cemmap

centre for microdata methods and practice

Discordant relaxations of misspecified models

Désiré Kédagni
Lixiong Li
Ismael Mourifié

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP19/21

Economic and Social Research Council

DISCORDANT RELAXATIONS OF MISSPECIFIED MODELS

DÉSIRÉ KÉDAGNI[†], LIXIONG LI,[‡] AND ISMAËL MOURIFIÉ[*]

Iowa State University, Johns Hopkins University and the University of Toronto

Abstract. In many set identified models, it is difficult to obtain a tractable characterization of the identified set, therefore, empirical works often construct confidence region based on an outer set of the identified set. Because an outer set is always a superset of the identified set, this practice is often viewed as conservative yet valid. However, this paper shows that, when the model is refuted by the data, a nonempty outer set could deliver conflicting results with another outer set derived from the same underlying model structure, so that the results of outer sets could be misleading in the presence of misspecification. We provide a sufficient condition for the existence of discordant outer sets which covers models characterized by intersection bounds and the Artstein (1983) inequalities. Furthermore, we develop a method to salvage misspecified models. We consider all minimum relaxations of a refuted model which restore data-consistency. We find that the union of the identified sets of these minimum relaxations is misspecification-robust and it has a new and intuitive empirical interpretation. Although this paper primarily focuses on discrete relaxations, our new interpretation also applies to continuous relaxations.

Keywords: Partial Identification, Identified/outer set, misspecification, nonconflicting hypothesis, Robust Identified set.

JEL subject classification: C12, C21, C26.

1. Introduction

A central challenge in structural estimation of economic models is that the hypothesized structure often fails to identify a single generating process for the data, either because of multiple equilibria or data observability constraints. In such a context, the econometrics of partially identified models has 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

Date: The present version is as of April 1, 2021. This paper subsumes a previously circulated (and now retired) working paper written by the same authors and titled: "Bounding Treatment effects using Unconditional Moment Restrictions." Mourifié thanks the support from Connaught and SSHRC Insight Grants # 435-2018-1273. All errors are ours.

[†] Department of Economics, Iowa State University. Email: dkedagni@iastate.edu

[‡] Correspondence address: Department of Economics, Johns Hopkins University. Email: lixiong.li@jhu.edu

[*] Department of Economics, University of Toronto. Email: ismael.mourifie@utoronto.ca.

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 confidence region for an *outer set*, i.e., 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 confidence region for an outer set has been entertained in various topic of studies where the parameters of interest are only partially identified, see for instance Blundell et al. (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), and Chesher and Rosen (2020), among many others.

In most of the empirical studies, obtaining a tight outer set is very often interpreted as an evidence for a small and then informative identified set; and this is because, under correct specification, any outer set contains the identified set. However, estimating the outer set could provide very misleading results when the initial model itself is misspecified.

The first main contribution of this paper is to characterize a class of partially identified models, for which when the initial model is misspecified we can always find at least two outer sets that fail to detect the violation of the initial model and are discordant to each other. We show that various models for which the identified sets are characterized by intersection bounds, conditional moment inequalities, or the Artstein (1983) inequalities share this property. This is a negative result for this class of models, since it shows that in presence of misspecification —which often cannot be directly tested because of the non-tractability of the identified set characterization, the result provided by an outer set is entirely driven by the set of restrictions the researcher choose to construct this outer set; and, we could always consider an alternative choice of restrictions that provides a result that conflicts with the one initially provided by the researcher. Before us, various papers have been concerned about misspecification in partially identified models. An important focus has been dedicated on 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 (2013) 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. When an identified set defined by moment inequalities is empty, Andrews and Kwon (2019) point out that standard confidence regions may fail to detect model specification and lead to results with spurious precision. In such a context, they develop valid confidence region that is robust to spurious precision. 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 no inference methods including

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

those developed in the pre-cited papers can fix the issue we are arising here, which is the potential discordance of two non-empty outer sets derived from the same underlying model structure. Therefore, we need to suggest a more primitive approach to deal with these discordant results.

This objective leads to our second main contribution which consists in providing a method to salvage models that are possibly misspecified because of the existence of discordant misspecified submodels —or discordant nonempty outer sets. Our 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 a new and intuitive empirical interpretation for it. Intuitively, we will say that an hypothesis is robust to misspecification if the hypothesis is compatible with all relaxed models that are data consistent.

Prior to us, Masten and Poirier (2020) also propose an approach to salvage misspecified models. They consider a continuous relaxation of the model up to the level where the model can no longer be falsified, denoted the *falsification adaptive set*. Our second contribution complements the work of Masten and Poirier (2020) by providing a general approach to salvage misspecified models using a sequence of discrete relaxations, and by providing an alternative interpretation of the falsification adaptive set introduced in Masten and Poirier (2020). In the case of a continuous relaxation, we provide sufficient conditions under which our misspecification robust bound is included in the falsification adaptive set. The use of discrete or continuous relaxation depends on the empirical application under scrutiny. Therefore, we see this contribution as complementary to Masten and Poirier (2020).

It is worth noting that discrete relaxations of misspecified models has been entertained in various existing papers, see for instance, Manski and Pepper (2000, 2009), Blundell et al. (2007), Kreider et al. (2012), Chen, Flores, and Flores-Lagunes (2018), Kédagni (2018), 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 potentially discordant minimum relaxations of the initial model.

We organize the rest of the paper as follows. Section 2 introduces our main idea using a simple illustrative example. Section 3 presents our general setting and main results on the characterization of discordant submodels. Section 4 introduce the misspecification robust bound used to salvage misspecified models. Section 5 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 example: Intersection bounds

Although the main idea of this paper can be applied to general models, let us start with a simple introductory model, where our main idea can be illustrated in a straightforward way. Let us consider a special case of the intersection bounds in Chernozhukov, Lee, and Rosen (2013) in which parameter θ is bounded by the conditional mean of an upper and lower bounds,

$$E[\underline{Y}|Z=z] \leq \theta \leq E[\overline{Y}|Z=z] \quad \text{almost surely,} \tag{2.1}$$

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

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

The identified set of θ is interval $[\underline{\gamma}, \overline{\gamma}]$ when $\underline{\gamma} \leq \overline{\gamma}$. 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 $E[\underline{Y}|Z] \leq E[\overline{Y}|Z]$ almost surely.*

This simple framework encompasses some important treatment effect models.

Example 1 (Discrete treatment model)**.** Consider a setting where $\mathcal{X} := \{x_1, ..., x_K\}$ 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 (2.1).

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 (2.1). 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,

$\underline{Y}(x) = Y + \sum_i \min\Big\{(x_i - X_i)\underline{\beta}_i, (x_i - X_i)\overline{\beta}_i\Big\}$ where the subscript i stands for the i^{th} dimension of the corresponding variables. Similarly, $\overline{Y}(x) = Y + \sum_i \max\Big\{(x_i - X_i)\underline{\beta}_i, (x_i - X_i)\overline{\beta}_i\Big\}$.

In practice, model (2.1) is sometimes implemented by solving its unconditional version,

$$E\big[h(Z)(\theta - \underline{Y})\big] \geq 0 \text{ and } E\big[h(Z)(\overline{Y} - \theta)\big] \geq 0, \tag{2.2}$$

where h is some nonnegative function mapping its input to $\mathbb{R}^m_+$ with $m < \infty$, and the inequalities in (2.2) are vector inequalities. The inference for (2.2) is typically much simpler than the inference for the original model (2.1), espectially when Z is multi-dimensional. Let $\widetilde{\Theta}(h)$ be the identified set for θ in model (2.2). As made explicit in the notation, $\widetilde{\Theta}(h)$ depends on the choice of instrumental function h. However, since (2.1) implies (2.2), 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 (2.1), i.e. $[\underline{\gamma}, \overline{\gamma}] \subseteq \widetilde{\Theta}(h)$. This inclusion relation is often used as a justification for using model (2.2), it may not be as informative as model (2.1), but its identification result $\widetilde{\Theta}(h)$ is often viewed as a conservative bound for $[\underline{\gamma}, \overline{\gamma}]$, the identified set for model (2.1).

Our first observation is that $\widetilde{\Theta}(h)$ is not always reliable. In fact, it can be very misleading when the original model (2.1) is refuted. To see this, define $\mathcal{H}^+_m$ to be the space of all nonnegative instrumental functions with dimension m. More formally, let $\mathcal{H}^+_m := \{h : \mathcal{Z} \mapsto \mathbb{R}^m_+$ such that $E\,\|h(Z)\| < \infty$, $E\,\|\underline{Y}h(Z)\| < \infty$, $E\,\|\overline{Y}h(Z)\| < \infty$ and $E|h_i(Z)| > 0$, $\forall\, i = 1, ..., m\}$. Then, we have the following theorem.

Theorem 1. *Suppose Assumption 1 holds. If the restriction in* (2.1) *is refuted, i.e.* $\underline{\gamma} > \overline{\gamma}$*, 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 (2.1) is refuted, Theorem 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. In the extreme case where the mean independence condition is so much violated that $[E\underline{Y}, E\overline{Y}] \subseteq (\overline{\gamma}, \underline{\gamma})$, this means that any point in the Manski worst-case bounds can be picked up as the point identification result by some choice of h.

Theorem 1 suggests that while often considered as conservative, $\widetilde{\Theta}(h)$ can be very misleading. The intuition is that the unconditional moment restriction can fail to detect the violation of (2.1) and, at the same time, provide a misleading informative identification result. The worst the violation of (2.1), the wider range of values this misleading result could take. Theorem 1 implies that when (2.1) is refuted, there exists two h and h' such that both $\widetilde{\Theta}(h)$ and $\widetilde{\Theta}(h')$ are nonempty but $\widetilde{\Theta}(h) \cap \widetilde{\Theta}(h')$ is empty. In other words, there exists two unconditional model (2.2) with two different choices of h, such that each of them is consistent with the data, but they are not compatible with each other. Since there is little reason to favour one h over the other, sticking to either of these would miss the whole picture and leads to misleading conclusion.

Remark 1. This result complements the findings in Andrews and Shi (2013). In Andrews, Kim, and Shi (2017), they propose an inference procedure and a Stata code named "**cmi-interval**" which constructs a confidence region for parameters θ in model like (2.1). Their inference approach transform (2.1) into (2.2) by selecting h in a sub-family of $\mathcal{H}_m^+$ 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 (2.1) could be misspecified. If the dimension of h is fixed, then the empirical result for (2.2) could be misleading even if the inference controls the size uniformly. The "**cmi-interval**" command gives the possibility to the researcher to fix m, in such a context, if (2.1) is indeed misspecified, two different researchers can chose two different pairs $(m, h) \neq (m', h')$ where each of them will provide a very tight outer set but with discordant information; and then raising the question on how to interpret the results under the model misspecification.

In the next section, we will show that the observation made here can be extended to more general settings. Indeed, whenever one starts from a full model as in (2.1) and implements it using some of its implications as in (2.2), one could get misleading results when the full model is refuted.

3. Misleading Submodels in a General Setting

Let A be some nonempty collection of assumptions. For any nonempty subset $A' \subseteq A$, define $\Theta_I(A')$ to be the identified set of θ when all assumptions $a \in A'$ hold. Let $\emptyset$ denote the empty set, and define $\Theta_I(\emptyset) = \Theta$ where Θ denote the parameter space. Throughout the paper, we assume Θ is some subset in a metric space. For each $a \in A$, we abbreviate $\Theta_I(\{a\})$ as $\Theta_I(a)$. We interpret A as the *full model*, and $A' \subsetneq A$ as a *submodel*. In the introductory example, the role of A is played by the set of all models (2.2) indexed by the instrumental function h. And, (2.1) holds if and only if all assumptions a in A hold.

We say model A' is *data-consistent* if $\Theta_I(A')$ is nonempty, and call it *refuted* if the reverse is true. By definition, $\Theta_I(A) \subseteq \Theta_I(A')$ for any $A' \subseteq A$. Therefore, if the full 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 other words, when the full model is data-consistent, each $\Theta_I(A')$ can be viewed as a conservative bound for $\Theta_I(A)$, and all of these submodels are mutally compatible.

We are interested in what would happen when the full model A is rejected. In particular, we want to find conditions under which there exists two data-consistent submodels A' and $A'' \in \mathcal{A}$ such that $\Theta_I(A') \cap \Theta_I(A'')$ is empty. When such sets A' and A'' exist, both of these two submodels fail to detect the failure of the full model, and could lead to discordant empirical results. Note that these sets A' and A'' do not always exist. Consider a simple example where $A = \{a_1, a_2, a_3\}$ with $\Theta_I(a_1) = \emptyset$ and $\Theta_I(\{a_2, a_3\}) \neq \emptyset$. Then, for any data-consistent submodel A', we have $\Theta_I(\{a_2, a_3\}) \subseteq \Theta_I(A')$. Hence, there does not exist two data-consistent submodels which are mutually incompatible. In the following, we provide a sufficient condition under which discordant submodels do exist.

Theorem 2. *Assume the two following conditions hold:*

(T2.C1) Whenever $\Theta_I(A) = \emptyset$, there exists some nonempty $A^* \subseteq A$ such that $\Theta_I(A^*) = \emptyset$ and $\Theta_I(a) \neq \emptyset$ for any $a \in A^*$.

(T2.C2) For any nonempty subset $B \subseteq A$, $\Theta_I(B) = \cap_{a \in B}\Theta_I(a)$.

And If A is not a finite set, assume, in addition, that the following condition holds

(T2.C3) $\Theta_I(a)$ is compact for each $a \in A$.

Then, the full model is refuted, i.e. $\Theta_I(A) = \emptyset$, if and only if there exist two finite subset $A', A'' \subseteq A$, such that both $\Theta_I(A)$ and $\Theta_I(A')$ are nonempty but $\Theta_I(A) \cap \Theta_I(A')$ is empty.

Moreover, when $\Theta_I(A) = \emptyset$, for any submodel $B \subseteq A$ with nonempty $\Theta_I(B)$, there exist two finite subset B', B'' of A, such that $\Theta_I(B \cup B') \neq \emptyset$, $\Theta_I(B'') \neq \emptyset$ and $\Theta_I(B \cup B') \cap \Theta_I(B'') = \emptyset$.

First, Theorem 2 shows that, under condition (T2.C1)-(T2.C3), whenever the full model is rejected there always exists a pair of submodels which are data-consistent and mutually discordant. Second, it suggests that, for every data-consistent submodel B, there exists another data-consistent submodel which is not compatible with some stronger version of B. Finally, it shows that even if A is not a finite set, we require only two finite subsets to reject A. In the partial identification literature, whenever the validity of a strong assumption is in question, some weaker assumption (or what we call a submodel here) is often considered as a better modelling choice, especially when this weak assumption still leads to informative results. Theorem 2 suggests that this top-to-down strategy could be misleading unless the researcher have a preference order on these mutually incompatible weaker assumptions ex ante, because one could get different conclusions by picking different submodels.

Let us now briefly discuss the three conditions in Theorem 2. Condition (T2.C1) is the key condition which generates misleading submodels. In condition (T2.C1), A^* can be A itself or some subset of it. A^* could also depend on the underlying data generating process. It requires the existence of a collection of assumptions which are data-consistent separately but are refuted jointly. The intuition is that if some data-consistent assumptions are refuted jointly, then some of them must be mutually incompatible. However, this condition alone does not guarantee the result of Theorem 2. In condition (T2.C2), we assume that the identified set of a set of assumptions is equal to the intersection of the identified sets of each assumption. This condition typically holds if every assumption a in A is imposed on observables and parameters, and it may not hold when some of the submodels involve assumptions on the unobservables. However, note that condition (T2.C2) can still covers the case where the full model involves assumptions on the unobservables. In most partially identified models, there exists a collection of assumptions A on the observables and parameters so that the original assumption which involves the unobservables holds if and only if assumption a holds for each $a \in A$. As shown later in Example 3, Theorem 2 is useful in these cases. In addition, condition (T2.C2) can be further weakened when A is a finite set. See Theorem $2'$ relegated to the Appendix for the sake of simplicity. Finally, condition (T2.C3) is satisfied in most empirical models where θ is of finite dimension and it is only needed when A contains an infinite number of assumptions. In Appendix B.2.1, we show the result of Theorem 2 can fail if any of these conditions is violated. In the following, we will explore the implications of Theorem 2 for different classes of partially identified models.

3.1. **Introductory example continued.** Before looking at those more advanced cases, it is probably helpful for us to revisit the introductory example to illustrate how these conditions can be verified in a simple model. In the introductory example, recall $\mathcal{H}_1^+$ stands for the set of all one dimensional nonnegative instrumental functions which satisfy some basic regularity conditions. The set A in the introductory example is the set of all assumptions indexed by $h \in \mathcal{H}_1^+$ with which (2.2) holds. And, A is indeed the full model because (2.1) holds if and only if (2.2) holds with each $h \in \mathcal{H}_1^+$. In addition, for all $h \in \mathcal{H}_1^+$, the identified set $\widetilde{\Theta}(h)$ for model (2.2) 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 whenever $E[\underline{Y}|Z] \leq E[\overline{Y}|Z]$ almost surely. Therefore, if we simply let $A^* = A$, then $\Theta_I(a) \neq \emptyset$ for all $a \in A^*$, thus condition (T2.C1) is satisfied. Since (2.2) are moment inequalities which only depend on observables and the parameter, condition (T2.C2) is also satisfied. Finally, since the identified set characterized by (2.2) is always a closed interval, condition (T2.C3) is satisfied under the regularity conditions imposed within the definition of $\mathcal{H}_1^+$.

