Estimating Economic Models with Testable Assumptions: Theory and Application to LATE

Moyu Liao

Pennsylvania State University

February 25, 2020

Abstract

This paper studies the identification and estimation problem in incomplete economic models with testable assumptions. Testable assumptions give strong and interpretable empirical content to the models but they also carry the possibility that our distribution of observed outcome may reject these assumptions. A natural way is to find a set of relaxed assumptions $\tilde{A}$ that cannot be rejected by any reasonable distribution of observed outcome and preserves identified set of parameter of interest. The main contribution of this paper is to characterize the property of such relaxed assumption $\tilde{A}$ using a generalized definition of refutability and confirmability. A general estimation and inference procedure is proposed, and can be applied to most incomplete economic models. As the key application, I study the Imbens and Angrist Monotonicity assumption in potential outcome framework. I give a set of relaxed assumptions $\tilde{A}$ that can never be rejected and always preserve the identified set of local average treatment effect (LATE). The LATE is point identified and easy to estimate under $\tilde{A}$.

Keywords — Incomplete Models; Refutability; Potential Outcome; LATE
1 Introduction

Empirical researchers often make convenient model assumptions in structural estimation. These assumptions usually come from economic theories or intuitions. For example, the 'No Defier' assumption in Imbens and Angrist (1994) assumes an instrument has a monotone effect on the decision of taking treatment; the 'Pure Strategy Nash Equilibrium' assumption in Bresnahan and Reiss (1991) assumes only pure strategy Nash Equilibrium are played in $2 \times 2$ entry game; the 'Self Selection' assumption in Roy (1951) assumes employees perfectly observe their salary and choose job sector optimally. Making such assumptions usually simplifies identification and estimation problems, and makes results easier to interpret. However, making assumptions comes with the cost that they may be rejected by some distribution of observed variables. As a result, empirical researchers are faced with the problem that their identification and estimation result may potentially fail to be consistent with some distribution of observed variables.

Formally, the goal of identification is to find a class of models that can rationalize the observed distribution of outcomes. Assumptions are restrictions on the model space, and they reflect empirical researchers’ prior on the true data generating process. In this sense, assumptions are imposed before any distribution of observables are seen, and hence should be comprehensive to incorporate all possible regular distribution of observables. Unfortunately, assumptions in the three examples above can be rejected by some reasonable distribution of observables (see (Kitagawa, 2015), (Mourifié and Wan, 2017) and (Mourifie, Henry, and Méango, 2018)). When an assumption $A$ can be potentially rejected, a natural way is to find a relaxed assumption $\tilde{A}$ so that no distribution of observables can reject $\tilde{A}$. This is the first criterion of choosing a relaxed assumption. On the other hand, we also want to keep some good feature of the old assumption, since it still reflect the economic theory and intuition behind it. Formally, given a parameter of interest $\theta$, we want to preserve the identified set $\Theta^{ID}$ if the original assumption $A$ cannot be rejected by the observed distribution of outcomes.

This paper aims to do four things: First, I extend the definition of refutability and confirmability of assumption in Breusch (1986) to incomplete economic structures, and use the similar notation to classify different types of assumptions. In incomplete structure environ-
ment, a structure can predict multiple distributions of observables. Analysis of economic structures faces an extra model uncertainty from the multiplicity of outcome distributions. As shown in the main paragraph, when we look at complete models, deciding whether there exists a structure in $A$ that rationalize data is equivalent to deciding whether the true structure is in the non-refutable set of $A$. I also provide sufficient conditions on $A$ under which the equivalence decision condition holds for incomplete structures. This is closely related to sub-vector inference problems (e.g. (Belloni, Bugni, and Chernozhukov (2018))) where researchers want to test whether there exists a structure $p$ whose sub-vector $\theta(p) = \theta_0$. The null hypothesis can then be written as an assumption $A$, and my result gives another interpretation to the null hypothesis.

Second, I characterize the property of relaxed assumption $\tilde{A}$ such that no distribution of observables can be rejected under $\tilde{A}$ and the identified set of parameter of interest $\Theta^{ID}$ does not change when $A$ can’t be rejected. The characterization of $\tilde{A}$ uses the generalized definition of refutability and confirmability. It is shown that when models are complete, a relaxed assumption $\tilde{A}$ that satisfies the two properties above always exists. I also characterize a sufficient condition for the existence of $\tilde{A}$ when we look at incomplete structures. The possible failure of finding $\tilde{A}$ in incomplete structures encourages researchers to find nature completion structures, and then find a nice relaxed assumption in the completed structures. However, relaxed assumptions under completed structures can be hard to rationalize by economic theory. This is illustrated by a $2 \times 2$ entry game example. Since there can be multiple relaxed assumptions, I give some criteria on how to choose $\tilde{A}$.

