Discordant relaxations of misspecified models

Lixiong Li
Department of Economics, Johns Hopkins University

Désiré Kédagni
Department of Economics, The University of North Carolina at Chapel Hill

Ismaël Mourifié
Department of Economics, University of Toronto, Department of Economics, Washington University in St. Louis, and NBER

In many set-identified models, it is difficult to obtain a tractable characterization of the identified set. Therefore, researchers often rely on nonsharp identification conditions, and empirical results are often based on an outer set of the identified set. This practice is often viewed as conservative yet valid because an outer set is always a superset of the identified set. However, this paper shows that when the model is refuted by the data, two sets of nonsharp identification conditions derived from the same model could lead to disjoint outer sets and conflicting empirical results. We provide a sufficient condition for the existence of such discordancy, which covers models characterized by conditional moment inequalities and the Artstein (1983) inequalities. We also derive sufficient conditions for the nonexistence of discordant submodels, therefore providing a class of models for which constructing outer sets cannot lead to misleading interpretations. In the case of discordancy, we follow Masten and Poirier (2021) by developing a method to salvage misspecified models, but unlike them, we focus on discrete relaxations. We consider all minimum relaxations of a refuted model that restores data consistency. We find that the union of the identified sets of these minimum relaxations is robust to detectable misspecifications and has an intuitive empirical interpretation.

Keywords. Partial identification, identified/outer set, misspecification, nonconflicting hypothesis.

JEL classification. C12, C21, C26.

1. Introduction

A central challenge in the structural estimation of economic models is that the hypothesized structure often fails to identify a single generating process for the data, either due to multiple equilibria or data observability constraints. In such a context, the econometrics of partially identified models have been trying to obtain a tractable characterization of parameters compatible with the available data and maintained assumptions (hereafter *identified set*). A question of particular relevance in applied work is that it is often very difficult to find a tractable characterization of the identified set and then to obtain a valid confidence region for it. To avoid this difficulty, a large part of the literature has been trying to provide a confidence region for an *outer set*, that is, a collection of values for the parameter of interest that contains the identified set but may also contain additional values.[1] Because of its tractability, constructing a confidence region for an outer set has been entertained in various topics of studies where the parameters of interest are only partially identified; see, for instance, Blundell, Gosling, Ichimura, and Meghir (2007), Ciliberto and Tamer (2009), Aucejo, Bugni, and Hotz (2017), Sheng (2020), De Paula, Richards-Shubik, and Tamer (2018), Dickstein and Morales (2018), Honoré and Hu (2020), Chesher and Rosen (2020), Gualdani (2021), and Berry and Compiani (2023), among many others.

In most empirical studies, obtaining a tight outer set is very often interpreted as evidence for a small and informative identified set.[2] This is because, under correct specification, any outer set contains the identified set. In this paper, we examine the implications of using outer sets for models that could be misspecified. We say a model is misspecified if the identified set of the model parameters is empty. We use refutation and misspecification interchangeably in this paper.

The first main contribution of this paper is to characterize a class of models for which outer sets based on nonsharp identification conditions may be discordant. For this class of models, as long as the model is misspecified, there always exist two sets of nonsharp identification conditions that fail to detect the violation of the model and at the same time yield outer sets that are disjoint with each other. Our result covers a large class of models studied in the partial identification literature, including models whose identified set is characterized by intersection bounds, conditional moment inequalities, or the Artstein (1983) inequalities. The discordancy that we find is a negative property because the result provided by an outer set could entirely be driven by the set of nonsharp identification conditions chosen by the researcher, and that we could always consider an alternative choice that provides a result that conflicts with the initial one.

Discordant outer sets only exist when the model is misspecified. In theory, a researcher could run a model specification test before using an outer set. However, in practice, although it is possible to construct a *nonsharp* specification test that only checks the sufficient conditions for model misspecification, constructing a *sharp* specification test, which checks the necessary and sufficient conditions for the emptiness of the iden-

---

[1] See Molinari (2020) for a detailed discussion.

[2] A tight outer set here refers to an outer set that is very small and informative.

tified set, is as challenging as obtaining a sharp characterization of the identified set. The nontractability of the latter is often the motivation to use outer sets in the first place.[3]

Therefore, our result shows that the usage of outer sets based on nonsharp identification conditions is an unreliable compromise. It suggests that looking for the sharp characterization of the identified set must not only be viewed as a theoretical exercise but also has important empirical relevance. The identified set not only exhausts all the identification restrictions in the model structure and assumptions but is also immune to the possible misleading conclusions of discordant submodels.

Our warning against the usage of outer sets based on nonsharp identification conditions should not discourage researchers from relaxing stringent primitive model assumptions and replacing them with weaker ones. Although the identified set of a weaker set of primitive model assumptions is necessarily an outer set of the stringent original model, it is different from an outer set based on nonsharp identification conditions of the original model. The identified set, derived from weaker primitive assumptions, has a clear and precise interpretation akin to the empirical content inherent in weaker primitive assumptions due to its sharpness. In contrast, the choice of nonsharp identification conditions is often driven by analytical or computational tractability. Researchers often lack primitive interpretations for the outer set based on these nonsharp identification conditions. These outer sets are only relevant to the empirical analysis when viewed as conservative bounds for the sharp results. Yet, this view of conservative bounds could be misleading because, as we show in this paper, the conclusion drawn from these outer sets can be driven entirely by the choice of nonsharp identification conditions instead of the empirical content of the model.

However, discordant outer sets do not exist in all refuted models, especially when they are based on weaker primitive assumptions instead of nonsharp identification conditions. We then derive sufficient conditions for the nonexistence of discordant submodels. This second result characterizes a class of models for which constructing outer sets cannot lead to misleading interpretations. In this case, outer sets would be conservative but always robust.

Prior to our work, various papers have been concerned about misspecification in partially identified models. An important focus has been dedicated to analyzing the impact of model misspecification on standard confidence regions used for set-identified models. Bugni, Canay, and Guggenberger (2012) analyze the behavior of usual inferential methods for moment inequality models under local model misspecification. Ponomareva and Tamer (2011) and Kaido and White (2012) consider the impact of misspecification on semiparametric partially identified models, respectively, in the linear regression model with an interval-valued outcome and in a framework where some nonparametric moment inequalities are correctly specified and misspecification is due to a parametric functional form. See also Allen and Rehbeck (2020) who propose a method for statistical inference on the minimum approximation error needed to explain aggregate data in quasilinear utility models. It is worth noting that if one tries to find a confidence

---

[3]See, for instance, Sheng (2020), Gualdani (2021), and the empirical application in Berry and Compiani (2023, Section 6, footnote 42).

region for an outer set, none of the inference methods, including those developed in the previously cited papers, can resolve the specific issue we are raising here. This is because two nonempty outer sets derived from the same underlying model structure can lead to discordant results. Adopting one of these outer sets without checking the validity of the underlying model could lead to misleading conclusions. Therefore, we need to suggest a more primitive approach to deal with these discordant results in this paper.

This objective leads to our second main contribution, which consists of providing a method to salvage models that are possibly misspecified because of the existence of discordant misspecified submodels or discordant nonempty outer sets. The main intuition is to construct some minimum relaxation of the full model by removing discordant submodels until all remaining submodels are compatible. Because there could be multiple ways to relax a model to restore data consistency, we take the union of the identified set of all these relaxed models. By doing so, we construct what we call the *misspecification robust bound*. We provide general sufficient conditions under which our misspecification robust bound exists and also provide an intuitive empirical interpretation for it. Intuitively, we will say that a hypothesis is robust to misspecification if the hypothesis is compatible with all relaxed models that are data consistent and is implied by at least one of those data-consistent relaxed models.

The misspecification robust bound concept is related to the minimally relaxed identified set introduced in Andrews and Kwon (2024), and to the falsification adaptive set concept introduced in Masten and Poirier (2021). The primary departure from Masten and Poirier (2021) lies in our emphasis on discrete relaxations, while Masten and Poirier (2021) focused exclusively on relaxing assumptions in a continuous manner. In general, the use of discrete or continuous relaxation depends on the empirical application under scrutiny. We explore various features of discrete relaxations beyond its formal definition.

It is worth noting that discrete relaxations of misspecified models have been entertained in various existing papers; see, for instance, Manski and Pepper (2000, 2009), Blundell et al. (2007), Kreider, Pepper, Gundersen, and Jolliffe (2012), Chen, Flores, and Flores-Lagunes (2018), Kédagni (2023), Mourifié, Henry, and Méango (2020), among many others. In these papers, when the initial model is too stringent, they suggested alternative weaker assumptions that are believed to be more compatible with the empirical application under scrutiny and for which the identified/outer set is not empty. However, some alternative reasonable relaxations may generate results that are discordant with what they suggested. To mitigate this issue, our misspecification robust bound approach suggests to collect information from all reasonable discordant minimum relaxations of the initial model.

We organize the rest of the paper as follows. Section 2 introduces two simple leading examples that will illustrate our main contributions. Section 3 presents our general setting and main results on the characterization of discordant submodels. Section 4 discusses a class of models for which constructing outer sets do not lead to misleading interpretations. Section 5 introduces the misspecification robust bound used to salvage misspecified models. Section 6 provides a numerical illustration of the discordancy issue by visiting the widely used entry game model studied in Ciliberto and Tamer (2009),

and also illustrates our misspecification robust bounds in a return to education example. The last section concludes, and additional results and proofs are relegated to the Appendix.

2. Introductory examples

Although the main idea of this paper can be applied to general models, we begin with these two straightforward examples to illustrate our main contributions.

2.1 *First leading example: Intersection bounds*

Let us consider a special case of the intersection bounds in Chernozhukov, Lee, and Rosen (2013) in which a parameter θ is bounded by the conditional mean of an upper and lower bounds,

$$E[\underline{Y}|Z = z] \le \theta \le E[\overline{Y}|Z = z] \quad \text{almost surely,} \tag{1}$$

where $\underline{Y}$ and $\overline{Y}$ are two observable random bounds and Z is a vector of instrumental variables. Let $\mathcal{Z}$ be the support of Z, and define[4]

$$\underline{\gamma} \equiv \sup_{z \in \mathcal{Z}} E[\underline{Y}|Z = z] \quad \text{and} \quad \overline{\gamma} \equiv \inf_{z \in \mathcal{Z}} E[\overline{Y}|Z = z]. \tag{2}$$

The identified set of θ is the interval $[\underline{\gamma}, \overline{\gamma}]$ when $\underline{\gamma} \le \overline{\gamma}$. We assume the following regularity condition holds in this example.

Assumption 1. *Assume $E|\underline{Y}| < \infty$ and $E|\overline{Y}| < \infty$. In addition, assume that the conditional expectations $E[\underline{Y}|Z]$, and $E[\overline{Y}|Z]$ exist and $E[\underline{Y}|Z] \le E[\overline{Y}|Z]$ almost surely.*

This simple framework encompasses some important treatment effect models, including discrete and continuous treatment models; see, for instance, Manski (1990, 1994) and Kim, Kwon, Kwon, and Lee (2018) among many others.[5] In practice, model (1) is sometimes implemented by solving its unconditional version,

$$E\big[h(Z)(\theta - \underline{Y})\big] \ge 0 \quad \text{and} \quad E\big[h(Z)(\overline{Y} - \theta)\big] \ge 0, \tag{3}$$

where h is some nonnegative function mapping its input to $\mathbb{R}_+^m$ with $m < \infty$, and the inequalities in (3) are vector inequalities. The inference for (3) is typically much simpler than the inference for the original model (1), especially when Z is multidimensional. Let $\widetilde{\Theta}(h)$ be the identified set for θ in model (3). As made explicit in the notation, $\widetilde{\Theta}(h)$ depends on the choice of instrumental function h. However, since (1) implies (3), we know that for every choice of h, $\widetilde{\Theta}(h)$ is always an outer set of the interval $[\underline{\gamma}, \overline{\gamma}]$, the identified set for θ in model (1), that is, $[\underline{\gamma}, \overline{\gamma}] \subseteq \widetilde{\Theta}(h)$. This inclusion relation is often

---

[4]The sup and inf operators in (2) should be understood as essential supremum and essential infimum respectively.

[5]For the sake of conciseness, those examples are discussed in more detail in Appendix A.1.

used as a justification for using model (3). Its identification result $\widetilde{\Theta}(h)$ is often viewed as a conservative bound for $[\underline{\gamma}, \overline{\gamma}]$, the identified set for model (1).

Our first observation is that the result based on $\widetilde{\Theta}(h)$ is not always reliable. Later in Section 3.1, we show that when the identified set of (1) is empty, that is, $\underline{\gamma} > \overline{\gamma}$, there always exist two h and h' such that both $\widetilde{\Theta}(h)$ and $\widetilde{\Theta}(h')$ are nonempty but $\widetilde{\overline{\Theta}}(h) \cap \widetilde{\Theta}(h')$ is empty. Thus, two researchers could apply the same model on the same data set and yet draw completely different conclusions from the outer sets by choosing different h functions. For example, if one observes $\widetilde{\Theta}(h) \subseteq (0, +\infty)$ for some h, one should not jump directly to the conclusion that the sign of θ is positive without verifying the nonemptiness of the identified set, since in the case of emptiness, there are circumstances under which another researcher may choose an alternative h' such that $\widetilde{\Theta}(h) \subseteq (-\infty, 0)$.

This caveat of outer sets is somewhat overlooked in the literature. As we listed some papers in the Introduction, it is common for researchers to construct a confidence interval for an outer set and draw conclusions based solely on its result. If the model is refutable and a researcher only studies an outer set in the empirical analysis without knowing whether the identified set is empty or not, results based on an outer set could be misleading in the intersection-bound model. In Section 3, we show that this caveat is indeed a concern for some widely used partial identification models, and in Section 6, we illustrate this discordancy issue using the entry game model studied in Ciliberto and Tamer (2009).

On the other hand, there exist models for which constructing outer sets does not lead to misleading interpretations. In the following, we will introduce a model that belongs to this category as our second leading example. We study this class of models in more detail later in Section 4.

2.2 *Second leading example: Adaptive Monotone IV (AMIV)*

Consider the following potential outcome model: $Y = \sum_{z \in \mathcal{Z}} \mathbb{1}(Z = z)[Y_{1z}D + Y_{0z}(1 - D)]$, where the treatment D is binary and the support $\mathcal{Z}$ of instrument Z is discrete and finite. Y_{dz} is the potential outcome when the treatment and the instrument are externally set to d and z, respectively. Assume Z is one-dimensional and assume, without loss of generality that $\mathcal{Z} = \{1, \ldots, k\}$. We are interested in the average potential outcomes $\theta_d = \sum_z P(Z = z)EY_{dz}$ for $d \in \{0, 1\}$, and then average treatment effect (ATE), that is, $\theta_1 - \theta_0$. In this framework, the seminal work of Manski (1990) derived sharp bounds on the ATE under three main assumptions: *E.1* the bounded support of the potential outcomes, *E.2* the exclusion restriction, that is, $EY_{dz} = EY_{dz'}$ for $z \neq z'$, and *E.3* the mean independence assumption, that is, $E[Y_{dz}|Z] = E[Y_{dz}]$ for all $z \in \mathcal{Z}$. However, there are some empirical evidences—for instance, Ginther (2000), where the identified for the ATE proposed by Manski is empty—suggesting a violation of the set of these assumptions. To be able to say something meaningful on the ATE in such a context, we introduce the following assumptions, which is one way to relax Manski's (1990) assumptions.

For any $z \in \{1, \ldots, k\}$, define assumption a_z to be the collection of the following assumptions:
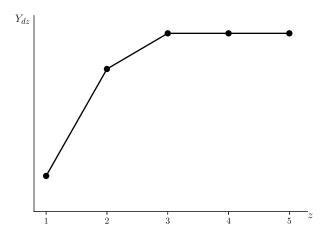
FIGURE 1. Illustration of restriction *E.3* when $z = 3$ and $k = 5$.

E.1 for each $d \in \{0, 1\}$ and any $t \in \{1, \ldots, k\}$, $P(Y_{dt} \in [\underline{y}_d, \overline{y}_d]) = 1$.

E.2 for each $d \in \{0, 1\}$ and any $t \in \{1, \ldots, k\}$, $E[Y_{dt}|Z] = E[Y_{dt}]$ almost surely.

E.3 for each $d \in \{0, 1\}$, $Y_{dt} \le Y_{dt'}$ for all $t \le t'$, and $Y_{dt} = Y_{dz}$ for all $t \ge z$.

Each assumption a_z has three parts. *E.1* requires the potential outcomes to have a bounded support. *E.2* is a mean independence assumption associated to the potential outcome Y_{dz}. The novelty here is *E.3*, which is an adaptive relaxation of the exclusion restriction. Indeed, in the extreme case when $z = 1$, *E.3* is equivalent to the full exclusion restriction, that is, $Y_{dz} = Y_{dz'}$ for all d, z and z', then *E.2* and *E.3* are equivalent to $E[Y_d|Z] = E[Y_d]$, which is the restriction under which Manski (1990) derived bounds on the ATE. On the other extreme, when $z = k$, *E.2* and *E.3* imply the MIV assumption introduced in Manski and Pepper (2000), that is, $z_1 < z_2 \Rightarrow E[Y_d|Z = z_1] \le E[Y_d|Z = z_2]$. However, when $1 < z < k$, we are in a middle-ground situation where the exclusion restriction is relaxed in such a way that $Y_{dz'}$ is monotone in z', but remains flat for $z' \ge z$. See Figure 1 for an illustration of how Y_{dz} depends on z under *E.3*.

Because the cut-off point z would be set based on the data through techniques elaborated in Section 5, we refer to this assumption as the Adaptive Monotone IV (AMIV) assumption. The economic rationality of the AMIV is that, even if Z is not a valid IV because it could positively affect the potential outcome, in some empirical contexts, it could be reasonable to consider that the marginal effect of the IV on the potential outcome becomes null after a certain cut-off point.

For example, let us consider the well-studied topic of measuring the returns to a college degree. In this empirical context, researchers have often used parental education as an instrumental variable for college education. Nonetheless, some argue that this instrument may not be valid. They point out that parents with higher education tend to have children with better unobserved skills that significantly impact potential earnings, despite the possibility that the marginal benefit of one extra year of a parent's education may be diminishing. The AMIV assumption introduced in this subsection combines these two features. In terms of our notation, Y is observed wage, D is an indicator for

college attendance, Z is parental education, and Y_{dz} is the potential wage when college attendance D and parental education Z are externally set to d and z, respectively. In the context here, the AMIV assumption acknowledges that parental education may positively influence children, while also accommodating the potential for its marginal impact to vanish beyond a certain threshold. We elaborate on this application in Section 6.2.

Notice that by construction, for all $z = 1, \ldots, k-1$, a_z implies a_{z+1}. Therefore, we have $\Theta(\{a_z\}) \subseteq \Theta(\{a_{z+1}\})$ for all $z \in \{1, \ldots, k-1\}$. In Section 4, we will show that this nested structure of the assumptions may lead to a situation where we cannot have discordant submodels. More generally, we will derive sufficient conditions for the nonexistence of discordant submodels.

When comparing these two leading examples, it is important to recognize that the motivation that leads to the construction of the outer sets or the relaxation of the initial model is different. In the first example, the outer sets are generated from nonsharp restrictions and are only used as a device aiming to provide useful information about the identified set. This situation usually occurs when one does not know how to conduct inference directly for the identified set or when the inference for the identified set is computationally intractable to implement. In Section 3 below, we will explore why such an approach could lead to unreliable results. In the second example, the outer sets are introduced in a more constructive manner aiming to relax a model refuted by the data. They are based on more primitive assumptions on the latent variables. In such a case, it is often possible to provide a series of outer sets that are not discordant with others. We will discuss this case in Section 4.

3. Misleading submodels

We view a model as a collection of constraints on the latent, observable variables, and the parameters. Throughout the paper, the parameter space Θ is assumed to be some subset in a metric space, which can be of finite or infinite dimensions. Let A be some nonempty collection of these constraints. We consider A as the *full model* (or simply model when there is no confusion) and, $A' \subsetneq A$ as a *submodel*. For any nonempty subset $A' \subseteq A$, we use $\Theta_I(A')$ to denote the set of parameter values that satisfy all the constraints in A'. For each $a \in A$, we abbreviate $\Theta_I(\{a\})$ as $\Theta_I(a)$. Let $\emptyset$ denote the empty set. For our purpose, $\Theta_I(\emptyset)$ can be an arbitrary nonempty subset of Θ such that $\Theta_I(a) \subseteq \Theta_I(\emptyset)$ for any $a \in A$. For the sake of simplicity, when there is no confusion, we will just use Θ to refer to $\Theta_I(\emptyset)$.

By definition, $\Theta_I(A') \subseteq \Theta_I(A'')$ if $A'' \subseteq A'$. As a result, for any $A' \subseteq A$, $\Theta_I(A')$ is an outer set of $\Theta_I(A)$. Moreover, we say a submodel A' is *data-consistent* if $\Theta_I(A')$ is nonempty, and call it *refuted* or *misspecified* if the reverse is true. Since $\Theta_I(A) \subseteq \Theta_I(A')$ for any $A' \subseteq A$, if the model A is data-consistent, we know that each $A' \subseteq A$ is also data consistent, and $\Theta_I(A') \cap \Theta_I(A'')$ is nonempty for any two submodels A' and A''.

In this section, we focus on A that consists of constraints that could be written in terms of the observable variables and model parameters only. This includes models where the identified sets are entirely characterized by a set of moment (in)equalities, including generalized method of moments (GMM) models, or Artstein (1983) inequalities

involving only observables and the parameter of interest. We refer to this type of constraints as identification conditions. These identification conditions are often derived from primitive assumptions involving the latent variables. In this restricted framework, we view A as the sharp identification conditions, and we call A' nonsharp identification conditions if $\Theta_I(A) \subsetneq \Theta_I(A')$. For instance, in the first leading example, the role of A is played by the set of moment inequalities (3) indexed by all instrumental functions h, with $\Theta_I(A) = [\underline{\gamma}, \overline{\gamma}]$. A submodel $A' \subseteq A$ could refer to (3) for a specific function h, with $\Theta_I(A') = \widetilde{\Theta}(h)$. If $[\underline{\gamma}, \overline{\gamma}] \subsetneq \widetilde{\Theta}(h)$, then A' is a nonsharp identification condition.

When identification conditions are written in terms of the observable variables and model parameters only, they satisfy the following assumption.

ASSUMPTION 2. *For any $A' \subseteq A$, $\Theta_I(A') = \bigcap_{a \in A'} \Theta_I(a)$.*

For example, all moment (in)equality models satisfy this assumption. This assumption may not hold when some $a \in A$ involves primitive restrictions on latent variables, as is the case in our second leading example.