Note that, in this example, if $B \subseteq A$ and B consists of m assumptions, then B refers to the submodel that (2.2) holds for some $h \in \mathcal{H}_m^+$. As a result, Theorem 2 implies that when (2.1) 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. This is a weaker result compared to what we found in Theorem 1, which is not surprising as Theorem 1 utilizes some model specific structures whereas Theorem 2 is built on more general properties.

In Appendix A.1, we provide sufficient conditions under which Theorem 2 applies to the following conditional moment inequalities:

$$E[m(X, Z; \theta)|Z] \leq 0 \text{ almost surely} \tag{3.1}$$

where $X \in \mathbb{R}^{k_1}$ 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.

3.2. **Random Sets and Choquet Capacity.** In this subsection, we consider a model with a vector of random variables Y and a vector of exogenous observable covariates X. Let $\mathcal{Y}(\theta)$ be some random closed set and assume $P(Y \in \mathcal{Y}(\theta)) = 1$. Then, Artstein (1983) shows that the conditional distribution of Y given X equals F if and only if for any compact set K in the support of Y, the following inequality holds:

$$P_F(Y \in K|X) \leq L(K, X; \theta) := P(\mathcal{Y}(\theta) \cap K \neq \emptyset|X) \text{ almost surely,} \tag{3.2}$$

where P_F refers to the probability measure corresponding to the distribution F. The quantity $L(\cdot, X; \theta)$ is often known as the Choquet capacity function. These type of models are studied in Galichon and Henry (2011). 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 that Y is observable and assume $L(K, X; \theta)$ is a known function of K and X given θ.

In general, one needs to check (3.2) for all compact sets in order to ensure the equivalence between the collection of inequalities in (3.2) and the distributional assumption on Y. In some circumstances, checking the inequalities for a collection of compact sets is equivalent to checking them for all compact sets, in which case, the collection is called the *core determining class* in the language of Galichon and Henry (2011). However, in practice, researchers typically pre-select some finite collection $\mathcal{K}$ of compact sets —which is not often a core determining class, and only check (3.2) for $K \in \mathcal{K}$. 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 Arstein 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 Arstein inequalities.[2]

We want to explore the consequence of this pre-selection procedure when the original model might be refuted. To fit this model into the general framework in Section 3, define $\mathscr{K}$ to be the collection of all nonempty compact sets in the support of Y. Define A as the set of all assumptions indexed by $K \in \mathscr{K}$ with which (3.2) holds. Then, for any finite subset $A' \subseteq A$, $\Theta_I(A')$ refers to the identified set when (3.2) is checked for only a finite number of different Ks. As in the previous section, condition (T2.C2) of Theorem 2 are satisifed by construction. We only need to verify condition (T2.C1) and (T2.C3).

Condition (T2.C3) is usually easy to check. It would hold, for example, when the parameter space Θ is compact and $L(K, X; \theta)$ is continuous in θ for all K and X. Condition (T2.C1) is more challenging. Without imposing more structures, the most natural choice for A^* is $A^* = A$. By construction, $\Theta_I(A^*) = \emptyset$ whenever $\Theta_I(A) = \emptyset$. However, in such a case, we need more regularity conditions to ensure $\Theta_I(a) \neq \emptyset$ for all $a \in A^*$. When the support of Y is discrete and finite, the following proposition provides a sufficient condition for Condition (T2.C1) to hold.

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

Lemma 1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be the support of X and Y, respectively. Suppose $\mathcal{Y}$ is discrete and finite and the parameter space $\Theta \subseteq \mathbb{R}^d$. In addition, suppose the following conditions hold for any $y \in \mathcal{Y}$,*

(L1.C1) $\inf_{x\in\mathcal{X}} P(Y = \{y\} \,|X = x) > 0$,
(L1.C2) there exists a sequence $\theta_1, \theta_2, ...$ in $\mathbb{R}^d$ 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 denotes the essential infimum. Then, there exists some compact set $\tilde{\Theta}$ such that condition (T2.C1) of Theorem 2 holds whenever $\tilde{\Theta} \subseteq \Theta$.*

Lemma 1 implies that the results in Theorem 2 hold in this example if (L1.C1) and (L1.C2) hold and if Θ is large enough to include the compact set $\tilde{\Theta}$. To see why we need Θ to be large enough, note that, if Θ is too small, it is possible that Θ is included in the identified set of some submodel satisfying (3.2) so that there will never exist discordent submodels even if the full model is misspecified. When the size of Θ is not a binding constrain, we get the following corollary as a result of Lemma 1 and Theorem 2.

Corollary 1. *Suppose the support of Y is discrete and finite and the parameter space $\Theta \subseteq \mathbb{R}^d$. Suppose (L1.C1) and (L1.C2) hold and $\tilde{\Theta} \subseteq \Theta$. For any collection $\mathcal{K}$ of compact sets, define $\Theta_I(\mathcal{K})$ as the identified set of the submodel satisfying* (3.2) *with each $K \in \mathcal{K}$. Then, the full model is refuted if and only if there exist two finite collections of compact sets $\mathcal{K}_1$ and $\mathcal{K}_2$ such that both $\Theta_I(\mathcal{K}_1)$ and $\Theta_I(\mathcal{K}_2)$ are nonempty, but these two identified sets have empty intersection.*

Moreover, whenever the full model is refuted, for any finite collection of compact sets $\widetilde{\mathcal{K}}$ with nonempty $\Theta_I(\widetilde{\mathcal{K}})$, there exist $\mathcal{K}_1$ and $\mathcal{K}_2$ such that $\Theta_I(\widetilde{\mathcal{K}} \cup \mathcal{K}_1) \neq \emptyset$ and $\Theta_I(\mathcal{K}_2) \neq \emptyset$, but $\Theta_I(\widetilde{\mathcal{K}} \cup \mathcal{K}_1) \cap \Theta_I(\mathcal{K}_2) = \emptyset$.

We conclude this subsection with an example of entry game model.

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

$$\pi_i = \gamma_i + X_i'\beta - \sum_{j\neq i} \delta_j Y_j + \epsilon_i$$

where the X_is are some covariates which might be player i specific, γ_i and β is the parameter coefficient, $Y_j \in \{0, 1\}$ stands for player j's entry decision, and $\delta_j > 0$ is the interaction parameter of player j. For simplicity, we assume that $Y = (Y_i : i = 1, ..., n)$ is always a pure strategy Nash equilibirum.

Assume $\epsilon = (\epsilon_1, ..., \epsilon_n)$ is independent of X and ϵ follows the normal distribution $N(0, \Sigma)$. Note that this is an assumption specifed on the unobservables. Let $\gamma = (\gamma_1, ..., \gamma_n)$ and $\theta = (\gamma, \beta, \delta, \Sigma)$ be the vector of all parameters. Let $\mathcal{Y} = \{y = (y_1, ..., y_n) : y_i \in \{0, 1\}, i = 1, ..., n\}$ be the set of all possible entry decisions, and let $2^{\mathcal{Y}}$ denote the set of all subsets of $\mathcal{Y}$. For any $K \in 2^{\mathcal{Y}}$, 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 simulation.

In Galichon and Henry (2011), the identified set of this model is shown to be the set of all θ which satisfies (3.2) for all nonempty set K of $\mathcal{Y}$. The number of these inequalities increases with n very quickly in the order of 2^{2^n}. Galichon and Henry (2011) provide some methods to reduce the number of inequalities by removing redundant inequalities in (3.2), but, in general, sharp characterization of the identified set involves a large number of inequalities. In practice for the sake of computational feasibility, emprical researchers often pre-select a finite collection $\mathcal{K}$ of subsets and only check (3.2) for each $K \in \mathcal{K}$. See, for example, Ciliberto and Tamer (2009), and Ciliberto, Murry, and Tamer (2020).

Let us now check conditions in Proposition 1. Condition (L1.C1) in Proposition 1 are low-level assumptions on the data. Since the Choquet capacity $L(K, x; \theta)$ is continuous in x in this example, Condition (L1.C1) can be implied by assuming X has a compact support and $P(Y = y|X) > 0$ almost surely for all y. Condition (L1.C2) in Proposition 1 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 $y_i = 0$.

4. Misspecification Robust Bounds

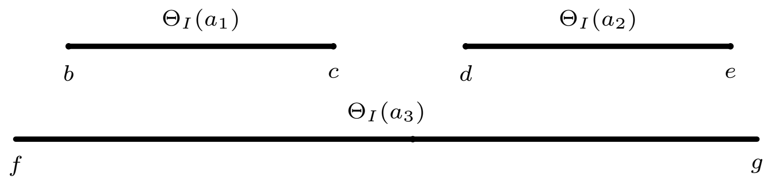
In this section, we explore what one could possibly do to salvage a refuted full model.

4.1. **Minimum data-consistent relaxation.** Our idea is to construct some minimum relaxation of the full model by removing incompatible submodels until all remaining submodels are compatible. Often, there are multiple ways to relax a model to restore data consistency. Since typically there is no reason to prefer one relaxation over the others, we take the union of the identified set of all these relaxed models. To illustrate our idea more clearly, let us consider a simple example where $A = \{a_1, a_2, a_3\}$ and the identified sets of each a_i are all closed intervals in $\mathbb{R}$ as shown in Figure 1 with $\Theta_I(a_1) = [b, c]$, $\Theta_I(a_2) = [d, e]$, $\Theta_I(a_3) = [f, g]$ and $f \leq b \leq c < d \leq e \leq g$. Assume also that $\Theta_I(\{a, a'\}) = \Theta_I(a) \cap \Theta_I(a')$ for $a, a' \in \{a_1, a_2, a_3\}$.

The full model A is refuted, but one can restore data-consistency by just removing a_1 or a_2 from A, which leads to $A' = \{a_2, a_3\}$ and $A'' = \{a_1, a_3\}$. A' and A'' are relaxations of A, since some restrictions have been removed. In some sense, they are also minimum relaxations since adding any removed assumptions to A' or A'' will be at odds with the data. More generally, we can define a minimum data-consistent relaxation as follows.

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.

FIGURE 1. The three-interval example



According to Definition 1, $A''' = \{a_3\}$ in the example of Figure 1 is not a minimum data-consistent relaxation, since it will remain data-consistent after including a_1 or a_2. The following theorem establishes the existence of minimum data-consistent relaxations in a general setting.

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 $B \subseteq A$, $\Theta_I(B) = \cap_{a\in B}\Theta_I(a)$.*

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

In general, there exist multiple minimum data-consistent relaxations. Indeed, whenever there exist two mutually incompatible data-consistent submodels, there exist at least two minimum data-consistent relaxations. If there is no reason to favor one submodel over another ex ante, it makes sense to take all of these relaxations into consideration.

Definition 2. Let $\mathscr{A}_R$ be the set of all minimum data-consistent relaxations. Define *misspecification robust bound* Θ_I^* as $\Theta_I^* := \cup_{\widetilde{A}\in\mathscr{A}_R}\Theta_I(\widetilde{A})$.

The next theorem shows that when the full model is data-consistent, Θ_I^* would be equal to the identified set of the full model. In fact, by the definition of minimum data-consistent relaxation, we know A is a minimum data-consistent relaxation. In addition if $\Theta_I(A) \neq \emptyset$, for any $A' \subsetneq A$, A' cannot be a minimum data-consistent relaxation, because there always exists some $a \in A\backslash A'$ with $\Theta_I(A' \cup \{a\}) \neq \emptyset$. Hence, A is the only minimum data-consistent relaxation so that $\Theta_I^* = \Theta_I(A)$. This result is summarized in the following theorem.

Theorem 4. *If $\Theta_I(A) \neq \emptyset$, then A is the unique minimum data-consistent relaxation and $\Theta_I^* = \Theta_I(A)$.*

Given Theorem 4, Θ_I^* can therefore be viewed as a way to generalize the identified set to possibly misspecified cases. Furthermore, in addition to its definition, Θ_I^* has some interesting interpretations.

Consider a hypothetical statement on θ that $\theta \in S$ for some subset S of Θ. Then, this hypothesis would be implied by some submodel A' if $\Theta_I(A') \subseteq S$. Reversely, it would be rejected by submodel A' if $\Theta_I(A') \cap S = \emptyset$. We say the hypothetical statement that $\theta \in S$ is *nonconflicting* if

(C1) $\theta \in S$ is implied by some data-consistent submodel. That is, there exists some submodel $A' \subseteq A$ such that $\Theta_I(A') \subseteq S$ and $\Theta_I(A') \neq \emptyset$.

(C2) $\theta \in S$ is not rejected by any data-consistent submodel. That is, there does not exist a submodel $A' \subseteq A$ with $\Theta_I(A') \neq \emptyset$ such that $\Theta_I(A') \cap S = \emptyset$.

We use Λ to denote the collection of all nonconflicting statements:

$$\Lambda \equiv \Big\{ S \subseteq \Theta : \ (C1) \text{ and } (C2) \text{ hold for } S \Big\}.$$

Nonconflicting statements are not unique. As can be seen, if $S \in \Lambda$, so are its supersets. 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 consider that being nonconflicting is a minimum requirement for a statement to be robust to model misspecification. If a statement fails to be nonconflicting, then either it is not implied by any of the data-consistent submodels, or it is rejected by some data-consistent submodels. In the following two theorems, we explore the linkage between nonconflicting statements and Θ_I^*.

Theorem 5. *Suppose the same assumptions in Theorem 3 hold. Suppose also that $\Theta_I(a) \neq \emptyset$ for at least one $a \in A$. Then $\Theta_I^* \in \Lambda$, so that for any arbitrary nonempty subset S of Θ, $\Theta_I^* \subseteq S$ implies $S \in \Lambda$.*

Theorem 5 states that $\theta \in \Theta_I^*$ is nonconflicting. It also provides a sufficient condition for being nonconflicting: any statement that is implied by $\theta \in \Theta_I^*$ is also nonconflicting. To see what this means, consider the simple case where θ is a scalar. Suppose we are interested in the sign of θ. Suppose Θ_I^* turns out to be within the positive real line, i.e., $\Theta_I^* \subseteq \mathbb{R}_{++}$. Then, Theorem 5 implies that the statement that θ is positive, i.e., $\theta \in \mathbb{R}_{++}$, is nonconflicting. 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.

We say S^* is the *smallest* element in Λ if $S^* \in \Lambda$ and $S^* \subseteq S$ for any $S \in \Lambda$. When Θ_I^* is the smallest element in Λ, then a statement that $\theta \in S$ is nonconflicting if and only if $\Theta_I^* \subseteq S$. In this case, Θ_I^* could have richer interpretation beyond Theorem 5. Consider the previous simple example again when θ is a scalar. 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 element in Λ, then we have the following conclusion: both θ is positive and θ is negative are conflicting 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. In the next theorem, we study when Θ_I^* is the smallest element of Λ.

Theorem 6. *Suppose the same assumptions in Theorem 5 hold. For the following three statements:*

(S1) Θ_I^* *is the smallest element in* Λ*,*

(S2) Λ *has a smallest element,*

(S3) at least one of the following two conditions is satisfied:

(T6.C1) there exists a unique minimum data-consistent relaxation,

(T6.C2) for any minimum data-consistent relaxation $\widetilde{A}$*,* $\Theta_I(\widetilde{A})$ *is a singleton,*

we have (S3) $\Rightarrow$ *(S1)* $\Rightarrow$ *(S2). Moreoever, if the following condition holds:*

(T6.C3) for any A' *and* A'' *with* $\Theta_I(A') \subseteq \Theta_I(A'')$ *and* $\Theta_I(A') \neq \emptyset$*, we have* $\Theta_I(A' \cup A'') \neq \emptyset$*,*

then we have (S3) $\Leftrightarrow$ *(S1)* $\Leftrightarrow$ *(S2).*

Theorem 6 provides (S3) as a sufficient condition for Θ_I^* being the smallest element in Λ. Condition (T6.C1) holds, for example, when the full model is correctly specified. It also holds in the binary IV model example we will study shortly in section 4.2.2. When the full model is misspecified, the identified set of a submodel would gradually shrink when imposing more and more assumptions in A, and eventually shrink to an empty set. Condition (T6.C2) means that, during this process, the identified set of a submodel will always become a singleton before it becomes empty. As we will see later, this condition holds in our introductory example. When Condition (T6.C3) holds, Theorem 6 states that (S3) is not only sufficient but also necessary for Θ_I^* being the smallest element in Λ. It also implies that whenever Λ has a smallest element, the smallest element must be Θ_I^*.

Condition (T6.C3) is implied by (T3.C2), so it always holds when all assumptions in A are imposed on the parameter θ and the observables. However, Condition (T6.C3) may fail when assumptions in A involve unobservables or nuisance parameters. And, the failure of (T6.C3) can lead to a discrepancy between Θ_I^* and the smallest element in Λ, but the smallest element in Λ is *not* an appropriate set to consider in this case. This is best illustrated by the following example. Consider a simple example illustrated in Figure 2, where θ is a scalar parameter and η is some nuisance parameter which could be the distribution of the unobservables or some other parameter not of interest. Suppose, there are only two assumptions $\{a_1\}$ and $\{a_2\}$ in A. Because the identified set of a_1 and a_2 for (θ, η) are disjoint, $\Theta_I(\{a_1, a_2\}) = \emptyset$, and the full model $A = \{a_1, a_2\}$ is refuted by the data. However, $\Theta_I(a_1) = [a, b] \subsetneq \Theta(a_2) = [c, d]$ so that Condition (T6.C3) is violated in this example. In this case, both a_1 and a_2 are minimum data-consistent relaxations and $\Theta_I^* = [c, d]$, but the smallest element in Λ is $[a, b]$. Although $[a, b]$ satisfy the minimum requirement to be a nonconflicting statement, we don't consider it an appropriate set to report in the empirical work because it implicitly assumes that $\eta > \eta^*$ which is not part of our original assumptions. In contrast, $\Theta_I^* = [c, d]$ is the tightest bounds one can put on θ without making extra assumptions on the nuisance parameters.