Third, I provide a general estimation and inference methods on parameters that can be applied to most economic structures. The method is a generalization of dilation approach in Galichon and Henry (2013). Their dilation is on the space of individual outcome variable, while my dilation is on the space of outcome distribution. Compared to their method, my inference procedure can deal with more complicated economic environment where the identified set of parameter of interest involves density of the distribution of outcomes. The idea of dilation method is to properly enlarge the set of distribution of outcome predicted by a structure, and see whether the empirical distribution of outcome falls in the enlarged set. The procedure is closely related to M-estimation based method in Chernozhukov et al.
where their method is to enlarge the value of population criterion function.

Fourth, as the key application, I look at the identification of local average treatment effect (LATE) in potential outcome framework. Kitagawa (2015) shows there is sharp testable implication of Imbens and Angrist Monotonicity (IA-M) assumption. However, there is no solution on what should be done when the IA-M assumption is rejected. I provide two generalization of Imbens and Angrist Monotonicity (IA-M) assumption that cannot be rejected by any distribution of observables and at the same time preserves the property of IA-M assumption. The first one is a relaxation of 'No Defier' assumption. The logic is to allow minimal measure of defier in the joint distribution of latent variables. This relaxed assumption can be rationalized as a two step model selection procedure used by empirical researchers. It can be shown that the minimal defier relaxed assumption not only preserve the identified set for LATE, but also preserve the identified set for all other parameter of interests. The second is a relaxation of independent instrument assumption. The logic is to allow minimal departure from independent instrument assumption for always takers and never takers while maintain the independent instrument condition for compliers. LATE is point identified and the value is the same under these two assumptions. Estimate of LATE can be constructed easily and has a normal limit distribution.

This paper is closely related to Masten and Poirier (2018), where they characterize the relaxation of baseline assumption $A$ using the falsification frontier. The key difference between this paper and Masten and Poirier (2018) is that this paper try to characterize a set of relaxed assumption $\tilde{A}$ that is not refutable before any distribution of observables are seen, while Masten and Poirier (2018) characterize the relaxation of $A$ after the distribution of observables are seen. These two approaches are the same when structure space is complete and there are no complication of multiplicity of outcome distributions. However, for incomplete models, the approach in Masten and Poirier (2018) is equivalent to first complete the models, and then work in the completed structure spaces.

The rest of the paper is organized as follows. Section 2 describes the generalized theory of refutability and confirmability, and binary decision problem in incomplete structures. Section 3 defines an identification problem and the theory on choosing relaxed assumption $\tilde{A}$. Section 4 describes the estimation and inference procedure using dilation method. Section 5
discusses the application to potential outcome framework. Section 6 concludes. Main proofs are collected in Appendixes.

2 A General Theory of Testable Assumption

Let’s call $X$ as observed random variables and let $F(X)$ be the distribution of it. We collect all possible regular distributions $F$ and call it the observation space. Here the term regular usually puts minimal assumption on the distribution, for example, only requires it to be continuous or discrete or the mixture.

**Definition 1.** The observation space $\mathcal{F}$ is the collection of all possible regular distribution of $F(X)$, where $X$ is the observed variables.

The distribution of observed variable $X$ is generated by some data generating process, which I call as a structural $p$.

**Definition 2.** An econometric structure $p = (G^p, M^p)$ consists of a distribution of latent variable $G^p(\epsilon)$ and a model $M^p$. Let $\mathcal{G}$ denote the space of all possible regular distribution of $G^p(\epsilon)$, a model is a correspondence $M^p : \mathcal{G} \Rightarrow \mathcal{F}$

The above definition of econometric structure is a reformulation of the economic structure defined in Jovanovic (1989). Since in most econometric problems, we focus on the distribution of outcomes $F$ instead of how each $X$ is related to $\epsilon$, we directly define a structure as a correspondence from $\mathcal{G}$ to $\mathcal{F}$.

**Definition 3.** A structure universe $\mathcal{P}$ is a collection of structures such that $\bigcup_{p \in \mathcal{P}} \in M^p(G^p) = \mathcal{F}$, and an assumption $A$ is a subset of $\mathcal{P}$.

Here I explicitly distinguish structure universe $\mathcal{P}$ and the assumption $A$, though both are just a collection of structures. Structure universe $\mathcal{P}$ should be thought as the common prior that can span across different empirical contexts. Typically, it is a commonly used framework on a topic. On the other hand, an assumption $A$ put restrictions that are suitable for the particular empirical contexts, or convenient for empirical analysis. The key requirement $\bigcup_{p \in \mathcal{P}} \in M^p(G^p) = \mathcal{F}$ requires that all possible distributions of outcome can be generated.
by some structure in the universe and no more other distribution beyond $\mathcal{F}$ can be generated. This is without loss of generality because we can always add structures to $\mathcal{G}$ or add distributions to $\mathcal{F}$.