In empirical works with partially identified models, researchers often use nonsharp identification conditions instead of the sharp ones. This is often motivated by two reasons: (i) sometimes the researchers may not even know the sharp characterization of the identified set, and (ii) the sharp identification conditions might be computationally intractable given the existing inferential methods.[6] When empirical results are based on an outer set obtained from nonsharp identification conditions, they are traditionally viewed as conservative yet valid because these outer sets are always supersets of the sharp identified set obtained using the sharp identification conditions. However, we are going to present a theorem that shows that outer sets obtained from nonsharp identification conditions are not always reliable and could potentially be misleading in the presence of model misspecification. We need the following assumption for this formal result.

ASSUMPTION 3. *There exists a collection $\mathscr{C}$ of subsets of A such that*:

(1) $\forall A' \in \mathscr{C}$, A' *is data-consistent and consists of finite elements in A,*

(2) $\Theta_I(\bigcup_{A' \in \mathscr{C}} A') = \Theta_I(A)$,

(3) *either $\mathscr{C}$ is finite or, for each $A' \in \mathscr{C}$, $\Theta_I(A')$ is compact.*

To clarify this assumption, we start with the simple case where all the conditions can be verified for $\mathscr{C} = \{\{a\} : a \in A\}$. In such a scenario, Assumption 3 breaks down into the following two parts: (*i*) every $a \in A$ is data-consistent, and (*ii*) either A is finite or $\Theta_I(a)$ is compact for each $a \in A$. Part (*i*) ensures that A will not be refuted because a specific $a \in A$ is refuted. When A is finite, this is enough to ensure that identification condtions are not mutually compatible when A is refuted. This reasoning can also apply to cases when A has infinite elements, given the compactness condition in part (*ii*) is

---

[6]See, for example, Berry and Compiani (2023, Section 6, footnote 42).

true. In Assumption 3, we permit $\mathscr{C}$ to be formulated in ways beyond $\mathscr{C} = \{\{a\} : a \in A\}$. By doing so, $\Theta_I(a)$ need not necessarily be compact for each $a \in A$ in infinite cases. We only require the identified set for some finite combinations of $a \in A$ is compact. This flexibility is helpful, for example, when some identification conditions determine the lower bound for parameters, while others define the upper bound.

We would verify Assumption 3 with $\mathscr{C} = \{\{a\} : a \in A\}$ for the first leading example in the next subsection. For an example of $\mathscr{C}$ being constructed in other ways, see Appendix A.2 where Assumption 3 is verified for conditional moment inequalities models. We are now ready to state the formal result.

THEOREM 1. *Under Assumptions 2 and 3, $\Theta_I(A) = \emptyset$ if and only if there exist two finite subsets A', $A'' \subseteq A$ such that both A' and A'' are data-consistent and $\Theta_I(A') \cap \Theta_I(A'') = \emptyset$.*

Moreover, when $\Theta_I(A) = \emptyset$, for any data-consistent $B \subseteq A$, there exists two finite subset B', $B'' \subseteq A$ such that both $B \cup B'$ and B'' are data-consistent and $\Theta_I(B \cup B') \cap \Theta_I(B'') = \emptyset$.

Theorem 1 tells us that when the model is refuted, an outer set derived from one set of nonsharp identification conditions could be completely different from an outer set obtained from a different set of nonsharp identification conditions. In such a case, the information delivered by an outer set depends mainly on which nonsharp identification conditions the researcher decides to use. Therefore, for this class of models, applied researchers must be very careful in interpreting outer sets based on nonsharp identification conditions. We will illustrate this point further with the first leading example in the next subsection. Later, in Section 6.1, we provide a numerical illustration of Theorem 1 when applied to an entry game example.

Moreover, Theorem 1 shows that, for any data-consistent nonsharp identification condition, there always exists another set of nonsharp identification conditions that are discordant with some of its strengthened versions. This suggests that the issue of discordancy is not confined to certain pairs of outer sets. When the model is refuted, we can start with any data-consistent outer set and further tighten its bounds by applying more restrictions. After incorporating just a finite number of additional restrictions, it will inevitably be in conflict with another data-consistent outer set, even if the cardinalities of A and $\mathscr{C}$ are uncountably infinite.

Although Theorem 1 seems to focus essentially on partially identified models, it also applies to point-identified models. Suppose that each outer set derived from each single identification condition is a singleton, that is, $\Theta_I(a)$ is a singleton for all $a \in A$. Then Theorem 1 says that the model is misspecified if and only if there exist two different identification conditions $a, a' \in A$ such that $\Theta_I(a) \neq \Theta_I(a')$. This is related to a long-existing observation in point-identified models: whenever a model is overidentified, one can test the model specification by comparing the point-estimates obtained from different identification conditions. In point-identified models, the issue of overidentification or misspecification is a direct concern for the researchers. In partially identified

models, however, applied researchers tend to believe that their results are more credible, and thus less sensitive to misspecification. Therefore, they often interpret the tightness of an outer set as a signal of an informative identified set, as discussed in Molinari (2020) regarding the "usefulness" of outer sets. Theorem 1 provides a different perspective that, for a certain class of models, outer sets could be misleading. Therefore, applied researchers should be very careful in interpreting outer sets based on nonsharp identification conditions, even if the model is only partially identified. In the following, we analyze the implications of Theorem 1 on our first leading example.

3.1 *Intersection bounds example continued*

To begin, let us construct A in this example. Define $\mathcal{H}_m^+$ to be the space of all nonnegative instrumental functions with dimension m. More formally, let $\mathcal{H}_m^+ \equiv \{h : \mathcal{Z} \mapsto \mathbb{R}_+^m$ such that $E\|h(Z)\| < \infty$, $E\|\underline{Y}h(Z)\| < \infty$, $E\|\overline{Y}h(Z)\| < \infty$, and $E[h_i(Z)] > 0, \forall i = 1, \ldots, m\}$. Let A be the set of all identification conditions (3) indexed by $h \in \mathcal{H}_1^+$. The set A constructed here represents sharp identification conditions, because (1) holds if and only if (3) holds for all $h \in \mathcal{H}_1^+$.

Next, we verify Assumptions 2 and 3 hold. Note that, for each $a \in A$, $\Theta_I(a)$ is nonempty and compact. This is because each $a \in A$ corresponds to an $h \in \mathcal{H}_1^+$, and because for all $h \in \mathcal{H}_1^+$, the identified set $\widetilde{\Theta}(h)$ for model (3) is equal to the following interval:

$$\left[\frac{E[h(Z)\underline{Y}]}{E[h(Z)]}, \frac{E[h(Z)\overline{Y}]}{E[h(Z)]}\right],$$

which is nonempty and compact under Assumption 1. We can let $\mathcal{C} = \{\{a\} : a \in A\}$. Then $\Theta_I(A')$ is nonempty and compact for all $A' \in \mathcal{C}$. Moreover, $\Theta_I(A) = \Theta_I(\bigcup_{A' \in \mathcal{C}} A')$ by the construction of $\mathcal{C}$. Thus, Assumption 3 is satisfied. Since (3) are moment inequalities, which only depend on observables and the parameter, Assumption 2 is also satisfied.

Note that, in this example, if $B \subseteq A$ and B consists of m assumptions, then B refers to the submodel that (3) holds for some $h \in \mathcal{H}_m^+$. As a result, Theorem 1 implies that when (1) is refuted, there must exist some $h_1 \in \mathcal{H}_{m_1}^+$ and $h_2 \in \mathcal{H}_{m_2}^+$, such that $\widetilde{\Theta}(h_1) \neq \emptyset$ and $\widetilde{\Theta}(h_2) \neq \emptyset$ but $\widetilde{\Theta}(h_1) \cap \widetilde{\Theta}(h_2)$ is empty. In fact, because of the specific structures of this example, we can even obtain the following stronger result.

PROPOSITION 1. *Suppose Assumption 1 holds. If the restriction in (1) is refuted, that is, $\underline{\gamma} > \overline{\gamma}$, then for any θ in $(\overline{\gamma}, \underline{\gamma})$, there exists some $h \in \mathcal{H}_2^+$ such that $\widetilde{\Theta}(h) = \{\theta\}$. Conversely, if there exists some integer m and some $h \in \mathcal{H}_m^+$ such that $\widetilde{\Theta}(h) = \{\theta\}$, then $\theta \in [\overline{\gamma}, \underline{\gamma}]$.*

When (1) is refuted, Proposition 1 shows that the unconditional moment restrictions can point identify any element in the crossed bound $(\overline{\gamma}, \underline{\gamma})$ with a properly chosen instrumental function. The width of $(\overline{\gamma}, \underline{\gamma})$ depends on the extent of the model violation: the worse the violation is, the wider this interval would be. In the extreme case where the mean independence condition is significantly violated such that $[E\underline{Y}, E\overline{Y}] \subseteq (\overline{\gamma}, \underline{\gamma})$, it implies that any point in the Manski worst-case bounds can be selected as the point-identification result by a suitable choice of h.

Proposition 1 also sheds some light on the implementation of the inference procedure in Andrews and Shi (2013), which is one of the most popular inference procedures used for conditional moment inequalities. Andrews and Shi (2013) converts the conditional moment inequalities into unconditional moment inequalities in the same way as we transformed, in the first leading example, (1) into (3). A notable distinction in Andrews and Shi's (2013) approach is that the number of instruments, that is, the dimension of h in our notation, increases to infinity as the sample size grows. Our results show that the usage of infinite number of instruments in the limit is crucial to achieve reliable inference results. In contrast, if researchers implement their inference but choose an instrumental function with a finite and fixed dimension, their results could be spuriously informative and misleading, as underscored in Proposition 1. See also the formal results in Appendix A.2.

It is worth noting that Theorem 1 applies to much more general frameworks. More precisely, in Appendices A.2 and A.3, we provide sufficient conditions under which Theorem 1 applies to two widely used classes of partially identified models. The first is a class of models where the identified set is characterized by the following type of conditional moment inequalities:

$$E\big[m(X, Z; \theta)|Z\big] \leq 0 \quad \text{almost surely.} \tag{4}$$

Here, $X \in \mathbb{R}^{k_1}$ and $Z \in \mathbb{R}^{k_2}$ are observable random variables, and $m(\cdot, \cdot; \theta) \in \mathbb{R}$ is some known integrable function for each θ. If researchers construct outer sets by transforming (4) into finite-dimensional unconditional moment inequalities, then a similar discordancy issue may happen, as the one we have seen in the first leading example. See Appendix A.2 for more details.

The second class of models that Theorem 1 could be applied to is the class for which the identified set can be characterized by Artstein's (1983) inequalities. In her recent survey, Molinari (2020) shows that this class of models includes simultaneous-move finite games with multiple equilibria, auction models with independent private values, network formation models, treatment effect models, etc. In many of those cases, the number of Artstein's (1983) inequalities that characterize the (sharp) identified set is extremely high (very often much higher than the sample size of the data under use). In practice, for the sake of computational feasibility, empirical researchers often pre-select a finite collection of Artstein's inequalities to obtain an outer set. As examples, we could cite Ciliberto and Tamer (2009), Haile and Tamer (2003), Sheng (2020), Chesher and Rosen (2020), and Berry and Compiani (2023), among many others. In Appendix A.3, we show that those pre-selected nonsharp moment inequalities suffer the same issue pointed out in Theorem 1.

In Section 6, we provide a numerical illustration of the discordancy issue by visiting the widely used entry game model studied in Ciliberto and Tamer (2009). We will explore in more detail the consequences of this pre-selection procedure when the original model might be refuted and illustrate the implication of our Theorem 1 in this widely used framework.

4. Compatible submodels and minimum data-consistent relaxation

As discussed in the previous section, there could be discordant submodels when the full model is refuted. However, the falsification of the full model does not necessarily lead to discordance of the submodels. Unlike in the previous section, we now consider A that consists of primitive assumptions on latent variables in addition to those on observable variables and the parameter of interest. In this wider class of models, we will present a sufficient condition that ensures that all data-consistent submodels are always compatible with each other. In this section, for the sake of simplicity, we focus on the case where A is finite. For the case where A is infinite, similar results could be derived under additional conditions. We elaborate on those conditions and results in Appendix A.4.

To state our result, we need to introduce a new concept. When the full model is refuted, we can obtain a data-consistent submodel by dropping or relaxing some of the assumptions. We say that a data-consistent submodel is a minimum relaxation if we relax the minimum number of assumptions needed to restore data consistency.

DEFINITION 1. Let $\widetilde{A}$ be a subset of A. We say $\widetilde{A}$ is a *minimum data-consistent relaxation* of A if $\Theta_I(\widetilde{A})$ is nonempty and for any $a \in A \backslash \widetilde{A}$, $\Theta_I(\widetilde{A} \cup a)$ is empty.

It is worth noting that the concept of a minimum data-consistent relaxation depends on how the researcher defines each of the simple assumptions a that constitute A. Therefore, all the subsequent results relying on the minimum data-consistent relaxation depend on the way the researcher constructs A. We will return to this point in Section 5.3. Furthermore, a minimum data-consistent relaxation always exists when A is finite; its existence requires additional conditions when A is infinite. See Appendix A.4. To illustrate this concept, let us consider a simple example where $A = \{a_1, a_2, a_3\}$. The identified sets of each a_i are all closed intervals in $\mathbb{R}$ as shown in Figure 2 with $\Theta_I(a_1) = [b, c]$, $\Theta_I(a_2) = [d, e]$, $\Theta_I(a_3) = [f, g]$, and $f \le b \le c < d \le e \le g$. Assume also for the purpose of illustration that $\Theta_I(\{a, a'\}) = \Theta_I(a) \cap \Theta_I(a')$ for $a, a' \in \{a_1, a_2, a_3\}$.

In this example, both $\{a_1, a_3\}$ and $\{a_2, a_3\}$ are minimum data-consistent relaxations. And, $\{a_3\}$ is not a minimum data-consistent relaxation, since it will remain data-consistent after including a_1 or a_2. In general, minimum data-consistent relaxations may or may not be unique. We will defer the discussion of multiple minimum data-consistent relaxations to the next section. In this section, we focus on the situation where there exists a unique minimum data-consistent relaxation. In fact, the uniqueness of the minimum data-consistent relaxation ensures the absence of discordancy issues discussed in the previous section. More precisely, in Appendix A.5, we show that the existence of two discordant submodels is equivalent to the existence of two minimum data-
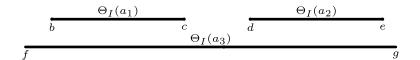

FIGURE 2. The three-interval example.

consistent relaxations that are also discordant each other under some regularity conditions. Therefore, the uniqueness of the minimum data-consistent relaxation ensures the absence of discordant submodels.

The following result describes the conditions under which there exists a unique minimum data-consistent relaxation.

Theorem 2. *Suppose A is finite. Then the following statements are equivalent*:

(T2.C1) *for any $A' \subseteq A$, A' is data-consistent if and only if all $a \in A'$ are data-consistent.*

(T2.C2) *There exists a unique minimum data-consistent relaxation A^*.*

Theorem 2 through its condition (T2.C1) provides a way to check whether there exists a unique minimum data-consistent relaxation A^*. Note that Condition (T2.C1) does not hold in every model: in a general model, each $a \in A'$ being data-consistent is necessary but not sufficient for A' to be data-consistent, because individual data-consistent assumptions might not be mutually compatible when combined. Condition (T2.C1) can be verified by investigating when each $A' \subseteq A$ is data-consistent, which could be done even without seeing the data. In terms of interpretation, condition (T2.C1) implies that all data-consistent submodels are compatible with each other. It also implies that the set of data-consistent submodels, that is, $\{A' \subseteq A : \Theta_I(A') \neq \emptyset\}$, is closed under the union operation: the union of data-consistent submodels remains data-consistent.

When the full model A is data-consistent, the unique minimum data-consistent relaxation A^* is just equal to A. When A is refuted, A^* can be viewed as the model learned from the data by removing all refuted assumptions in A while keeping all the data-consistent ones. Indeed, condition (T2.C1) suggests that $A^* = \{a \in A : \Theta_I(a) \neq \emptyset\}$. The interpretation of A^* and its role as the unique minimum data-consistent relaxation will be studied further in Appendix A.6. One way to illustrate condition (T2.C1) is to consider our second leading example.

4.1 *AMIV example continued*

Recall that by construction, for all $z = 1, \ldots, k - 1$, a_z implies a_{z+1}. In addition, define $a^\dagger$ as the collection of *E.1* and *E.2*. Let $A = \{a_1, \ldots, a_k, a^\dagger\}$ be the collection of all assumptions. Then the full model A is the classic mean independence assumption considered in Manski (1990). Within this second leading example, all assumptions are nested. That is, for any two $a, a' \in A$, either a implies a' or a' implies a. Therefore, the data-consistency of a set of assumptions is equal to the data-consistency of the strongest assumption in that set, which implies the validity of (T2.C1).[7] Therefore, (T2.C1) holds in this example. Theorem 2 then implies that all data-consistent submodels will be compatible with each other and there exists a unique minimum data-consistent relaxation A^*. The following result characterizes the identified set of A^*. To state the result, we use the following notation: $\underline{Y}_d = Y\mathbb{1}(D = d) + \underline{y}_d\mathbb{1}(D \neq d)$, $\overline{Y}_d = Y\mathbb{1}(D = d) + \overline{y}_d\mathbb{1}(D \neq d)$, $\underline{q}_{dz} = E[\underline{Y}_d | Z = z]$, and $\overline{q}_{dz} = E[\overline{Y}_d | Z = z]$.

---

[7]However, it is worth noting this nested structure is not necessary for condition (T2.C1) to hold.

PROPOSITION 2. *Assume that $P(Y \in [\underline{y}_d, \overline{y}_d] | D = d) = 1$ for any $d \in \{0, 1\}$. Let $\theta = (\theta_1, \theta_0)$ be the parameter of interest. Then model A always has a unique minimum data-consistent relaxation A^*, and A^* always contains $a^\dagger$. In addition, for any $z = 1, \ldots, k$, $a_z \in A^*$ if and only if the following two conditions hold for each $d \in \{0, 1\}$:*

$$\forall z' < z, \quad \max(\underline{q}_{dt} : t \le z') \le \min(\overline{q}_{dt} : t \ge z') \tag{5}$$

and

$$\max(\underline{q}_{dt} : t = 1, \ldots, k) \le \min(\overline{q}_{dt} : t \ge z). \tag{6}$$

Hence, $a_z \in A^*$ implies that $a_{z'} \in A^*$ for all $z' > z$. Moreover, if $\{z : a_z \in A^*\}$ is nonempty, define $z^* = \min\{z : a_z \in A^*\}$ and

$$\Gamma_{d,z^*} = \Bigg[\sum_{z < z^*} P(Z = z) \max(\underline{q}_{dt}, t \le z) + \sum_{z \ge z^*} P(Z = z) \max(\underline{q}_{dt} : t = 1, \ldots, k),$$

$$\sum_{z < z^*} P(Z = z) \min(\overline{q}_{dt} : t \ge z) + \sum_{z \ge z^*} P(Z = z) \min(\overline{q}_{dt} : t \ge z^*) \Bigg]. \tag{7}$$

Then $\Theta_I(A^*) = \Gamma_{1,z^*} \times \Gamma_{0,z^*}$. If $\{z : a_z \in A^*\}$ is empty, then $\Theta_I(A^*) = [E[\underline{Y}_1], E[\overline{Y}_1]] \times [E[\underline{Y}_0], E[\overline{Y}_0]]$.

REMARK 1. It is worth noting that, for simplicity, we impose the cut-off z^* to be the same for all potential outcomes in *E.3*; however, we do not need to do so. We can let the data determine the cut-offs for each potential outcome separately.

5. MISSPECIFICATION ROBUST BOUNDS

In this section, we consider cases where there are multiple data-consistent relaxations. According to Theorem 2, the multiplicity of minimum data-consistent relaxations is a necessary condition for the existence of discordant submodels. Indeed, whenever there are two mutually incompatible data-consistent submodels, there are at least two minimum data-consistent relaxations. If there is no reason to favor one submodel over another ex ante, it is reasonable to consider all of these relaxations.

DEFINITION 2. Let $\mathscr{A}_R$ be the set of all minimum data-consistent relaxations. The *misspecification robust bound* Θ_I^* is defined as $\Theta_I^* \equiv \bigcup_{\widetilde{A} \in \mathscr{A}_R} \Theta_I(\widetilde{A})$.

The misspecification robust bound concept is similar to the falsification adaptive set concept introduced in Masten and Poirier (2021). However, a distinctive feature of this section is that we focus on discrete relaxations, where an assumption is either dropped or kept, while Masten and Poirier (2021) focuses exclusively on relaxing assumptions in a continuous way. In general, the type of relaxation depends on the empirical question under study. In the following, we derive the misspecification robust bound for our two leading examples.

5.1 *Intersection bounds example continued*

For the model (1), the misspecification robust bound is given in the following result.

PROPOSITION 3. *Suppose Assumption 1 holds, then*

$$
\Theta_I^* = \begin{cases}
[\underline{\gamma}, \overline{\gamma}] & \textit{if } \underline{\gamma} \leq \overline{\gamma}, \\
[\overline{\gamma}, \underline{\gamma}] & \textit{if } \overline{\gamma} < \underline{\gamma},\ P\big(E[\underline{Y}|Z] \leq \overline{\gamma}]\big) > 0 \textit{ and } P\big(E[\overline{Y}|Z] \geq \underline{\gamma}]\big) > 0, \\
(\overline{\gamma}, \underline{\gamma}] & \textit{if } \overline{\gamma} < \underline{\gamma},\ P\big(E[\underline{Y}|Z] \leq \overline{\gamma}]\big) = 0 \textit{ and } P\big(E[\overline{Y}|Z] \geq \underline{\gamma}]\big) > 0, \\
[\overline{\gamma}, \underline{\gamma}) & \textit{if } \overline{\gamma} < \underline{\gamma},\ P\big(E[\underline{Y}|Z] \leq \overline{\gamma}]\big) > 0 \textit{ and } P\big(E[\overline{Y}|Z] \geq \underline{\gamma}]\big) = 0, \\
(\overline{\gamma}, \underline{\gamma}) & \textit{if } \overline{\gamma} < \underline{\gamma},\ P\big(E[\underline{Y}|Z] \leq \overline{\gamma}]\big) = 0 \textit{ and } P\big(E[\overline{Y}|Z] \geq \underline{\gamma}]\big) = 0.
\end{cases}
\tag{8}
$$

A direct implication of Proposition 3 is that if $P(E[\underline{Y}|Z] \leq \overline{\gamma}) > 0$ and $P(E[\overline{Y}|Z] \geq \underline{\gamma}) > 0$ hold, which are mild technical requirements, the misspecification robust bound simplifies to $\Theta_I^* = [\min(\underline{\gamma}, \overline{\gamma}), \max(\underline{\gamma}, \overline{\gamma})]$ whether or not the full model is refuted.

