the Analysis and Verification of Hardware and Software Systems," *Science of Computer Programming*, 2008, *72* (1–2), 3–21.

Bar, H. and F. Molinari, "Computational Methods for Partially Identified Methods via Data Augmentation and Support Vector Machine," *Private communication*, 2018.

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

Berry, S., "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 1992, *60* (4), 889–917.

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

Borwein, J. and Q. Zhu, *Techniques of Variational Analysis*, Springer, 2005.

Bresnahan, T. and P. Reiss, "Empirical Models of Discrete Games," *Journal of Econometrics*, 1991, *48* (1-2), 57–81.

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

Chesher, A., A. Rosen, and K. Smolinski, "An Instrumental Variable Model of Multiple Discrete Choice," *Quantitative Economics*, 2013, *4* (2), 157–196.

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

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

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

_ **, C. Murry, and E. Tamer**, "Market Structure and Competition in Airline Markets," *Working Paper*, 2018.

Clarkson, K., "More Output-sensitive Geometric Algorithms," *Proceedings 35th Annual Symposium on Foundations of Computer Science*, 1994, pp. 695–702.

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

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

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

Gao, W. and M. Li, "Robust Semiparametric Estimation in Panel Multinomial Choice Models," *Working Paper*, 2018.

Ho, K. and A. Pakes, "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review*, 2014, *104* (12), 3841–3884.

Jia, P., "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 2008, *76* (6), 1263–1316.

Khachiyan, L., E. Boros, K. Borys, V. Gurvich, and K. Elbassioni, "Generating All Vertices of a Polyhedron Is Hard," *06 Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithm*, 2009, pp. 758–765.

Luo, Y. and H. Wang, "Core Determining Class: Construction, Approximation and Inference," *Working Paper*, 2017.

Manski, C., "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 1975, *3* (3), 205–228.

—, "Identification of Binary Response Models," *Journal of the American Statistical Association*, 1988, *83* (403), 729–738.

—, *Identification Problems in the Social Sciences*, Harvard University Press, 1999.

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

—, *Identification for Prediction and Decision*, Harvard University Press, 2008.

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

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

Pakes, A. and J. Porter, "Moment Inequalities for Multinomial Choice with Fixed Effects," NBER Working Paper No. 21893 2016.

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

Reguant, M., "Bounding Outcomes in Counterfactual Analysis," *Working Paper*, 2016.

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

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

Shi, X., M. Shum, and W. Song, "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," *Econometrica*, 2018, *86* (2), 737–761.

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

Uetake, K. and Y. Watanabe, "Entry by Merger: Estimates from a Two-Sided Matching Model with Externalities," *Working Paper*, 2017.