**Example 1. (Binary Entry Game)** Consider a binary entry game in [Bresnahan and Reiss (1991)](https://doi.org/10.1086/259902). Suppose there are two firms who make decisions $Y_{im}$ on whether to enter a market $m$, where $i = 1, 2$ is the firm index, $Y_{im}$ is binary and $Y_{im} = 1$ means firm $i$ enters market $m$. Profit functions is

$$
\pi_{1m} = \alpha_1 + \beta_1 X_{1m} + \delta_1 Y_{2m} + \epsilon_{1m}
$$

$$
\pi_{2m} = \alpha_2 + \delta_2 Y_{1m} + \epsilon_{2m}
$$

where $(\epsilon_{1m}, \epsilon_{2m})$ is a market shock observed by both firms. To simplify the discussion, let’s consider $X_{1m}$ only takes two values: $X_{1m} \in \{0, 1\}$. Let’s focus on competitive entry so $\delta_1 \leq 0$ and $\delta_2 \leq 0$.

Suppose the econometrician only observe the number of players in market $m$, denoted by $Y_m \in \{0, 1, 2\}$. So $Y_m = Y_{1m} + Y_{2m}$. Since $Y_m$ and $X$ are discrete, the observation space can be described by a collection of six-dimensional probability vectors

$$
\mathcal{F} = \left\{ (a_1, \ldots, a_6) \mid \sum_{j=1}^{6} a_j = 1, \ a_j \geq 0 \right\}
$$

where each $F \in \mathcal{F}$ corresponds to the probability vector:

$$
F = \begin{pmatrix}
Pr_F(Y_m = 2, X_{1m} = 0) \\
Pr_F(Y_m = 1, X_{1m} = 0) \\
Pr_F(Y_m = 0, X_{1m} = 0) \\
Pr_F(Y_m = 2, X_{1m} = 1) \\
Pr_F(Y_m = 1, X_{1m} = 1) \\
Pr_F(Y_m = 0, X_{1m} = 1)
\end{pmatrix}
\quad (2.1)
$$

Both $\epsilon_m = (\epsilon_{1m}, \epsilon_{2m})$ and $X_{1m}$ are latent variable, and $G^p$ is a joint distribution of them. A model $M^p = (\theta^p, S^p)$ consists of two parts: $\theta^p = (\alpha^p_1, \alpha^p_2, \beta^p_1, \delta^p_1, \delta^p_2)$ is the profit parameter, and $S^p \in \{\text{PSNE, L2R}\}$ is the solution concept of the game, where PSNE stands for pure strategy Nash Equilibrium and L2R stands for level-2 rationality.
Given the profit parameter $\theta^p$ and value of $X_{1m}$, the space of $(\epsilon_{1m}, \epsilon_{2m})$ is partitioned into 5 parts as figure 1.

If $\epsilon_m \in A(y|X_{1m})$, where $y = (Y_1, Y_2) \in \{(1,1), (1,0), (0,1), (0,0)\}$, and $S^p \in \{PSNE, L2R\}$, the solution concept predicts $Y_m = \sum_i Y_{im}$ as the unique outcome. If $\epsilon_m \in A(M|X_{1m})$, and $S^p = PSNE$, then the solution concept predicts $Y_m = 1$ as the outcome. If $\epsilon_m \in A_X(M|X_{1m})$, and $S^p = L2R$, then the solution concept does not put any restriction on outcome $Y_m$. 

Figure 1: Partition of the space of $(\epsilon_{1m}, \epsilon_{2m})$
Following Aradillas-Lopez and Tamer (2008), if \( S^p = L2R \), \( M^p(G^p) \) is the collection of

\[
\begin{align*}
G^p(\epsilon_m \in A((1,1)X_{1m} = 0), X_{1m} = 0) + c_0^2 \\
G^p(\epsilon_m \in A((0,1)U (1,0)X_{1m} = 0), X_{1m} = 0) + c_0^1 \\
G^p(\epsilon_m \in A((0,0)X_{1m} = 0), X_{1m} = 0) + c_0^0 \\
G^p(\epsilon_m \in A((1,1)X_{1m} = 1), X_{1m} = 1) + c_1^2 \\
G^p(\epsilon_m \in A((1,0)U (0,1)X_{1m} = 1), X_{1m} = 1) + c_1^1 \\
G^p(\epsilon_m \in A((0,0)X_{1m} = 1), X_{1m} = 1) + c_0^1 \\
\end{align*}
\]

\[c_i^2 + c_i^1 + c_i^0 = G^p(\epsilon_m \in A(M), X_{1m} = i) \quad c_i^j \geq 0 \text{ for } i, j = 0, 1, 2\] (2.2)