When the full model is refuted, Theorem 2 (and its alternative version Theorem 2′ relegated to appendix) establishes the existence of a and a' in A such that $\Theta_I(a) \neq \emptyset$, $\Theta_I(a') \neq \emptyset$ and

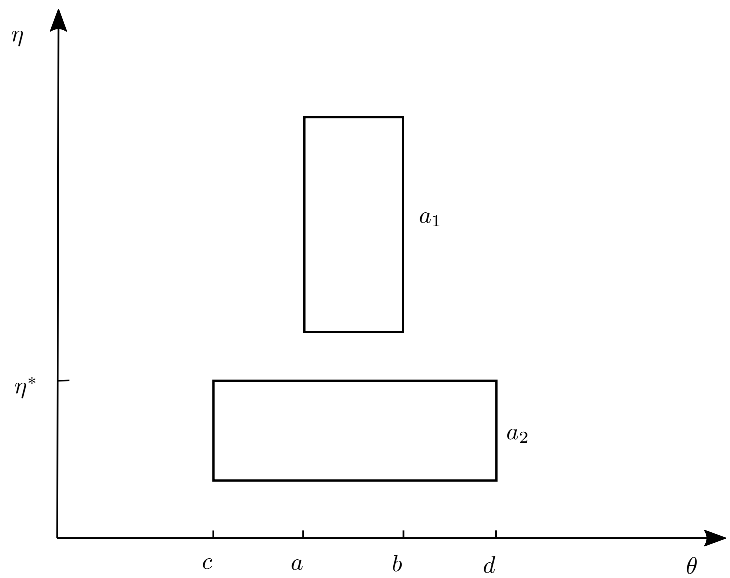


FIGURE 2. Illustration when Condition (T6.C3) is violated

$\Theta_I(a') \cap \Theta_I(a) = \emptyset$. However, they do not say anything on how different $\Theta_I(a')$ and $\Theta_I(a)$ are. Interestingly, the "size" of Θ_I^*, which can be gauged using its alternative characterization proposed in Theorem 6, may provide more accurate information on the type of discordance that happens between $\Theta_I(a)$ and $\Theta_I(a')$. Indeed in such a context, when Θ_I^* is small, it means that the data-consistent submodels though discordant are not that different, in other terms it suggests that the Hausdorff distance between $\Theta_I(a')$ and $\Theta_I(a)$ is always very small. Conversely, when Θ_I^* is large it means that the mutually discordant submodels deliver very different results or each of them is large.

Now, we illustrate what our misspecification robust bound Θ_I^* looks like in three examples.

4.2. **Some examples of misspecification robust bounds.**

4.2.1. *Introductory example continued.* For the model (2.1), the misspecification robust bound is given in the following result:

Proposition 1. *Suppose Assumption 1 holds, then:*

$$\Theta_I^* = \begin{cases} [\underline{\gamma}, \overline{\gamma}] & \textit{if } \underline{\gamma} \le \overline{\gamma}, \\ [\overline{\gamma}, \underline{\gamma}] & \textit{if } \overline{\gamma} < \underline{\gamma}, \quad P(E[\underline{Y}|Z] \le \overline{\gamma}]) > 0 \textit{ and } P(E[\overline{Y}|Z] \ge \underline{\gamma}]) > 0, \\ (\overline{\gamma}, \underline{\gamma}] & \textit{if } \overline{\gamma} < \underline{\gamma}, \quad P(E[\underline{Y}|Z] \le \overline{\gamma}]) = 0 \textit{ and } P(E[\overline{Y}|Z] \ge \underline{\gamma}]) > 0, \\ [\overline{\gamma}, \underline{\gamma}) & \textit{if } \overline{\gamma} < \underline{\gamma}, \quad P(E[\underline{Y}|Z] \le \overline{\gamma}]) > 0 \textit{ and } P(E[\overline{Y}|Z] \ge \underline{\gamma}]) = 0, \\ (\overline{\gamma}, \underline{\gamma}) & \textit{if } \overline{\gamma} < \underline{\gamma}, \quad P(E[\underline{Y}|Z] \le \overline{\gamma}]) = 0 \textit{ and } P(E[\overline{Y}|Z] \ge \underline{\gamma}]) = 0. \end{cases} \tag{4.1}$$

Moreover, Condition (T6.C2) holds so that Θ_I^* is the smallest element in Λ, i.e., $\Theta_I^* = \underset{S \in \Lambda}{\cap} S$.

A direct implication of Proposition 1 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 $\Theta_I^* = [\min(\underline{\gamma}, \overline{\gamma}), \max(\underline{\gamma}, \overline{\gamma})]$.

4.2.2. *Binary instrumental variable (IV) model.* Consider the following potential outcome model: $Y = [Y_{11}Z + Y_{10}(1-Z)]D + [Y_{01}Z + Y_{00}(1-Z)](1-D)]$ where Y, D, and Z are all binary. Y_{dz} represents the potential outcome when D and Z are externally set to d and z, respectively. The parameters of interest are the average potential outcomes, i.e., $\theta_{dz} := E[Y_{dz}]$ with $\theta := (\theta_{11}, \theta_{10}, \theta_{01}, \theta_{00})$. Let us denote $q_{ij}(z) := P(Y = i, D = j|Z = z)$ for $i, j \in \{0, 1\}$. Let consider the following set of assumptions:

- $a_1 := Y_{dz} \perp Z$, $d, z \in \{0, 1\}$,
- $a_2 := Y_{11} \geq Y_{10}$,
- $a_3 := Y_{11} \leq Y_{10}$,
- $a_4 := Y_{01} \geq Y_{00}$,
- $a_5 := Y_{01} \leq Y_{00}$.

Manski (1990) has studied the identification of the average potential outcomes under the IV independence and exclusion restrictions, i.e., $A = \{a_1, a_2, a_3, a_4, a_5\} = \{Y_{dz} \perp Z \;\&\; Y_{d1} = Y_{d0}, d \in \{0, 1\}\}$. Pearl (1994) has shown that model A is falsifiable by deriving the so called Pearl's instrumental inequality, and Kitagawa (2009) proved that the instrumental inequality are the most informative to detect observable violations of model A. The instrumental inequality are the following:

$$q_{11}(1) + q_{01}(0) \leq 1, \tag{4.2}$$

$$q_{11}(0) + q_{01}(1) \leq 1, \tag{4.3}$$

$$q_{10}(1) + q_{00}(0) \leq 1, \tag{4.4}$$

$$q_{10}(0) + q_{00}(1) \leq 1. \tag{4.5}$$

If at least one of these inequalities is violated model A is rejected. In such a scenario, it is useful to construct our misspecification robust bound. Before doing so, notice that any assumption a_i, $i \in \{1, ..., 5\}$ has a clear economic interpretation. For instance, a_2 implies that the average causal direct effect (ACDE) associated to the treatment $D = 1$ is non-negative, i.e., $ACDE(d) = \theta_{d1} - \theta_{d0} \geq 0$ for $d = 1$.[3] In other terms, the instrument Z is not necessarily excluded from the outcome equation, but has a direct non-negative causal effect on the outcome when the treatment is externally set to $D = 1$. We can prove that:

[3]Please refer to Pearl (2001) and Cai et al. (2008) for a detailed discussion on the ACDE.

$$\Theta_I^* = \begin{cases} \Theta_I(A) & \text{if Eqs 4.2-4.5 hold,} \\ \Theta_I(\{a_1,a_2,a_4,a_5\}) & \text{if Eqs 4.3-4.5 hold, but Eq 4.2 is violated,} \\ \Theta_I(\{a_1,a_3,a_4,a_5\}) & \text{if Eqs 4.2, 4.4, 4.5 hold, but Eq 4.3 is violated,} \\ \Theta_I(\{a_1,a_2,a_3,a_4\}) & \text{if Eqs 4.2, 4.3, 4.5 hold, but Eq 4.4 is violated,} \\ \Theta_I(\{a_1,a_2,a_3,a_5\}) & \text{if Eqs 4.2, 4.3, 4.4 hold, but Eq 4.5 is violated,} \\ \Theta_I(\{a_1,a_2,a_5\}) & \text{if Eqs 4.3,4.5 hold, but Eqs 4.2, 4.4 are violated,} \\ \Theta_I(\{a_1,a_2,a_4\}) & \text{if Eqs 4.3,4.4 hold, but Eqs 4.2, 4.5 are violated,} \\ \Theta_I(\{a_1,a_3,a_5\}) & \text{if Eqs 4.3,4.5 hold, but Eqs 4.2, 4.4 are violated,} \\ \Theta_I(\{a_1,a_3,a_4\}) & \text{if Eqs 4.3,4.4 hold, but Eqs 4.2, 4.5 are violated,} \end{cases} \tag{4.6}$$

where

$$\Theta_I(A)) = \left\{ \theta : \begin{cases} \sup_z q_{11}(z) \le \theta_{10} = \theta_{11} \le 1 - \sup_z q_{01}(z) \\ \sup_z q_{10}(z) \le \theta_{01} = \theta_{00} \le 1 - \sup_z q_{00}(z). \end{cases} \right\}$$

$$\Theta_I(\{a_1,a_2,a_4,a_5)\}) = \left\{ \theta : \begin{cases} q_{11}(0) \le \theta_{10} \le 1 - q_{01}(0), \\ q_{11}(1) \le \theta_{11} \le 1 - q_{01}(1), \\ \theta_{11} - \theta_{10} \ge \max\{0, q_{11}(1) + q_{01}(0) - 1\}, \\ \sup_z q_{10}(z) \le \theta_{01} = \theta_{00} \le 1 - \sup_z q_{00}(z). \end{cases} \right\}$$

For the sake of conciseness, the remaining sets from $\Theta_I(\{a_1,a_3,a_4,a_5\})$ to $\Theta_I(\{a_1,a_3,a_4\})$ are relegated to Appendix A.2.

Notice that in this model, there exists a unique minimum data-consistent relaxation, so that Theorem 6 implies that Θ_I^* is the smallest element in Λ. Moreover, the uniqueness of the minimum data-consistent relaxation implies that we can infer the reason of why the initial model is violated from the minimum data-consistent relaxation and Θ_I^*. For instance, if $\{a_1,a_2,a_4,a_5\}$ is the minimum data-consistent relaxation, we know that A was violated because $ACDE(1) = 0$ is a conflicting statement while $ACDE(1) > 0$ is a non-conflicting statement.

4.2.3. *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 $\mathcal{Z} = \{1,...,k\}$. We are interested in the average potential outcome $\theta_d = \sum_z P(Z=z)EY_{dz}$ for $d \in \{0,1\}$.

For any $z \in \{1,...,k\}$, define assumption a_z to be the collection of the following assumptions:

E.1 for each $d \in \{0,1\}$ and any $z' \in \{1,...,k\}$, $P(Y_{dz'} \in [\,\underline{y}_d\,,\ \overline{y}_d\,]) = 1$.
E.2 for each $d \in \{0,1\}$ and any $z' \in \{1,...,k\}, E[Y_{dz'}|Z] = E[Y_{dz'}]$ almost surely.
E.3 for each $d \in \{0,1\}$, $Y_{dz'} \le Y_{dz''}$ for all $z' \le z''$, and $Y_{dz'} = Y_{dz}$ for all $z' \ge z$.

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, on one 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* is equivalent to the mean independence assumption invoked in Manski (1990), i.e., $E[Y_d|Z] = E[Y_d]$. On the other extreme, when $z = k$, *E.2* and *E.3* implies the MIV assumption introduced in Manski and Pepper (2000), i.e., $z_1 < z_2 \Rightarrow E[Y_d|Z = z_1] \leq E[Y_d|Z = z_2]$. 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' and remain flat for $z' \geq z$. We call this assumption the AMIV assumption because we allow this cut-off point z to be determined by the data. 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 context 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. See Figure 3 for an illustration.

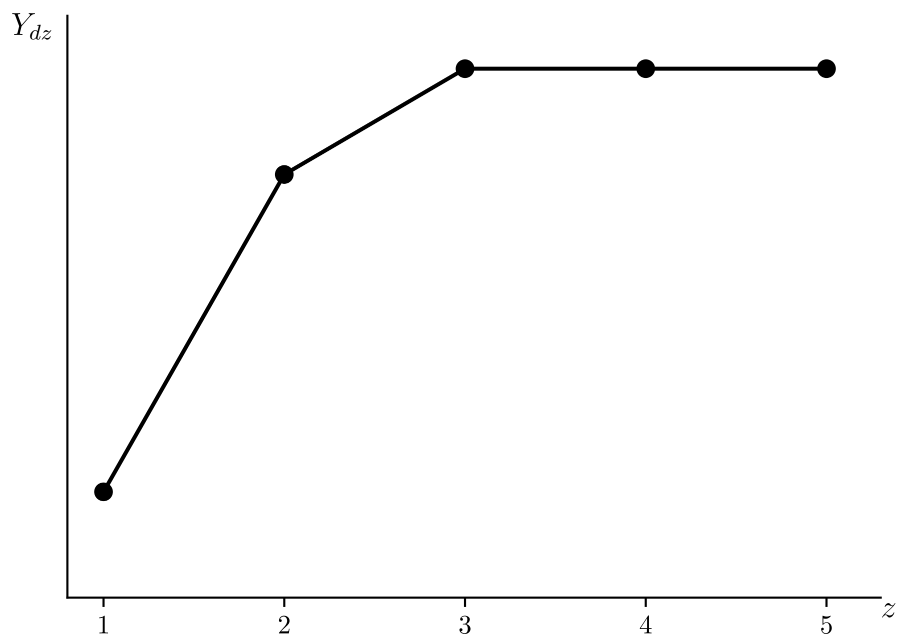


FIGURE 3. Illustration of restriction *E.3* when $z = 3$ and $k = 5$.

By construction, for all $z = 1, ..., k-1$, a_z implies a_{z+1}. In addition, define $a^\dagger$ as the collection of *E.1* and *E.2*. Let $A = \{a_1, ..., a_k, a^\dagger\}$ be the collection of all assumptions. Then, the full model A is the classic mean independence assumption considered in Manski (1990). Define $\underline{Y}_d = Y\mathbb{1}(D = d) + \underline{y}_d\mathbb{1}(D \neq d)$, $\overline{Y}_d = Y\mathbb{1}(D = d) + \overline{y}_d\mathbb{1}(D \neq d)$, $\underline{q}_{dz} = E[\underline{Y}_d|Z = z]$, and $\overline{q}_{dz} = E[\overline{Y}_d|Z = z]$. Then, we have the following result:

Proposition 2. *Assume that* $P(Y \in [\underline{y}_d, \overline{y}_d] \big| D = d) = 1$ *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* $z = 1, ..., k$, $a_z \in A^*$ *if and only if the following two conditions hold for each* $d \in \{0, 1\}$:

$$\forall z' < z, \quad \max(\underline{q}_{dt} : t \leq z') \leq \min(\overline{q}_{dt} : t \geq z')$$

and

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

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*

$$\Gamma_{d,z^*} = \left[\sum_{z < z^*} P(Z = z) \max(\underline{q}_{dt}, t \leq z) + \sum_{z \geq z^*} P(Z = z) \max(\underline{q}_{dt} : t = 1, ..., k), \right.$$
$$\left. \sum_{z < z^*} P(Z = z) \min(\overline{q}_{dt} : t \geq z) + \sum_{z \geq z^*} P(Z = z) \min(\overline{q}_{dt} : t \geq z^*) \right].$$

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

Remark 2. It is worth noting that while, for simplicity, we impose the cut-off z^* to be the same for all potential outcomes in *E.3*, we do not need to do so. We can let the data determine the cut-offs that are compatible with it and this for each potential outcome separately.

4.3. **Relation between Misspecification Robust Bound and Falsification Adaptive Set.**

Recently, Masten and Poirier (2020) has introduced an approach to salvage misspecified models. The idea is to continuously perturb a refuted full model into a model which is no longer falsified by the data. In the following, we are showing how both approaches are related.

For any $\epsilon \geq 0$ and any $a \in A$, let a_ϵ denote the assumption after relaxing and perturbating assumption a. The degree of pertubation is measured by ϵ: when $\epsilon = 0$, $a_\epsilon = a$; when $\epsilon = 1$, a_ϵ is a null assumption which does not impose any restrictions so that $\Theta_I(a_\epsilon) = \Theta$; when $\epsilon \in (0, 1)$, the assumption a is partially relaxed and the exact form of a_ϵ would depend on the specific way of perturbation which is chosen by the researcher. For any $\delta : A \to [0, 1]$, define $A(\delta) \coloneqq \{a_{\delta(a)} : a \in A\}$ as the purturbed full model. For any two $\delta_1 : A \to [0, 1]$ and $\delta_2 : A \to [0, 1]$, 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* in Masten and Poirier (2020) can be defined as $FF = \{\delta : A \to [0, 1] \ :\ \Theta_I(A(\delta)) \neq \emptyset$ 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 modified slightly the definition of the falsification frontier of Masten and Poirier (2020) to allow for non-continuous cases.[4] Then, the *falsification adaptive set* $\Theta_I^\dagger$ is defined as $\Theta_I^\dagger = \cup_{\delta \in FF} \Theta_I(A(\delta))$.