5.2 *AMIV example continued*

A direct implication of Proposition 2 is that the AMIV model has a unique minimum data-consistent relaxation A^*, which can be summarized as follows: $A^* = \{a^\dagger\} \cup \{a_z :$ equations (5) and (6) hold\}. Therefore, the misspecification robust bound for the AMIV model is

$$
\Theta_I^* = \begin{cases}
\Gamma_{1,z^*} \times \Gamma_{0,z^*} & \text{if } A^* \neq \{a^\dagger\}, \\
[E[\underline{Y}_1], E[\overline{Y}_1]] \times [E[\underline{Y}_0], E[\overline{Y}_0]] & \text{if } A^* = \{a^\dagger\},
\end{cases}
\tag{9}
$$

where $z^* = \min\{z : a_z \in A^*\}$.

5.3 *Empirical interpretation of the misspecification robust bound*

As we pointed out earlier, Θ_I^* will depend on how the researcher decides to define the assumptions a that constitute A. Different constructions of A correspond to different ways to relax a refuted model. It is inevitable that there almost always exist multiple ways to relax a stringent and refuted model, and different ways of relaxations would lead to different results. Instead of drawing a general conclusion about which relaxation approach is superior, we believe it is important to offer an empirical interpretation of Θ_I^* for a given A. In this way, even if different researchers may construct A based on their own economic interpretations of the model, they would have a clear interpretation of their results.

In Theorem 5 in Appendix A.6, we show that the misspecification robust bound Θ_I^* is both *rationalizable* and *nonconflicting* in the following sense:

- (Rationalizable) The statement that Θ_I^* contains the true parameter is implied by some data-consistent submodel. That is, there exists some data-consistent submodel $A' \subseteq A$ such that $\Theta_I(A') \subseteq \Theta_I^*$.

- (Nonconflicting) The statement that Θ_I^* contains the true parameter is not rejected by any data-consistent submodel. That is, there does not exist a data-consistent submodel $A' \subseteq A$ such that $\Theta_I(A') \cap \Theta_I^* = \emptyset$.

When the full model is refuted, different data-consistent submodels can imply different and potentially discordant statements on θ. Among all possible statements on θ, we think that being rationalizable and nonconflicting is a minimum requirement for a statement to be robust to model misspecification. If a statement fails to be rationalizable, then it is not implied by any of the data-consistent submodels. If a statement is not nonconflicting, then it is rejected by some data-consistent submodels.

The fact that Θ_I^* is both rationalizable and nonconflicting gives it an interesting empirical interpretation. Consider the simple case where θ is a scalar. Suppose we are interested in the sign of θ. And, suppose Θ_I^* turns out to be within the positive real line, that is, $\Theta_I^* \subseteq \mathbb{R}++$. Then it means that some submodels identify the sign of θ to be positive, and whenever the sign of θ can be identified by a submodel, the sign of θ is always positive.

In some cases, Θ_I^* is the *smallest* set that is both rationalizable and nonconflicting. That is, for any $\widetilde{\Theta} \subseteq \Theta$, $\widetilde{\Theta}$ is both rationalizable and nonconflicting if and only if $\Theta_I^* \subseteq \widetilde{\Theta}$. In this case, Θ_I^* could have richer interpretations. Consider the previous simple example again. Suppose it turns out that $\Theta_I^* \cap \mathbb{R}_{++} \neq \emptyset$ and $\Theta_I^* \cap \mathbb{R}_{--} \neq \emptyset$ so that $\theta \in \Theta_I^*$ does not imply the sign of θ. If we know Θ_I^* is the smallest rationalizable and nonconflicting set, then we have the following conclusion: neither θ is positive nor θ is negative are rationalizable and nonconflicting statements. In other words, the value of Θ_I^* in this case implies that the data and the model cannot provide a clear statement on the sign of θ. Finally, in Theorem 7 in Appendix A.6, we show that Θ_I^* would be the smallest rationalizable and nonconflicting set if there exists a unique minimum data-consistent relaxation or the identified set for each minimum data-consistent relaxation is a singleton.

5.4 *Discrete relaxation versus continuous relaxation*

As can be seen, the misspecification robust bound relaxes a refuted model in a discrete way: an assumption is either fully kept or dropped during the relaxation. There are many other ways to relax and salvage a refuted model. One can also relax assumptions continuously as in Masten and Poirier (2021). In general, different relaxations will lead to different results, and it is hard to compare all the possible approaches. However, there does exist a special case where discrete relaxation always leads to more informative results than any other ways of relaxations.

In order to make an adequate comparison, we need to introduce the terminology used in Masten and Poirier (2021). For any $\epsilon \in [0, 1]$ and any $a \in A$, let a_ϵ denote the assumption after relaxing assumption a. The degree of relaxation is measured by ϵ: when $\epsilon = 0$, $a_\epsilon = a$; when $\epsilon \in (0, 1)$, the assumption a is partially relaxed but the exact form of a_ϵ would depend on the specific way of relaxation chosen by the researcher; when $\epsilon = 1$, the assumption a is completely relaxed and a_ϵ is a null assumption, which does not impose any restriction. Assume the relaxation is monotone: if $\epsilon_1 \leq \epsilon_2$, a_{ϵ_1} is stronger than a_{ϵ_2}, in the sense that a_{ϵ_1} implies a_{ϵ_2}. For any $\delta : A \to [0, 1]$, define

$A(\delta) \equiv \{a_{\delta(a)} : a \in A\}$ as the perturbed full model. For any two $\delta_1 : A \rightarrow [0, 1]$ and $\delta_2 : A \rightarrow [0, 1]$, we write $\delta_1 < \delta_2$ if $\delta_1(a) \leq \delta_2(a)$ for all $a \in A$ and $\delta_1(a) < \delta_2(a)$ for some $a \in A$. Then the *falsification frontier* (FF) in Masten and Poirier (2021) can be defined as FF $= \{\delta : A \rightarrow [0, 1] : \Theta_I(A(\delta)) \neq \emptyset$ and there does not exist δ' such that $\Theta_I(A(\delta')) \neq \emptyset, \Theta_I(A(\delta')) \subsetneq \Theta_I(A(\delta))$ and $\delta' < \delta\}$. We slightly modified the definition of the falsification frontier of Masten and Poirier (2021) to ensure the nonemptiness of FF in some special cases.[8] Then the *falsification adaptive set* $\Theta_I^\dagger$ is defined as $\Theta_I^\dagger = \bigcup_{\delta \in \mathrm{FF}} \Theta_I(A(\delta))$.

Note that $\Theta_I^\dagger$ depends on the specific way that one chooses to relax the assumptions. If one chooses to relax them discretely, that is, if $a_\epsilon = a_{\mathbb{1}(\epsilon > 0)}$ for any ϵ and a, then $\Theta_I^\dagger$ is equal to the minimum data-consistent relaxation Θ_I^*. If one chooses a different way of relaxation, the $\Theta_I^\dagger$ is generally different. In some special cases, however, Θ_I^* is always included in $\Theta_I^\dagger$ no matter which way of relaxation is chosen. More precisely, whenever for any minimum data-consistent relaxation $\widetilde{A}$, $\Theta_I(\widetilde{A})$ is a singleton, it can be shown that $\Theta_I^* \subseteq \Theta_I^\dagger$ for any type of relaxation chosen by the researcher. We formally state and prove this result, respectively, in Theorem 8 in Appendix A.7.

6. Numerical and empirical illustrations

In this section, through numerical exercises, we illustrate two of our main theoretical results. In Section 6.1, we consider the entry game model studied in Ciliberto and Tamer (2009, CT). We simulate a misspecified entry game model such that the sharp identification conditions deliver an empty identified set. In such a context, we show that it is possible to generate multiple nonempty conflicting outer sets by just selecting different sets of nonsharp identification conditions. This illustrates the discordancy issue raised in Theorem 1. In Section 6.2, we revisit a return to college application and report the estimated identified set of the minimum data-consistent relaxation derived under the assumption that parental education satisfies the AMIV assumption.

6.1 *Numerical illustration of discordant outer sets in an entry game model*

Consider an entry game model with m players. Each player i chooses $Y_i \in \{0, 1\}$ to maximize its payoff:

$$\pi_i = Y_i \left(\alpha_i + X_i \beta - \sum_{j \neq i} \delta_j^i Y_j + \epsilon_i \right), \tag{10}$$

where δ_j^i is player j's competition impact on player i, $\epsilon = (\epsilon_1, \ldots, \epsilon_m) \sim N(0, I_m)$, and I_m is the identity matrix. Denote $\alpha = (\alpha_i : i = 1, \ldots, m)$, $\delta = (\delta_j^i : i \neq j)$, and, let $\theta = (\alpha, \beta, \delta)$ collect all the parameters. We model the players' behavior as pure-strategy Nash equilibrium. Therefore, we restrict the parameter space Θ to be the set of parameters where pure-strategy equilibria exist with probability 1. This is the same empirical model used

---

[8]The original definition in Masten and Poirier (2021), written in our notation, is FF $= \{\delta : A \rightarrow [0, 1] : \Theta_I(A(\delta)) \neq \emptyset$ and there does not exist δ' such that $\Theta_I(A(\delta')) \neq \emptyset$ and $\delta' < \delta\}$. With our modified definition, we do not need to worry about the possibility that there is a sequence of $\{\delta_i : i \geq 1\}$ such that $\delta_n \rightarrow \delta^*$, $\Theta_I(A(\delta^*)) = \emptyset$, $\Theta_I(A(\delta_n)) = \Theta_I(A(\delta_1)) \neq \emptyset$, and $\delta_{n+1} < \delta_n$ for all $n \geq 1$.

in Ciliberto and Tamer (2009), except for two simplifications: (i) we assume that β is the same for all players; (ii) we assume that the distribution of ϵ is known.

Let $\mathcal{K}$ be the collection of all subsets of $\mathcal{Y} \equiv \{0, 1\}^m$. Let F denote the joint distribution of (Y, X) in the data. Define $\Gamma(x, \epsilon; \theta)$ as the set of all pure-strategy Nash equilibria given (x, ϵ), the parameter θ and the payoff function in (10). As shown in Galichon and Henry (2011) and equivalently in Beresteanu, Molchanov, and Molinari (2011), θ belongs to the identified set $\Theta_I(F)$ if and only if

$$\forall K \in \mathcal{K}, \quad \mathbb{P}_F(Y \in K | X) \leq \mathbb{P}(\Gamma(X, \epsilon; \theta) \cap K \neq \emptyset | X), \quad X\text{-a.s.} \tag{11}$$

Equation (11) characterized the Artstein (1983) inequalities associated to the entry game model under study. For each fix covariate x, we have 2^{2^m} inequalities to be checked. In this case, our full model A is defined by the whole set of Artstein's inequalities, and then its cardinality is $2^{2^m} \times Card(X)$. This becomes easily nontractable even for a relatively small number of firms, that is, for instance, for a fixed x, and 5 firms we have $2^{2^5} = 4{,}294{,}967{,}296$ inequalities to be checked. Therefore, in practice, outer sets are almost always used. Let A' be a subset of $\mathcal{K}$. Then the outer set associated with A' is

$$\Theta_I(F, A') \equiv \{\theta \in \Theta : \forall K \in A', \mathbb{P}_F(Y \in K | X) \leq \mathbb{P}(\Gamma(X, \epsilon; \theta) \cap K \neq \emptyset | X), X\text{-a.s.}\}$$

A' is a nonsharp identification condition whenever $\Theta_I(F, A') \neq \Theta_I(F, \mathcal{K}) \equiv \Theta_I(F)$. In CT, for each fixed x, they considered the outer set associated with A_{ct} defined as $A_{\text{ct}} = \{\{y\} : y \in \mathcal{Y}\} \cup \{\{y\}^c : y \in \mathcal{Y}\}$ where $\{y\}^c$ stands for the complement set of $\{y\}$ in $\mathcal{Y}$. It is worth noting that whenever $m > 2$, A_{ct} is a nonsharp identification condition.

6.1.1 *Data generating process (DGP)*   In order to illustrate the issue of nonreliability of the outer sets in presence of misspecification, we generate a joint distribution F from a game that might be different from the model (10). Assume that, in the data generating process, each player i chooses $Y_i \in \{0, 1\}$ to maximize the following payoff instead of (10):

$$\pi_i = Y_i\left(\alpha_i + X_i\beta - \sum_{j \neq i}\delta_j^i Y_j - \sum_{j_1, j_2 \neq i}\gamma_{j_1, j_2}^i Y_{j_1}Y_{j_2} + \epsilon_i\right). \tag{12}$$

The extra vector of parameter $\gamma \equiv \{\gamma_{j_1, j_2}^i : \text{where } (i, j_1, j_2) \text{ are mutually different}\}$ captures the second-order competition effect. When $\gamma = 0$, the model in (10) is correctly specified. If $\gamma \neq 0$, the model in (10) is misspecified. To complete the model, we assume that, whenever there are multiple pure-strategy Nash equilibria, players will choose each equilibrium with the same probability. We assume that the support of X is a bounded interval $\mathcal{X}$ in $\mathbb{R}^{\dim(X)}$. Without loss of generality, we assume X is distributed uniformly in its support.

In the simulation, we focus on the simple case where $\gamma_{j_1, j_2}^1 = \gamma^*$ for all (j_1, j_2), and $\gamma_{j_1, j_2}^i = 0$ for all $(i \neq 1, j_1, j_2)$. Then the joint distribution of (Y, X) generated from this data generating process is indexed by (θ, γ^*), which we write as F_{θ, γ^*}. Note that γ^* measures the degree of misspecification: the larger the value of γ^* is, the larger the degree of misspecification of model in (10) is. In the simulation design, we impose that $m = 3$,
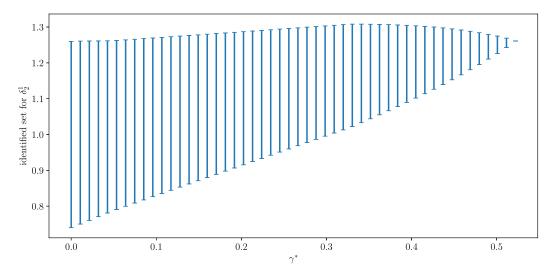
FIGURE 3. $\Lambda(F_{\theta,\gamma^*}, A_{\mathrm{ct}})$ at various values of γ^*.

$\dim X = 1$, $X \sim U[-1, 1]$, $\beta = 0.1$, $\alpha = [1, 1, 1]$, and $\delta^i_j = 1$ for all (i, j). In the following, we construct outer sets both for the parameter δ and also for a counterfactual outcome in a counterfactual experiment. In both cases, we will show that we will be able to generate three outer sets (including the CT outer set) that are discordant with each other and none of them contains the true value.

6.1.2 *Discordant nonsharp identification conditions for* δ Our first objective is to illustrate the existence of discordant nonsharp identification conditions. Here, we focus on the (projected) identified set for δ^1_2. Given the DGP with (θ, γ^*), the outer set for δ^1_2 associated with the nonsharp identification conditions in A' is characterized as follows: $\Lambda(F_{\theta,\gamma^*}, A') = \{t : \exists(\alpha, \beta, \delta) \in \Theta_I(F_{\theta,\gamma^*}, A') \text{ s.t. } \delta^1_2 = t\}$. Now, let us denote by $\overline{\gamma}^*_{\mathrm{ct}}$ the maximum degree of misspecification that could not be detected by the submodel A_{ct}. More precisely, we define $\overline{\gamma}^*_{\mathrm{ct}} := \sup\{\gamma^* : \Lambda(F_{\theta,\gamma^*}, A_{\mathrm{ct}}) \neq \varnothing\}$, therefore, $\Lambda(F_{\theta,\gamma^*}, A_{\mathrm{ct}})$ is a nonempty outer set when $\gamma^* \leq \overline{\gamma}^*_{\mathrm{ct}}$, and it becomes empty whenever $\gamma^* > \overline{\gamma}^*_{\mathrm{ct}}$. In Figure 3, we plot in blue the CT outer set, that is, $\Lambda(F_{\theta,\gamma^*}, A_{\mathrm{ct}})$, at various degree of model misspecification, for $\gamma^* \in [0, \overline{\gamma}^*_{\mathrm{ct}}]$, where 0 corresponds to no misspecification and $\overline{\gamma}^*_{\mathrm{ct}}$ corresponds to the maximum degree of misspecification that is not detectable with A_{ct}.

A first remark is that the CT outer set shrinks when the degree of misspecification increases and at some point it no longer contains the true value. This illustrates that the tightness of the outer set should not systematically be interpreted as a signal of an informative identified set but it could just signal a presence of misspecification.

In Appendix A.3, we explain why the findings of Theorem 1 apply to the entry game example. So, according to Theorem 1, there must exist some other A' nonsharp identification condition, which is discordant with A_{ct}. Indeed, we are able to find two sets of nonsharp identification conditions, denoted as A_1 and A_2, which result in discordant identification results with A_{ct}. In Figure 4, we plot $\Lambda(F_{\theta,\gamma^*}, A_1)$ in orange and
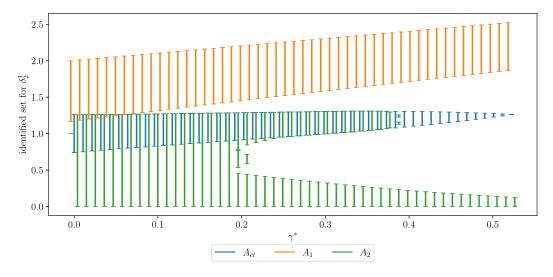
FIGURE 4. $\Lambda(F_{\theta,\gamma^*}, A_1)$ and $\Lambda(F_{\theta,\gamma^*}, A_2)$ at various values of γ^*.

$\Lambda(F_{\theta,\gamma^*}, A_2)$ in green. $\Lambda(F_{\theta,\gamma^*}, A_1)$ suggests values for δ_2^1 that are higher than those suggested by $\Lambda(F_{\theta,\gamma^*}, A_{\mathrm{ct}})$ while $\Lambda(F_{\theta,\gamma^*}, A_2)$ suggests values that are lower. When the degree of misspecification is higher than 0.4, these three outer sets have no overlap.

6.1.3 *Discordant counterfactual predictions* In many of the empirical games applications, applied researchers are very often interested in implementing counterfactual analyses. Interestingly, we observe not only discordant results for the parameters, but we also observe this phenomenon for counterfactuals. Below, we illustrate a scenario where different outer sets lead to discordant counterfactual outcomes when the full model is misspecified. Therefore, we illustrate the fact that the discordancy issue also applies to counterfactual outcomes. Let us consider a counterfactual where firm 3 is no longer a potential entrant of the market. This type of counterfactuals would arise, for example, when firm 3 is a foreign firm and is banned from the home market due to a trade policy, or when firm 3 is merged with other firms.

Figure 5 plots the different counterfactual predictions of submodels A_{ct}, A_1, and A_2 for the probability that only one firm enters the market, that is, the probability of the presence of a monopoly in a market with characteristics $x_0 = (0, 0, 0)$. As we can see clearly, three submodels give counterfactual predictions that are discordant with each other.

6.2 *An empirical illustration for compatible submodels*

6.2.1 *Context and data* Estimating the causal impact of college education on later earnings has always been troublesome for economists because of the endogeneity of the level of education. To evaluate the returns to schooling, different approaches have been proposed, and most of them rely on the validity of instruments such as parental education, tuition fees, quarter of birth, distance to college, etc. The validity of all these
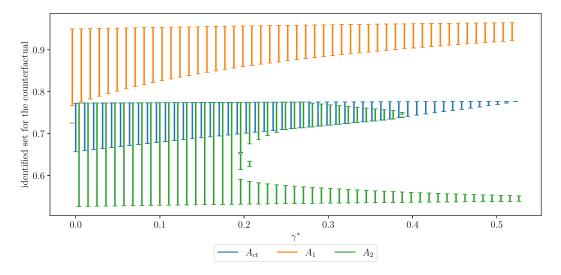
FIGURE 5. $\mathbb{P}(Y_1 + Y_2 = 1 | X = x_0)$ at various γ^*.

IVs has been widely criticized because of their potential correlation with the children's unobserved skills.

In order to accommodate potentially invalid instruments, Manski and Pepper (2000, 2009) introduced the monotone IV (MIV) that does not require the IV to be valid but only imposes a positive dependence relationship between the IV and potential earnings. For instance, parental education may not be independent of potential wages, but plausibly does not negatively affect future earnings. In such a context, bounds on the average return to education can be derived.

In this application, we will consider the AMIV assumption introduced in Section 4.1. We consider that parental education can have a positive effect on children's future earnings, but this marginal positive effect could plausibly become null after some cut-off. The particularity of our method is to let this cut-off be determined by the data using our misspecification robust bounds.

We consider the data used in Heckman, Tobias, and Vytlacil (2001, HTV). The data consist of a sample of 1230 white males taken from the National Longitudinal Survey of Youth of 1979 (NSLY79).[9] The data contain information on the log weekly wage, college education, father's education, mother's education, among many other variables. Following HTV, we consider the college enrollment indicator as the treatment: it is equal to 1 if the individual has completed at least 13 years of education and 0 otherwise. In this empirical exercise, we use the maximum of parental education as the candidate instrumental variable. Some summary statistics are reported in Table 1.

---

[9] The NLSY79 survey is sponsored and directed by the U.S. Bureau of Labor Statistics, and managed by the Center for Human Resource Research (CHRR) at The Ohio State University. Interviews are conducted by the National Opinion Research Center (NORC) at the University of Chicago. See Bureau of Labor Statistics, U.S. Department of Labor (2019) for more details.

Table 1. Summary statistics.

	Total
Observations	1230
log wage	2.4138 (0.5937)
college	0.4325 (0.4956)
father's education	12.44715 (3.2638)
mother's education	12.1781 (2.2781)
max(father's education, mother's education)	13.1699 (2.7123)

Note: Average and standard deviation (in the parentheses)

6.2.2 *Methodology and results*   We start by constructing the 95% confidence region for the identified sets of the average structural functions $E[Y_d]$, $d \in \{0, 1\}$ and the average treatment effect $E[Y_1 - Y_0]$ under the Manski (1990) mean independence assumption, denoted as $\Theta_I(\text{MI})$, and under the MIV assumption, denoted as $\Theta_I(\text{MIV})$. In addition, we construct an estimate of our misspecification robust bounds under the AMIV assumption, denoted as $\Theta_I^*(\text{AMIV})$, using the following steps:

1. The support of our instrument is $\mathcal{Z} = \{0, 1, \ldots, 20\}$. For each $z \in \{0, 1, \ldots, 20\}$, we test the implications (5) and (6) using the intersection bounds method of Chernozhukov, Lee, and Rosen (2013), implemented in Chernozhukov, Kim, Lee, and Rosen (2015). For each $d \in \{0, 1\}$, we set z_d^* as the smallest z for which we do not reject (5) and (6). The use of Chernozhukov et al.'s (2015) Stata package yields a 95% confidence set for $[\max(\underline{q}_{dt} : t \leq z), \min(\overline{q}_{dt} : t \geq z)]$ for each $z < z_d^*$, and $[\max(\underline{q}_{dt} : t = 1, \ldots, k), \min(\overline{q}_{dt} : t \geq z_d^*)]$.