On the other hand, if \( S^p = PSNE \), \( M^p(G^p) \) is a singleton:

\[
\begin{align*}
G^p(\epsilon_m \in A((1,1)X_{1m} = 0), X_{1m} = 0) \\
G^p(\epsilon_m \in A((0,1)U (1,0)U M X_{1m} = 0), X_{1m} = 0) \\
G^p(\epsilon_m \in A((0,0)X_{1m} = 0), X_{1m} = 0) \\
G^p(\epsilon_m \in A((1,1)X_{1m} = 1), X_{1m} = 1) \\
G^p(\epsilon_m \in A((1,0)U (0,1)U M X_{1m} = 1), X_{1m} = 1) \\
G^p(\epsilon_m \in A((0,0)X_{1m} = 1), X_{1m} = 1) \\
\end{align*}
\]

\[c_i^2 + c_i^1 + c_i^0 = G^p(\epsilon_m \in A(M), X_{1m} = i) \quad c_i^j \geq 0 \text{ for } i, j, k = 0, 1\] (2.3)

Level 2 Rationality can be used as a reference model since it can generate player’ behavior in various way. Pure strategy Nash Equilibrium model on the other hand predict cleaner individual behavior and hence makes identification problem easier to interpret. We follow Ciliberto and Tamer (2009) to further consider continuously distributed \( \epsilon \) that is independent of \( X_{1m} \). One can consider a structure universe \( \mathcal{P} \) as

\[
\mathcal{P} = \left\{ p \mid \theta^p \in \mathbb{R}^3 \times \mathbb{R}^2, \quad S^p \in \{PSNE, L2R\} \right\}
\]

\[
\epsilon_m \perp X_{1m} \text{ holds for } G^p \quad \text{(2.4)}
\]

\[
G^p(\epsilon_m) \text{ has a density } g^p
\]

The following proposition states that \( \mathcal{P} \) is a structure universe.

**Proposition 1.** The \( \mathcal{F} \) and \( \mathcal{P} \) defined in the 2 \( \times \) 2 entry game satisfies \( \cup_{p \in \mathcal{P}} M^p(G^p) = \mathcal{F} \)

8
A commonly used empirical assumption is that the solution concept is PSNE, so the assumption set is

$$A = \{ p \mid \theta^p \in \mathbb{R}^3 \times \mathbb{R}^2 \cup \mathbb{R}^2_+, S^p = \text{PSNE} \}$$

(2.5)

$$\epsilon_m \perp X_{1m} \text{ holds for } G^p$$

$$G^p(\epsilon_m|X_{1m}) \text{ has density } g^p \text{ over } \mathbb{R}^2$$

Example 2. (Potential Outcome Framework) Consider the potential outcome framework in Imbens and Angrist (1994) where the observed outcome \(Y_i\) is generated by the potential treatment decision \(D_i(1), D_i(0)\), potential outcome \(Y_i(1), Y_i(0)\) and binary instrument \(Z_i\) through

\[
Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i) \\
D_i = D_i(1)Z_i + D_i(0)(1 - Z_i)
\]

(2.6)

The observed variable is \((Y_i, D_i, Z_i)\). To fix the idea, we consider \(Y_i\) to be a continuous variable such as income, and \(D_i\) and \(Z_i\) to be binary. So the observation space is

$$\mathcal{F} = \left\{ F(Y_i, D_i, Z_i) \mid F(Y_i|D_i, Z_i) \text{ has density } f_{dz} \text{ over } \mathbb{R} \right\}$$

(2.7)

$$D_i, Z_i \text{ are binary}$$

The latent variable includes the potential outcome \(Y_i(1), Y_i(0)\), potential treatment decision \(D_i(1), D_i(0)\) and \(Z\), and \(G\) is the joint distribution of them:

$$G = \left\{ G(Y_i(1), Y_i(0), D_i(1), D_i(0), Z_i) \mid D_i, Z_i \text{ are binary} \right\}$$

(2.8)

$$G(Y_i(1), Y_i(0)|D_i(1), D_i(0), Z_i) \text{ has density } g \text{ over } \mathbb{R}^2$$

There is no model parameter in equation (2.6), and thus all models \(M^p\) are the same. The structure predicted distribution of observables is just the push forward measure of \(G\) under mapping (2.6). The structure universe \(P\) is defined as

$$P = \{ p \mid G^p \in \mathcal{G}, M^p \text{ satisfies } (2.6) \}$$

(2.9)

The Imbens and Angrist Monotonicity assumption is the exogeneity and monotonicity of
For the two examples above, the key difference is that in binary entry game, model mapping is potentially multi-valued, while in potential outcome framework, model mapping is single-valued. This gives rise to the difference between complete and incomplete structure.