Theorem 7. *Suppose Condition (S3) in Theorem 6 hold. Then, $\Theta_I^* \subseteq \Theta_I^\dagger$ for any type of perturbation chosen by the researcher.*

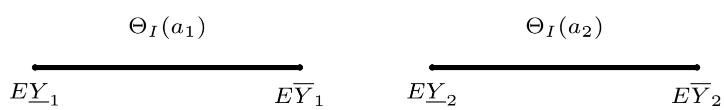
Whenever Condition (S3) holds, Theorem 7 shows that our misspecification robust bound is always included in the falsification adaptive set. Moreover, because of Theorem 5, the falsification adaptive set is also an element of Λ, and therefore carries a similar type of interpretation.

In our illustrative example case, whenever Assumption 1 holds and $P(E[\underline{Y}|Z] \leq \overline{\gamma}]) > 0$ and $P(E[\overline{Y}|Z] \geq \underline{\gamma}]) > 0$, we can prove that there exists one way of perturbation, so that our misspecification robust bounds and the falsification adaptive set coincide; more precisely, $\Theta_I^* = \Theta_I^\dagger = [\min(\underline{\gamma}, \overline{\gamma}), \max(\underline{\gamma}, \overline{\gamma})]$.

However, when Condition (S3) fails to hold, the misspecification robust bounds and the falsification adaptive set may deliver fairly different type of results depending on the type of relaxation considered in the Masten and Poirier (2020) approach.

Consider $\mathcal{A} = \{a_1, a_2\}$ where a_1 refers to the assumption that $E\underline{Y}_1 \leq \theta \leq E\overline{Y}_1$, and a_2 refers to the assumption that $E\underline{Y}_2 \leq \theta \leq E\overline{Y}_2$. Suppose $E\underline{Y}_1 \leq E\overline{Y}_1 < E\underline{Y}_2 \leq E\overline{Y}_2$. The sets $\Theta_I(a_1)$ and $\Theta_I(a_2)$ are illustrated in Figure 4. In this simple case, $\Theta_I^* = \left[E\underline{Y}_1, E\overline{Y}_1\right] \cup \left[E\underline{Y}_2, E\overline{Y}_2\right]$. The

FIGURE 4. The two-interval example



falsification adaptive set in Masten and Poirier (2020) depends on how the full model is perturbed. One way to perturb the full model is the following:

$$E\underline{Y}_1 - \delta_1 \leq \theta \leq E\overline{Y}_1 + \delta_2 \quad \text{and} \quad E\underline{Y}_2 - \delta_3 \leq \theta \leq E\overline{Y}_2 + \delta_4. \tag{4.7}$$

The falsification adaptive set $\Theta_I^\dagger$ in this example is the set of all θ which satisfy the following two conditions: (i) there exist some $\delta_1, ..., \delta_4 \in \mathbb{R}_+$ such that θ satisfies (4.7), and (ii) there does not exist $\tilde{\theta}$ and $\tilde{\delta}_1, ..., \tilde{\delta}_4 \in \mathbb{R}_+$ satisfying the following two conditions: (a) $\tilde{\delta}_i \leq \delta_i$ for all i and $\tilde{\delta}_j < \delta_j$ for some j, and that (b) (4.7) holds with $(\theta, \delta_1, ..., \delta_4)$ replaced by $(\tilde{\theta}, \tilde{\delta}_1, ..., \tilde{\delta}_4)$. One can show $\Theta_I^\dagger = [E\overline{Y}_1, E\underline{Y}_2]$.

[4]With this definition, we do not need to exclude the possibility that there is a sequence of $\{\delta_i : i \geq 1\}$ such that $\delta_n \to \delta^*$, $\Theta_I(A(\delta^*)) = \emptyset$, $\Theta_I(A(\delta_n)) = \Theta_I(A(\delta_1)) \neq \emptyset$ and $\delta_{n+1} < \delta_n$ for all $n \geq 1$.

As we can see, Θ_I^* and $\Theta_I^\dagger$ are very different and they do not intersect except at the endpoints $E\overline{Y}_1$ and $E\underline{Y}_2$. This difference is due to the different ideas behind these two constructions. When Θ_I^* is constructed, we view each a in $\mathcal{A}$ as separate assumption. When the full model is refuted, we conclude that at least one of a_1 and a_2 must be wrong. Because we do not know which one is violated and each of them is data-consistent, we report $\Theta_I(a_1) \cup \Theta_I(a_2)$. The idea behind $\Theta_I^\dagger$ is that all assumptions could potentially be wrong, and all assumptions are subject to small perturbations. Which one should an empirical researcher choose mainly depends on her understanding of why the full model is refuted.

A more practical difference between our two approaches is illustrated in the binary IV model discussed earlier in subsection 4.2.2. Our misspecification robust bound is fairly different from the falsification adaptive set proposed by Masten and Poirier (2020) for the same model. We propose a discrete relaxation of the model while they proposed a continuous one. In general, the type of relaxation depends on the empirical question under study. When the instrumental inequality are violated while the IV is known to be randomly assigned, the discrete relaxation we consider looks easier to interpret and more in line with the relaxation previously entertained in the causal inference literature, see Cai et al. (2008), Wang, Robins, and Richarson (2017), Kédagni and Mourifié (2020) and references therein.

5. An empirical illustration

5.1. **Context and Data.** Estimating the causal impact of college education on later earnings has always been troublesome for economists because of 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 IVs have been widely criticized because of their potential correlation with children 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 impose a positive dependence relationship between the IV and the 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 4.2.3. Therefore, we consider that parental education can have a positive effect on children future earnings, but this marginal positive effect could plausibly becomes 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 consists of a sample of 1,230 white males taken from the National Longitudinal Survey of Youth of 1979 (NSLY). The data contains information on the log weekly wage, college education, father's education, mother's education, among many other variables. Following HTV, we consider the college enrolment 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 the parental education as the candidate instrumental variable. Some summary statistics are reported in Table 1.

TABLE 1. Summary Statistics

	Total
Observations	1,230
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)

Average and standard deviation (in the parentheses)

5.2. **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, i.e., $\Theta_I(MI)$, and under the MIV assumption, i.e. $\Theta_I(MIV)$. In addition we construct an estimate of our misspecification robust bounds under the AMIV assumption. Recall that when z^* is equal to the lowest value in the support of Z, the AMIV equals to the MI assumption, while when z^* is equal to the highest value in the support of Z, the AMIV equals to the MIV assumption.

The results are summarized in Table 2. In column (1), we compute the 95% confidence region for the ATE under the mean independence assumption, which turns out to be empty, meaning that the data shows clear evidence against the use of parental education as a valid IV. In column (4), we consider the MIV assumption, we first test the validity of the MIV using the test proposed by Hsu, Liu, and Shi (2019), we do not reject the MIV assumption even at 10 % level, then estimates of $\Theta_I(MIV)$ are presented in column (4). As can be seen, when using the Manski and Pepper (2000) MIV relaxation, we move from an empty identification region to a wide and non-informative identification region. However, our misspecification robust bounds provide relatively smaller set estimate for the ATE. Column (3) shows estimates of our misspecification robust bounds when the same cut-off is imposed for both potential outcomes as in Proposition 2, while column (2) shows estimates when we allow the cut-off to differ from potential outcomes as discussed in Remark 2.[5] In the latter case, we see that the proposed approach almost identifies the sign of the ATE.

[5]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.

Table 2. Results

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

Conf.: confidence MI: Mean Independence, MIV: Monotone IV

6. Discussion

In this paper, we show that there could exist 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 it is possible: the identified set not only exhaust all the identification restriction in the model structure and assumption, 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 sharp characterization of the identified set is not tractable, our results suggest that empirical researchers should be more careful when working with outer sets, especially when the bounds that they get are very tight. For example, as a robustness check, one could construct the outer sets in different ways and see if there is any discordance between these outer sets.

Finally, we propose the misspecification robust bound as a way to salvage a refuted model. By construction, it is the smallest set implied by all minimum data-consistent relaxations and compatible with all data-consistent submodels. 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 which always covers the misspecification robust bound proposed in this paper. This type of outer set will be immune to the issue raised in this paper. It remains unclear how to construct such outer sets, but this could be one plausible step beyond the findings in this paper.

References

Allen, R. and J. Rehbeck. 2020. "Satisficing, Aggregation, and Quasilinear Utility." *Unpublished manuscript* .

Andrews, D. W. K., W. Kim, and X. Shi. 2017. "Stata Commands for Testing Conditional Moment Inequalities/Equalities." *Stata Journal* 17 (1):56–72.

Andrews, D. W. K. and S. Kwon. 2019. "Inference in Moment Inequality Models That Is Robust to Spurious Precision under Model Misspecification." *Working paper* .

Andrews, D. W. K. and X. Shi. 2013. "Inference Based on Conditional Moment Inequalities." *Econometrica* 81:609–666.

Artstein, Z. 1983. "Distributions of random sets and random selections." *Israel Journal of Mathematics* 46:313–324.

Aucejo, E. M., F. A. Bugni, and V. J. Hotz. 2017. "Identification and Inference on Regressions with Missing Covariate Data." *Econometric Theory* 33 (1):196–241.

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 (2):pp. 323–363.

Cai, Z., M. Kuroki, J. Pearl, and J. Tian. 2008. "Bounds on direct effects in the presence of confounded intermediate variables." *Biometrics* 64:695–701.

Chen, X., C.A. Flores, and A. Flores-Lagunes. 2018. "Going Beyond LATE: Bounding Average Treatment Effects of Job Corps Training." *Journal of Human Resources* 53 (4):1050–1099.

Chernozhukov, Victor, Sokbae Lee, and Adam M. Rosen. 2013. "Intersection Bounds: Estimation and Inference." *Econometrica* 81 (2):667–737. URL `http://dx.doi.org/10.3982/ECTA8718`.

Chesher, A. and A. M. Rosen. 2020. "Econometric Modeling of Interdependent Discrete Choice with Applications to Market Structure." *cemmap working paper CWP25/20* .

Ciliberto, F., C. Murry, and E. Tamer. 2020. "Market Structure and Competition in Airline Markets." *Working paper* .

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

de Paula, A., S. Richards-Shubik, and E. Tamer. 2018. "Identifying Preferences in Networks With Bounded Degree." *Econometrica* 86 (1):263–288.

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

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

Haile, P. A. and E. Tamer. 2003. "Inference with an Incomplete Model of English Auctions." *Journal of Political Economy* 111 (1):1–51.

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—-23.

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

Hsu, Yu-Chin, Chu-An Liu, and Xiaoxia Shi. 2019. "Testing Generalized Regression Monotonicity." *Econometric Theory, Forthcoming* 35 (6):1146–1200.

Kaido, H. and H. White. 2013. *Estimating Misspecified Moment Inequality Models.* in Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White Jr, ed. by X. Chen, and N. R. Swanson. New York, NY: Springer.

Kédagni, D. 2018. "Identifying Treatment Effects in the Presence of Confounded Types." *Economics Working Papers, Iowa State University* 18014.

Kédagni, D. and I. Mourifié. 2020. "Generalized Instrumental Inequalities: Testing the Instrumental Variable Independence Assumption." *Biometrika* 107 (3):661–675.

Kim, W., K. Kwon, S. Kwon, and S. Lee. 2018. "The identification power of smoothness assumptions in models with counterfactual outcomes." *Quantitative Economics* 9:617–642.

Kitagawa, Toru. 2009. "Identification Region of the Potential Outcome Distributions under Instrument Independence." *CeMMAP working paper CWP30/09* .

Kreider, B., J. V. Pepper, C. Gundersen, and D. Jolliffe. 2012. "Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported." *Journal of American Statistical Association* 107:958–975.

Manski, C. F. 1994. "The Selection Problem." *in Advances Economics, Sixth World Congress, C. Sims (ed.), Cambridge University Press* 1:143–170.

Manski, C. F. and J. Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica* 68:997–1010.

———. 2009. "More on Monotone Instrumental Variables." *Econometrics Journal* 12:S200–S216.

Manski, Charles F. 1990. "Nonparametric Bounds on Treatment Effects." *American Economic Reviews, Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association* 80 (2):319–323.

Masten, M. A. and A. Poirier. 2020. "Salvaging Falsified Instrumental Variable Models." *Working Paper* .

Molinari, F. 2020. "Microeconometrics with Partial Identification." *Handbook of Econometrics, Volume 7A (forthcoming)* .

Mourifié, I., M. Henry, and R. Méango. 2020. "Sharp Bounds and Testability of a Roy Model of STEM Major Choices." *Journal of Political Economy* 8 (128):3220–3283.

Pearl, J. 2001. "Direct and indirect effects." *In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence* :411–420.

Pearl, Judea. 1994. "On the Testability of Causal Models with Latent and Instrumental Variables." *Uncertainty in Artificial Intelligence* 11:435–443.

Peterson, A. V. 1976. "Bounds for a Joint Distribution Function with Fixed Sub-Distribution Functions: Application to Competing Risks." *Proceedings of the National Academy of Sciences of the United States of America* 73 (1):11–13.

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

Sheng, Shuyang. 2020. "A structural econometric analysis of network formation games through subnetworks." *Econometrica* 88 (5):1829–1858.

Wang, L., J. M. Robins, and T. S. Richarson. 2017. "On falsification of the Binary Instrumental Variable Model." *Biometrika* 104 (1):229–236.

Appendix A. Additional Results

A.1. **Conditional Moment Inequalities.** Let us now consider a more general setting than the introductory example. Assume the full model is a conditional moment inequality,

$$E[m(X,Z;\theta)|Z] \leq 0 \text{ almost surely} \tag{A.1}$$

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 for each θ. In practice, empirical researchers sometimes use the following unconditional model instead:

$$E[w(Z)m(X,Z;\theta)] \leq 0, \tag{A.2}$$

where $w(\cdot)$ is some nonnegative weighting funciton. We want to understand what would happen when one conduct empirical analysis based on (A.2) when (A.1) happens to be refuted.

To answer this question, define $\mathcal{W}_m^+$ to be the set of all m-dimensional nonnegative function w which satisfies $0 < E\|w(Z)\|^2 < \infty$ and $E\|w(Z)m(X,Z;\theta)\| < \infty$ for all $\theta \in \Theta$. Define A to be the set of all submodels indexed by $w \in \mathcal{W}_1^+$. Then, each $B \subseteq A$ with m elements corresponds to the submodel which (A.2) hold for some $w \in \mathcal{W}_m^+$. By construnction, Condition (T2.C2) are satisfied. What is left to check is whether Condition (T2.C1) and (T2.C3) are satisified.

To verify Condition (T2.C1), we need to construct a A^* so that $\Theta_I(A^*) = \emptyset$ whenever the full model is refuted and $\Theta_I(a) \neq \emptyset$ for all $a \in A^*$. Suppose for any z_0 in the support of Z, there exists some $\theta_0 \in \Theta$ and some $\delta(z_0) > 0$ such that $E[m(X,Z;\theta_0)|Z] \leq 0$ for almost every Z with $\|Z - z_0\| \leq \delta(z_0)$. Then, we know the following submodel:

$$E[\mathbb{1}(\|(Z - z_0\| < \delta(z_0))m(X,Z;\theta)] \leq 0$$

is always data-consistent. Therefore, if we define function $\delta_{z_0,\epsilon}$ as $\delta_{z_0,\epsilon}(z) = \mathbb{1}(\|z - z_0\| < \epsilon)$ and the collection of functions $\mathcal{W}^*$ as $\mathcal{W}^* := \{\delta_{z,\epsilon} : \epsilon \in [0,\delta(z)), z \in \mathcal{Z}\}$ where $\mathcal{Z}$ is the support of Z, then the A^* can be constructed as the set of submodels indexed by $w \in \mathcal{W}^*$ with which (A.2) hold. Under the regularity conditions listed in the following proposition, we can also ensure $\Theta_I(A^*) = \emptyset$ whenever $\Theta_I(A) = \emptyset$.

Proposition 3. *Let $\mathcal{Z}$ be the support of Z. Suppose*

(a) $E\|m(X,Z;\theta)\| < \infty$ for each $\theta \in \Theta$.
(b) either $\mathcal{Z}$ is discrete, or $E[m(X,Z;\theta)|Z = z]$ is continuous[6]in z.
(c) for any z_0 in the support of Z, there exists some neighborhood Ω_0 of z_0 and some $\theta_0 \in \Theta$ such that $E[m(X,Z;\theta_0)|Z] \leq 0$ for all most every Z in Ω_0.

Then, Condition (T2.C1) of Theorem 2 is satisfied.

Note that a sufficient condition for Condition (c) in the above proposition is that for any z_0 in the support, there exists some $\theta_0 \in \Theta$ satisfying $E[m(X,Z;\theta_0)|Z = z_0] < 0$.

Finally, let us verify Condition (T2.C3) of Theorem 2. Suppose Θ is compact. Then, for any $a \in A$, we can ensure $\Theta_I(a)$ is compact if $E[w(Z)m(X,Z;\theta)]$ is a continuous function for every $w \in \mathcal{W}_1^+$. A simple sufficient condition for $E[w(Z)m(X,Z;\theta)]$ to be continuous is that $E[m(X,Z;\theta)|Z]$ is continuous in θ almost surely and it is dominated by some integrable function.