2. We then plug the 95% confidence bounds obtained from step (1) into the bounds in Equation (7), where we replace $\mathbb{P}(Z = z)$ by its sample analog. This procedure leads to an estimate of the identified set Γ_{d,z^*} for each $d \in \{0, 1\}$, which yields an estimate for the identified set $\Theta_I^*(\text{AMIV})$.[10]

The same procedure is applied to get an estimate for $\Theta_I(\text{MIV})$ except that z_d^* is set to 20 for each $d \in \{0, 1\}$. Finally, since the identified set for the ATE under the mean independence assumption denoted as $\Theta_I(\text{MI})$ takes the form of standard intersection bounds, we use the Chernozhukov et al. (2015) package to obtain its 95% confidence bounds.

The results are summarized in Table 2. Column (1) shows that the 95% confidence region for $\Theta_I(\text{MI})$ is empty. In other words, the data shows clear evidence against the use of parental education as a valid IV. On the other hand, column (4) shows the result for $\Theta_I(\text{MIV})$.[11] As can be seen, we move from an empty identification region to a wide and noninformative identification region. In contrast, our misspecification robust bounds provide a nonempty yet relatively smaller set estimate for the ATE. Column (2) shows

---

[10]Because our primarily focus in this paper is about identification, we do not attempt to study the statistical issues related to the derivation of a valid confidence region for the misspecification robust bound. We leave this open question for future research.

[11]We even test the validity of the MIV using the test proposed by Hsu, Liu, and Shi (2019), and we do not reject the MIV assumption even at 10 % level.

TABLE 2. Results.

Set Estimates/ 95% Conf. Bounds	(1) Θ_I(MI)	(2) Θ_I^*(AMIV) $(z_1^*, z_0^*) = (0, 11)$	(3) Θ_I^*(AMIV) $(z_1^*, z_0^*) = (11, 11)$	(4) Θ_I(MIV)
$\theta_1 \equiv \mathbb{E}[Y_1]$	[2.535, 2.815]	[2.535, 2.815]	[2.412, 2.816]	[0.933, 2.815]
$\theta_0 \equiv \mathbb{E}[Y_0]$	Empty	[2.547, 2.591]	[2.547, 2.591]	[2.548, 2.814]
ATE $\equiv \mathbb{E}[Y_1 - Y_0]$	Empty	[−0.056, 0.268]	[−0.179, 0.269]	[−1.881, 0.267]

[1]All values in column (1) are the 95% confidence intervals.

[2]All values in columns (2)–(4) are set estimates based on the 95% confidence interval of $\Theta_I(a_z)$.

estimates of our misspecification robust bounds Θ_I^*(AMIV) when we allow the cut-offs to differ across potential outcomes as discussed in Remark 1, while column (3) shows estimates where the cut-offs are restricted to be the same for both potential outcomes as in Proposition 2. In the former case, we see that our proposed approach almost identifies the sign of the ATE.

7. Discussion

In this paper, we demonstrate the existence of discordant submodels in a wide range of models in the presence of model misspecification. This provides another reason why one should use the sharp characterization of the identified set whenever possible. The identified set not only exhausts all the identification restrictions in the model structure and assumptions but also is immune to the possible misleading conclusions of discordant submodels. Unlike an outer set, the identified set will be empty when the model is refuted by the data.

In empirical applications where a sharp characterization of the identified set is not tractable, our results suggest that empirical researchers should exercise caution when working with nonsharp identification conditions, especially when the bounds they obtain are very tight. For example, as a robustness check, one could construct the outer sets in different ways and check for any discordance between them.

Salvaging a refuted model is usually a challenging task, as it often involves some arbitrariness in how the model gets relaxed, and it could sometimes be computationally intractable. However, things get much easier when the minimum data-consistent relaxation is unique. In this case, it is apparent which assumptions are consistent with the data and which assumptions are not because all the data-consistent assumptions are compatible with each other (Theorems 2 and 4). Moreover, the identified set of any data-consistent submodel can be viewed as a conservative bound for the misspecification robust bound in this case, making the computation a lot easier.

When the uniqueness of the minimum data-consistent relaxation is beyond reach, one can still choose to find the misspecification robust bound we proposed in this paper. It always leads to rationalizable and nonconflicting statements (Theorem 5), and it is sometimes the most informative rationalizable and nonconflicting statement (Theorem 7). We work out the misspecification robust bound in some simple examples, but

its exact solution could be too complicated to solve when the underlying model involves many structures. In those challenging cases, it might be possible to construct an outer set that always covers the misspecification robust bound proposed in this paper. This type of outer sets will be immune to the issue raised in this paper. It remains unclear how to construct such outer sets, but this could be one reasonable step beyond the findings in this paper.

Appendix A: Additional results

This Appendix collects some additional theoretical results. We put all the proofs in Appendix C except for very short ones.

A.1 *Example of intersection bounds*

EXAMPLE 1 (Discrete Treatment Model). Consider a setting where $\mathcal{X} \equiv \{x_1, \ldots, x_K\}$ is the set of all possible treatments. Let Y_k be the potential outcome when the treatment is externally set to x_k. The observed outcome Y is defined as follows: $Y = \sum_k \mathbb{1}(X = x_k) Y_k$. Let us define $\theta_k \equiv E[Y_k]$ and assume that Y_k has a bounded support $[\underline{y}_k, \overline{y}_k]$. The random bound for Y_k can be constructed as follows: $\underline{Y}_k \equiv Y \mathbb{1}(X = x_k) + \underline{y}_k \mathbb{1}(X \neq x_k)$ and $\overline{Y}_k \equiv Y \mathbb{1}(X = x_k) + \overline{y}_k \mathbb{1}(X \neq x_k)$. If we assume the mean independence assumption $E[Y_k|Z] = E[Y_k]$, we obtain a special case of (1).

Discrete treatment models with bounded potential outcomes are usually considered in Manski's work. See, for instance, Manski (1990, 1994) among many others.

EXAMPLE 2 (Smooth Treatment Model). Consider a smooth treatment model as in Kim et al. (2018). When the treatment is x, the potential outcome is $Y(x) = g(x, \epsilon)$ where g is an unknown function, and ϵ is individual heterogeneous characterization. Assume $g(x, \epsilon)$ is Lipschitz continuous in x with Lipschitz constant equal to τ. Suppose we are interested in $\theta_x = E[Y(x)]$. The lower and upper bounds can be constructed as $\underline{Y}(x) = Y - \|X - x\|\tau$ and $\overline{Y}(x) = Y + \|X - x\|\tau$. As in the discrete treatment case, if we assume $E[Y(x)|Z] = E[Y(x)]$, we obtain model (1). As a special case, one can also consider a linear model with heterogeneous coefficient, $Y = X'\beta + \epsilon$ where β is a vector of an unobserved random coefficient. Suppose the coefficient space for β is $[\underline{\beta}, \overline{\beta}]$. Then $\underline{Y}(x) = Y + \sum_i \min\{(x_i - X_i)\underline{\beta}_i, (x_i - X_i)\overline{\beta}_i\}$ where the subscript i stands for the ith dimension of the corresponding variables. Similarly, $\overline{Y}(x) = Y + \sum_i \max\{(x_i - X_i)\underline{\beta}_i, (x_i - X_i)\overline{\beta}_i\}$.

A.2 *Conditional moment inequalities*

Let us now consider a more general setting than the introductory example. Assume the full model is a conditional moment inequality,

$$E[m(X; \theta)|Z] \leq 0 \quad \text{almost surely,} \tag{13}$$

where $X \in \mathbb{R}^{k_1}$ and $Z \in \mathbb{R}^{k_2}$ are observable random variables and $m(\cdot, \cdot; \theta)$ is some known integrable function with $E\|m(X; \theta)\| < \infty$ for each θ. We focus on the case where

Z are continuous random variables. Random variables X and Z could have overlaps. In practice, empirical researchers sometimes use the following unconditional model instead:

$$E\big[w(Z)m(X;\theta)\big]\le 0, \tag{14}$$

where $w(\cdot)$ is some nonnegative weighting function. We want to understand what would happen when one conduct empirical analysis based on (14) when (13) happens to be refuted.

To answer this question, define $\mathcal{W}_m^+$ to be the set of all m-dimenstional nonnegative function w, which satisfies $0 < E\|w(Z)\|^2 < \infty$ and $E\|w(Z)m(X;\theta)\| < \infty$ for all $\theta \in \Theta$. Define A as the collection of condition (14) for all $w \in \mathcal{W}_1^+$, that is,

$$A := \big\{(14)\ \text{with}\ w : w \in \mathcal{W}_1^+\big\}.$$

With this definition, any subset B of A with m elements corresponds to the condition, which (14) holds for some $w \in \mathcal{W}_m^+$. By the construction of A, Assumption 2 is satisfied.

To verify Assumption 3, we need to construct a $\mathscr{C}$. Let $\mathcal{Z}$ be the support of Z. For any $z \in \mathcal{Z}$ and any $\epsilon > 0$, define function $h_{z,\epsilon}$ as $h_{z,\epsilon}(Z) = \mathbb{1}(\|Z - z\| < \epsilon)$. Suppose that, for any $z \in \mathcal{Z}$, there exists some $\theta \in \Theta$ and some $\delta(z) > 0$ such that $E[m(X;\theta)|Z] \le 0$ for almost every Z with $\|Z - z\| \le \delta(z)$. Then, for each $z \in \mathcal{Z}$ and each $\epsilon \in (0, \delta(z))$, there exists some $\theta \in \Theta$ such that $E[h_{z,\epsilon}(Z)m(X,\theta)] \le 0$. Define the collection of functions $\mathcal{W}^*$ as $\mathcal{W}^* := \{h_{z,\epsilon} : \epsilon \in (0, \delta(z)), z \in \mathcal{Z}\}$. Then we can construct $\mathscr{C}$ as

$$\mathscr{C} := \big\{\{a\} : a \in A^*\big\}, \quad \text{where}\ A^* := \big\{(14)\ \text{with}\ w : w \in \mathcal{W}^*\big\}. \tag{15}$$

The following proposition shows that this $\mathscr{C}$ satisfies Assumption 3 under some regularity conditions.

PROPOSITION 4. *Assume that*

(a) *there exists a function $g(z;\theta)$ such that* (i) *for every $\theta \in \Theta$, $E[m(X, Z;\theta)|Z] = g(Z;\theta)$ almost surely;* (ii) *$g(z;\theta)$ is continuous in z for any given θ;* (iii) *$g(z;\theta)$ is continuous in θ for any given z.*

(b) *for any z in the support of Z, there exists some $\delta(z) > 0$ and some $\theta \in \Theta$ such that $E[m(X;\theta)|Z] \le 0$ for almost every Z satisfying $\|Z - z\| \le \delta(z)$.*

(c) *there exists some function $\gamma(\cdot)$ such that $\sup_{\theta \in \Theta} \|E[m(X;\theta)|Z]\| \le \gamma(Z)$ almost surely and $E|\gamma(Z)|^2 < \infty$.*

(d) *Θ is compact.*

Then Assumption 3 is satisfied for $\mathscr{C}$ constructed in (15).

As a result, Theorem 1 can be applied here. In the context of moment inequalities, the result in Theorem 1 means that (13) is refuted if and only if there exists $w_1 \in \mathcal{W}_{m_1}^+$ and $w_2 \in \mathcal{W}_{m_2}^+$ such that both (14) with $w = w_1$ and (14) with $w = w_2$ are not refuted but

the identified sets of these two sets of identification conditions have empty intersection. Moreover, whenever (13) is refuted, for any $\tilde{w} \in \mathcal{W}_{\tilde{m}}^+$ with which (14) is data-consistent, there exists some $w_1 \in \mathcal{W}_{m_1}^+$ and $w_2 \in \mathcal{W}_{m_2}^+$ such that both the (14) with $w = (\tilde{w}, w_1)$ and (14) with $w = w_2$ are data-consistent but their identified sets have empty intersection.

This result complements the findings in Andrews and Shi (2013). In Andrews and Shi (2013), they propose an inference procedure for models like (13). Their inference transform (13) into (14) by selecting w in a subfamily of $\mathcal{W}_m^+$ and letting $m \to \infty$ as the sample size increases. Our result shows that increasing m to infinity is crucial to ensure the robustness of the result if (13) could be misspecified. If the dimension of w is fixed, then the empirical result for (14) could be misleading even if the inference controls the size uniformly.

A.3 *Random sets and Choquet capacity*

In this section, we consider models whose identified set can be described with random sets and Choquet capacity functions. Let Y be a vector of endogenous random variables, and let X be a vector of exogenous observable covariates. Let $\mathcal{Y}$ and $\mathcal{X}$ denote the support of Y and X, respectively. Here, the parameter θ of interest could be of infinite dimensions, and its parameter space Θ need not be compact.

Let $\Gamma(\theta)$ be some random closed set in $\mathcal{Y}$, which could depend on θ, X, and some latent random variables. Assume $P(Y \in \Gamma(\theta)) = 1$. Artstein (1983) shows that the conditional distribution of Y given X equals $F_{Y|X}$ almost surely if and only if for any compact subset K of $\mathcal{Y}$, the following inequality holds:

$$P_F(Y \in K|X) \leq L(K, X; \theta) \quad \text{almost surely, where}$$
$$L(K, X; \theta) := P\big(\Gamma(\theta) \cap K \neq \emptyset | X\big), \tag{16}$$

where P_F refers to the probability measure corresponding to $F_{Y|X}$. The $L(\cdot, X; \theta)$ is often known as the *Choquet capacity function*. This type of models plays an important role in the partial identification literature. We refer to Molinari (2020) for more background introductions. Often in practice, either $P_F(Y \in K|X)$ or $L(K, X; \theta)$ can be identified from the data, and the other one can typically be derived or simulated from some additional assumptions. For the purpose of illustration, we consider the case where Y and X are observable so that $P_F(Y \in K|X)$ can be identified from the data, and assume $L(K, X; \theta)$ is a known function of K and X given θ.

In general, one needs to check (16) for all compact sets of $\mathcal{Y}$ in order to ensure this collection of moment inequalities is a sharp identification condition. In some circumstances, checking the inequalities for all compact sets is equivalent to checking the inequalities only for a subcollection of compact sets, in which case, this subcollection is called the *core determining class* in the language of Galichon and Henry (2011). However, in practice, researchers often pre-select some finite collection $\mathcal{K}$ of compact sets that are not core-determining, and they only check (16) for compact sets in this $\mathcal{K}$. For instance, in the treatment effect literature, the well-known Manski (1994) bounds on the potential outcome distributions implemented in various applications such as in Blundell et al. (2007), or Peterson (1976) bounds on competing risk, use only a finite and not

sufficient collection of Artstein inequalities. See, respectively, Molinari (2020) and Mourifié, Henry, and Méango (2020) for a detailed discussion. In empirical games, auction, and network applications we can also cite Ciliberto and Tamer (2009), Haile and Tamer (2003), Sheng (2020), Chesher and Rosen (2020), among many others who also focused on a finite and not sufficient collection of Artstein inequalities.[12]

We want to explore the consequences of this pre-selection procedure when the original model might be refuted by the data. For simplicity, we focus on the case where Y only takes a finite number of possible values. In this case, the support $\mathcal{Y}$ of Y is a finite set and the collection of all compact sets in $\mathcal{Y}$ is simply the power set of $\mathcal{Y}$, that is, the collection of all subsets of $\mathcal{Y}$.

To fit this model into the general framework in Section 3, define A as the collection of all Artsein's inequality, that is,

$$A := \big\{(16) \text{ with } K : K \subseteq \mathcal{Y}\big\}.$$

With this definition, any subset A' of A corresponds to testing (16) only for a pre-selected collection of compact sets. In order to apply Theorem 1, we need to verify Assumptions 2 and 3. By the construction of A, Assumption 2 is satisfied. To verify Assumption 3, let us construct $\mathscr{C}$ as

$$\mathscr{C} := \big\{\{a\} : a \in A\big\}. \tag{17}$$

The following proposition provides a sufficient condition under which Assumption 3 is satisfied with this choice of $\mathscr{C}$.

PROPOSITION 5. *Let $\mathcal{X}$ and $\mathcal{Y}$ be the supports of X and Y, respectively. Suppose $\mathcal{Y}$ is a finite set. The parameter space Θ may or may not be compact. Suppose that, for each $y \in \mathcal{Y}$, the following assumptions hold*:

  (L5.C1) $\inf_{x \in \mathcal{X}} P(Y = y | X = x) > 0,$

  (L5.C2) *there exists a sequence $\theta_1, \theta_2, \ldots$ in Θ such that $\inf_{x \in \mathcal{X}} L(\{y\}, x; \theta_k) \to 1$ as $k \to \infty$,*

where the inf *in the above two conditions refers to the essential infimum. Then Assumption 3 holds for $\mathscr{C}$ defined in* (17).

As a result, Theorem 1 could be applied here. For any pre-selected collection $\mathcal{K}$ of compact subsets, define $\Theta_I(\mathcal{K})$ as the set of parameters, which satisfy (16) for all $K \in \mathcal{K}$. In this context, the result in Theorem 1 means that the model is refuted if and only if there exists two $\mathcal{K}_1$ and $\mathcal{K}_2$ such that $\Theta_I(\mathcal{K}_1) \neq \emptyset$, $\Theta_I(\mathcal{K}_2) \neq \emptyset$ and $\Theta_I(\mathcal{K}_1) \cap \Theta_I(\mathcal{K}_2) = \emptyset$. Moreover, for any pre-selected collection $\mathcal{K}$ of compact sets with $\Theta_I(\mathcal{K}) \neq \emptyset$, there always exist two finite collections $\mathcal{K}_1$ and $\mathcal{K}_2$ such that $\Theta_I(\mathcal{K} \cup \mathcal{K}_1) \neq \emptyset$, $\Theta_I(\mathcal{K}_2) \neq \emptyset$, and $\Theta_I(\mathcal{K} \cup \mathcal{K}_1) \cap \Theta_I(\mathcal{K}_2) = \emptyset$.

We conclude this subsection with the entry game model as an example. We are going to verify all the conditions in Proposition 5 for this example.

---

[12]See Molinari (2020) for a detailed discussion.

EXAMPLE 3 (Entry Game). Consider an m-player complete information entry game as in Ciliberto and Tamer (2009). Assume there are m players, where player i's payoff function is specified as

$$\pi_i = Y_i \left(\gamma_i + X_i' \beta_i - \sum_{j \neq i} \delta_j^i Y_j + \epsilon_i \right),$$

where the X_is are some covariates, which might be player i specific, $Y_i \in \{0, 1\}$ stands for player i's entry decision, and Y_j stands for the decision of player j. Here, γ_i and β_i are player-specific parameter coefficient, and $\delta_j^i > 0$ is the parameter that describes the strategic interaction between player i and j. We assume that $Y = (Y_i : i = 1, \ldots, m)$ is always a pure-strategy Nash equilibrium.

Assume $\epsilon = (\epsilon_1, \ldots, \epsilon_m)$ is independent of X and ϵ follows the normal distribution $N(0, \Sigma)$. Let $\gamma = (\gamma_1, \ldots, \gamma_m)$, $\beta = (\beta_1, \ldots, \beta_m)$, and $\delta = (\delta_j^i : i \neq j)$. Let $\theta = (\gamma, \beta, \delta, \Sigma)$ be the vector of all parameters. Let $\mathcal{Y} = \{y = (y_1, \ldots, y_m) : y_i \in \{0, 1\}, i = 1, \ldots, m\}$ be the set of all possible entry decisions. For any $K \subseteq \mathcal{Y}$, define $L(K, X, \theta)$ to be the probability that at least one $y \in K$ is a pure-strategy Nash equilibrium given X and θ. In practice, $L(K, X, \theta)$ can often be solved from numerical simulations.

In Galichon and Henry (2011), the identified set of this model is shown to be the set of all θ, which satisfies (16) for every subset K of $\mathcal{Y}$. The number of these inequalities increases with m very quickly in the order of 2^{2^m}. Galichon and Henry (2011) provide some methods to reduce the number of inequalities by removing redundant inequalities in (16), but in general, sharp characterization of the identified set involves a large number of inequalities. In practice for the sake of computational feasibility, empirical researchers often pre-select a finite collection $\mathcal{K}$ of subsets and only check (16) for each $K \in \mathcal{K}$. See, for example, Ciliberto and Tamer (2009) and Ciliberto, Murry, and Tamer (2021).

Let us now check conditions in Proposition 5. Let $\Theta = \mathbb{R}^{\dim(\theta)}$. Condition (L5.C1) in Proposition 5 is a low-level condition that can be directly verified by the data. In theory, this condition would hold, for example, if the true data generating process has the following properties: (*i*) the support of ϵ is $\mathbb{R}^m$ conditional on almost every X, and (*ii*) for each i, player i's payoff function is

$$\pi_i = Y_i \big(g(X_i, Y_{-i}) + \epsilon_i \big),$$

where $Y_{-i} = \{Y_j : j \neq i\}$ and g can be an arbitrary function of (X_i, Y_{-i}) that is bounded in their support. This class of data generating processes nests the model that we imposed above, but the true data generating process need not be the same as our model. That is, Condition (L5.C1) in Proposition 5 holds even if the model is misspecified. Condition (L5.C2) in Proposition 5 also holds, because for each $y \in \mathcal{Y}$, one can have $L(\{y\}, x; \theta_k) \to 1$ by simply fixing $\beta = 0$, $\delta = 0$ and let $\gamma \to \gamma^*$ where $\gamma_i^* = \infty$ if $y_i = 1$ and $\gamma_i^* = -\infty$ if $y_i = 0$.

A.4 *Existence and uniqueness of minimum data-consistent relaxation*

THEOREM 3. *Suppose one of the following two conditions is satisfied*:

(T3.C1) *A is a finite set.*

(T3.C2) *For any $a \in A$, $\Theta_I(a)$ is compact. Moreover, for any $B \subseteq A$, $\Theta_I(B) = \bigcap_{a \in B} \Theta_I(a)$.*

Then there exists some minimum data-consistent relaxation of A. Moreover, for any data-consistent $A' \subseteq A$, there exists some minimum data-consistent relaxation $\widetilde{A}$ such that $A' \subseteq \widetilde{A}$.

Theorem 3 not only establishes the existence of a minimum data-consistent relaxation, but also shows that any data-consistent subset $A' \subseteq A$ can be further strengthened into a minimum data-consistent relaxation by including additional assumptions. It is worth noting that, when A is a finite set, the result of Theorem 3 does not require any additional conditions. When A is an infinite set, we need $\Theta_I(a)$ to be compact for each $a \in A$. In addition, we need that, for any $B \subseteq A$, $\Theta_I(B) = \bigcap_{a \in B} \Theta_I(a)$, which would hold, for example, if θ fully describes the distribution of both observed and latent random variables as in a maximum likelihood setting.