**Definition 4.** A structure \( p \) is called **complete** if \( M^p(G^p) \) is a singleton. Otherwise it is called **incomplete**. The universe \( \mathcal{P} \) is called **complete universe** if all structure \( p \) in \( \mathcal{P} \) is complete, otherwise it is incomplete.

I now give several definitions that helps to classify assumptions. These definitions are extension of the definition of refutability and confirmability in Breusch (1986) to incomplete models.

**Definition 5.** (non-Refutable set) The **strongly non-refutable set** of \( A \) under \( \mathcal{P} \) is defined as

\[
\mathcal{H}^{snf}_{\mathcal{P}}(A) = \{ p \in \mathcal{P} : M^p(G^p) \subset \bigcup_{p^* \in A} M^{p^*}(G^{p^*}) \}
\]

The **weakly non-refutable set** of \( A \) under \( \mathcal{P} \) is defined as

\[
\mathcal{H}^{wnf}_{\mathcal{P}}(A) = \{ p \in \mathcal{P} : M^p(G^p) \cap \big( \bigcup_{p^* \in A} M^{p^*}(G^{p^*}) \big) \neq \emptyset \}
\]

Strongly non-refutable set is the collection of structures that can be observationally equivalent to some structure in \( A \) for all observed distribution \( F \). Weakly non-refutable set is the collection of structures that can be observationally equivalent to some structure in \( A \) for some observed distribution \( F \). In incomplete structures, we do not specify how the distribution of observables are selection, it is possible when \( p^* \in \mathcal{H}^{wnf}_{\mathcal{P}}(A) \setminus A \), some \( F \in M^{p^*}(G^{p^*}) \) allows us to say \( p^* \notin A \) because \( F \notin \bigcup_{p \in A} M^p(G^p) \), while for another \( \tilde{F} \in M^{p^*}(G^{p^*}) \) we cannot tell whether \( p^* \in A \) or not. In other words, there is no \( F \) can help us to decide whether \( p \) is in \( A \) or in \( \mathcal{H}^{snf}_{\mathcal{P}}(A) \setminus A \); some but not all \( F \) can tell us whether \( p \) is in \( A \) or in \( \mathcal{H}^{wnf}_{\mathcal{P}}(A) \setminus A \); and all \( F \) can tell us whether \( p \) is in \( A \) or in \( \mathcal{H}^{wnf}_{\mathcal{P}}(A)^c \). In particular

\[
A \subset \mathcal{H}^{snf}_{\mathcal{P}}(A) \subset \mathcal{H}^{wnf}_{\mathcal{P}}(A)
\]
Definition 6. (Confirmable set) The strongly confirmable set of $A$ under $\mathcal{P}$ is defined as

$$\mathcal{H}_{\mathcal{P}}^{scon}(A) = \{ p \in \mathcal{P} : M^p(G^p) \subset \bigcap_{p^* \in A^c} (M^{p^*}(G^{p^*})^c) \}$$

The weakly confirmable set of $A$ under $\mathcal{P}$ is defined as

$$\mathcal{H}_{\mathcal{P}}^{wcon}(A) = \{ p \in \mathcal{P} : M^p(G^p) \cap \bigcap_{p^* \in A^c} (M^{p^*}(G^{p^*})^c) \neq \emptyset \}$$

Strongly confirmable set is the collection of structures that cannot be observationally equivalent to any structures outside $A$ for all observed distribution $F$. Weakly confirmable set is the collection of structures that cannot be observationally equivalent to all structures outside $A$ for some observed distribution $F$. In particular,

$$\mathcal{H}_{\mathcal{P}}^{scon}(A) \subset \mathcal{H}_{\mathcal{P}}^{wcon}(A) \subset A$$

The relation of the set can be characterized by the following figure.

![Diagram showing the relations between $\mathcal{H}_{\mathcal{P}}^{scon}(A)$, $\mathcal{H}_{\mathcal{P}}^{wcon}(A)$, and $A$]

The following proposition helps to interpret the confirmable set of $A$ as the non-refutable set of $A^c$.

Proposition 2. The following operation holds:

1. $[\mathcal{H}_{\mathcal{P}}^{snf}(A)]^c = \mathcal{H}_{\mathcal{P}}^{wcon}(A^c)$

2. $[\mathcal{H}_{\mathcal{P}}^{wnf}(A)]^c = \mathcal{H}_{\mathcal{P}}^{scon}(A^c)$
3. $\mathcal{H}_P^{wcon}(\mathcal{H}_P^{wcon}(A)) = \mathcal{H}_P^{wcon}(A)$

4. $\mathcal{H}_P^{snf}(\mathcal{H}_P^{snf}(A)) = \mathcal{H}_P^{snf}(A)$

Example 1 (Continued). We now derive the non-refutable set and confirmable set of Pure Strategy Nash Equilibrium assumption (2.5). The result is stated in the following proposition.