Proposition 4. *Suppose*

(a) $E[m(X,Z;\theta)|Z = z]$ is continuous in θ almost surely.
(b) there exists some function $g(\cdot)$ such that $\sup_{\theta\in\Theta}\|E[m(X,Z;\theta)|Z]\| \leq g(Z)$ almost surely and $E|g(Z)|^2 < \infty$.
(c) Θ is compact.

Then, Condition (T2.C3) of Theorem 2 is satisfied.

Combine these two propositions, we have the following result as a corollary of Theorem 2.

Corollary 2. *Suppose*

(a) either $\mathcal{Z}$ is discrete, or $E[m(X,Z;\theta)|Z = z]$ is continuous in z.
(b) $E[m(X,Z;\theta)|Z = z]$ is continuous in θ almost surely.

[6]A more precise statement is that there exists some $g(z;\theta)$ such that $g(z;\theta)$ is continuous in z and $E[m(X,Z;\theta)|Z = z] = g(z;\theta)$ almost surely.

(c) for any z_0 in the support of Z, there exists some neighborhood Ω_0 of z_0 and some $\theta_0 \in \Theta$ such that $E[m(X, Z; \theta_0)|Z] \leq 0$ for all most every Z in Ω_0.
(d) there exists some function $g(\cdot)$ such that $\sup_{\theta\in\Theta} \|E[m(X, Z; \theta)|Z]\| \leq g(Z)$ almost surely and $E|g(Z)|^2 < \infty$.
(e) Θ is compact.

Then, (A.1) *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 the submodel* (A.2) *with $w = w_1$ and the submodel* (A.2) *with $w = w_2$ are not refuted but the identified sets of these two submodels have empty intersection.*

Moreover, whenever (A.1) *is refuted, for any $\tilde{w} \in \mathcal{W}^+_{m}$ with which submodel* (A.2) *is refuted, there exists some $w_1 \in \mathcal{W}^+_{m_1}$ and $w_2 \in \mathcal{W}^+_{m_2}$ such that both the submodel* (A.2) *with $w = (\tilde{w}, w_1)$ and the submodel* (A.2) *with $w = w_2$ are data-consistent but the identified sets of these two submodels 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 (A.1). Their inference transform (A.1) into (A.2) by selecting w in a sub-family of $\cup_{m\geq 1}\mathcal{W}^+_m$ and letting $m \to \infty$ as the sample size increases. Our result shows that increasing m to infinity is crutial to ensure the robustness of the result if (A.1) could be misspecified. If the dimension of w is fixed, then the empirical result for (A.2) could be misleading even if the inference controls the size uniformly. Since a finite m is often used in the finite sample, our result also raise the question how to interpret the result of a test under possible model misspecification.

A.2. **Binary IV example continued.**

$$\Theta_I(\{a_1, a_3, a_4, a_5\}) = \left\{ \theta : \begin{cases} q_{11}(0) \leq \theta_{10} \leq 1 - q_{01}(0), \\ q_{11}(1) \leq \theta_{11} \leq 1 - q_{01}(1), \\ \theta_{11} - \theta_{10} \leq \min\{0, 1 - q_{01}(1) - q_{11}(0)\}, \\ \sup_z q_{10}(z) \leq \theta_{01} = \theta_{00} \leq 1 - \sup_z q_{00}(z). \end{cases} \right\}$$

$$\Theta_I(\{a_1, a_2, a_3, a_4\}) = \left\{ \theta : \begin{cases} q_{10}(0) \leq \theta_{00} \leq 1 - q_{00}(0), \\ q_{10}(1) \leq \theta_{01} \leq 1 - q_{00}(1), \\ \max\{0, q_{10}(1) + q_{00}(0) - 1\} \leq \theta_{01} - \theta_{00} \\ \sup_z q_{11}(z) \leq \theta_{10} = \theta_{11} \leq 1 - \sup_z q_{01}(z). \end{cases} \right\}$$

$$\Theta_I(\{a_1, a_2, a_3, a_5\}) = \left\{ \theta : \begin{cases} q_{10}(0) \leq \theta_{00} \leq 1 - q_{00}(0), \\ q_{10}(1) \leq \theta_{01} \leq 1 - q_{00}(1), \\ \theta_{01} - \theta_{00} \leq \min\{0, 1 - q_{00}(1) - q_{10}(0)\}, \\ \sup_z q_{11}(z) \leq \theta_{10} = \theta_{11} \leq 1 - \sup_z q_{01}(z). \end{cases} \right\}$$

$$\Theta_I(\{a_1, a_2, a_5\}) = \left\{ \theta : \begin{cases} q_{11}(0) \leq \theta_{10} \leq 1 - q_{01}(0), \\ q_{11}(1) \leq \theta_{11} \leq 1 - q_{01}(1), \\ \max\{0, q_{11}(1) + q_{01}(0) - 1\} \leq \theta_{11} - \theta_{10} \leq 1 - q_{01}(1) - q_{11}(0), \\ q_{10}(0) \leq \theta_{00} \leq 1 - q_{00}(0), \\ q_{10}(1) \leq \theta_{01} \leq 1 - q_{00}(1), \\ \theta_{01} - \theta_{00} \leq \min\{0, 1 - q_{00}(1) - q_{11}(0)\}. \end{cases} \right\}$$

$$\Theta_I(\{a_1, a_2, a_4\}) = \left\{ \theta : \begin{cases} q_{11}(0) \leq \theta_{10} \leq 1 - q_{01}(0), \\ q_{11}(1) \leq \theta_{11} \leq 1 - q_{01}(1), \\ \max\{q_{11}(1) + q_{01}(0) - 1, 0\} \leq \theta_{11} - \theta_{10}, \\ q_{10}(0) \leq \theta_{00} \leq 1 - q_{00}(0), \\ q_{10}(1) \leq \theta_{01} \leq 1 - q_{00}(1), \\ \max\{0, q_{10}(1) + q_{00}(0) - 1\} \leq \theta_{01} - \theta_{00}. \end{cases} \right\}$$

$$\Theta_I(\{a_1, a_3, a_5\}) = \left\{ \theta : \begin{cases} q_{11}(0) \leq \theta_{10} \leq 1 - q_{01}(0), \\ q_{11}(1) \leq \theta_{11} \leq 1 - q_{01}(1), \\ \theta_{11} - \theta_{10} \leq \min\{1 - q_{01}(1) - q_{11}(0), 0\}, \\ q_{10}(0) \leq \theta_{00} \leq 1 - q_{00}(0), \\ q_{10}(1) \leq \theta_{01} \leq 1 - q_{00}(1), \\ \theta_{01} - \theta_{00} \leq \min\{1 - q_{00}(1) - q_{11}(0), 0\}. \end{cases} \right\}, \Theta_I(\{a_1, a_3, a_4\}) = \left\{ \theta : \begin{cases} q_{11}(0) \leq \theta_{10} \leq 1 - q_{01}(0), \\ q_{11}(1) \leq \theta_{11} \leq 1 - q_{01}(1), \\ \theta_{11} - \theta_{10} \leq \min\{1 - q_{01}(1) - q_{11}(0), 0\}, \\ q_{10}(0) \leq \theta_{00} \leq 1 - q_{00}(0), \\ q_{10}(1) \leq \theta_{01} \leq 1 - q_{00}(1), \\ \max\{0, q_{10}(1) + q_{00}(0) - 1\} \leq \theta_{01} - \theta_{00}. \end{cases} \right\}$$

Appendix B. Proof of the main results

B.1. **Proof of Theorem 1.** Theorem 1 is an immediate result of the following two lemmas.

Lemma 2. *Suppose Assumption 1 hold and $\overline{\gamma} < \underline{\gamma}$. Define the interval $\mathcal{W}$ as the following:*

$$\mathcal{W} := \begin{cases} [\overline{\gamma}, \underline{\gamma}] & \text{if } P(E[\overline{Y}|Z] = \overline{\gamma}) > 0 \text{ and } P(E[\underline{Y}|Z] = \underline{\gamma}) > 0 \\ [\overline{\gamma}, \underline{\gamma}) & \text{if } P(E[\overline{Y}|Z] = \overline{\gamma}) > 0 \text{ and } P(E[\underline{Y}|Z] = \underline{\gamma}) = 0 \\ (\overline{\gamma}, \underline{\gamma}] & \text{if } P(E[\overline{Y}|Z] = \overline{\gamma}) = 0 \text{ and } P(E[\underline{Y}|Z] = \underline{\gamma}) > 0 \\ (\overline{\gamma}, \underline{\gamma}) & \text{if } P(E[\overline{Y}|Z] = \overline{\gamma}) = 0 \text{ and } P(E[\underline{Y}|Z] = \underline{\gamma}) = 0 \end{cases} \tag{B.1}$$

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

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

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

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

Since $E[\underline{Y}|Z]-\underline{\theta} \leq -\delta$ and $h_{i'}$ 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 $P(E[\overline{Y}|Z] = \overline{\gamma}) = 0$. These result then implies that $\widetilde{\Theta}(h) \cap \mathcal{W} \neq \emptyset$ whenever $\widetilde{\Theta}(h) \neq \emptyset$. □

Lemma 3. *Suppose Assumption 1 hold and $\overline{\gamma} < \underline{\gamma}$. Let $\mathcal{W}$ be the interval defined as in* (B.1)*. Then, for any $\theta \in \mathcal{W}$, there exists some $h \in \mathcal{H}_2^+$ such that $\widetilde{\Theta}(h) = \{\theta\}$.*

Proof of Lemma 3. 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 $P(E[\overline{Y}|Z] = \overline{\gamma}) > 0$, for any $\vartheta \geq \overline{\gamma}$, we also have $P(\vartheta \geq E[\overline{Y}|Z]) > 0$. Since $\theta \in \mathcal{W}$, we conclude that $P(Z \in \overline{S}^-) > 0$. Similarly, that $\theta \in \mathcal{W}$ 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

$$\begin{aligned} & E[h_1(Z)(\tilde{\theta} - \underline{Y})] \geq 0 \\ \Leftrightarrow \quad & E[h_1(Z)]\tilde{\theta} \geq E[h_1(Z)\underline{Y}] \\ \Leftrightarrow \quad & \tilde{\theta} \geq \theta \end{aligned}$$

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

$$\begin{aligned} E[h_2(Z)\theta] &= \theta \\ &= E[h_2(Z)\overline{Y}] \\ &\geq E[h_2(Z)\underline{Y}] \end{aligned}$$

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.2. **Proof of Theorem 2.** To prove Theorem 2, we need the following lemma.

Lemma 4. *Suppose Θ is a subset in a metric space. Suppose also that at least one of the following conditions hold:*

(1) A is a finite set.
(2) For any $a \in A$, $\Theta_I(a)$ is compact. Moreover, for any $B \subseteq A$, $\Theta_I(B) = \cap_{a\in B}\Theta_I(a)$.

Then, for any submodel $A_0 \subseteq A$ with $\Theta_I(A_0) \neq \emptyset$, there exists some $\tilde{A} \subseteq A$ such that (i) $A_0 \subseteq \tilde{A}$, (ii) $\Theta_I(\tilde{A}) \neq \emptyset$, (iii) for any $a \in A\backslash\tilde{A}$, $\Theta_I(\tilde{A} \cup \{a\}) = \emptyset$.

Proof of Lemma 4. Let A_0 be an arbitrary subset of A with $\Theta_I(A_0) \neq \emptyset$. If $\Theta_I(A) \neq \emptyset$, then we can let $\tilde{A} = A$ and the results hold trivially. In the following, we focus on cases where $\Theta_I(A) = \emptyset$.

Suppose A is finite. Then, A_0 is also finite, enumerate it as $A_0 = \{a_1, ..., a_k\}$. Moreover, $\{a \in A : \Theta_I(a) \neq \emptyset\}$ is also finite and enumerate it as $\{a_1, ..., a_k, a_{k+1}, ..., a_n\}$. Construct B_i, $i = k, k+1, ..., n$ recursively as follows. Let $B_k = A_0$. For any $i \geq k$, let B_{i+1} equal B_i if $\Theta_I(B_i \cup \{a_{i+1}\}) = \emptyset$ and equal $B_i \cup \{a_{i+1}\}$ if otherwise. By construction, $A_0 \subseteq B_i$ and $\Theta_I(B_i) \neq \emptyset$ for each $i = k, ..., n$. Moreover, for any $a \in A\backslash B_n$, either $\Theta_I(a) = \emptyset$ which implies that $\Theta_I(B_n \cup \{a\}) = \emptyset$, or $\Theta_I(a) \neq \emptyset$ which implies that there must exist some $i \in \{k, ..., n\}$ with $\Theta_I(B_i \cup \{a\}) = \emptyset$ so that $\Theta_I(B_n \cup \{a\}) = \emptyset$. In other words, $\Theta_I(B_n \cup \{a\}) = \emptyset$ for any $a \in A\backslash B_n$. Hence, let $\tilde{A} = B_n$ and we get the desired result.

Suppose A is not finite. Define $\mathscr{A} = \{A' \subseteq A : A_0 \subseteq A' \text{ and } \Theta_I(A') \neq \emptyset\}$. $\mathscr{A}$ is not empty because $A_0 \in \mathscr{A}$. By the construction of $\mathscr{A}$, there exists a submodel $\tilde{A}$ which satisfies all three desired conditions in the lemma if and only if $\mathscr{A}$ has a maximum element in terms of partial order $\subseteq$, i.e. there exists some $\tilde{A} \in \mathscr{A}$ such that there does not exist an $A' \in \mathscr{A}$ with $\tilde{A} \subsetneq A'$. To show $\mathscr{A}$ does have such a maximum element, we are going to apply Zorn's lemma. Let $\mathscr{Z}$ be an arbitrary nonempty chain in terms of $\subseteq$, i.e. let $\mathscr{Z}$ be an arbitrary nonempty subset of $\mathscr{A}$ such that for any A' and A'' in $\mathscr{Z}$, either $A'' \subseteq A'$ or $A' \subseteq A''$. Define $A^\dagger = \cup_{A'\in\mathscr{Z}}A'$. Then, $\Theta_I(A^\dagger) = \cap_{A'\in\mathscr{Z}}\Theta_I(A')$.

We claim $\Theta_I(A^\dagger)$ is a compact set. To see why this is so, note that $\Theta_I(A^\dagger) = \cap_{a\in A^\dagger}\Theta_I(a)$. Since for any $a \in A$, $\Theta_I(a)$ is compact and since Θ is a metric space, $\Theta_I(A^\dagger)$ is also compact. Similarly, we can prove that, for any $A' \in \mathcal{Z}$, $\Theta_I(A')$ is a compact set.

We further claim that $\Theta_I(A^\dagger)$ is nonempty. We prove this claim by contradiction. Suppose $\Theta_I(A^\dagger)$ is empty, then $\Theta = (\Theta_I(A^\dagger))^C = \cup_{A'\in\mathcal{Z}}\Theta_I(A')^C$ by De Morgan's laws. For any $A' \in \mathcal{Z}$, because $\Theta_I(A')$ is compact, $\Theta_I(A')^C$ is an open set. Fix an arbitrary element B in $\mathcal{Z}$. Then, $\Theta_I(B) \subseteq \cup_{A'\in\mathcal{Z}}\Theta_I(A')^C$. Because $\Theta_I(B)$ is compact and nonempty, there exists a finite number of $A_1, ..., A_K$ such that $\Theta_I(B) \subseteq \cup_{k\in\{1,...,K\}}\Theta_I(A_k)^C$. This implies that $\Theta_I(B) \cap \Theta_I(A_1) \cap \cdots \cap \Theta_I(A_K)$ is empty, which contradicts to the fact that $\mathcal{Z}$ is a chain in terms of $\subseteq$ and that $\Theta_I(A')$ is nonempty for each $A' \in \mathscr{A}$.

Because $\Theta_I(A^\dagger)$ is nonempty, $A^\dagger \in \mathscr{A}$. Moreover, by the construction of $A^\dagger$, $A' \subseteq A^\dagger$ for any $A' \in \mathscr{Z}$. By Zorn's lemma, we conclude that $\mathscr{A}$ contains a maximum element with partial order $\subseteq$. This completes the proof. □

Proof of Theorem 2. Whenever there exists two submodels A_1 and A_2 with $\Theta_I(A_1) \neq \emptyset$, $\Theta_I(A_2) \neq \emptyset$ and $\Theta_I(A_1) \cap \Theta_I(A_2) = \emptyset$, $\Theta_I(A_1 \cup A_2) \subseteq \Theta_I(A_1) \cap \Theta_I(A_2)$ implies that $\Theta_I(A_1 \cup A_2) = \emptyset$ so that $\Theta_I(A) = \emptyset$. Moreover, the existence of A^* when $\Theta_I(A) = \emptyset$ implies that there always exists some nonempty submodel $B \subseteq A$ with $\Theta_I(B) \neq \emptyset$. Therefore, to prove all the results in Theorem 2, we only need to prove that when $\Theta_I(A) = \emptyset$, for any nonempty $B \subseteq A$ with $\Theta_I(B) \neq \emptyset$, there exists two finite set B' and B'' such that $\Theta_I(B \cup B') \neq \emptyset$, $\Theta_I(B'') \neq \emptyset$ and $\Theta_I(B \cup B') \cap \Theta_I(B'') = \emptyset$.