The following theorem studies the uniqueness of the minimum data-consistent relaxation, and is a generalized version of Theorem 2 in the main text.

THEOREM 4. *Statement (T2.C1) implies (T2.C2). If either (T3.C1) or (T3.C2) holds, then (T2.C2) implies (T2.C1).*

A.5 *Discordancy and multiplicity of minimum data-consistent relaxations*

PROPOSITION 6. *Whenever Condition (T3.C1) or (T3.C2) hold, we have the following result: There exists two data-consistent $A_1, A_2 \subseteq A$ with $\Theta_I(A_1) \cap \Theta_I(A_2) = \emptyset$ if and only if there exists two minimum data-consistent relaxations $\widetilde{A}_1$ and $\widetilde{A}_2$ such that $\Theta_I(\widetilde{A}_1) \cap \Theta_I(\widetilde{A}_2) = \emptyset$.*

PROOF FOR PROPOSITION 6. We first prove the **if** part. Suppose there exists two minimum data-consistent relaxations $\widetilde{A}_1$ and $\widetilde{A}_2$ such that $\Theta_I(\widetilde{A}_1) \cap \Theta_I(\widetilde{A}_2) = \emptyset$. By the definition of minimum data-consistent relaxation, both $\widetilde{A}_1$ and $\widetilde{A}_2$ are data-consistent subsets of A. Thus, this proves the existence of two data-consistent $A_1, A_2 \subseteq A$ with $\Theta_I(A_1) \cap \Theta_I(A_2) = \emptyset$.

Next, we want to prove the **only if** part. Suppose that there exists two data-consistent $A_1, A_2 \subseteq A$ with $\Theta_I(A_1) \cap \Theta_I(A_2) = \emptyset$. By Theorem 3, there exists two minimum data-consistent relaxations $\widetilde{A}_1$ and $\widetilde{A}_2$ such that $A_1 \subseteq \widetilde{A}_1$ and $A_2 \subseteq \widetilde{A}_2$. Because $\widetilde{A}_1 \subseteq A_1$, we have $\Theta_I(\widetilde{A}_1) \subseteq \Theta_I(A_1)$. Similarly, we have $\Theta_I(\widetilde{A}_2) \subseteq \Theta_I(A_2)$. Because $\Theta_I(A_1) \cap \Theta_I(A_2) = \emptyset$, we must have $\Theta_I(\widetilde{A}_1) \cap \Theta_I(\widetilde{A}_2) = \emptyset$. □

A.6 *Empirical interpretation of misspecification robust bound*

As discussed in Section 5.3, a rationalizable and nonconflicting set has rich interpretations. The following theorem shows that the misspecification robust bound Θ_I^* is both rationalizable and nonconflicting.

THEOREM 5. *Suppose either* $(T3.C1)$ *or* $(T3.C2)$ *holds. Then* Θ_I^* *is both rationalizable and nonconflicting. That is,*

- *(rationalizable) there exists some submodel* $A' \subseteq A$ *such that* $\Theta_I(A') \subseteq \Theta_I^*$ *and* $\Theta_I(A') \neq \emptyset$.

- *(nonconflicting) there does not exist a submodel* $A' \subseteq A$ *with* $\Theta_I(A') \neq \emptyset$ *such that* $\Theta_I(A') \cap \Theta_I^* = \emptyset$.

As discussed in Section 5.3, Θ_I^* have even richer explanations when it is the smallest rationalizable and nonconflicting set. Recall that a set S^* is the smallest rationalizable and nonconflicting set, if S^* is rationalizable and nonconflicting, and $S^* \subseteq S$ for every rationalizable and nonconflicting set S. The smallest rationalizable and nonconflicting set does not always exists. However, the following theorem shows that, under mild conditions, whenever the smallest rationalizable and nonconflicting set exists, it is equal to Θ_I^*.

THEOREM 6. *Suppose either* $(T3.C1)$ *or* $(T3.C2)$ *holds. Assume*

(T6.C1) *there does not exist two different minimum data-consistent relaxation* $\tilde{A}_1$ *and* $\tilde{A}_2$ *such that* $\Theta_I(\tilde{A}_1) \subsetneq \Theta_I(\tilde{A}_2)$.

Then, whenever the smallest rationalizable and nonconflicting set exists, it is equal to Θ_I^*.

Condition (T6.C1) could be verified from the data. Note that, for any two different minimum data-consistent relaxations $\tilde{A}_1$ and $\tilde{A}_2$, we always have $\Theta_I(\tilde{A}_1 \cup \tilde{A}_2) = \emptyset$, that is, $\tilde{A}_1$ and $\tilde{A}_2$ are not compatible with each other. As it is unlikely that for two sets $\tilde{A}_1$ and $\tilde{A}_2$ to satisfy $\Theta_I(\tilde{A}_1) \subsetneq \Theta_I(\tilde{A}_2)$ while being incompatible with each other, we consider (T6.C1) as a mild condition. Finally, Condition (T6.C1) would hold, if there is a unique minimum data-consistent relaxation, or if the identified set of every minimum data-consistent relaxation is a singleton set. In fact, these two conditions are also sufficient conditions for Θ_I^* being the smallest rationalizable and nonconflicting set, as shown in the following theorem.

THEOREM 7. *Suppose either* $(T3.C1)$ *or* $(T3.C2)$ *holds. Assume one of the following conditions is satisfied*:

(T7.C1) *there exists a unique minimum data-consistent relaxation,*

(T7.C2) *for any minimum data-consistent relaxation* $\widetilde{A}$, $\Theta_I(\widetilde{A})$ *is a singleton.*

Then Θ_I^* *is the smallest set that are both rationalizable and nonconflicting.*

A.7 *Discrete relaxation versus continuous relaxation*

THEOREM 8. *Suppose* (*T7.C2*) *holds. Then* $\Theta_I^* \subseteq \Theta_I^\dagger$ *for any type of relaxation chosen by the researcher.*

APPENDIX B: PROOF OF THE MAIN RESULTS

B.1 *Proof of Theorem 1*

First of all, note that Assumption 2 implies that for any $A', A'' \subseteq A$, $\Theta_I(A' \cup A'') = \Theta_I(A') \cap \Theta_I(A'')$.

If there exists $A', A'' \subseteq A$ such that $\Theta_I(A') \neq \emptyset$, $\Theta_I(A'') \neq \emptyset$, and $\Theta_I(A') \cap \Theta_I(A'') = \emptyset$, then $\Theta_I(A) \subseteq \Theta_I(A' \cup A'') = \Theta_I(A') \cap \Theta_I(A'') = \emptyset$. Hence, $\Theta_I(A) = \emptyset$ if there exists $A', A'' \subseteq A$ such that $\Theta_I(A') \neq \emptyset$, $\Theta_I(A'') \neq \emptyset$, and $\Theta_I(A') \cap \Theta_I(A'') = \emptyset$.

Reversely, if $\Theta_I(A) = \emptyset$, we want to show that there exists two finite subsets $A', A'' \subseteq A$ such that $\Theta_I(A') \neq \emptyset$, $\Theta_I(A'') \neq \emptyset$, and $\Theta_I(A') \cap \Theta_I(A'') = \emptyset$. More specifically, we are going to show the following statement:

$$\text{When } \Theta_I(A) = \emptyset, \text{ there exists } A' \in \mathscr{C} \text{ and } \{A_1, \ldots, A_n\} \subseteq \mathscr{C} \text{ for some finite } n$$

$$\text{such that both } A' \text{ and } A'' = \bigcup_{i=1}^n A_i \text{ are data-consistent, but } \Theta_I(A') \cap \Theta_I(A'') = \emptyset. \quad (18)$$

To show (18), we consider two cases.

Case 1: The $\mathscr{C}$ in Assumption 3 has infinite elements. In this case, $\Theta_I(A')$ is compact for all $A' \in \mathscr{C}$. Define $\mathscr{D} := \{B : \Theta_I(B) \neq \emptyset, \text{ and } \exists \mathscr{C}' \subseteq \mathscr{C}, B = \bigcup_{A' \in \mathscr{C}'} A'\}$. Because $\Theta_I(A') \neq \emptyset$ for all $A' \in \mathscr{C}$, $\mathscr{C} \subseteq \mathscr{D}$. Hence, $\mathscr{D}$ is nonempty. Moreover, because the intersection of compact sets is compact, we know $\Theta_I(A')$ is compact for any $A' \in \mathscr{D}$.

Note that $\subseteq$ can be viewed as a partial order for elements within $\mathscr{D}$. We are going to show $\mathscr{D}$ has a maximal element in terms of $\subseteq$, that is, there exists some $A' \in \mathscr{C}$ such that you cannot find an $A'' \in \mathscr{D}$ satisfying $A' \subseteq A''$ and $A' \neq A''$.

To show that $\mathscr{D}$ has a maximal element in terms of $\subseteq$, we are going to invoke the Zorn's lemma. Let $\mathscr{X}$ be an arbitrary nonempty chain in $\mathscr{D}$. That is, $\mathscr{X} \neq \emptyset$, $\mathscr{X} \subseteq \mathscr{D}$ and, for any $A', A'' \in \mathscr{X}$, either $A' \subseteq A''$ or $A'' \subseteq A'$. Define $A^\dagger := \bigcup_{A' \in \mathscr{X}} A'$. Then $\Theta_I(A^\dagger) = \bigcap_{A' \in \mathscr{X}} \Theta_I(A')$. Because $\mathscr{X}$ is a chain, $\{\Theta_I(A') : A' \in \mathscr{X}\}$ is also a chain in terms of $\subseteq$. Because $\mathscr{X} \subseteq \mathscr{D}$, $\Theta_I(A')$ is nonempty and compact for any $A' \in \mathscr{X}$. Hence, Lemma 1 (stated and proved below) implies that $\Theta_I(A^\dagger)$ is nonempty. As a result, $A^\dagger \in \mathscr{D}$. Moreover, for any $A' \in \mathscr{X}$, $A' \subseteq A^\dagger$. Thus, $\mathscr{D}$, as a partially ordered set in terms of $\subseteq$, has the following property: every nonempty chain $\mathscr{X}$ in $\mathscr{D}$ has an upper bound $A^\dagger$ in $\mathscr{D}$. By Zorn's lemma, this implies that $\mathscr{D}$ has a maximal element in terms of $\subseteq$.

Let A^* be a maximal element of $\mathscr{D}$ in terms of $\subseteq$. Because $A^* \in \mathscr{D}$, $\Theta_I(A^*) \neq \emptyset$. Because we have $\bigcap_{\tilde{A} \in \mathscr{C}} \Theta_I(\tilde{A}) = \Theta_I(\bigcap_{\tilde{A} \in \mathscr{C}} \tilde{A}) = \emptyset$ when $\Theta_I(A) = \emptyset$, $A^* \neq \bigcup_{\tilde{A} \in \mathscr{C}} \tilde{A}$. Therefore, there must exist some $A' \in \mathscr{C}$ such that A' is not a subset of A^*. Moreover, because $\mathscr{C} \subseteq \mathscr{D}$, and because A^* is a maximal element of $\mathscr{D}$ in terms of $\subseteq$, $A^* \cup A' \notin \mathscr{D}$. This implies that $\Theta_I(A^* \cup A') = \emptyset$. Because $\Theta_I(A^* \cup A') = \Theta_I(A^*) \cap \Theta_I(A')$, $\Theta_I(A^*) \cap \Theta_I(A') = \emptyset$. Because $A^* \in \mathscr{D}$, there exists $\mathscr{C}' \subseteq \mathscr{C}$ such that $A^* = \bigcup_{\tilde{A} \in \mathscr{C}'} \tilde{A}$, and hence,

$\Theta_I(A^*) = \bigcap_{\widetilde{A} \in \mathscr{C}'} \Theta_I(\widetilde{A})$. Because $\Theta_I(\widetilde{A})$ is compact for each $\widetilde{A} \in \mathscr{C}'$, Lemma 2 (shown and proved below) implies that there exists $\{A_1, \ldots, A_n\} \subseteq \mathscr{C}' \subseteq \mathscr{C}$ for some finite n such that $\Theta_I(\bigcup_{i=1}^n A_i) \cap \Theta_I(A') = \emptyset$ and $\Theta_I(\bigcup_{i=1}^n A_i) \neq \emptyset$. This proves (18).

Case 2: the $\mathscr{C}$ in Assumption 3 has finite elements. Enumerate $\mathscr{C}$ as $\mathscr{C} = \{A_1, \ldots, A_K\}$. For any $k \in \{1, \ldots, K\}$, define $B_k = \bigcup_{i=1}^k A_i$. For any k, $\Theta_I(B_k) = \bigcap_{i=1}^k \Theta_I(A_i)$. Define $\mathcal{N} = \{k \in \{1, \ldots, K\} : \Theta_I(B_k) \neq \emptyset\}$. Because $\Theta_I(A_1) \neq \emptyset$, $\mathcal{N}$ is nonempty and $1 \in \mathcal{N}$. Let n be the largest element in $\mathcal{N}$. By construction, $\Theta_I(B_n) \neq \emptyset$. By Assumption 3, $\Theta_I(B_K) = \Theta_I(\bigcup_{i=1}^K A_i) = \emptyset$. Therefore, we know $n < K$ and $\Theta_I(B_{n+1}) = \emptyset$. Because $B_{n+1} = B_n \cup A_{n+1}$, we know that $\Theta_I(B_n \cup A_{n+1}) = \Theta_I(B_n) \cap \Theta_I(A_{n+1}) = \emptyset$. This proves (18) with $A' = A_{n+1}$.

We have proven (18) in both above cases, which completes the proof for the first result of the theorem.

For the second part of the results, when $\Theta_I(A) = \emptyset$, for any $B \subseteq A$ with $\Theta_I(B) \neq \emptyset$, we want to show there exists two finite subsets $B', B'' \subseteq A$ such that $\Theta_I(B \cup B') \neq \emptyset$, $\Theta_I(B'') \neq \emptyset$ and $\Theta_I(B \cup B') \cap \Theta_I(B'') = \emptyset$. Let A' and A'' be the two finite subsets of A stated in (18). Consider the following two cases:

1. Suppose $\Theta_I(A') \cap \Theta_I(B) = \emptyset$. Then let $B' = \emptyset$, $B'' = A'$. We have that $\Theta_I(B \cup B') \neq \emptyset$, $\Theta_I(B'') \neq \emptyset$, and $\Theta_I(B \cup B') \cap \Theta_I(B'') = \emptyset$.

2. Suppose $\Theta_I(A') \cap \Theta_I(B) \neq \emptyset$. Then $\Theta_I(A' \cup B) = \Theta_I(A') \cap \Theta_I(B) \neq \emptyset$. Moreover, $\Theta_I(A' \cup B) \cap \Theta_I(A'') = \Theta_I(B) \cap (\Theta_I(A') \cap \Theta_I(A'')) = \emptyset$. Let $B' = A'$ and $B'' = A''$. Then we have that $\Theta_I(B \cup B') \neq \emptyset$, $\Theta_I(B'') \neq \emptyset$, and $\Theta_I(B \cup B') \cap \Theta_I(B'') = \emptyset$.

This completes the proof of Theorem 1.

LEMMA 1. *Let $\mathscr{B}$ be a collection of nonempty compact sets within metric space $\mathcal{T}$. Moreover, suppose $\mathscr{B}$ is a nonempty chain in terms of $\subseteq$, that is, for any $B, B' \in \mathscr{B}$, either $B \subseteq B'$ or $B' \subseteq B$. Then $\bigcap_{B \in \mathscr{B}} B$ is nonempty.*

PROOF. For the purpose of contradiction, suppose $\bigcap_{B \in \mathscr{B}} B$ is empty. For any $B \in \mathscr{B}$, let B^C denote the complement of B. Because the complement of $\bigcap_{B \in \mathscr{B}} B$ is $\bigcup_{B \in \mathscr{B}} B^C$, empty $\bigcap_{B \in \mathscr{B}} B$ implies that $\bigcup_{B \in \mathscr{B}} B^C = \mathcal{T}$. Pick an arbitrary $B' \in \mathscr{B}$. The fact that $\bigcup_{B \in \mathscr{B}} B^C = \mathcal{T}$ implies that $\{B^C : B \in \mathscr{B}\}$ is an open cover of B'. Because B' is compact, there exists a finite $\{B_1, \ldots, B_n\} \subseteq \mathscr{B}$ with $n < \infty$ such that $B' \subseteq \bigcup_{i=1}^n B_i^C$. This is equivalent to $B' \cap (\bigcap_{i=1}^n B_i) = \emptyset$. In other words, we can find $n + 1$ elements in $\mathscr{B}$ whose intersection is empty. This contradicts the assumption that $\mathscr{B}$ is a chain in terms of $\subseteq$ and that every set in $\mathscr{B}$ is nonempty. □

LEMMA 2. *Let $\mathscr{B}$ be a collection of nonempty compact sets within metric space $\mathcal{T}$ and $\bigcap_{B \in \mathscr{B}} B \neq \emptyset$. Let B' be a nonempty compact set in $\mathcal{T}$ such that $B' \cap (\bigcap_{B \in \mathscr{B}} B) = \emptyset$. Then there exists $\{B_1, \ldots, B_n\} \subseteq \mathscr{B}$ for some finite n such that $\bigcap_{i=1}^n B_i \neq \emptyset$ and $B' \cap (\bigcap_{i=1}^n B_i) = \emptyset$.*

PROOF. Because $B' \cap (\bigcap_{B \in \mathscr{B}} B) = \emptyset$, we know $\{B^C : B \in \mathscr{B}\}$ is an open cover of B'. Because B' is compact, there must exist a finite subcover $\{B_1^C, \ldots, B_n^C\} \subseteq \{B^C : B \in \mathscr{B}\}$ for B'. This implies that $B' \cap (\bigcap_{i=1}^n B_i) = \emptyset$. Finally, because $\bigcap_{i=1}^n B_i \supseteq (\bigcap_{B \in \mathscr{B}} B)$, $\bigcap_{i=1}^n B_i \neq \emptyset$. □

B.2 *Proof of Proposition* 1

Proposition 1 is an immediate result of the following two lemmas.

LEMMA 3. *Suppose Assumption* 1 *holds and* $\overline{\gamma} < \underline{\gamma}$. *Define the interval* $\mathcal{W}$ *as the following*:

$$
\mathcal{W} \equiv \begin{cases}
[\overline{\gamma}, \underline{\gamma}] & \text{if } P\big(E[\overline{Y}|Z] = \overline{\gamma}\big) > 0 \text{ and } P\big(E[\underline{Y}|Z] = \underline{\gamma}\big) > 0, \\
[\overline{\gamma}, \underline{\gamma}) & \text{if } P\big(E[\overline{Y}|Z] = \overline{\gamma}\big) > 0 \text{ and } P\big(E[\underline{Y}|Z] = \underline{\gamma}\big) = 0, \\
(\overline{\gamma}, \underline{\gamma}] & \text{if } P\big(E[\overline{Y}|Z] = \overline{\gamma}\big) = 0 \text{ and } P\big(E[\underline{Y}|Z] = \underline{\gamma}\big) > 0, \\
(\overline{\gamma}, \underline{\gamma}) & \text{if } P\big(E[\overline{Y}|Z] = \overline{\gamma}\big) = 0 \text{ and } P\big(E[\underline{Y}|Z] = \underline{\gamma}\big) = 0.
\end{cases}
\tag{19}
$$

For any integer m *and any* $h \in \mathcal{H}_m^+$, *if* $\widetilde{\Theta}(h)$ *is nonempty, then* $\widetilde{\Theta}(h) \cap \mathcal{W}$ *is nonempty.*

PROOF OF LEMMA 3. Since h has m dimensions, we can write $h = (h_1, \ldots, h_m)$. Then $\widetilde{\Theta}(h)$ can be characterized as $\widetilde{\Theta}(h) = [\underline{\theta}, \overline{\theta}]$, where

$$
\underline{\theta} = \max_i \frac{E\big[h_i(Z)\underline{Y}\big]}{E\big[h_i(Z)\big]} \quad \text{and} \quad \overline{\theta} = \min_i \frac{E\big[h_i(Z)\overline{Y}\big]}{E\big[h_i(Z)\big]}.
$$

Let us first prove $\underline{\theta} \leq \underline{\gamma}$. Suppose, for the purpose of contradiction, $\delta \equiv \underline{\theta} - \underline{\gamma} > 0$. Let $i' \in \arg\max_i E[h_i(Z)\underline{Y}]/E[h_i(Z)]$. Then we have

$$
E\big[h_{i'}(Z)\big(E[\underline{Y}|Z] - \underline{\theta}\big)\big] = 0.
\tag{20}
$$

Because $\delta \equiv \underline{\theta} - \underline{\gamma} > 0$, $E[\underline{Y}|Z] - \underline{\theta} \leq -\delta$. In addition, because $h_{i'}$ is nonnegative, we have $E[h_{i'}]\delta \leq 0$, which contradicts to the fact that $\delta > 0$ and $E[h_{i'}(Z)] > 0$. Moreover, if $P(E[\underline{Y}|Z] = \underline{\gamma}) = 0$, then $E[h_i(Z)\underline{Y}] < \underline{\gamma} \cdot E[h_i(Z)]$ for all i so that $\underline{\theta} < \underline{\gamma}$.

Similarly, we can show $\overline{\theta} \geq \overline{\gamma}$, and that $\overline{\theta} > \overline{\gamma}$ if $P(E[\overline{Y}|Z] = \overline{\gamma}) = 0$. This result then implies that $\widetilde{\Theta}(h) \cap \mathcal{W} \neq \emptyset$ whenever $\widetilde{\Theta}(h) \neq \emptyset$. □

LEMMA 4. *Suppose Assumption* 1 *holds and* $\overline{\gamma} < \underline{\gamma}$. *Let* $\mathcal{W}$ *be the interval defined as in* (19). *Then, for any* $\theta \in \mathcal{W}$, *there exists some* $h \in \mathcal{H}_2^+$ *such that* $\widetilde{\Theta}(h) = \{\theta\}$.