**Proposition 3.** The pure strategy Nash Equilibrium assumption $A$ defined above satisfies:

1. $\mathcal{H}_P^{scon}(A) = \mathcal{H}_P^{wcon}(A) = \emptyset$

2. Define two difference between conditional probability

   $$
   \Delta_{11} \equiv (Pr_F(Y_m = 2|X_{1m} = 1) - Pr_F(Y_m = 2|X_{1m} = 0))
   $$

   $$
   \Delta_{00} \equiv (Pr_F(Y_m = 0|X_{1m} = 1) - Pr_F(Y_m = 0|X_{1m} = 0))
   $$

   $\mathcal{H}_P^{snf}(A)$ is the collection of $p$ such that $\forall F \in M^p(G^p)$, $F$ satisfies

   $$
   \Delta_{11}\Delta_{00} \leq 0
   $$

3. $\mathcal{H}_P^{wnf}(A) = \mathcal{P}$

The first and third statement of the proposition comes from the fact that if two structures only differs in the solution concept, i.e. $p_1 = (G, \theta, PSNE)$ and $p_2 = (G, \theta, L2R)$, the set of predicted outcome under $p_1$ is nested in $p_2$, i.e. $M^{p_1}(G) \subset M^{p_2}(G)$. Therefore, we can never confirm the solution concept is PSNE, and L2R can never be rejected. The intuition of the second implication can be seen from figure 2 below where $\beta_1 > 0$.
When $\beta_1 > 0$, the blue shaded region is where $X_1 = 1$ implies both enter while $X_1 = 0$ does not, while the orange shaded region is where $X_1 = 0$ implies neither enters while $X_1 = 1$ does not. So the change of $\Delta_{11}$ and $\Delta_{00}$ should always move in different direction.

**Example 2** (Continued). Kitagawa (2015) derives the non-refutable and confirmable set of exogenous and monotone instrument assumption (2.10). The testable implication is stated using the following two quantities:

$$P(B, d) = Pr_{F}(Y_i \in B, D_i = d | Z_i = 1)$$
$$Q(B, d) = Pr_{F}(Y_i \in B, D_i = d | Z_i = 0)$$

The result is summarized in the following proposition.

**Proposition 4.** The Imbens and Angrist exogenous and monotone instrument assumption $A$ satisfies:

1. $\mathcal{H}_{P}^{scon}(A) = \mathcal{H}_{P}^{wcon}(A) = \emptyset$
2. $\mathcal{H}_{\mathcal{P}}^{\text{snf}}(A) = \mathcal{H}_{\mathcal{P}}^{\text{scon}}(A)$ is the collection of structures $p$ such that if $F = M^p(G^p)$, then for all set $B$ that is measurable:

$$P(B, 1) \geq Q(B, 1)$$
$$Q(B, 0) \geq P(B, 0)$$

(2.14)

In the end of this subsection, I provide a characterization of an assumption set $A$

Lemma 2.1. The following three conditions are equivalent:

1. $\mathcal{H}_{\mathcal{P}}^{\text{snf}}(A) = \mathcal{H}_{\mathcal{P}}^{\text{scon}}(A)$

2. $A = \mathcal{H}_{\mathcal{P}}^{\text{snf}}(A)$

3. $A = \mathcal{H}_{\mathcal{P}}^{\text{scon}}(A)$

2.1 Binary Decision on $A$

When an empirical researcher makes assumption $A$, a commonly asked question is whether we can tell $A$ is indeed appropriate. Such binary decision are usually made after the researcher observe the distribution of outcomes $F$. However, given the feature of $A$, not all possible distribution of outcome $F$ can help us make the binary decision. When some $F$ is observed, we may still have ambiguity on whether $A$ is true or not. I provide two definitions to classify an assumption $A$ based on whether we can make precise binary decision on $A$ for some or all distributions $F$.

Definition 7. An assumption $A$ is called weakly binary detectable under $\mathcal{P}$ if there exists $F \in \mathcal{F}$ such that any of the following condition holds:

1. $F \notin \bigcup_{p^* \in A^c} M^p(G^p)$;

2. $A \cap \mathcal{P}^{-1}(F) = \emptyset$

where

$$\mathcal{P}^{-1}(F) = \{p \in \mathcal{P} : F \in M^p(G^p)\}$$

And we say $A$ can be detected by $F$. 
If condition 1 holds, it implies the true structure \( p \) that generates \( F \) must be in \( A \), since \( F \) cannot be predicted by \( A^c \), and this confirms \( p \in A \); if condition 2 holds, it implies the true structure \( p \) cannot be in \( A \), since \( F \) cannot be predicted by \( A \), and this refutes \( p \in A \). If both conditions fail, it means \( F \) can be predicted by both structures inside and outside \( A \), which creates an ambiguity on the decision. If \( A \) can be detected by any \( F \) in the structure space \( \mathcal{P} \), we say it is strongly binary detectable.