Now, suppose $\Theta_I(A) = \emptyset$ and let B be an arbitrary submodel with $\emptyset \neq B \subseteq A$ and $\Theta_I(B) \neq \emptyset$. Lemma 4 implies that there exists some $\tilde{A} \subseteq A$ such that *(i)* $B \subseteq \tilde{A}$, *(ii)* $\Theta_I(\tilde{A}) \neq \emptyset$, *(iii)* for any $a \in A\backslash\tilde{A}$, $\Theta_I(\tilde{A} \cup \{a\}) = \emptyset$. Since $\cap_{a\in A^*}\Theta_I(a) = \emptyset$, there must exists some $a^* \in A^*$ such that $\Theta_I(a^*) \cap \Theta_I(\tilde{A}) \subsetneq \Theta_I(\tilde{A})$, because otherwise $\Theta_I(\tilde{A}) \subseteq \cap_{a\in A^*}\Theta_I(a)$ which conflicts to $\cap_{a\in A^*}\Theta_I(a) = \emptyset$. Becuase $\Theta_I(a^*) \cap \Theta_I(\tilde{A}) \subsetneq \Theta_I(\tilde{A})$ implies that $a \notin \tilde{A}$, the construction of $\tilde{A}$ implies that $\Theta_I(\{a^*\} \cup \tilde{A}) = \emptyset$. Under Condition (T2.C2), we know $\Theta_I(a^*) \cap \Theta_I(\tilde{A}) = \emptyset$.

Suppose A is a finite set. Then, $\tilde{A}\backslash B$ is also a finite set. Therefore, we get the desired result with $B' = \tilde{A}\backslash B$, and $B'' = \{a^*\}$.

Suppose A is an infinite set. Then, Condition (T2.C3) hold in this case. Define $\mathcal{T} \coloneqq \{\Theta_I(B \cup \{a, a^*\}) : a \in \tilde{A}\}$. Then, $\mathcal{T}$ is not empty because $\Theta_I(B \cup \{a^*\}) \in \mathcal{T}$ and $B \subseteq \tilde{A}$. Because $\Theta_I(a^*) \cap \Theta_I(\tilde{A}) = \emptyset$, and because Condition (T2.C2) hold, $\cap_{S\in\mathcal{T}}S = \Theta_I(\tilde{A} \cup \{a^*\}) = \emptyset$. Hence, $\Theta_I(B) \subseteq (\cap_{S\in\mathcal{T}}S)^C$ so that $\Theta_I(B) \subseteq \cup_{S\in\mathcal{T}}S^C$. Because any $S \in \mathcal{T}$ is compact under Condition (T2.C3) and hence closed, $\{S^C : S \in \mathcal{T}\}$ is an open covering for $\Theta_I(B)$. Because $\Theta_I(B)$ is compact under Condition (T2.C3), there must exist a finite number of $S_1, ..., S_n \in \mathcal{T}$ such that $\Theta_I(B) \subseteq S_1^C \cup ... \cup S_n^C = (S_1 \cap ... \cap S_n)^C$, or equivalently, $\Theta_I(B) \cap S_1 \cap ... \cap S_n = \emptyset$. This implies that there exist $a_1, a_2, ..., a_n \in \tilde{A}$ such that $\Theta_I(B) \cap \Theta_I(B \cup \{a^*, a_1\}) \cap ... \cap \Theta_I(B \cup \{a^*, a_n\}) = \emptyset$. Under Condition (T2.C2) this is equivalent to $\Theta_I(B \cup \{a_1, ..., a_n\}) \cap \Theta_I(a^*) = \emptyset$.

Since $a_1, ..., a_n \in \tilde{A}$, we know $\Theta_I(B \cup \{a_1, ..., a_n\}) \neq \emptyset$. Therefore, we get the desired result with $B' = \{a_1, ..., a_n\}$ and $B'' = \{a^*\}$. $\square$

B.2.1. *When Theorem 2 fails to hold.* In this section, we present four counterexamples to show the result in Theorem 2 could fail if any of its four conditions are violated.

- Suppose only Condition (T2.C1) is violated. Consider the following example, where $\mathcal{A} = \{a_1, a_2, a_3\}$ with $\Theta_I(a_1) \subseteq \Theta_I(a_2)$, $\Theta_I(a_1) \neq \emptyset$ and $\Theta_I(a_3) = \emptyset$. Then, the full model A is refuted, but A only has three data-consistent submodels, namely, $\{a_1\}$, $\{a_2\}$ and $\{a_1, a_2\}$. And, $\Theta_I(a_1) \cap \Theta_I(a_2) \cap \Theta_I(\{a_1, a_2\}) = \Theta_I(a_1) \neq \emptyset$.
- Suppose only Condition (T2.C2) is violated. Consider the following example. Let ϵ be some unobservable random variable. Assumption a_1 assumes $E[\epsilon^2] = 1$ and assumption a_2 assumes $E[\epsilon^2] = 2$. Let θ be the variance of ϵ. If $\mathcal{A} = \{a_1, a_2\}$, we know A will always be refuted, since a_1 and a_2 are never compatible, but $\Theta_I(a_1) \cap \Theta_I(a_2) = [0, 1]$.
- Suppose only Condition (T2.C3) is violated and A is infinite. Consider the following example, let $\mathcal{A} = \{a_1, a_2, ...\}$ with $\Theta_I(a_i) = (0, 1/i)$. Then, $\Theta_I(A) = \cap_{i \geq 1} \Theta_I(a_i) = \emptyset$, but for any $i \neq j$, $\Theta_I(a_i) \cap \Theta_I(a_j) = (0, 1/\max(i, j)) \neq \emptyset$. Since the identified set for each a_i is nested, we conclude that for any A_1, A_2 with $\Theta_I(A_1) \neq \emptyset$, $\Theta_I(A_2) \neq \emptyset$, we would have $\Theta_I(A_1) \cap \Theta_I(A_2) \neq \emptyset$.

B.2.2. *An alternative version of Theorem 2.*

Theorem 2$'$. *Suppose A is a finite set. In addition, assume*

(1) whenever $\Theta_I(A) = \emptyset$, there exists a nonempty subset A^* of A such that $\Theta_I(A^*) = \emptyset$ and $\Theta_I(a) \neq \emptyset$ for all $a \in A^*$.
(2) for any A', $A'' \subseteq A$, $\Theta_I(A') \cap \Theta_I(A'') \neq \emptyset$ implies $\Theta_I(A' \cup A'') \neq \emptyset$.

Then,

(a) when $\Theta_I(A) = \emptyset$, for any $B \subseteq A$ with $\Theta_I(B) \neq \emptyset$, there exists some B' and $B'' \subseteq A$ such that $B \subseteq B'$, $\Theta_I(B') \neq \emptyset$, $\Theta_I(B'') \neq \emptyset$ and $\Theta_I(B') \cap \Theta_I(B'') = \emptyset$.
(b) $\Theta_I(A) = \emptyset$ if and only if there exists A' and $A'' \subseteq A$ such that $\Theta_I(A') \neq \emptyset$, $\Theta_I(A'') \neq \emptyset$ and $\Theta_I(A') \cap \Theta_I(A'') = \emptyset$.

Proof of Theorem 2$'$. Let us first prove the first part of the theorem. Suppose $\Theta_I(A) = \emptyset$. Fix an arbitrary $B \subseteq A$ with $\Theta_I(B) \neq \emptyset$. Suppose there exists some $a \in A^*$ such that $\Theta_I(B) \cap \Theta_I(a) = \emptyset$, then we are done. Suppose, instead, that $\Theta_I(B) \cap \Theta_I(a) \neq \emptyset$ for any $a \in A^*$, define $\mathscr{A} = \{B \cup \{a\} : a \in A^*\}$. Since $\Theta_I(B) \cap \Theta_I(a) \neq \emptyset$ for any $a \in A^*$, we know $\Theta_I(B') \neq \emptyset$ for any $B' \in \mathscr{A}$. Moreover, $\mathscr{A}$ is also a finite set and $\Theta_I(\cup_{B' \in \mathscr{A}} B') = \emptyset$. Enumerate $\mathscr{A} = \{B_1, ..., B_n\}$ and define $\tilde{B}_k = \cup_{i=1,...,k} B_i$. Then, $\Theta_I(\tilde{B}_1) \neq \emptyset$ and $\Theta_I(\tilde{B}_n) = \emptyset$. Define $k^* = \min\{k : \Theta_I(\tilde{B}_k) = \emptyset\}$. Note $1 < k^* \leq n$. and Define $B' = \tilde{B}_{k^*-1}$ and $B'' = B_{k^*}$. Then, $\Theta_I(B') \neq \emptyset$ and $\Theta_I(B'') \neq \emptyset$. Moreover, $\Theta_I(B' \cup B'') = \emptyset$ implies that $\Theta_I(B') \cap \Theta_I(B'') = \emptyset$. Finally, by construction, $B \subseteq B'$, which completes the proof for the first part.

Let us now prove the second part of the theorem. If $\Theta_I(A) = \emptyset$, we know from the first result that there must exists some A' and $A'' \subseteq A$ such that $\Theta_I(A') \neq \emptyset$, $\Theta_I(A'') \neq \emptyset$ and $\Theta_I(A') \cap \Theta_I(A'') = \emptyset$. Reversely, if there exists some A' and $A'' \subseteq A$ such that $\Theta_I(A') \neq \emptyset$, $\Theta_I(A'') \neq \emptyset$ and $\Theta_I(A') \cap \Theta_I(A'') = \emptyset$, then we must have $\Theta_I(A' \cup A'') = \emptyset$ which implies that $\Theta_I(A) = \emptyset$. $\square$

B.3. **Proof for Theorem 3.** Suppose for any nonempty $B \subseteq A$, $\Theta_I(B) = \emptyset$. Recall that $\Theta_I(\emptyset) = \Theta$. Then, the empty set is the only minimum data-consistent relaxation of A and we are done.

Suppose there exists some $B \subseteq A$ such that $\Theta_I(B) \neq \emptyset$. Then, the desired result follows from Lemma 4.

B.4. **Proof of Theorem 4.** When $\Theta_I(A) \neq \emptyset$, A is a minimum data-consistent relaxation. Moreover, fix an arbitrary minimum data-consistent relaxation A'. Then, the definition of A' assumes that for any $A \in A \backslash A'$, $\Theta_I(A' \cup \{a\}) = \emptyset$. Since $\Theta_I(A) \neq \emptyset$, we conclude that $A = A'$. Hence, A is the unique minimum data-consistent relaxation. By the definition of Θ_I^*, we have $\Theta_I^* = \Theta_I(A)$.

B.5. **Proof of Theorem 5.** Let us first prove the first result. To prove $\Theta_I^* \in \Lambda$, we need to check (C1) and (C2) for Θ_I^*. To prove (C1) for Θ_I^*, note that there exists some nonempty minimum data-consistent relaxation A^* by Theorem 3. Then, $\Theta_I(A^*) \neq \emptyset$ and $\Theta_I(A^*) \subseteq \Theta_I^*$. To prove (C2) for Θ_I^*, note that by Theorem 3, for any $A' \subseteq A$ with $\Theta_I(A') \neq \emptyset$, there exists some minimum data-consistent relaxation A^* such that $A' \subseteq A^*$. Therefore, $\Theta_I(A^*) \subseteq \Theta_I(A') \cap \Theta_I^*$ so that $\Theta_I(A') \cap \Theta_I^* \neq \emptyset$. This proves that $\Theta_I^* \in \Lambda$. Finally, for any S with $\Theta_I^* \subseteq S$, the fact that (C1) and (C2) holds for Θ_I^* implies that (C1) and (C2) also hold for S.

B.6. **Proof of Theorem 6.** We first prove (S3) $\Rightarrow$ (S1) $\Rightarrow$ (S2). Then, we prove (S2) $\Rightarrow$ (S3) under Condition (T6.C3).

Prove (S3) $\Rightarrow$ (S1): To show Θ_I^* is the smallest element in Λ, note that $\Theta_I^* \in \Lambda$ by Theorem 5. We only need to show that for any $S \in \Lambda$, $\Theta_I^* \subseteq S$.

Suppose there exists a unique minimum data-consistent relaxation A^*. Then, $\Theta_I^* = \Theta_I(A^*)$ by the definition of Θ_I^*. For any $S \in \Lambda$, there must exists some A' such that $\Theta_I(A') \subseteq S$ and $\Theta_I(A') \neq \emptyset$. By Theorem 3, we know $A' \subseteq A^*$ so that $\Theta_I(A^*) \subseteq \Theta(A') \subseteq S$. Hence, $\Theta_I^* \subseteq S$ for any $S \in \Lambda$.

Suppose for any minimum data-consistent relaxation $\widetilde{A}$, $\Theta_I(\widetilde{A})$ is a singleton set. For any $S \in \Lambda$, because (C2) holds for S, for any minimal data-consistent relaxation $\widetilde{A}$, we have $\Theta_I(\widetilde{A}) \cap S \neq \emptyset$. Since $\Theta_I(\widetilde{A})$ is a singleton set, this implies that $\Theta_I(\widetilde{A}) \subseteq S$ for all minimal data-consistent relaxation $\widetilde{A}$. Hence, $\Theta_I^* \subseteq S$.

Prove (S1) $\Rightarrow$ (S2): This result is immediate. If Θ_I^* is the smallest element in Λ, Λ of course has a smallest element.

Prove (S2) $\Rightarrow$ (S3) under Condition (T6.C3): To show (S2) $\Rightarrow$ (S3), we only need to show that when there exists at least two minimum data-consistent relaxations, (S2) implies that for any minimum data-consistent relaxation $\widetilde{A}$, $\Theta_I(\widetilde{A})$ is a singleton set.

To show this result, we need the following two lemmas, the proof of which will be shown later in Section B.6.1 and B.6.2.

Lemma 5. *Let $S_1, ..., S_n$ be a finite number of sets which satisfies: (i) for any $i = 1, ..., n$, $S_i \neq \emptyset$; and (ii) for any $i, j = 1, ..., n$ with $i \neq j$, $S_i \nsubseteq S_j$ and $S_j \nsubseteq S_i$. Define $T = \cup_{i=1}^n S_i$ and define $\mathcal{W} = \{S :$ the following conditions hold for S, (i) $\exists i \in \{1, ..., n\}$ such that $S_i \subseteq S$; (ii) $\forall j \in \{1, ..., n\},\ S_j \cap S \neq \emptyset\}$. Let $W = \cap_{S \in \mathcal{W}} S$. Then, we have the following results:*

- *(1) If $n = 1$, $W \in \mathcal{W}$.*
- *(2) If $n \geq 2$, $W \in \mathcal{W}$ if and only if S_i is singleton for all $i = 1, ..., n$.*
- *(3) $W \in \mathcal{W}$ implies $W = T$.*

Lemma 6. *Let $\mathscr{T}$ be a nonempty collection of sets which satisfies (i) for any $S \in \mathscr{T}$, $S \neq \emptyset$; and (ii) for any two $S, S' \in \mathscr{T}$, $S \cap S' = \emptyset$. Define $T = \cup_{S \in \mathscr{T}} S$ and define $\mathcal{W} = \{S :$ the following conditions hold for S, (i) $\exists S' \in \mathscr{T}$ such that $S' \subseteq S$; (ii) $\forall S' \in \mathscr{T},\ S' \cap S \neq \emptyset\}$. Let $W = \cap_{S \in \mathcal{W}} S$. Then, we have the following results:*

- *(1) If $\mathscr{T}$ only contains one set, then $W \in \mathcal{W}$.*
- *(2) If $\mathscr{T}$ contains at least two sets, then $W \in \mathcal{W}$ if and only if S is singleton for all $S \in \mathscr{T}$.*
- *(3) $W \in \mathcal{W}$ implies $W = T$.*

Suppose (S2) hold, i.e. suppose $\cap_{S \in \Lambda} S \in \Lambda$. Suppose also that there exists at least two minimum data-consistent relaxations. We are going to use the above two lemmas to show that for any minimum data-consistent relaxation $\widetilde{A}$, $\Theta_I(\widetilde{A})$ is a singleton set.

- When A is a finite set, we can enumerate all the minimum data-consistent relaxation as A_1, ..., A_n with $n \geq 2$. Define $S_i = \Theta_I(A_i)$ for $i = 1, ..., n$. By the definition of minimum data-consistent relaxation, $S_i \neq \emptyset$ for each $i = 1, ..., n$. Moreover, Condition (T6.C3) implies that there does not exists $i, j \in \{1, 2, ..., n\}$ with $i \neq j$ such that $S_i \subseteq S_j$, because, if otherwise, $\Theta_I(A_i \cup A_j) \neq \emptyset$ so that A_i and A_j will not be minimum data-consistent relaxations. Define $\mathcal{W} = \{S :$ the following conditions hold for S, (i) $\exists i \in \{1, ..., n\}$ such that $S_i \subseteq S$; (ii) $\forall j \in \{1, ..., n\},\ S_j \cap S \neq \emptyset\}$. Then, $\mathcal{W} = \Lambda$ due to the results in Theorem 3. Because $n \geq 2$, the second result in Lemma 5 implies that S_i is singleton for each $i = 1, ..., n$.
- Suppose A is an infinite set. Because we assume all the conditions in Theorem 3, we know that for any $a \in A$, $\Theta_I(a)$ is compact. Moreover, for any $B \subseteq A$, $\Theta_I(B) = \cap_{a \in B} \Theta_I(a)$. Then, for any two minimum data-consistent relaxation A' and A'', we must have $\Theta_I(A') \cap \Theta_I(A'') = \emptyset$, because, if otherwise, $\Theta_I(A' \cup A'') = \Theta_I(A') \cap \Theta_I(A'') \neq \emptyset$ so that A' and A'' will not be minimum data-consistent relaxations. Define $\mathscr{T} = \{\Theta_I(A') : A'$ is a minimum data-consistent relaxation. $\}$ and define $\mathcal{W} = \{S :$ the following conditions hold for S, (i) $\exists S' \in \mathscr{T}$ such that $S' \subseteq S$; (ii) $\forall S' \in \mathscr{T},\ S' \cap S \neq \emptyset\}$. Then, $\mathcal{W} = \Lambda$ due to the results in Theorem 3. Because $\mathscr{T}$ has at least two elements, The second result in Lemma 6 implies that all sets in $\mathscr{T}$ are singleton sets.