PROOF OF LEMMA 4. Fix any $\theta \in \mathcal{W}$. Define $\underline{S}^+ = \{z : E[\underline{Y}|Z = z] \geq \theta\}$, $\underline{S}^- = \{z : E[\underline{Y}|Z = z] \leq \theta\}$, $\overline{S}^+ = \{z : E[\overline{Y}|Z = z] \geq \theta\}$, and $\overline{S}^- = \{z : E[\overline{Y}|Z = z] \leq \theta\}$. Note that, for any $\vartheta > \overline{\gamma}$, the definition of $\overline{\gamma}$ implies that $P(\vartheta \geq E[\overline{Y}|Z]) > 0$. When $P(E[\overline{Y}|Z] = \overline{\gamma}) > 0$, we also have $P(\vartheta \geq E[\overline{Y}|Z]) > 0$ for any $\vartheta \geq \overline{\gamma}$. Since $\theta \in \mathcal{W}$, we conclude that $P(Z \in \overline{S}^-) > 0$. Similarly, $\theta \in \mathcal{W}$ also implies that $P(Z \in \underline{S}^+) > 0$. Moreover, since $E[\underline{Y}|Z] \leq E[\overline{Y}|Z]$ almost surely, we know $\underline{S}^+ \subseteq \overline{S}^+$ and $\overline{S}^- \subseteq \underline{S}^-$ almost surely. Therefore, $P(Z \in \underline{S}^-) > 0$ and $P(Z \in \overline{S}^+) > 0$.

Next, we show there exists some nonnegative function h_1, which satisfies $E[\underline{Y}h_1(Z)] = \theta$ and $E[h_1(Z)] = 1$. Define $h_1^+(z) = \mathbb{1}(z \in \underline{S}^+)/P(Z \in \underline{S}^+)$ and $h_1^-(z) = \mathbb{1}(z \in \underline{S}^-)/P(Z \in \underline{S}^-)$. By construction, h_1^+ and h_1^- are nonnegative, and $E[h_1^+(Z)] = 1$ and $E[h_1^-(Z)] = 1$. Moreover, $E[\underline{Y}h_1^+(Z)] \geq \theta \geq E[\underline{Y}h_1^-(Z)]$. Hence, there must exists some $q \in [0, 1]$ such that $E[\underline{Y}(qh_1^-(Z) + (1 - q)h_1^+(Z))] = \theta$. Let $h_1 = qh_1^-(Z) + (1 -$

$q)h_1^+(Z)$. Then such h_1 satisfies $E[\underline{Y}h_1(Z)] = \theta$ and $E[h_1(Z)] = 1$. Similarly, there exists some nonnegative function h_2, which satisfies $E[\overline{Y}h_2(Z)] = \theta$ and $E[h_2(Z)] = 1$.

Then $E[h_1(Z)(\tilde{\theta} - \underline{Y})] \geq 0$ is equivalent to $\tilde{\theta} \geq \theta$. To see this, note that

$$E\big[h_1(Z)(\tilde{\theta} - \underline{Y})\big] \geq 0$$
$$\Leftrightarrow \quad E\big[h_1(Z)\big]\tilde{\theta} \geq E\big[h_1(Z)\underline{Y}\big]$$
$$\Leftrightarrow \quad \tilde{\theta} \geq \theta,$$

where the second equivalence follows from $E[\underline{Y}h_1(Z)] = \theta$ and $E[h_1(Z)] = 1$. Similarly, we can show $E[h_2(Z)(\overline{Y} - \tilde{\theta})] \geq 0$ is equivalent to $\tilde{\theta} \leq \theta$. Let $h = (h_1, h_2)$. These equivalence relation implies that if $\tilde{\theta} \in \widetilde{\Theta}(h)$, then $\tilde{\theta} = \theta$.

Moreover, we have

$$E\big[h_2(Z)\theta\big] = \theta$$
$$= E\big[h_2(Z)\overline{Y}\big]$$
$$\geq E\big[h_2(Z)\underline{Y}\big],$$

where the first equality follows from $E[h_2(Z)] = 1$, and the second equality follows from $\theta = E[h_2(Z)\overline{Y}]$, and the last inequality comes from $E[\underline{Y}|Z] \leq E[\overline{Y}|Z]$ almost surely. Similarly, we can show $E[h_1(Z)\theta] \leq E[h_1(Z)\overline{Y}]$. Therefore, $\theta \in \widetilde{\Theta}(h)$. As a result, $\widetilde{\Theta}(h) = \{\theta\}$. □

B.3 *Proof of Theorem 2*

Theorem 2 is a corollary of Theorem 4, which is proved below in Section C.4.

B.4 *Proof of Proposition 2*

Recall the notation used in this example: $\underline{Y}_d \equiv Y\mathbb{1}(D = d) + \underline{y}_d\mathbb{1}(D \neq d)$, $\overline{Y}_d \equiv Y\mathbb{1}(D = d) + \overline{y}_d\mathbb{1}(D \neq d)$, $\underline{q}_{dt} \equiv E[\underline{Y}_d|Z = t]$, and $\overline{q}_{dt} \equiv E[\overline{Y}_d|Z = t]$. Proposition 2 is an immediate corollary of the following two lemmas.

LEMMA 5. *In model $Y = \sum_{z \in \mathcal{Z}} \mathbb{1}(Z = z)[Y_{1z}D + Y_{0z}(1 - D)]$ where $\mathcal{Z} = \{1, 2, \ldots, k\}$, fix an arbitrary $z^* = 1, \ldots, k$. Let Θ_{I,z^*} be the identified set of a_{z^*}, that is, the identified set of E.1, E.2, and E.3 for $z = z^*$. Then:*

1. *$\Theta_{I,z^*} \neq \emptyset$ if and only if the following two conditions hold for each $d \in \{0, 1\}$:*

$$\forall z < z^*, \quad \max(\underline{q}_{dt} : t \leq z) \leq \min(\overline{q}_{dt} : t \geq z) \tag{21}$$

　　　and

$$\max(\underline{q}_{dt} : t = 1, \ldots, k) \leq \min(\overline{q}_{dt} : t \geq z^*). \tag{22}$$

2. *if $\Theta_{I,z^*} \neq \emptyset$, then $\Theta_{I,z^*} = \Gamma_{1,z^*} \times \Gamma_{0,z^*}$.*

LEMMA 6. *In model* $Y = \sum_{z \in \mathcal{Z}} \mathbb{1}(Z = z)[Y_{1z}D + Y_{0z}(1 - D)]$ *where* $\mathcal{Z} = \{1, 2, \ldots, k\}$, *let* Θ_I *be the identified set of* $a^\dagger$, *that is, the identified set of* E.1 *and* E.2. *Then:*

1. $\Theta_I \neq \emptyset$ *if and only if* $P(Y \in [\underline{y}_d, \overline{y}_d] | D = d) = 1$ *for any* $d \in \{0, 1\}$.

2. *when* $\Theta_I \neq \emptyset$, $\Theta_I = [E[\underline{Y}_1], E[\overline{Y}_1]] \times [E[\underline{Y}_0], E[\overline{Y}_0]]$.

PROOF OF LEMMA 5. The results of this lemma can be divided into the following two parts:

1. For any $z^* = 1, \ldots, k$, $\Theta_{I,z^*} \neq \emptyset$ only if that (21) and (22) hold for each $d = 0, 1$. Moreover, $\Theta_{I,z^*} \subseteq \Gamma_{1,z^*} \times \Gamma_{0,z^*}$.

2. If (21) and (22) hold, then $\Theta_{I,z^*} \neq \emptyset$ and $\Theta_{I,z^*} \supseteq \Gamma_{1,z^*} \times \Gamma_{0,z^*}$.

Let us now prove these two parts one by one.

B.4.0.1 *Part 1*   Fix any $d \in \{0, 1\}$. Suppose assumption a_{z^*} holds, that is Assumptions *E.1*, *E.2*, and *E.3* hold for $z = z^*$. Assumption *E.3* implies that for any $z' < z^*$ and $t \leq z'$ we have $Y_{dt} \leq Y_{dz'}$ so that $E[Y_{dt} | Z = z'] \leq E[Y_{dz'} | Z = z']$. Due to *E.2*, we know $E[Y_{dt} | Z = z'] = E[Y_{dt} | Z = t]$, so that $E[Y_{dt} | Z = t] \leq E[Y_{dz'} | Z = z']$. Since $\underline{q}_{dt} \leq E[Y_{dt} | Z = t]$, we conclude that $\max_{t \leq z'} \underline{q}_{dt} \leq E[Y_{dz'} | Z = z']$. Similarly, *E.3* implies that for any $z' < z^*$ and $t \geq z'$, we have $Y_{dz'} \leq Y_{dt}$ so that $E[Y_{dz} | Z = z'] \leq E[Y_{dt} | Z = z']$. Because of *E.2*, and because $\overline{q}_{dt} \geq E[Y_{dt} | Z = t]$, we know that $E[Y_{dz'} | Z = z'] \leq \min_{t \geq z} \overline{q}_{dt}$. Hence, for any $d \in \{0, 1\}$,

$$\forall z' < z^*, \quad \max(\underline{q}_{dt} : t \leq z') \leq E[Y_{dz'} | Z = z'] \leq \min(\overline{q}_{dt} : t \geq z'). \tag{23}$$

Now, for any $z' \geq z^*$, *E.3* implies that $Y_{dt} \leq Y_{dz'}$ for any $t \in \{1, \ldots, k\}$. Hence, $E[Y_{dt} | Z = z'] \leq E[Y_{dz'} | Z = z']$ for all t. Because *E.2* implies that $E[Y_{dt} | Z = t] = E[Y_{dt} | Z = z']$, we have $E[Y_{dt} | Z = t] \leq E[Y_{dz'} | Z = z']$ for all t, so that $\max(\underline{q}_{dt} : t = 1, \ldots, k) \leq E[Y_{dz'} | Z = z']$. For any $z' \geq z^*$, Assumption *E.3* implies that $Y_{dt} \geq Y_{dz'}$ for all $t \geq z^*$. Hence, $E[Y_{dt} | Z = z] \geq E[Y_{dz} | Z = z]$ for all $t \geq z^*$. Assumption *E.2* then implies that $E[Y_{dt} | Z = t] \geq E[Y_{dz'} | Z = z']$ for all $t \geq z^*$, so that $\min(\overline{q}_{dt} : t \geq z^*) \geq E[Y_{dz} | Z = z]$. Hence, we conclude that for any $d \in \{0, 1\}$:

$$\forall z' \geq z^*, \quad \max(\underline{q}_{dt} : t = 1, \ldots, k) \leq E[Y_{dz'} | Z = z'] \leq \min(\overline{q}_{dt} : t \geq z^*). \tag{24}$$

Combine (23) and (24), we conclude that for any d, $\theta_d \in \Gamma_{d,z^*}$, so that $\Theta_{I,z^*} \subseteq \Gamma_{1,z^*} \times \Gamma_{0,z^*}$. Moreover, because Assumptions *E.1*, *E.2*, and *E.3* imply (23) and (24), the violation of (21) and (22) implies that $\Theta_{I,z^*} = \emptyset$. Equivalently, $\Theta_{I,z^*} \neq \emptyset$ only if (21) and (22) hold for any $d \in \{0, 1\}$.

B.4.0.2 *Part 2*   We want to prove that (21) and (22) implies that $\Theta_{I,z^*} \neq \emptyset$ and $\Theta_{I,z^*} \supseteq \Gamma_{1,z^*} \times \Gamma_{0,z^*}$. Fix an arbitrary $d \in \{0, 1\}$. First of all, we are going to prove that one can construct Y_{dz}, which achieves the lower bound in Γ_{d,z^*}, satisfies Assumptions *E.1–E.3*, and is compatible with the data at the same time.

Define γ_z for each $z = 1, \ldots, k$ as follows:

- for $z < z^*$, let γ_z be the value which solves

$$\max(\underline{q}_{dt} : t \leq z) = E\big[\mathbb{1}(D = d)Y|Z = z\big] + E\big[\mathbb{1}(D \neq d)Y|Z = z\big]\gamma_z.$$

  Then $\gamma_z \in [\underline{y}_d, \overline{y}_d]$ if $\overline{q}_{dz} \geq \max(\underline{q}_{dt} : t \leq z)$, which is implied by (21).

- for $z \geq z^*$, let γ_z be the value, which solves

$$\max(\underline{q}_{dt} : t = 1, \ldots, k) = E\big[\mathbb{1}(D = d)Y|Z = z\big] + E\big[\mathbb{1}(D \neq d)Y|Z = z\big]\gamma_z.$$

  Then $\gamma_z \in [\underline{y}_d, \overline{y}_d]$ if $\max(\underline{q}_{dt} : \forall t) \leq \overline{q}_{dz}$, which is implied by (22).

Define $W_{dz} \equiv \mathbb{1}(D = d, Z = z)Y + \mathbb{1}(D \neq d, Z = z)\gamma_z$. Then, by construction,

$$E[W_{dz}|Z = z] = \begin{cases} \max(\underline{q}_{dt} : t \leq z) & \text{if } z < z^*, \\ \max(\underline{q}_{dt} : t = 1, \ldots, k) & \text{if } z \geq z^*, \end{cases} \tag{25}$$

which implies that

$$\forall z \leq t, \quad E[W_{dz}|Z = z] \leq E[W_{dt}|Z = t], \tag{26}$$

$$\forall z \geq z^*, \quad E[W_{dz}|Z = z] = \max(\underline{q}_{dt} : t = 1, \ldots, k). \tag{27}$$

Moreover, because $\gamma_z \in [\underline{y}_d, \overline{y}_d]$ for any $z \in \{1, \ldots, k\}$, we know $P(W_{dz} \in [\underline{y}_z, \overline{y}_z]) = 1$ for all $z \in \{1, \ldots, k\}$. And, $P(W_{dz} = Y|D = d, Z = z) = 1$ for any d and z.

Now, for any $t \in \{1, \ldots, k\}$, define, $\phi_{dt}(\alpha) \equiv (1 - \alpha)W_{dt} + \alpha\overline{y}_d$ and $\psi_{dt}(\alpha) \equiv (1 - \alpha)\underline{y}_d + \alpha W_{dt}$. We claim that, for any $t \neq z$, there exists $\alpha_{tz} \in [0, 1]$, which solves the following equations:

$$\begin{aligned} \forall t < z, \quad & E[W_{dz}|Z = z] = E\big[\phi_{dt}(\alpha_{tz})|Z = t\big], \\ \forall t > z, \quad & E[W_{dz}|Z = z] = E\big[\psi_{dt}(\alpha_{tz})|Z = t\big]. \end{aligned} \tag{28}$$

To see why it is so, note that

$$\begin{aligned} \forall t < z, \quad & E[W_{dt}|Z = t] = E\big[\phi_{dt}(0)|Z = t\big] \quad \text{and} \quad E\big[\phi_{dt}(1)|Z = t\big] = \overline{y}_d, \\ \forall t > z, \quad & \underline{y}_d = E\big[\phi_{dt}(0)|Z = t\big] \quad \text{and} \quad E\big[\phi_{dt}(1)|Z = t\big] = E[W_{dt}|Z = t]. \end{aligned} \tag{29}$$

These results, combined with (26), imply that

$$\begin{aligned} \forall t < z, \quad & E\big[\phi_{dt}(0)|Z = t\big] \leq E[W_{dz}|Z = z] \leq E\big[\phi_{dt}(1)|Z = t\big], \\ \forall t > z, \quad & E\big[\psi_{dt}(0)|Z = t\big] \leq E[W_{dz}|Z = z] \leq E\big[\phi_{dt}(1)|Z = t\big], \end{aligned}$$

which implies the existence of $\alpha_{tz} \in [0, 1]$ satisfying (28) for all $t \neq z$.

In addition, $(\alpha_{tz} : t \neq z)$ has some extra properties. Because (26) holds and $E[\phi_{dt}(\alpha)|Z = t]$ is an increasing function of α,

$$\forall t < z < z', \quad \alpha_{tz} \leq \alpha_{tz'}. \tag{30}$$

Because (26) holds and $E[\psi_t(\alpha)|Z = t]$ is an increasing function of α,

$$\forall z < z' < t, \quad \alpha_{tz} \le \alpha_{tz'}. \tag{31}$$

Construct $Y_{dz} \equiv \sum_{t<z} \phi_{dt}(\alpha_{tz}) + W_{dz} + \sum_{t>z} \psi_{dt}(\alpha_{tz})$. Because $P(W_{dz} \in [\underline{y}_d, \overline{y}_d]) = 1$, Assumption *E.1* holds for this Y_{dz}. Because of (28), Assumption *E.2* holds for this Y_{dz}, that is, $E[Y_{dz}|Z = t] = E[Y_{dz}|Z = z]$ for any t, z with $t \ne z$.

To show Assumption *E.3* also holds for this Y_{dz}, note that, for any z_1, z_2 with $1 \le z_1 < z_2 \le k$:

- If $Z < z_1$, $Y_{dz_1} = \phi_{dZ}(\alpha_{Zz_1}) \le \phi_{dZ}(\alpha_{Zz_2}) = Y_{dz_2}$ because of (30) and because $\phi_{dZ}(\alpha)$ is increasing in α.

- If $Z = z_1$, $Y_{dz_1} = W_{dz_1} \le \phi_{dZ}(\alpha_{Zz_2}) = Y_{dz_2}$ because of the definition of $\phi_{dZ}(\alpha)$.

- If $z_1 < Z < z_2$, $Y_{dz_1} = \psi_{dZ}(\alpha_{Zz_1}) \le W_{dZ} \le \phi_{dZ}(\alpha_{Zz_2}) = Y_{dz_2}$ because of the definition of $\phi_{dZ}(\alpha)$ and $\psi_{dZ}(\alpha)$.

- If $Z = z_2$, $Y_{dz_1} = \psi_{dZ}(\alpha_{Zz_1}) \le W_{dZ} = Y_{dz_2}$ because of the definition of $\psi_{dZ}(\alpha)$.

- If $z_2 < Z$, $Y_{dz_1} = \psi_{dZ}(\alpha_{Zz_1}) \le \psi_{dZ}(\alpha_{Zz_2}) = Y_{dz_2}$ because of (31) and because $\psi_{dZ}(\alpha)$ is increasing in α.

As a result, $Y_{dz_1} \le Y_{dz_2}$ almost surely for any $z_1 \le z_2$. Moreover, because of (27), $\alpha_{tz} = \alpha_{tz'}$ for any t, z, and z' with $t < \min(z, z')$ and $z^* \le \min(z, z')$. Because of (27) and (29), $\alpha_{tz} = 0$ for any $z^* \le t < z$, and $\alpha_{tz} = 1$ for any $t > z \ge z^*$. Given these results, one can show that for any $z' \ge z^*$, $Y_{dz'} = \sum_{t<z^*} \phi_{dt}(\alpha_{tz^*}) + \sum_{t=z^*}^{k} W_{dt}$. This implies that Assumption *E.3* also holds. So far, we have shown that Y_{dz} constructed above satisfies assumption a_{z^*}.

Finally, because $E[Y_{dz}] = E[Y_{dz}|Z = z] = E[W_{dz}|Z = z]$ and because of (25), we know $\sum_z P(Z = z)E[Y_{dz}]$ achieves the lower bound in $\Gamma_{d,z}$. Moreover, because $P[Y_{dz} = Y|D = d, Z = z] = 1$, this construction of Y_{dz} is consistent with the data. Combine all the above results for an arbitrary $d \in \{0, 1\}$, we have constructed Y_{dz}, which satisfies assumption a_{z^*} and, at the same time, $\sum_z P(Z = z)E[Y_{dz}]$ achieves the lower bound of $\Gamma_{d,z}$.

Similarly, one can construct Y_{dz}, which satisfies assumption a_{z^*} and $\sum_z P(Z = z)E[Y_{dz}]$ achieves the upper bound of $\Gamma_{d,z}$, by defining γ'_z as follows:

- for $z < z^*$, let γ'_z be the value, which solves

$$\min(\overline{q}_{dt} : t \ge z) = E\big[\mathbb{1}(D = d)Y|Z = z\big] + E\big[\mathbb{1}(D \ne d)Y|Z = z\big]\gamma_z.$$

- for $z \ge z^*$, let γ'_z be the value, which solves

$$\min(\overline{q}_{dt} : t \ge z^*) = E\big[\mathbb{1}(D = d)Y|Z = z\big] + E\big[\mathbb{1}(D \ne d)Y|Z = z\big]\gamma_z.$$

Following the same steps as before except replacing γ_z with γ'_z, one can show that the constructed Y_{dz} satisfies *E.1–E.3* and $\sum_z P(Z = z)E[Y_{dz}]$ achieves the upper bound of $\Gamma_{d,z}$.

Taking convex combinations of the constructions, which achieve the upper and lower bound, every point in $\Gamma_{d,z}$ can be achieved under Assumptions *E.1–E.3*. This completes the proof. □

PROOF OF LEMMA 6. Suppose *E.1* and *E.2* hold. For any $z \in \{1, 2, \ldots, k\}$ and any $d \in \{0, 1\}$, we have

$$\mathbb{1}(Z = z, D = d)Y + \mathbb{1}(Z \neq z \text{ or } D \neq d)\underline{y}_d$$

$$\leq Y_{dz} \leq \mathbb{1}(Z = z, D = d)Y + \mathbb{1}(Z \neq z \text{ or } D \neq d)\overline{y}_d.$$

Therefore, $\underline{q}_{dz} \leq E[Y_{dz}|Z = z] \leq \overline{q}_{dz}$. Because of *E.2*, this implies that $\underline{q}_{dz} \leq E[Y_{dz}] \leq \overline{q}_{dz}$. As a result, $E[\underline{Y}_d] \leq \sum_z P(Z = z)EY_{dz} \leq E[\overline{Y}_d]$, which proves that $\Theta_I \subseteq [E[\underline{Y}_1], E[\overline{Y}_1]] \times [E[\underline{Y}_0], E[\overline{Y}_0]]$. Moreover, when $P(Y \in [\underline{y}_d, \overline{y}_d]|D = d) = 1$ for any $d \in \{0, 1\}$ fails to hold, *E.1* will fail to hold. Hence, $\Theta_I \neq \emptyset$ only if $P(Y \in [\underline{y}_d, \overline{y}_d]|D = d) = 1$ for any $d \in \{0, 1\}$.

Suppose that $P(Y \in [\underline{y}_d, \overline{y}_d]|D = d) = 1$ for any $d \in \{0, 1\}$ hold. Then we know that for each $z = 1, \ldots, k$ and each d, $\underline{y}_d \leq \underline{q}_{dz} \leq \overline{q}_{dz} \leq \overline{y}_d$. Construct Y_{dz} as the following for each z and d:

$$Y_{dz} = \mathbb{1}(Z = z, D = d)Y + \mathbb{1}(Z = z, D \neq d)\underline{y}_d + \mathbb{1}(Z \neq z)\underline{q}_{dz}.$$

By construction, $\theta_d = \sum_z P(Z = z)EY_{dz} = \sum_z P(Z = z)\underline{q}_{dz} = E[\underline{Y}_d]$. Moreover, one can check that this construction also satisfies Assumptions *E.1* and *E.2*. Similarly, for each d, we can construct Y'_{dz} as

$$Y'_{dz} = \mathbb{1}(Z = z, D = d)Y + \mathbb{1}(Z = z, D \neq d)\overline{y}_d + \mathbb{1}(Z \neq z)\overline{q}_{dz}.$$

Again, Y'_{dz} satisfies Assumptions *E.1* and *E.2* by construction. In addition, $\theta_d = \sum_z P(Z = z)EY'_{dz} = E[\overline{Y}_d]$. By considering (Y_{1z}, Y'_{0z}), (Y'_{1z}, Y_{0z}), (Y_{1z}, Y_{0z}), and (Y'_{1z}, Y'_{0z}), we conclude that Θ_I is nonempty and $\Theta_I = [E[\underline{Y}_1], E[\overline{Y}_1]] \times [E[\underline{Y}_0], E[\overline{Y}_0]]$. □

B.5 *Proof of Proposition* 3

Recall the A in the introductory example is the set of all conditions (3) indexed by $h \in \mathcal{H}_1^+$.