**Definition 8.** An assumption \( A \) is called strongly binary detectable under \( \mathcal{P} \) if it can be detected by any \( F \in \mathcal{F} \).

When an assumption is a strong hypothesis, then for all possible observed distribution \( F \), we can determine without ambiguity whether the true \( p \) that generates \( F \) is in \( A \). The following provides a characterization of strongly binary detectable assumption \( A \).

**Proposition 5.** An assumption \( A \) is a strong hypothesis if and only if \( H_{scon}^\mathcal{P}(A) = H_{wnf}^\mathcal{P}(A) \).

Binary decision on a set is closely related to hypothesis testing. In statistical testing, we only have a sample from the observed distribution and we make binary decision based on the sample. In binary decision problem, we get full access to the observed distribution, and we don’t have Type I and II error due to sampling error. In this sense, with precisely stated null hypothesis, we want no ambiguities under our decision rules. Otherwise, these two problems are similar in their nature. In many cases, a hypothesis can be stated as a subset of \( \mathcal{P} \). Let’s consider two types of null hypotheses:

\[
H_0 : \exists p \in A \cap \mathcal{P}^{-1}(F) \quad \text{vs} \quad H_1 : \mathcal{H}_0^c
\]

and

\[
\mathcal{H}_0 : \mathcal{P}^{-1}(F) \subset A \quad \text{vs} \quad \mathcal{H}_1 : \mathcal{H}_0^c
\]

For both problems, null set and alternative set do not intersect. The first type problem try to decide the existence of a structure in \( A \) that rationalize observation, this is the non-refutable rule; the second problem try to decide whether \( F \) can only be rationalized by structures inside \( A \), this is the confirmable rule. However, in both these two decision problems, the statements are not related to the true structure \( p_0 \) that generates \( F \). The following proposition interprets non-refutable and confirmation rule as a test on the true structure \( p_0 \).
Proposition 6. (Equivalent Decision) Let \(\theta^0\) be the true structure that generates \(F\). If \((\mathcal{H}_P^{\text{wnf}}(A) \setminus \mathcal{H}_P^{\text{snf}}(A)) \cap \mathcal{P}^{-1}(F) = \emptyset\) then the following two conditions are equivalent:

1. \(\exists p \in A \cap \mathcal{P}^{-1}(F)\)

2. The true structure \(\theta^0 \in \mathcal{H}_P^{\text{snf}}(A)\)

If \((\mathcal{H}_P^{\text{wcon}}(A) \setminus \mathcal{H}_P^{\text{scon}}(A)) \cap \mathcal{P}^{-1}(F) = \emptyset\) then the following two conditions are equivalent:

1. \(\mathcal{P}^{-1}(F) \subset A\)

2. The true structure \(\theta^0 \in \mathcal{H}_P^{\text{scon}}(A)\)

Remark 2.1. A key application of the proposition above is sub-vector decision. In many cases we are only interested in a certain feature of \(\theta(p)\), called \(\theta(p)\), and we want to test whether \(\theta(p) = \theta_0\) for some value \(\theta_0\). In this case the set \(A\) can be formulated as \(A = \{p \in A : \theta(p) = \theta_0\}\). Testing existence of \(p \in A\) such that \(\theta(p) = \theta_0\) is equivalent to the true structure \(\theta^0\) is in the strong non-refutable set of \(A\), whenever \(\mathcal{H}_P^{\text{wnf}}(A) = \mathcal{H}_P^{\text{snf}}(A)\).

Corollary 2.1. If \((\mathcal{H}_P^{\text{wcon}}(A) \setminus \mathcal{H}_P^{\text{scon}}(A)) \cap \mathcal{P}^{-1}(F) \neq \emptyset\), there exists some \(F \in \mathcal{F}\) and \(\theta^0\) that generates \(F\) such that the following two conditions are not the same:

1. \(\exists p \in A \cap \mathcal{P}^{-1}(F)\)

2. The true structure \(\theta^0 \in \mathcal{H}_P^{\text{snf}}(A) \cap \mathcal{P}^{-1}(F)\)

Note that for complete model, the condition that \((\mathcal{H}_P^{\text{wnf}}(A) \setminus \mathcal{H}_P^{\text{snf}}(A)) \cap \mathcal{P}^{-1}(F) = \emptyset\) holds automatically. Testing existence is equivalent to testing whether the true structure \(\theta^0\) is in the strongly non-refutable set. For incomplete model, if \((\mathcal{H}_P^{\text{wnf}}(A) \setminus \mathcal{H}_P^{\text{snf}}(A)) \cap \mathcal{P}^{-1}(F) \neq \emptyset\), it is possible that \(\exists p \in A \cap \mathcal{P}^{-1}(F)\) and \(\theta^0 \notin \mathcal{H}_P^{\text{snf}}(A)\) both hold. Similar situation also holds for \(\mathcal{H}_0\).