Therefore, in both cases, we have shown that for any minimum data-consistent relaxation $\widetilde{A}$, $\Theta_I(\widetilde{A})$ is a singleton set.

B.6.1. *Proof of Lemma 5.* The proof of Lemma 5 builds on the following lemma, whose proof will be provided later in this section.

Lemma 7. *Let $S_1, ..., S_n$ be a finite number of sets which satisfies: (i) for any $i = 1, ..., n$, $S_i \neq \emptyset$; and (ii) for any $i, j = 1, ..., n$ with $i \neq j$, $S_i \nsubseteq S_j$ and $S_j \nsubseteq S_i$, and (iii) there exists $i = 1, ..., n$ such that S_i contains at least two* *elements. Define* $\mathcal{B} = \{S : \forall i = 1, ..., n,\ S \cap S_i \neq \emptyset\}$. *We say* S *is a minimal element in* $\mathcal{B}$ *if* $S \in \mathcal{B}$ *and for any* $S' \subsetneq S$, $S' \notin \mathcal{B}$. *Then,* $\mathcal{B}$ *has at least two different minimal elements.*

Proof of Lemma 5. First of all, note that $T \in \mathcal{W}$ by the definition of $\mathcal{W}$.

Suppose $n = 1$. Then, $T = S_1$ by the definition of T. Moreover, $S_1 \in \mathcal{W}$ by the definition of $\mathcal{W}$. Finally, for any $S \in \mathcal{W}$, we must have $S_1 \subseteq S$ by the first condition in the definition of $\mathcal{W}$. Hence, $S_1 = W$ and $W = T$.

Suppose $n \geq 2$. We are going to show that, for the following statements:

(L7.S1) for each $i = 1, ..., n$, S_i is singleton.
(L7.S2) $W \in \mathcal{W}$
(L7.S3) $W = T$

the following relations hold: (L7.S1)$\Rightarrow$ $\big($(L7.S2) and (L7.S3)$\big)$ and that (L7.S2)$\Rightarrow$ (L7.S1).

To show (L7.S1)$\Rightarrow$ $\big($(L7.S2) and (L7.S3)$\big)$. Suppose (L7.S1) is true. Note that if S is singleton, then for any other set S', $S' \cap S \neq \emptyset$ if and only if $S \subseteq S'$. Now, for any $S \in \mathcal{W}$, the second condition in the definition of $\mathcal{W}$ implies that $S_i \subseteq S$ for all $i = 1, ..., n$ so that $T \subseteq S$. Since $T \in \mathcal{W}$, we know $T = W$ and $W \in \mathcal{W}$.

To show (L7.S2)$\Rightarrow$ (L7.S1). Suppose (L7.S2) is true. We prove S_i is singleton for all $i = 1, ..., n$ using the following steps:

Step 1: We show that if there exists some $i = 1, ..., n$ such that S_i is singleton, then S_j is singleton for all $j = 1, ..., n$. . To see why this is so, let $S^* \in \{S_i : i = 1, ..., n\}$ be singleton. Suppose, for the purpose of contradiction, there exists $j = 1, ..., n$ such that S_j is not a singleton. Then, apply Lemma 7 and we know that there exists two different minimal elements S' and S'' of $\mathcal{B} = \{S : \forall i = 1, ..., n,\ S \cap S_i \neq \emptyset\}$ so that (i) for any $i = 1, ..., n$, $S' \cap S_i \neq \emptyset$ and $S'' \cap S_i \neq \emptyset$; and (ii) $S' \cap S'' \notin \mathcal{B}$. Because $S^* \in \{S_i : i = 1, ..., n\}$ is singleton, we know $S^* \subseteq S'$ and $S^* \subseteq S''$ so that $S' \in \mathcal{W}$ and $S'' \in \mathcal{W}$. Moreover, since $S' \cap S'' \notin \mathcal{B}$, we know $S' \cap S'' \notin \mathcal{W}$. Because $W \subseteq (S' \cap S'')$ by the definition of W, we conclude that $W \notin \mathcal{W}$, which leads to the contradiction.

Step 2: We show that there must exist some $i = 1, ..., n$ such that S_i is singleton. Suppose, for the purpose of contradiction, that S_i is not singleton for all $i = 1, ..., n$. Since $W \in \mathcal{W}$, there must exists some $S^* \in \{S_1, ..., S_n\}$ such that $S^* \subseteq W$. Define $\mathcal{P} = \{S_i : S_i \neq S^*, S_i \cap S^* \neq \emptyset\}$ and $\mathcal{Q} = \{S_i : S_i \cap S^* = \emptyset\}$.

- When $\mathcal{Q} \neq \emptyset$, apply Lemma 7 to the sets in $\mathcal{Q}$, then there exists two different minimal elements S' and S'' of $\mathcal{B} = \{S : \forall S' \in \mathcal{Q},\ S \cap S' \neq \emptyset\}$ so that (i) for any $S \in \mathcal{Q}$, $S' \cap S \neq \emptyset$ and $S'' \cap S_i \neq \emptyset$; and (ii) $S' \cap S'' \notin \mathcal{B}$. By the definition of $\mathcal{W}$ and the construction of $\mathcal{P}$ and $\mathcal{Q}$, we know that $S^* \cup S' \in \mathcal{W}$ and $S^* \cup S'' \in \mathcal{W}$. However, since $S' \cap S'' \notin \mathcal{B}$, we know $(S^* \cup S') \cap (S^* \cup S'') = S^* \cup (S' \cap S'') \notin \mathcal{W}$ because the second condition in the definition of $\mathcal{W}$ is violated for some set in $\mathcal{Q}$. Since $W \subseteq S^* \cup (S' \cap S'')$, we conclude that $W \notin \mathcal{W}$, which leads to contradiction.
- When $\mathcal{Q} = \emptyset$, we know $\mathcal{P} \neq \emptyset$ (if otherwise, $n = 1$) and $S^* \in \mathcal{W}$. By the definition of W, we know $W \subseteq S^*$. Because $S^* \subseteq W$ by the definition of S^*, we conclude that $W = S^*$. To proceed, we consider two sub-cases:
  - when $S^* \nsubseteq \cup_{S \in \mathcal{P}} S$: Construct W' as $W' = \cup_{S \in \mathcal{P}} S$. By the construction of W', we know $W' \in \mathcal{W}$. This implies that $W \subseteq W'$ by the definition of W. However, $W = S^* \nsubseteq W'$, which leads to the contradiction.
  - when $S^* \subseteq \cup_{S \in \mathcal{P}} S$: Then, $\mathcal{P}$ contains at least two sets. If otherwise, $S^* \subseteq \cup_{S \in \mathcal{P}} S$ implies that there exists some $S \in \{S_i : i = 1, ..., n\}$ such that $S^* \subsetneq S$ which is ruled out the assumption of this lemma. Fix an arbitrary $\tilde{S} \in \mathcal{P}$. For any $S \in \mathcal{P}$ with $S \neq \tilde{S}$, pick arbitrary $\theta \in S \backslash S^*$. Let $S^\dagger$ be the collection of these θs. By the construction of $S^\dagger$, we know $S^\dagger \cap S^* = \emptyset$ and that, for any $S \in \mathcal{P}$ with $S \neq \tilde{S}$, $S^\dagger \cap S \neq \emptyset$. Define $\widetilde{W} = \tilde{S} \cup S^\dagger$. Then, $\widetilde{W} \in \mathcal{W}$, because $S^* \cap \widetilde{W} = S^* \cap \tilde{S} \neq \emptyset$ (recall $\tilde{S} \in \mathcal{P}$), $\tilde{S} \subseteq \widetilde{W}$ and that for any $S \in \mathcal{P}$ with $S \neq \tilde{S}$, $\emptyset \neq S^\dagger \cap S \subseteq \widetilde{W} \cap S$. However,

$$\begin{aligned} W \cap \widetilde{W} &= S^* \cap \widetilde{W} \\ &= S^* \cap (\tilde{S} \cup S^\dagger) \\ &= (S^* \cap \tilde{S}) \cup (S^* \cap S^\dagger) \\ &= S^* \cap \tilde{S} \quad \text{(because } S^* \cap S^\dagger = \emptyset) \\ &\subsetneq S^* \qquad \text{(because } S^* \nsubseteq \tilde{S} \text{ and } \tilde{S} \nsubseteq S^*). \end{aligned}$$

    Because $W = S^*$, this implies that $W \cap \widetilde{W} \subsetneq W$. But, this means that $W \nsubseteq \widetilde{W}$, which contradicts to that $\widetilde{W} \in \mathcal{W}$ and that $W = \cap_{S \in \mathcal{W}} S$.

  Therefore, there is always a contradiction when $\mathcal{Q} = \emptyset$.

Since we get contradiction in all the above cases, we conclude that there must exist some $i = 1, ..., n$ such that S_i is singleton. This result combines the Step 1 shows that S_i is singleton for all $i = 1, ..., n$. Combine all the above results, we have shown that (L7.S1) $\Leftrightarrow$ (L7.S2) when $n \geq 2$.

Finally, we need to show $W \in \mathcal{W}$ implies that $W = T$. Suppose $W \in \mathcal{W}$. When $n = 1$, we have shown that $W = T$. When $n \geq 2$, we have shown that (L7.S2) implies (L7.S1) which further implies (L7.S3). Hence, $W = T$ in both cases. Therefore, $W \in \mathcal{W}$ implies $W = T$. $\square$

Proof of Lemma 7. Without loss of generality, let us assume that S_1 contains at least two elements. Pick two arbitrary points θ_1 and θ_1' in S_1 with $\theta_1 \neq \theta_1'$. Since we have $S_i \nsubseteq S_1$ for any $i = 2, ..., n$, we can pick $\theta_i \in S_i \backslash S_1$ for any $i = 2, ..., n$. By construction, $S_1 \cap \{\theta_2, ..., \theta_n\} = \emptyset$, $\{\theta_1, \theta_2, ..., \theta_n\} \in \mathcal{B}$ and $\{\theta_1', \theta_2, ..., \theta_n\} \in \mathcal{B}$. It's possible that there exists some $\theta_i \in \{\theta_1, ..., \theta_n\}$ such that $\{\theta_1, ..., \theta_n\}\backslash\theta_i \in \mathcal{B}$. Keep removing these redundant elements until there does not exist any redundant element. In this way, one can find a subset $S^* \subseteq \{\theta_1, ..., \theta_n\}$ such that S^* is a minimal element of $\mathcal{B}$. Similarly, one can find a subset $S' \subseteq \{\theta_1', \theta_2, ..., \theta_n\}$ such that S' is a minimal element of $\mathcal{B}$. Since $\{\theta_2, ..., \theta_n\} \cap S_1 = \emptyset$, we must have $\theta_1 \in S^*$, because $S^* \notin \mathcal{B}$ if otherwise. Similarly, we know $\theta_1' \in S'$. Since $\theta_1 \neq \theta_1'$, we know that $S^* \neq S'$. Therefore, $\mathcal{B}$ has at least two different minimal elements. $\square$

B.6.2. *Proof of Lemma 6.* Suppose $\mathscr{T}$ only contains one set, say $\mathscr{T} = \{S\}$. Then, $T = S$ by the definition of T. Moreover, for any $S' \in \mathcal{W}$, we know $S \subseteq S'$ by the first condition in the definition of $\mathcal{W}$. In addition, $S \in \mathcal{W}$. Therefore, $W = S = T$ and $W \in \mathcal{W}$.

Suppose $\mathscr{T}$ contains at least two sets. We are going to show that, for the following statements:

(L6.S1) for each $S \in \mathscr{T}$, S is singleton.
(L6.S2) $W \in \mathcal{W}$
(L6.S3) $W = T$

the following relations hold: (L6.S1)$\Rightarrow$ $\big($(L6.S2) and (L6.S3)$\big)$ and that (L6.S2)$\Rightarrow$ (L6.S1).

Suppose (L6.S1) is true. For any $S \in \mathcal{W}$, the second condition in the definition of $\mathcal{W}$ implies that $T \subseteq S$. Moreover, $T \in \mathcal{W}$ by the definition of $\mathcal{W}$. Therefore, $T = W$ by the definition of W and $W \in \mathcal{W}$. Hence, (L6.S1) implies $\big($(L6.S2) and (L6.S3)

Suppose (L6.S2) is true. Suppose, for the purpose of contradiction, (L6.S1) is not true. Then, there exists some $S^* \in \mathscr{T}$ such that S^* contains at least two element. Pick θ and θ' in S^* with $\theta \neq \theta'$. Construct $W_1 = \{\theta\} \cup (\cup_{S \in \mathscr{T}: S \neq S^*} S)$ and $W_2 = \{\theta'\} \cup (\cup_{S \in \mathscr{T}: S \neq S^*} S)$. Then, $W_1 \in \mathcal{W}$ and $W_2 \in \mathcal{W}$. Moreover, $W_1 \cap W_2 = \cup_{S \in \mathscr{T}: S \neq S^*} S$. Because that for any $S \in \mathscr{T}$ with $S \neq S^*$, $S \cap S^* = \emptyset$, we know $S^* \cap (W_1 \cap W_2) = \emptyset$ which implies that $W_1 \cap W_2 \notin \mathcal{W}$. Because $W \subseteq W_1 \cap W_2$, we conclude that $W \notin \mathcal{W}$, which leads to the contradiction. Therefore, (L6.S2) implies that (L6.S1).

The above result implies that (L6.S1) $\Leftrightarrow$ (L6.S2) when $\mathscr{T}$ contains at least two sets.

Finally, we need to prove that $W \in \mathcal{W}$ implies that $W = T$. Suppose $W \in \mathcal{W}$ hold. When $\mathscr{T}$ only contains one set, we have already shown that $W = T$. When $\mathscr{T}$ contains at least two sets, (L6.S2) implies (L6.S1) which further implies that (L6.S3) i.e. $W = T$. Hence, $W = T$ whenever $W \in \mathcal{T}$.

B.7. **Proof of Theorem 7.** Let $\widetilde{A}$ be any minimal data-consistent relaxation. Define δ as $\delta(a) = 1 - \mathbb{1}(a \in \widetilde{A})$. Since, for any $a \in \widetilde{A}$, $a_{\delta(a)} = a$ and for any $a \notin \widetilde{A}$, $a_{\delta(a)}$ is a null assumption which does not have any restriction, we know $\Theta_I(A(\delta)) = \Theta_I(\widetilde{A})$. Moreover, for any $\delta' < \delta$ with $\Theta_I(A(\delta')) \neq \emptyset$, we must have that $\Theta_I(\widetilde{A}) = \Theta_I(A(\delta'))$, because $\Theta_I(A(\delta')) \subseteq \Theta_I(A(\delta)) = \Theta_I(\widetilde{A})$ and $\Theta_I(\widetilde{A})$ is singleton. This implies that $\Theta_I(\widetilde{A}) \subseteq \Theta_I^\dagger$. Since this result holds for any minimal data-consistent relaxation $\widetilde{A}$, we conclude that $\Theta_I^* \subseteq \Theta_I^\dagger$.

B.8. **Results in Section A.1.**

Proof of Proposition 3. Recall $\mathcal{Z}$ is the support of Z. Define $B(z, \epsilon) = \{z' : \|z - z'\| < \epsilon\}$. For any $\theta \in \Theta$, let $g(z; \theta)$ be the continuous version of $E[m(X, Z; \theta)|Z = z]$. That is, $g(z; \theta)$ is continuous in z and $g(z; \theta) = E[m(X, Z; \theta)|Z = z]$ almost surely.

By the second condition in the hypothesis of this proposition, for any z_0 in the support of Z, there exists some $\theta_0(z_0) \in \Theta$ and $\delta(z_0) > 0$ such that for almost every Z in $B(z_0, \delta(z_0))$, $E[m(X, Z; \theta_0)|Z] \leq 0$. Therefore, if we define function $\delta_{z_0,\epsilon}$ as $\delta_{z_0,\epsilon}(z) = \mathbb{1}(\|z - z_0\| < \epsilon)$ and the collection of functions $\mathcal{W}'$ as $\mathcal{W}' \coloneqq \{\delta_{z,\epsilon} : \epsilon \in [0, \delta(z)), z \in \mathcal{Z}\}$ where $\mathcal{Z}$ is the support of Z, then the A^* can be constructed as the set of submodels indexed by $w \in \mathcal{W}'$ with which (A.2) hold. Then, for each $a \in A^*$, $\Theta_I(a) \neq \emptyset$.