Let us first show Θ_I^* is equal the interval specified in (8). By Lemma 4, we know that for each $\theta \in (\overline{\gamma}, \gamma)$, there exists some $A' \subseteq A$ such that $\Theta_I(A') = \{\theta\}$. By Theorem 3, there exists some minimum data-consistent relaxation A^* such that $A' \subseteq A^*$. Since $\Theta_I(A')$ is singleton, we know $\Theta_I(A^*) = \Theta_I(A') = \{\theta\}$. Therefore, $(\overline{\gamma}, \gamma) \subseteq \Theta_I^*$.

We claim that if $P(E[\underline{Y}|Z] \leq \overline{\gamma}) > 0$, then $\overline{\gamma} \in \Theta_I^*$ and there exists some $A' \subseteq A$ with $\Theta_I(A') = \{\overline{\gamma}\}$. To see why it is so, suppose $P(E[\underline{Y}|Z] \leq \overline{\gamma}) > 0$. Then define $S_1 = \{z : E[\underline{Y}|Z = z] \leq \overline{\gamma}\}$ and $S_2 = \{z : E[\underline{Y}|Z = z] \geq \overline{\gamma}\}$. Since $P(E[\underline{Y}|Z] \leq \overline{\gamma}) > 0$, we know $P(Z \in S_1) > 0$. Since $\gamma > \overline{\gamma}$, we know $P(Z \in S_2) > 0$. Now, define $h_1(z) = \mathbb{1}(z \in S_1)/P(Z \in S_1)$ and $h_2(z) = \mathbb{1}(z \in S_2)/P(Z \in S_2)$. Then $Eh_1(Z) = 1$, $Eh_2(Z) = 1$, $E[h_1(Z)\underline{Y}] \leq \overline{\gamma}$, and $E[h_2(Z)\overline{Y}] \geq \overline{\gamma}$. Therefore, there must exist $\underline{h}$ as a convex combination of h_1 and h_2 such that $E\underline{h}(Z) = 1$ and $E[\underline{h}(Z)\underline{Y}] = \overline{\gamma}$. Hence, $\overline{\gamma} \in \widetilde{\Theta}(\underline{h})$ and $\widetilde{\Theta}(\underline{h}) \cap (-\infty, \overline{\gamma}) = \emptyset$. Moreover, for each $i = 1, 2, \ldots$, construct $\overline{h}_i(z)$ as $\overline{h}_i(z) = \mathbb{1}(E[\overline{Y}|Z = z] \in [\overline{\gamma}, \overline{\gamma} + 1/i))$. By the definition of $\overline{\gamma}$, we know $E\overline{h}_i(Z) > 0$ for each $i \geq 1$. Note that the identified set of (3) of $\overline{h}_i$, $\widetilde{\Theta}(\overline{h}_i)$ is

$$\left[\frac{E[\overline{h}_i(Z)\underline{Y}]}{E\overline{h}_i(Z)}, \frac{E[\overline{h}_i(Z)\overline{Y}]}{E\overline{h}_i(Z)}\right].$$

Because $E[\underline{Y}|Z] \leq E[\overline{Y}|Z]$ almost surely, the law of iterated expectation implies that $\overline{\gamma} \in \widetilde{\Theta}(\overline{h}_i)$. Moreover, by construction, for any $\theta > \overline{\gamma}$, $\theta \notin \bigcap_i \widetilde{\Theta}(\overline{h}_i)$. Therefore, if we define $\mathcal{H}' = \{\overline{h}_i : i \geq 1\} \cup \{\underline{h}\}$, we have $\bigcap_{h \in \mathcal{H}'} \widetilde{\Theta}(h) = \{\overline{\gamma}\}$. This implies that $\overline{\gamma} \in \Theta_I^*$ and there exists some $A' \subseteq A$ with $\Theta_I(A') = \{\overline{\gamma}\}$.

Next, we claim that if $P(E[\underline{Y}|Z] \leq \overline{\gamma}) > 0$, then $\Theta_I^* \cap (-\infty, \overline{\gamma}) = \emptyset$. To see this, note that Lemma 3 implies that for any $a \in A$, $\Theta_I(a) \cap [\overline{\gamma}, \underline{\gamma}] \neq \emptyset$. Let A' be an arbitrary minimum data-consistent relaxation A' of A. Because Assumption 2 holds in this example, the preceding result implies that for any minimum data-consistent relaxation A' of A, we know $\Theta_I(A') \cap [\overline{\gamma}, \underline{\gamma}] \neq \emptyset$. Our claim will be verified if we can prove $\Theta_I(A') \cap (-\infty, \overline{\gamma}) = \emptyset$. Suppose not, that is, suppose $\Theta_I(A') \cap (-\infty, \overline{\gamma}) \neq \emptyset$. Because $\Theta_I(A')$ is a closed interval, the fact that $\Theta_I(A') \cap [\overline{\gamma}, \underline{\gamma}] \neq \emptyset$ implies that $\overline{\gamma} \in \Theta_I(A')$. Because we have proven that $\bigcap_{h \in \mathcal{H}'} \widetilde{\Theta}(h) = \{\overline{\gamma}\}$, and because A' is a minimum data-consistent relaxation, we know $\Theta_I(A') = \{\overline{\gamma}\}$, which leads to contradiction.

Next, we claim that if $P(E[\underline{Y}|Z] \leq \overline{\gamma}) = 0$, then $\Theta_I^* \cap (-\infty, \overline{\gamma}] = \emptyset$. To see this, note that $P(E[\underline{Y}|Z] \leq \overline{\gamma}) = 0$ implies $P(E[\underline{Y}|Z] > \overline{\gamma}) = 1$. Therefore, for any $h \in \mathcal{H}_1^+$, $E[h(Z)(\theta - \underline{Y})] \geq 0$ implies that

$$E[h(Z)(\theta - \underline{Y})] \geq 0$$
$$\Rightarrow \quad E[h(Z)\underline{Y}] \leq E[h(Z)]\theta$$
$$\Leftrightarrow \quad E[h(Z)E[\underline{Y}|Z]] \leq E[h(Z)]\theta$$
$$\Rightarrow \quad E[h(Z)]\overline{\gamma} < E[h(Z)]\theta,$$

where the last inequality follows from the fact that $P(E[\underline{Y}|Z] > \overline{\gamma}) = 1$. Therefore, we know for any $h \in \mathcal{H}_1^+$, $\widetilde{\Theta}(h) \cap (-\infty, \overline{\gamma}] = \emptyset$. This implies $\Theta_I^* \cap (-\infty, \overline{\gamma}] = \emptyset$.

Following similar steps as above, we can also prove the following results:

- If $P(E[\overline{Y}|Z] \geq \underline{\gamma}) > 0$, then $\underline{\gamma} \in \Theta_I^*$ and there exists some $A' \subseteq A$ with $\Theta_I(A') = \{\underline{\gamma}\}$.

- If $P(E[\overline{Y}|Z] \geq \underline{\gamma}) > 0$, then $\Theta_I^* \cap (\underline{\gamma}, +\infty) = \emptyset$.

- If $P(E[\overline{Y}|Z] \geq \underline{\gamma}) = 0$, then $\Theta_I^* \cap [\underline{\gamma}, +\infty) = \emptyset$.

Combining these results and that $(\overline{\gamma}, \underline{\gamma}) \subseteq \Theta_I^*$, we conclude that Θ_I^* is equal to the interval specified in (8).

Appendix C: Proofs for additional results

C.1 *Proof for Proposition 4*

We need to verify that the $\mathscr{C}$ constructed in (15) satisfies all three requirements in Assumption 3.

First, we are going to show that $\forall A' \in \mathscr{C}$, A' is data-consistent and consists of finite elements in A. Fix an arbitrary $A' \in \mathscr{C}$. By the construction of $\mathscr{C}$, A' only contains one element in A: (14) holds for some $w = h_{z,\epsilon} \in \mathcal{W}^*$. By assumption, there exists some $\theta \in \Theta$ such that $E[m(X; \theta)|Z] \leq 0$ for all most every Z satisfying $\|Z - z\| \leq \delta(z)$. This implies that this θ must also satisfy (14) with $w = h_{z,\epsilon}$ for all $0 < \epsilon < \delta(z)$. This proves $\Theta_I(A') \neq \emptyset$.

Second, we need to prove that $\Theta_I(\bigcup_{A' \in \mathscr{C}} A') = \Theta_I(A)$. Because $\mathcal{W}^* \subseteq \mathcal{W}_1^+$, we know $\bigcup_{A' \in \mathscr{C}} A' \subseteq A$ so that $\Theta_I(\bigcup_{A' \in \mathscr{C}} A') \supseteq \Theta_I(A)$. Hence, we only need to show $\Theta_I(\bigcup_{A' \in \mathscr{C}} A') \subseteq \Theta_I(A)$. By assumption, there exists a function $g(z; \theta)$ such that (*i*) for every $\theta \in \Theta$, $E[m(X, Z; \theta)|Z] = g(Z; \theta)$ almost surely; (*ii*) $g(z; \theta)$ is continuous in z for any given θ; (*iii*) $g(z; \theta)$ is continuous in θ for any given z. Because $\theta \in \Theta_I(A)$ if and only if θ satisfy (13), we know that

$$\theta \notin \Theta_I(A) \quad \text{if and only if} \quad P(Z \in \mathcal{Z}_\theta) > 0 \quad \text{where } \mathcal{Z}_\theta := \big\{z \in \mathcal{Z} : g(z; \theta) > 0\big\}.$$

Fix an arbitrary $\theta \notin \Theta_I(A)$. Because $g(\cdot, \theta)$ is a continuous function of z given θ, $\mathcal{Z}_\theta$ is an open set of z. Therefore, there must exist some $z \in \mathcal{Z}$ and $\epsilon \in (0, \delta(z))$ such that $\{z' : \|z' - z\| < \epsilon\} \subseteq \mathcal{Z}_\theta$. As a result, the law of iterated expectation implies that

$$E\big[h_{z,\epsilon}(Z)g(Z; \theta)\big] = E\big[h_{z,\epsilon}(Z)m(X; \theta)\big] > 0.$$

This means that $\theta \notin \Theta_I(\bigcup_{A' \in \mathscr{C}} A')$. Thus, we have proven that $\theta \notin \Theta_I(\bigcup_{A' \in \mathscr{C}} A')$ if $\theta \notin \Theta_I(A)$, which is equivalent to $\Theta_I(\bigcup_{A' \in \mathscr{C}} A') \subseteq \Theta_I(A)$.

Finally, we need to show that $\Theta_I(A')$ is compact for any $A' \in \mathscr{C}$. Fix an arbitrary $A' \in \mathscr{C}$. By the construction of $\mathscr{C}$, A' only contains one element in A: (14) holds for some $w \in \mathcal{W}^*$. Define $\kappa(\theta) := E[w(Z)g(Z; \theta)]$. Because $g(z; \theta)$ is continuous in θ for any given z, and because, by assumption, $\sup_{\theta \in \Theta} \|g(z; \theta)\| = \sup_{\theta \in \Theta} \|E[m(X; \theta)|Z = z]\| \leq \gamma(z)$ and $E|\gamma(Z)| < \infty$, and because $w(\cdot)$ is a bounded function, the dominated convergence theorem implies that $\kappa(\theta)$ is a continuous at any $\theta \in \Theta$. The law of iterated expectations implies that $E[w(Z)m(X; \theta)] = E[w(Z)g(Z; \theta)] := \kappa(\theta)$. Therefore, $\Theta_I(A') = \{\theta \in \Theta : \kappa(\theta) \leq 0\}$. Because κ is continuous in θ, we know $\Theta_I(A')$ is a closed set. Because Θ is compact by assumption, $\Theta_I(A')$ is compact.

C.2 *Proof of Proposition* 5

We need to verify that $\mathscr{C}$ constructed in (17) satisfies all three requirements in Assumption 3.

First, we are going to show that $\forall A' \in \mathscr{C}$, A' is data-consistent and consists of finite elements in A. Fix an arbitrary $A' \in \mathscr{C}$. By the construct of $\mathscr{C}$ in (17), A' is a singleton set, which only contains one element in A: (16) holds for some $K \subseteq \mathcal{Y}$. If $K = \mathcal{Y}$, (16) holds for any $\theta \in \Theta$ because $L(K, X; \theta) = 1$ almost surely in this case. If $K = \emptyset$, (16) holds for any $\theta \in \Theta$ because $P_F(Y \in K|X) = 0$ almost surely in this case.

If $K \subsetneq \mathcal{Y}$ and $K \neq \emptyset$, pick an arbitrary $y' \in \mathcal{Y} \backslash K$. By (L5.C1), we know $\inf_{x \in \mathcal{X}} P(Y = y'|X = x) > 0$. Therefore, we know

$$\sup_{x \in \mathcal{X}} P_F(Y \in K|X = x) = 1 - \inf_{x \in \mathcal{X}} P_F(Y \notin K|X = x) \leq 1 - \inf_{x \in \mathcal{X}} P_F\big(Y = y'|X = x\big) < 1.$$

On the other hand, pick an arbitrary $y'' \in K$, (L5.C2) implies that there exists some sequence $\theta_k \in \Theta$ such that

$$\inf_{x \in \mathcal{X}} L\big(\{y''\}, x; \theta_k\big) \to 1 \quad \text{as } k \to \infty.$$

Therefore, there must exist some $\theta^* \in \Theta$ such that $\inf_{x \in \mathcal{X}} L(\{y''\}, x; \theta^*) \geq \sup_{x \in \mathcal{X}} P_F(Y \in K|X = x)$. This implies that

$$\sup_{x \in \mathcal{X}} P_F(Y \in K|X = x) \leq \inf_{x \in \mathcal{X}} L(\{y''\}, x; \theta^*) \leq \inf_{x \in \mathcal{X}} L(K, x; \theta^*)$$

the last equality holds because $y'' \in K$ implies that $L(\{y''\}, x; \theta^*) \leq L(K, x; \theta^*)$ for any x. The above inequality implies that $\theta^* \in \Theta_I(A')$, because

$$P_F(Y \in K|X) \leq \sup_{x \in \mathcal{X}} P_F(Y \in K|X = x) \leq \inf_{x \in \mathcal{X}} L(K, x; \theta^*) \leq L(K, X; \theta^*) \quad \text{almost surely.}$$

Hence, $\Theta_I(A')$ is nonempty. That is, we have shown that every $A' \in \mathscr{C}$ is data-consistent.

Second, we need to prove $\Theta_I(\bigcup_{A' \in \mathscr{C}} A') = \Theta_I(A)$. This is trivial, because $\bigcup_{A' \in \mathscr{C}} A' = A$ by the construction of $\mathscr{C}$ in (17).

Finally, because $\mathcal{Y}$ is a finite set, A is a finite set. By the construction of $\mathscr{C}$ in (17), $\mathscr{C}$ is also a finite set. This completes the proof.

C.3 *Proof for Theorem* 3

Let us start with two trivial cases: (*i*) suppose $\Theta_I(A') = \emptyset$ for any $A' \subseteq A$. Then $\emptyset$ is the minimum data-consistent relaxation in this case; (*ii*) suppose $\Theta_I(A) \neq \emptyset$. Then A is the minimum data-consistent relaxation. Next, let us consider the following nontrivial case.

Suppose $\Theta_I(A) = \emptyset$ and there exits some $A_0 \subseteq A$ such that $\Theta_I(A_0) \neq \emptyset$. We are going to show that there exists some minimum data-consistent relaxation $\widetilde{A}$ such that $A_0 \subseteq \widetilde{A}$. Let us consider two cases:

C.3.0.1 *Case 1: (T3.C1) holds*   Because $\Theta_I(A) = \emptyset$ while $\Theta_I(A') \neq \emptyset$, A_0 cannot be A so that $A \setminus A_0$ is nonempty. Because of (T3.C1), $A \setminus A_0$ is a finite set. Enumerate it as $A \setminus A_0 = \{a_1, \ldots, a_k\}$. Construct $A_1, \ldots, A_k$ iteratively as follows: For any $i = 1, \ldots, k$, define $A_i = A_{i-1} \cup \{a_i\}$ if $\Theta_I(A_{i-1} \cup \{a_i\}) \neq \emptyset$, and define $A_i = A_{i-1}$ if otherwise. By construction, for each $i = 1, \ldots, k$, $\Theta_I(A_i) \neq \emptyset$. Moreover, if $a_i \notin A_k$, we must have $\Theta_I(A_k \cup \{a_i\}) = \emptyset$ because $\Theta_I(A_i \cup \{a_i\}) = \emptyset$ and $\Theta_I(A_k \cup \{a_i\}) \subseteq \Theta_I(A_i \cup \{a_i\})$. Therefore, A_k must be a minimum data-consistent relaxation, and $A_0 \subseteq A_k$ by construction.

C.3.0.2 *Case 2: (T3.C2) holds*   Define $\mathscr{A} = \{A' : A' \subseteq A, A_0 \subseteq A' \text{ and } \Theta_I(A') \neq \emptyset\}$. $\mathscr{A}$ is not empty because $A_0 \in \mathscr{A}$. We are going to prove that there exists some minimum data-consistent relaxation $\widetilde{A}$ such that $A_0 \subseteq \widetilde{A}$, which is equivalent to show the following statement:

$$\mathscr{A} \text{ has a maximum element } \widetilde{A} \in \mathscr{A} \text{ in terms of partial order } \subseteq,$$

$$\text{that is, there is no } A' \in \mathscr{A} \text{ such that } \widetilde{A} \subsetneq A'. \tag{32}$$

To prove (32), we are going to invoke Zorn's lemma. Let $\mathscr{L}$ be an arbitrary nonempty chain in $\mathscr{A}$ in terms of $\subseteq$. Because $\mathscr{L}$ is a chain, for any A' and A'' in $\mathscr{L}$, there is either $A' \subseteq A''$ or $A'' \subseteq A'$, so that there is either $\Theta_I(A') \subseteq \Theta_I(A'')$ or $\Theta_I(A'') \subseteq \Theta_I(A')$. Define $A^\dagger = \bigcup_{A' \in \mathscr{L}} A'$. Because of (T3.C2), $\Theta_I(A^\dagger) = \bigcap_{A' \in \mathscr{L}} \Theta_I(A')$. By Lemma 1, $\Theta_I(A^\dagger)$ is

nonempty. This implies that $A^\dagger \in \mathscr{A}$. Moreover, by construction, $A' \subseteq A^\dagger$ for any $A' \in \mathscr{X}$. Therefore, we have shown that there exists an upper bound $A^\dagger$ in $\mathscr{A}$ in terms of partial order $\subseteq$ for any nonempty chain $\mathscr{X}$ in $\mathscr{A}$. Zorn's lemma then implies (32).

C.4 *Proof of Theorem* 4

We first prove the first part of the theorem.

C.4.0.1 *(T2.C1)* $\Rightarrow$ *(T2.C2)*   Construct $A^* = \{a \in A : \Theta_I(a) \neq \emptyset\}$. We are going to show that A^* is the only minimum data-consistent relaxation. For any $a \notin A^*$, we have $\Theta_I(a) = \emptyset$ by the construction of A^*. Hence, for any $a \notin A^*$, $\Theta_I(A^* \cup \{a\}) = \emptyset$ because $\Theta_I(A^* \cup \{a\}) \subseteq \Theta_I(a)$. Moreover, (T2.C1) implies that $\Theta_I(A^*) \neq \emptyset$ because every $a \in A^*$ is data-consistent. Therefore, A^* is a minimum data-consistent relaxation.

Suppose, for the purpose of contradiction, there exists another minimum data-consistent relaxation A' different from A^*. Because A' is a minimum data-consistent relaxation, there is no $A' \subseteq A^*$. Therefore, $A' \backslash A^*$ must be nonempty. Pick an arbitrary $a' \in A' \backslash A^*$. Because $a' \notin A^*$, we know $\Theta_I(a') = \emptyset$ by the construction of A^*. Therefore, $\Theta_I(A') = \emptyset$ because $\Theta_I(A') \subseteq \Theta_I(a')$. This contradicts to the fact that A' is a minimum data-consistent relaxation.

C.4.0.2 *If either (T3.C1) or (T3.C2) holds, (T2.C2)* $\Rightarrow$ *(T2.C1)*   Suppose either (T3.C1) or (T3.C2) hold. Let A^* denote the unique minimum data-consistent relaxation. First of all, we are going to prove the following statement:

$$A^* = \{a \in A : \Theta_I(a) \neq \emptyset\}. \tag{33}$$

Because A^* is a minimum data-consistent relaxation, $\Theta_I(A^*) \neq \emptyset$. Because $\Theta_I(A^*) \subseteq \Theta_I(a)$ for every $a \in A^*$, we know $\Theta_I(a)$ is nonempty for every $a \in A^*$. This implies that $A^* \subseteq \{a \in A : \Theta_I(a) \neq \emptyset\}$. To show, $A^* \supseteq \{a \in A : \Theta_I(a) \neq \emptyset\}$, note that

- when $A^* = A$, we have $A^* \supseteq \{a \in A : \Theta_I(a) \neq \emptyset\}$ trivially.

- when $A^* \subsetneq A$. Pick an arbitrary $a' \notin A^*$. Suppose, for the purpose of contradiction, $\Theta_I(a') \neq \emptyset$. Then Theorem 3 implies that there exists some minimum data-consistent relaxation $\widetilde{A}$ with $a' \in \widetilde{A}$. Because $a' \in \widetilde{A}$ and $a' \notin A^*$, we must have $\widetilde{A} \neq A^*$, which contradicts to A^* being the unique minimum data-consistent relaxation. Hence, $\Theta_I(a') = \emptyset$ for any $a' \notin A^*$, which is equivalent to $A^* \supseteq \{a \in A : \Theta_I(a) \neq \emptyset\}$.

This proves (33).

Because of (33), $a \in A$ is data-consistent if and only if $a \in A^*$. Therefore, for any $A' \subseteq A$, all $a \in A'$ are data-consistent if and only if $A' \subseteq A^*$. As a result, we can show (T2.C1) if the following statement is true:

$$A' \subseteq A^* \quad \text{if and only if} \quad \Theta_I(A') \neq \emptyset. \tag{34}$$

To see why (34) is indeed true, note that:

- Because $\Theta_I(A^*) \neq \emptyset$, and because $\Theta_I(A^*) \subseteq \Theta_I(A')$ for any $A' \subseteq A^*$, we know $\Theta_I(A') \neq \emptyset$ if $A' \subseteq A^*$.