Next, we need to show that $\Theta_I(A^*) = \emptyset$ whenever $\Theta_I(A) = \emptyset$. Suppose $\Theta_I(A) = \emptyset$. Fix an arbitrary $\theta \in \Theta$. Since $\theta \notin \Theta_I(A)$, there must exists some $\mathcal{Z}' \subseteq \mathcal{Z}$ with $P(Z \in \mathcal{Z}') > 0$ such that $E[m(X, Z; \theta)|Z] > 0$ for all $Z \in \mathcal{Z}'$.

We claim there exists some z_0 and $\epsilon_0 \geq 0$ such that $E[\mathbb{1}(z \in B(z_0, \epsilon_0))m(X, Z; \theta)] > 0$. If $\mathcal{Z}$ is discrete, then there exists some $z^* \in \mathcal{Z}'$ such that $P(Z = z^*) > 0$. Hence, our claim can be verified for $z_0 = z^*$ and $\epsilon_0 = 0$. If $E[m(X, Z; \theta)|Z = z]$ is continuous in z, for all $z' \in \mathcal{Z}'$, there must exist some $\epsilon(z') > 0$ such that $E[m(X, Z; \theta)|Z] > 0$ for almost every Z in $B(z', \epsilon(z'))$. Since $\mathcal{Z}$ is a subset of real space $\mathbb{R}^{k_2}$, $\mathcal{Z}'$ must be separable. That is, there exists a countable subset $\mathcal{D}$ of $\mathcal{Z}'$ such that $\mathcal{D}$ is dense in $\mathcal{Z}'$. Since $\mathcal{D} \subseteq \mathcal{Z}'$, $E[m(X, Z; \theta)|Z] > 0$ for almost every Z in $\cup_{z \in \mathcal{D}} B(z, \epsilon(z))$. Moreover, since $\mathcal{D}$ is dense in $\mathcal{Z}'$, we know $\mathcal{Z}' \subseteq \cup_{z \in \mathcal{D}} B(z, \epsilon(z))$ so that $P(Z \in \cup_{z \in \mathcal{D}} B(z, \epsilon(z))) > 0$. Since $\mathcal{D}$ is countable, we have $P(Z \in B(z^*, \epsilon(z^*)) > 0$ for some $z^* \in \mathcal{D}$. Hence, our claim is verified for $z_0 = z^*$ and $\epsilon_0 = \epsilon(z^*)$.

Finally, the claim in the previous paragraph implies that there exists some $a \in A^*$ such that $\theta \notin \Theta_I(a)$. Hence, $\theta \notin \Theta_I(A^*)$. This completes the proof that $\Theta_I(A^*) = \emptyset$ whenever $\Theta_I(A) = \emptyset$. □

Proof of Proposition 4. Because Θ is compact, to show $\Theta_I(a)$ is compact for all $a \in A$, we only need to show that for any $w \in \mathcal{W}_1^+$, $E[w(Z)m(X, Z; \theta)]$ is continuous in θ. For any $\theta \in \Theta$, and for any θ_n converges to θ, we know $E[m(X, Z; \theta_n)|Z] \to E[m(X, Z; \theta)|Z]$ almost surely as $n \to \infty$. Since the norm of $w(Z)E[m(X, Z; \theta)|Z]$ is dominated by $w(Z)g(Z)$ and $E\|w(Z)g(Z)\| \leq \sqrt{E[\|w(Z)\|^2]}\sqrt{E[\|g(Z)\|^2}$, the dominated convergence theorem implies that

$$E[w(Z)m(X, Z; \theta_n)] \to E[w(Z)m(X, Z; \theta)]$$

as $n \to \infty$. This completes the proof. □

B.9. **Results in Section 3.2.**

Proof for Proposition 3.2. Let $\mathcal{Y}$ be the support of Y. Define $\delta = \min_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} P(Y = y|X = x)$ where the inf stands for the essential infimum. Then, $\delta > 0$. For each $y \in \mathcal{Y}$, there exists a sequence $\theta_1(y)$,..., $\theta_n(y)$, ... in $\mathbb{R}^d$, such that $L(\{y\}, x; \theta_n(y)) \to 1$ as $n \to 1$. Therefore, there must exists some n such that $L(\{y\}, x; \theta_n(y)) > 1 - \delta$ for any $y \in \mathcal{Y}$. Then, for any nonempty subset K of $\mathcal{Y}$ with $K \subsetneq \mathcal{Y}$, and any $y \in K$,

$$P(Y \in K|X) \leq 1 - \delta < L(\{y\}, x; \theta_n(y)) \leq L(K, X; \theta_n(X, y)) \text{ almost surely}$$

Therefore, define $\Theta_U = \{\theta_i(y) : i \leq n, y \in \mathcal{Y}\}$. Therefore, for any subset K of Y, there exists some $\theta \in \Theta_U$ such that

$$P(Y \in K|X) \leq L(K, X; \theta) \text{ almost surely}$$

This completes the proof. □

B.10. **Proof of Proposition 1.** Recall the A in the introductory example is the set of all assumption indexed by $h \in \mathcal{H}_1^+$ with which (2.2) holds.

Let us first show Θ_I^* is equal the interval specified in (4.1). By Lemma 3, we know that for each $\theta \in (\overline{\gamma}, \underline{\gamma})$, 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, $(\overline{\gamma}, \underline{\gamma}) \subseteq \Theta_I^*$.

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 $\underline{\gamma} > \overline{\gamma}$, 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 exists $\underline{h}$ 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 (2.2) 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 $\theta > \overline{\gamma}$, $\theta \notin \cap_i \widetilde{\Theta}(\overline{h}_i)$. Therefore, if we define $\mathcal{H}' = \{\overline{h}_i : i \geq 1\} \cup \{\underline{h}\}$, we have $\cap_{h \in \mathcal{H}'} \widetilde{\Theta}(h) = \{\overline{\gamma}\}$. 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 2 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 Condition (T2.C2) 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, i.e 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've proven that $\cap_{h \in \mathcal{H}'} \widetilde{\Theta}(h) = \{\overline{\gamma}\}$, 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

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

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 equals to the interval specified in (4.1). Moreover, we have also shown that for any $\theta \in \Theta_I^*$, there exists some $A' \subseteq A$ such that $\Theta_I(A') = \{\theta\}$. This implies that for any minimum data-consistent relaxation $\widetilde{A}$, $\Theta_I(\widetilde{A})$ is a singleton set. By Theorem 6, we know Θ_I^* is the smallest element in Λ.

B.11. **Proof of Proposition 2.** Recall the notation used in this example: $\underline{Y}_d := Y\mathbb{1}(D = d) + \underline{y}_d\mathbb{1}(D \neq d)$, $\overline{Y}_d := Y\mathbb{1}(D = d) + \overline{y}_d\mathbb{1}(D \neq d)$, $\underline{q}_{dt} := E[\underline{Y}_d|Z = t]$ and $\overline{q}_{dt} := E[\overline{Y}_d|Z = t]$. Proposition 2 is an immediate corollary of the following two lemmas.

Lemma 8. *In model $Y = \sum_{z \in \mathcal{Z}} \mathbb{1}(Z = z)[Y_{1z}D + Y_{0z}(1 - D)]$ where $\mathcal{Z} = \{1, 2, ..., k\}$. Fix an arbitrary $z^* = 1, ..., k$. Let Θ_{I,z^*} be the identified set of a_{z^*}, i.e. 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 $d \in \{0, 1\}$:

$$\forall z < z^*, \quad \max(\underline{q}_{dt} : t \leq z) \leq \min(\overline{q}_{dt} : t \geq z) \tag{B.2}$$

and

$$\max(\underline{q}_{dt} : t = 1, ..., k) \leq \min(\overline{q}_{dt} : t \geq z^*) \tag{B.3}$$

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

Lemma 9. *In model $Y = \sum_{z \in \mathcal{Z}} \mathbb{1}(Z = z)[Y_{1z}D + Y_{0z}(1 - D)]$ where $\mathcal{Z} = \{1, 2, ..., k\}$. Let Θ_I be the identified set of $a^\dagger$, i.e. the identified set of* E.1 *and* E.2*. Then,*

(1) $\Theta_I \neq \emptyset$ if and only if $P\big(Y \in [\underline{y}_d, \overline{y}_d] | D = d\big) = 1$ for any $d \in \{0, 1\}$.
(2) when $\Theta_I \neq \emptyset$, $\Theta_I = \Big[E[\underline{Y}_1], E[\overline{Y}_1]\Big] \times \Big[E[\underline{Y}_0], E[\overline{Y}_0]\Big]$.

Proof of Lemma 8. The results of this lemma can be divided into the following two parts:

(1) For any $z^* = 1, ..., k$, $\Theta_{I,z^*} \neq \emptyset$ only if that (B.2) and (B.3) hold for each $d = 0, 1$. Moreover, $\Theta_{I,z^*} \subseteq \Gamma_{1,z^*} \times \Gamma_{0,z^*}$.
(2) if (B.2) and (B.3) 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.

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

$$\forall z' < z^*, \quad \max(\underline{q}_{dt} : t \le z') \le E[Y_{dz'}|Z=z'] \le \min(\overline{q}_{dt} : t \ge z') \tag{B.4}$$

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

$$\forall z' \ge z^*, \quad \max(\underline{q}_{dt} : t=1,...,k) \le E[Y_{dz'}|Z=z'] \le \min(\overline{q}_{dt} : t \ge z^*). \tag{B.5}$$

Combine (B.4) and (B.5), 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 Assumption *E.1*, *E.2* and *E.3* imply (B.4) and (B.5), the violation of (B.2) and (B.3) implies that $\Theta_{I,z^*} = \emptyset$. Equivalently, $\Theta_{I,z^*} \ne \emptyset$ only if (B.2) and (B.3) hold for any $d \in \{0,1\}$.

Part 2. We want to prove that (B.2) and (B.3) implies that $\Theta_{I,z^*} \ne \emptyset$ 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 $z = 1,...,k$ as follows:

- for $z < z^*$, let γ_z be the value which solves
$$\max(\underline{q}_{dt} : t \le z) = E[\mathbb{1}(D=d)Y|Z=z] + E[\mathbb{1}(D \ne d)Y|Z=z]\gamma_z$$
Then, $\gamma_z \in [\underline{y}_d, \overline{y}_d]$ if $\overline{q}_{dz} \ge \max(\underline{q}_{dt} : t \le z)$, which is implied by (B.2).
- for $z \ge z^*$, let γ_z be the value which solves
$$\max(\underline{q}_{dt} : t=1,...,k) = E[\mathbb{1}(D=d)Y|Z=z] + E[\mathbb{1}(D \ne d)Y|Z=z]\gamma_z$$
Then, $\gamma_z \in [\underline{y}_d, \overline{y}_d]$ if $\max(\underline{q}_{dt} : \forall t) \le \overline{q}_{dz}$ which is implied by (B.3).

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

$$E[W_{dz}|Z=z] = \begin{cases} \max(\underline{q}_{dt} : t \le z) & \text{if } z < z^* \\ \max(\underline{q}_{dt} : t=1,...,k) & \text{if } z \ge z^* \end{cases} \tag{B.6}$$

which implies that

$$\forall z \le t, \qquad E[W_{dz}|Z=z] \le E[W_{dt}|Z=t] \tag{B.7}$$

$$\forall z \ge z^*, \qquad E[W_{dz}|Z=z] = \max(\underline{q}_{dt} : t=1,...,k) \tag{B.8}$$

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

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

$$\begin{aligned} \forall t < z, &\quad E[W_{dz}|Z=z] = E[\phi_{dt}(\alpha_{tz})|Z=t] \\ \forall t > z, &\quad E[W_{dz}|Z=z] = E[\psi_{dt}(\alpha_{tz})|Z=t]. \end{aligned} \tag{B.9}$$

To see why it is so, note that

$$\begin{aligned} \forall t < z, &\quad E[W_{dt}|Z=t] = E[\phi_{dt}(0)|Z=t] \text{ and } E[\phi_{dt}(1)|Z=t] = \overline{y}_d, \\ \forall t > z, &\quad \underline{y}_d = E[\phi_{dt}(0)|Z=t] \text{ and } E[\phi_{dt}(1)|Z=t] = E[W_{dt}|Z=t]. \end{aligned} \tag{B.10}$$

These results, combined with (B.7), imply that

$$\forall t < z, \qquad E[\phi_{dt}(0)|Z=t] \le E[W_{dz}|Z=z] \le E[\phi_{dt}(1)|Z=t],$$

$$\forall t > z, \qquad E[\psi_{dt}(0)|Z=t] \le E[W_{dz}|Z=z] \le E[\phi_{dt}(1)|Z=t].$$

which implies the existence of $\alpha_{tz} \in [0,1]$ satisfying (B.9) for all $t \neq z$.

In addition, $(\alpha_{tz} : t \neq z)$ has some extra properties. Because (B.7) holds and $E[\phi_{dt}(\alpha)|Z = t]$ is an increasing function of α,

$$\forall t < z < z', \quad \alpha_{tz} \leq \alpha_{tz'}. \tag{B.11}$$

Because (B.7) holds and $E[\psi_t(\alpha)|Z = t]$ is an increasing function of α,

$$\forall z < z' < t, \quad \alpha_{tz} \leq \alpha_{tz'}. \tag{B.12}$$

Construct $Y_{dz} := \sum_{t<z} \phi_{dt}(\alpha_{tz}) + W_{dz} + \sum_{t>z} \psi_{dt}(\alpha_{tz})$. Because $P(W_{dz} \in [\underline{y}_d, \overline{y}_d]) = 1$, assumption *E.1* holds for this Y_{dz}. Because of (B.9), assumption *E.2* holds for this Y_{dz}, i.e. $E[Y_{dz}|Z = t] = E[Y_{dz}|Z = z]$ for any t, z with $t \neq z$.

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

- If $Z < z_1$, $Y_{dz_1} = \phi_{dZ}(\alpha_{Zz_1}) \leq \phi_{dZ}(\alpha_{Zz_2}) = Y_{dz_2}$ because of (B.11) and because $\phi_{dZ}(\alpha)$ is increasing in α.
- If $Z = z_1$, $Y_{dz_1} = W_{dz_1} \leq \phi_{dZ}(\alpha_{Zz_2}) = Y_{dz_2}$ because of the definition of $\phi_{dZ}(\alpha)$.
- If $z_1 < Z < z_2$, $Y_{dz_1} = \psi_{dZ}(\alpha_{Zz_1}) \leq W_{dZ} \leq \phi_{dZ}(\alpha_{Zz_2}) = Y_{dz_2}$ because of the definition of $\phi_{dZ}(\alpha)$ and $\psi_{dZ}(\alpha)$.
- If $Z = z_2$, $Y_{dz_1} = \psi_{dZ}(\alpha_{Zz_1}) \leq W_{dZ} = Y_{dz_2}$ because of the definition of $\psi_{dZ}(\alpha)$.
- If $z_2 < Z$, $Y_{dz_1} = \psi_{dZ}(\alpha_{Zz_1}) \leq \psi_{dZ}(\alpha_{Zz_2}) = Y_{dz_2}$ because of (B.12) and because $\psi_{dZ}(\alpha)$ is increasing in α.

As a result, $Y_{dz_1} \leq Y_{dz_2}$ almost surely for any $z_1 \leq z_2$. Moreover, because of (B.8), $\alpha_{tz} = \alpha_{tz'}$ for any t, z and z' with $t < \min(z, z')$ and $z^* \leq \min(z, z')$. Because of (B.8) and (B.10), $\alpha_{tz} = 0$ for any $z^* \leq t < z$, and $\alpha_{tz} = 1$ for any $t > z \geq z^*$. Given these results, one can show that for any $z' \geq z^*$, $Y_{dz'} = \sum_{t<z^*} \phi_{dt}(\alpha_{tz^*}) + \sum_{t=z^*}^{k} W_{dt}$. 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 (B.6), 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
$$\min(\overline{q}_{dt} : t \geq z) = E[\mathbb{1}(D = d)Y|Z = z] + E[\mathbb{1}(D \neq d)Y|Z = z]\gamma_z$$
- for $z \geq z^*$, let γ'_z be the value which solves
$$\min(\overline{q}_{dt} : t \geq z^*) = E[\mathbb{1}(D = d)Y|Z = z] + E[\mathbb{1}(D \neq d)Y|Z = z]\gamma_z$$

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 assumption *E.1-E.3*. This completes the proof. $\square$

Proof of Lemma 9. Suppose *E.1* and *E.2* hold. For any $z \in \{1, 2, ..., k\}$ 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 $\Theta_I \subseteq \Big[E[\underline{Y}_1], E[\overline{Y}_1]\Big] \times \Big[E[\underline{Y}_0], E[\overline{Y}_0]\Big]$. Moreover, when $P\big(Y \in [\underline{y}_d, \overline{y}_d]|D = d\big) = 1$ for any $d \in \{0, 1\}$ fails to hold, *E.1* will fail to hold. Hence, $\Theta_I \neq \emptyset$ only if $P\big(Y \in [\underline{y}_d, \overline{y}_d]|D = d\big) = 1$ for any $d \in \{0, 1\}$.

Suppose that $P\big(Y \in [\underline{y}_d, \overline{y}_d]|D = d\big) = 1$ for any $d \in \{0, 1\}$ hold. Then, we know that for each $z = 1, ..., k$ 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 $\Theta_I = \Big[E[\underline{Y}_1], E[\overline{Y}_1]\Big] \times \Big[E[\underline{Y}_0], E[\overline{Y}_0]\Big]$. $\square$