- Fix an arbitrary set A' with $\Theta_I(A') \neq \emptyset$. Because $\Theta_I(A') \subseteq \Theta_I(a)$ for every $a \in A'$, we know $\Theta_I(a) \neq \emptyset$ for every $a \in A'$. Because of (33), this implies that $a \in A^*$ for every $a \in A'$, that is, $A' \subseteq A^*$. Thus, we have shown that $A' \subseteq A^*$ if $\Theta_I(A') \neq \emptyset$.

This completes the proof.

C.5 *Proof of Theorem 5*

Recall that $\mathscr{A}_R$ denote the collection of all minimum data-consistent relaxations. Because either (T3.C1) or (T3.C2) holds, $\mathscr{A}_R$ is nonempty.

First of all, we are going to prove Θ_I^* is rationalizable. Because $\Theta_I^* = \bigcup_{A' \in \mathscr{A}_R} \Theta_I(A')$, for any minimum data-consistent relaxation $\widetilde{A}$, we must have $\Theta_I(\widetilde{A}) \subseteq \Theta_I^*$. Because $\mathscr{A}_R$ is nonempty, we know Θ_I^* is rationalizable.

Second, we are going to prove Θ_I^* is nonconflicting. Fix an arbitrary data-consistent subset A' of A. By Theorem 3, there exists a minimum data-consistent relaxation $\widetilde{A}$ such that $A' \subseteq \widetilde{A}$. Because $A' \subseteq \widetilde{A}$, we know $\Theta_I(\widetilde{A}) \subseteq \Theta_I(A')$, so that $\Theta_I(A') \cap \Theta_I(\widetilde{A}) \neq \emptyset$. Because $\widetilde{A} \in \mathscr{A}_R$ and $\Theta_I^* = \bigcup_{\widetilde{A}' \in \mathscr{A}_R} \Theta_I(\widetilde{A}')$, we know $\Theta_I(A') \cap \Theta_I^* \neq \emptyset$.

C.6 *Proof for Theorem 6*

Suppose the smallest rationalizable and nonconflicting set exists. Denote it as S^*. We are going to show that $S^* = \Theta_I^*$ when (T6.C1) is true. By Theorem 5, Θ_I^* is both rationalizable and nonconflicting. Therefore, $S^* \subseteq \Theta_I^*$. What is left to show is $\Theta_I^* \subseteq S^*$.

Define $\mathscr{A}_R$ to be the collection of all minimum data-consistent relaxations. Define $\mathscr{A}_1 = \{A' \in \mathscr{A}_R : \Theta_I(A') \subseteq S^*\}$ and $\mathscr{A}_2 = \mathscr{A}_R \backslash \mathscr{A}_1$. Because S^* is rationalizable, there exists some data-consistent $A' \subseteq A$ such that $\Theta_I(A') \subseteq S^*$. By Theorem 3, there exists some $\tilde{A} \in \mathscr{A}_R$ such that $A' \subseteq \tilde{A}$. Therefore, $\Theta_I(\tilde{A}) \subseteq \Theta_I(A') \subseteq S^*$. Hence, $\mathscr{A}_1$ is not empty.

Because $\mathscr{A}_R = \mathscr{A}_1 \cup \mathscr{A}_2$, we know

$$\Theta_I^* = \left(\bigcup_{A' \in \mathscr{A}_1} \Theta_I(A') \right) \cup \left(\bigcup_{A' \in \mathscr{A}_2} \Theta_I(A') \right).$$

Therefore, to show $\Theta_I^* \subseteq S^*$, we only need to show $\mathscr{A}_2$ is an empty set. We discuss two cases:

- Suppose there exists some $A^* \in \mathscr{A}_1$ such that $\Theta_I(A^*)$ contains at least two different elements. We are going to show that $\mathscr{A}_2 = \emptyset$ in this case. Suppose, for the purpose of contradiction, $\mathscr{A}_2$ is nonempty. Pick an arbitrary $A^\dagger$ within $\mathscr{A}_2$. Fix also an arbitrary element a^* in $\Theta_I(A^*)$. Define S' as follows:

$$S' = \begin{cases} \emptyset & \text{if } \Theta_I(A^*) \cap \Theta_I(A^\dagger) \neq \emptyset, \\ \{a^*\} & \text{if } \Theta_I(A^*) \cap \Theta_I(A^\dagger) = \emptyset. \end{cases}$$

Define $S^\dagger$ as

$$S^\dagger := \Theta_I(A^\dagger) \cup \left(\bigcup_{A' \in \mathscr{A}_R} \Theta_I(A') \backslash \Theta_I(A^*) \right) \cup S'.$$

Because $\Theta_I(A^\dagger) \subseteq S^\dagger$, $S^\dagger$ is rationalizable. Because of (T6.C1), $\Theta_I(A') \backslash \Theta_I(A^*) \neq \emptyset$ for each $A' \in \mathscr{A}_R$ with $\Theta_I(A') \neq \Theta_I(A^*)$; otherwise, we would have $\Theta_I(A') \subsetneq \Theta_I(A^*)$ for some $A' \in \mathscr{A}_R$, which violates (T6.C1). As a result, $S^\dagger \cap \Theta_I(A') \neq \emptyset$ for any $A' \in \mathscr{A}_R$ with $\Theta_I(A') \neq \Theta_I(A^*)$. Next, because $\Theta_I(A^\dagger) \cup S'$ and $\Theta_I(A^*)$ has a nonempty intersection by the construction of S', we know that $S^\dagger \cap \Theta_I(A') \neq \emptyset$ for any $A' \in \mathscr{A}_R$ with $\Theta_I(A') = \Theta_I(A^*)$. In total, we know $S^\dagger \cap \Theta_I(A') \neq \emptyset$ for any $A' \in \mathscr{A}_R$. Because of this result, and because by Theorem 3 for any $A \subseteq A$, there exists some $A' \in \mathscr{A}_R$ such that $\Theta_I(A') \subseteq \Theta_I(A'')$; we know $S^\dagger$ is nonconflicting. So far, we have shown that $S^\dagger$ is both rationalizable and nonconflicting.

Then we claim that there is no $\Theta_I(A^*) \subseteq S^\dagger$. By the construction of $S^\dagger$, $S^\dagger \cap \Theta_I(A^*) = (\Theta_I(A^\dagger) \cup S') \cap \Theta_I(A^*)$. So, this claim is also equivalent to that there is no $\Theta_I(A^*) \subseteq \Theta_I(A^\dagger) \cup S'$. To see why this claim is true, we discuss two cases:

– when $\Theta_I(A^*) \cap \Theta_I(A^\dagger) \neq \emptyset$, $S' = \emptyset$. Because of (T6.C1), there is no $\Theta_I(A^*) \subsetneq \Theta_I(A^\dagger)$. Because $A^* \in \mathscr{A}_1$ and $A^\dagger \in \mathscr{A}_2$, there is no $\Theta_I(A^*) = \Theta_I(A^\dagger)$. In total, there is no $\Theta_I(A^*) \subseteq \Theta_I(A^\dagger)$. Because $S' = \emptyset$, we know there is no $\Theta_I(A^*) \subseteq \Theta_I(A^\dagger) \cup S'$ is this case.

– when $\Theta_I(A^*) \cap \Theta_I(A^\dagger) = \emptyset$, $(\Theta_I(A^\dagger) \cup S') \cap \Theta_I(A^*) = S' \cap \Theta_I(A^*) = \{a^*\}$. Because $\Theta_I(A^*)$ contains at least two different elements, there is no $\Theta_I(A^*) \subseteq S'$, which implies there is no $\Theta_I(A^*) \subseteq \Theta_I(A^\dagger) \cup S'$ is this case.

In total, we have to verify the claim that there is no $\Theta_I(A^*) \subseteq S^\dagger$. However, because S^* is the smallest rationalizable and nonconflicting set, and because $S^\dagger$ is both rationalizable and nonconflicting, there must be $S^* \subseteq S^\dagger$, which further implies $\Theta_I(A^*) \subseteq S^\dagger$ because $A^* \in \mathscr{A}_1$. This leads to the contradiction. As a result, $\mathscr{A}_2$ must be empty in this case.

• Suppose that, for any $A' \in \mathscr{A}_1$, $\Theta_I(A')$ only contains one element. We are going to show that $\mathscr{A}_2 = \emptyset$ in this case. Suppose, for the purpose of contradiction, that $\mathscr{A}_2$ is nonempty. By the construction of $\mathscr{A}_1$ and $\mathscr{A}_2$, there is no $\Theta_I(A_1) = \Theta_I(A_2)$ for any $A_1 \in \mathscr{A}_1$ and $A_2 \in \mathscr{A}_2$. Because of (T6.C1), there is no $\Theta_I(A_1) \subsetneq \Theta_I(A_2)$ for any $A_1 \in \mathscr{A}_1$ and $A_2 \in \mathscr{A}_2$. In total, there is no $\Theta_I(A_1) \subseteq \Theta_I(A_2)$ for any $A_1 \in \mathscr{A}_1$ and $A_2 \in \mathscr{A}_2$. Because $\Theta_I(A')$ only contains one element for any $A' \in \mathscr{A}_1$, for any $S \subseteq \Theta$, $\Theta_I(A') \cap S \neq \emptyset$ would mean $\Theta_I(A') \subseteq S$. Therefore, we must have $\Theta_I(A') \cap \Theta_I(A'') = \emptyset$ for any $A' \in \mathscr{A}_1$ and $A'' \in \mathscr{A}_2$. Define S_1 as

$$S_1 = \bigcup_{A' \in \mathscr{A}_1} \Theta_I(A').$$

And, define S_2 as

$$S_2 = \left(\bigcup_{A' \in \mathscr{A}_2} \Theta_I(A') \right) \cap S^*.$$

Because $\Theta_I(A') \cap \Theta_I(A'') = \emptyset$ for any $A' \in \mathscr{A}_1$ and $A'' \in \mathscr{A}_2$, $S_1 \cap S_2 = \emptyset$. Moreover, because S^* is nonconflicting, $S^* \cap \Theta_I(A') \neq \emptyset$ for each $A' \in \mathscr{A}_2$. Therefore, $S_2 \cap \Theta_I(A') \neq \emptyset$ for any $A' \in \mathscr{A}_2$. This implies that $S_1 \cup S_2$ is nonconflicting. Moreover, because $\mathscr{A}_1$ is nonempty, S_1 is rationalizable so that $S_1 \cup S_2$ is also rationalizable. As a result, $S_1 \cup S_2$ is both nonconflicting and rationalizable. Next, define

$$S_3 = \left(\bigcup_{A' \in \mathscr{A}_2} \Theta_I(A') \right) \backslash S^*.$$

By the definition of $\mathscr{A}_1$, $S_1 \subseteq S^*$. Therefore, $S_1 \cap S_3 = \emptyset$. Also, we have $S_2 \cap S_3 = \emptyset$ by the construction of S_2 and S_3. In addition, by the construction of $\mathscr{A}_2$, for each $A' \in \mathscr{A}_2$, $\Theta_I(A') \backslash S^* \neq \emptyset$. Therefore, $S_3 \cap \Theta_I(A') \neq \emptyset$ for each $A' \in \mathscr{A}_2$. This implies that $S_1 \cup S_3$ is nonconflicting. Moreover, because $\mathscr{A}_1$ is nonempty, S_1 is rationalizable so that $S_1 \cup S_3$ is also rationalizable. As a result, $S_1 \cup S_3$ is both nonconflicting and rationalizable.

So far, we have shown that $S_1 \cup S_2$ is both rationalizable and nonconflicting. And, we have shown that $S_1 \cup S_3$ is both rationalizable and nonconflicting. Because S^* is the smallest rationalizable and nonconflicting set, we must have $S^* \subseteq S_1 \cup S_2$ and $S^* \subseteq S_1 \cup S_3$. In other words,

$$S^* \subseteq (S_1 \cup S_2) \cap (S_1 \cup S_3).$$

However, because $S_1 \cap S_2 = \emptyset$, $S_1 \cap S_3 = \emptyset$, and $S_2 \cap S_3 = \emptyset$, we know $(S_1 \cup S_2) \cap (S_1 \cup S_3) = S_1$. As a result, we have

$$S^* \subseteq S_1. \tag{35}$$

We have already shown that we must have $\Theta_I(A') \cap \Theta_I(A'') = \emptyset$ for any $A' \in \mathscr{A}_1$ and $A'' \in \mathscr{A}_2$. Therefore, $S_1 \cap \Theta_I(A'') = \emptyset$ for any $A'' \in \mathscr{A}_2$. Because $\mathscr{A}_2$ is nonempty, this means that S_1 is not nonconflicting. Because of (35), this implies that S^* is not nonconflicting, which contradicts to the fact that S^* is both rationalizable and nonconflicting.

We have shown that $\mathscr{A}_2$ must be empty in both of the above cases. This completes the proof.

C.7 *Proof for Theorem 7*

By Theorem 5, Θ_I^* is both rationalizable and nonconflicting. If we could show $\Theta_I^* \subseteq S$ for an arbitrary set S that is both rationalizable and nonconflicting, then Θ_I^* would be the smallest rationalizable and nonconflicting set. Fix an arbitrary set S that is both rationalizable and nonconflicting.

We first prove that (T7.C1) implies $\Theta_I^* \subseteq S$. Because S is rationalizable, there exists some data-consistent $A' \subseteq A$ such that $\Theta_I(A') \subseteq S$. By Theorem 3, there exists a minimum data-consistent relaxation $\tilde{A}$ such that $A' \subseteq \tilde{A}$. Because $A' \subseteq \tilde{A}$, we know $\Theta_I(\tilde{A}) \subseteq \Theta_I(A')$. Because of (T7.C1), $\Theta_I^* = \Theta_I(\tilde{A})$. Therefore, $\Theta_I^* = \Theta_I(\tilde{A}) \subseteq \Theta_I(A') \subseteq S$.

Next, we prove that (T7.C2) implies $\Theta_I^* \subseteq S$. Let $\mathscr{A}_R$ denote the collection of all minimum data-consistent relaxations. Because S is nonconflicting, we must have $\Theta_I(A') \cap S \neq \emptyset$ for each $A' \in \mathscr{A}_R$. For each $A' \in \mathscr{A}_R$, because $\Theta_I(A')$ only contains one element under (T7.C2), $\Theta_I(A') \cap S \neq \emptyset$ is equivalent to $\Theta_I(A') \subseteq S$. As a result, we know $\Theta_I^* \subseteq S$ because $\Theta_I^* = \bigcup_{A' \in \mathscr{A}_R} \Theta_I(A')$.

C.8 *Proof for Theorem 8*

Let $\mathscr{A}_R$ be the set of all minimum data-consistent relaxations. Suppose for each $\tilde{A} \in \mathscr{A}_R$, $\Theta_I(\tilde{A})$ is singleton. Then we want to show that $\Theta_I^* = \bigcup_{\tilde{A} \in \mathscr{A}_R} \Theta_I(\tilde{A})$ is a subset of $\Theta_I^\dagger$ no matter which type of relaxation is chosen by the researcher.

To show this result, pick an arbitrary $\tilde{A} \in \mathscr{A}_R$. Define $\delta^* : A \to [0, 1]$ as follows: $\delta^*(a) = 0$ if $a \in \tilde{A}$, and $\delta^*(a) = 1$ if $a \notin \tilde{A}$. By construction, $\Theta_I(A(\delta^*)) = \Theta_I(\tilde{A})$ only contains one element. Therefore, we must have $\delta^* \in$ FF, because there cannot exist some $\delta < \delta^*$ with $\Theta_I(A(\delta)) \neq \emptyset$ and $\Theta_I(A(\delta)) \subsetneq \Theta_I(A(\delta^*))$. Because this holds for any $\tilde{A} \in \mathscr{A}_R$, we know $\Theta_I^* \subseteq \Theta_I^\dagger$.

References

Allen, Roy and John Rehbeck (2020), "Satisficing, aggregation, and quasilinear utility." Available at SSRN 3180302. [333]

Andrews, Donald W. K. and Soonwoo Kwon (2024), "Misspecified moment inequality models: Inference and diagnostics." *Review of Economic Studies*, 91 (1), 45–76. [334]

Andrews, Donald W. K. and Xiaoxia Shi (2013), "Inference based on conditional moment inequalities." *Econometrica*, 81 (2), 609–666. [342, 357]

Artstein, Zvi (1983), "Distributions of random sets and random selections." *Israel Journal of Mathematics*, 46, 313–324. [331, 332, 338, 342, 349, 357]

Aucejo, Esteban M., Federico A. Bugni, and V. Joseph Hotz (2017), "Identification and inference on regressions with missing covariate data." *Econometric Theory*, 33 (1), 196–241. [332]

Berry, Steven T. and Giovanni Compiani (2023), "An instrumental variable approach to dynamic models." *The Review of Economic Studies*, 90 (4), 1724–1758. [332, 333, 339, 342]

Blundell, Richard, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir (2007), "Changes in the distribution of male and female wages accounting for employment composition using bounds." *Econometrica*, 75, 323–363. [332, 334, 357]

Bugni, Federico A., Ivan A. Canay, and Patrik Guggenberger (2012), "Distortions of asymptotic confidence size in locally misspecified moment inequality models." *Econometrica*, 80 (4), 1741–1768. [333]

Bureau of Labor Statistics, U.S. Department of Labor (2019), "National longitudinal survey of youth 1979 cohort, 1979–2016 (rounds 1–27)." Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University, Columbus, OH. https://www.nlsinfo.org/content/cohorts/nlsy79. [352]

Chen, Xuan, Carlos A. Flores, and Alfonso Flores-Lagunes (2018), "Going beyond LATE: Bounding average treatment effects of job corps training." *Journal of Human Resources*, 53 (4), 1050–1099. [334]

Chernozhukov, Victor, Wooyoung Kim, Sokbae Lee, and Adam M. Rosen (2015), "Implementing intersection bounds in Stata." *Stata Journal*, 15 (1), 21–44. [353]

Chernozhukov, Victor, Sokbae Lee, and Adam M. Rosen (2013), "Intersection bounds: Estimation and inference." *Econometrica*, 81 (2), 667–737. [335, 353]

Chesher, Andrew and Adam M. Rosen (2020), "Econometric modeling of interdependent discrete choice with applications to market structure." Cemmap Working Paper. [332, 342, 358]

Ciliberto, Federico, Charles Murry, and Elie Tamer (2021), "Market structure and competition in airline markets." *Journal of Political Economy*, 129 (11), 2995–3038. [359]

Ciliberto, Federico and Elie Tamer (2009), "Market structure and multiple equilibria in airline markets." *Econometrica*, 77 (6), 1791–1828. [332, 334, 336, 342, 348, 349, 358, 359]

De Paula, Áureo, Seth Richards-Shubik, and Elie Tamer (2018), "Identifying preferences in networks with bounded degree." *Econometrica*, 86 (1), 263–288. [332]

Dickstein, Michael J. and Eduardo Morales (2018), "What do exporters know?" *The Quarterly Journal of Economics*, 133 (4), 1753–1801. [332]

Galichon, Alfred and Marc Henry (2011), "Set identification in models with multiple equilibria." *The Review of Economic Studies*, 78 (4), 1264–1298. [357, 359]

Ginther, Donna K. (2000), "Alternative estimates of the effect of schooling on earnings." *Review of Economics and Statistics*, 82 (1), 103–116. [336]

Gualdani, Cristina (2021), "An econometric model of network formation with an application to board interlocks between firms." *Journal of Econometrics*, 224, 345–370. [332, 333]

Haile, Philip A. and Elie Tamer (2003), "Inference with an incomplete model of English auctions." *Journal of Political Economy*, 111 (1), 1–51. [342, 358]

Heckman, James J., Justin L. Tobias, and Edward J. Vytlacil (2001), "Four parameters of interest in the evaluation of social programs." *Southern Economic Journal*, 68 (2), 210–223. [352]

Honoré, Bo E. and Luojia Hu (2020), "Selection without exclusion." *Econometrica*, 88 (3), 1007–1029. [332]

Hsu, Yu-Chin, Chu-An Liu, and Xiaoxia Shi (2019), "Testing generalized regression monotonicity." *Econometric Theory*, 35 (6), 1146–1200. [353]

Kaido, Hiroaki and Halbert White (2012), *Estimating Misspecified Moment Inequality Models*. Springer. [333]

Kédagni, Désiré (2023), "Identifying treatment effects in the presence of confounded types." *Journal of Econometrics*, 234 (2), 479–511. [334]

Kim, Wooyoung, Koohyun Kwon, Soonwoo Kwon, and Sokbae Lee (2018), "The identification power of smoothness assumptions in models with counterfactual outcomes." *Quantitative Economics*, 9 (2), 617–642. [335, 355]

Kreider, Brent, John V. Pepper, Craig Gundersen, and Dean Jolliffe (2012), "Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported." *Journal of the American Statistical Association*, 107 (499), 958–975. [334]

Manski, Charles F. (1990), "Nonparametric bounds on treatment effects." *The American Economic Review*, 80 (2), 319–323. [335, 336, 337, 344, 353, 355]

Manski, Charles F. (1994), "The selection problem." In *Advances Economics, Sixth World Congress*, Vol. 1 (Sims, C., ed.), 143–170, Cambridge University Press. [335, 355, 357]

Manski, Charles F. and John V. Pepper (2000), "Monotone instrumental variables: With an application to the returns to schooling." *Econometrica*, 68 (4), 997–1010. [334, 337, 352]

Manski, Charles F. and John V. Pepper (2009), "More on monotone instrumental variables." *The Econometrics Journal*, 12 (*suppl_1*), S200–S216. [334, 352]

Masten, Matthew A. and Alexandre Poirier (2021), "Salvaging falsified instrumental variable models." *Econometrica*, 89 (3), 1449–1469. [331, 334, 345, 347, 348]

Molinari, Francesca (2020), "Microeconometrics with partial identification." *Handbook of econometrics*, 7, 355–486. [332, 341, 342, 357, 358]

Mourifié, Ismael, Marc Henry, and Romuald Méango (2020), "Sharp bounds and testability of a Roy model of STEM major choices." *Journal of Political Economy*, 128 (8), 3220–3283. [334, 358]

Peterson, Arthur V. Jr. (1976), "Bounds for a joint distribution function with fixed subdistribution functions: Application to competing risks." *Proceedings of the National Academy of Sciences*, 73 (1), 11–13. [357]

Ponomareva, Maria and Elie Tamer (2011), "Misspecification in moment inequality models: Back to moment equalities?" *The Econometrics Journal*, 14 (2), 186–203. [333]

Sheng, Shuyang (2020), "A structural econometric analysis of network formation games through subnetworks." *Econometrica*, 88 (5), 1829–1858. [332, 333, 342, 358]