3 Identification under an Assumption

In most empirical studies, the goal is to find the value of parameter of interest rather than a class of structures that is consistent with data. Usually, these parameters of interest
directly reveal some interpretable aspects of the environment. Imposing assumptions helps
to restrict the set of data-consistent parameters, but there is a potential problem: an imposed
assumption \( A \) may be rejected by the distribution of observables. So in many empirical
studies, researchers first present some summary statistics that justifies the assumption. There
are two major problems with this approach. First, such justifications consist heuristic pre-
testing procedures of assumption \( A \), and any subsequent inference on the parameter of
interest may have incorrect size control due to pre-testing. Second, researchers do not specify
what they will do if \( A \) is rejected. Most likely they will choose the assumption that cannot
be rejected by data. This is an ex post way of choosing an assumption \( A \). Philosophically,
an assumption reflects a researcher’s prior understanding of the economic environment and
should thus independent of the data he sees. In potential outcome framework, researchers
impose monotonicity of instrument assumption because they believe such instruments should
motivate agents to choose treatment, rather than testing the assumption first and then decide
to use it.

In this section, I formalize the definition of identification problem and discuss how to
deal with an existing situation that an assumption \( A \) may be rejected by data.

**Definition 9.** An parameter of interest \( \theta \) under \((\mathcal{P}, A)\) is a function \( \theta : (G, M) \to \Theta \). An
identified set \( \Theta^{ID}(F) \) of \( \theta \) under observed distribution \( F \) is a correspondence \( \Theta^{ID} : F \rightrightarrows \Theta \)
such that

\[
\Theta^{ID}(F) = \{ \theta(p) : p \in A \cap \mathcal{P}^{-1}(F) \}
\]

We call \((\mathcal{P}, A, \theta, \Theta^{ID})\) an identification problem.

The parameter of interest can of very general form, it can be the whole structure \( p \) or a
sub-vector of \( p \) and can include counterfactual outcome. For example, if \( M^p \) is characterized
by a vector \( \beta \), and our counterfactual analysis is to find the predicted distribution of outcome
when \( \beta_1 = 0 \). Then the parameter of interest \( \theta \) can be defined as

\[
\theta(p) = M^{p^*}(G^p) \quad \text{where} \quad \beta_1^* = 0 \quad \text{and} \quad \beta_i^* = \beta_i \quad \forall i \neq 1
\]

where \( M^{p^*} \) is characterized by \( \beta^* \).
Remark 3.1. We should note that it is possible that for some $F$, $\mathcal{P}^{-1}(F) \cap A = \emptyset$, so the identified set is not well defined. In particular if $\mathcal{H}_p^{mf}(A) = \mathcal{P}$, there always exists some $F^*$ such that $\mathcal{P}^{-1}(F^*) \cap A = \emptyset$.

Example 1 (Continued). Suppose we only want to know whether $X_{1m}$ has positive or not on $\pi_{1m}$, the mapping $\theta^{ID}(p) = \beta_1^p$. The identified set of $\beta_1$ under PSNE assumption $A$ can be shown to be

$$
\Theta^{ID}(F) = \begin{cases} 
[0, \infty) & \text{if } \Delta_{11} \geq 0 \text{ and } \Delta_{00} \leq 0 \\
(-\infty, 0] & \text{if } \Delta_{11} \leq 0 \text{ and } \Delta_{00} \geq 0 \\
\emptyset & \text{o.w.}
\end{cases} \quad (3.1)
$$

There is a difference between using non-refutable assumptions in an identification problem and falsifiability in economic theory. Popper’s falsifiability criteria says any kind of scientific theory should be able to be rejected under some observed outcomes. For example, pure Nash solution concept in a one-shot $2 \times 2$ battle of sexes game can be falsified if we observed the couple do not make the same choices. By the falsifiability criteria, pure Nash solution concept is a scientific theory in battle of sexes. However, if in the lab experiment, we only observe one choice of the couple, then it is never possible to tell whether pure Nash is used or not.

Definition 10. (Popper’s Falsifiability) Let $\mathcal{K}$ be the space of joint distribution of $(\epsilon, X)$. For any $p \in \mathcal{P}$, let $H(p) : \mathcal{P} \rightarrow \mathcal{K}$ be the joint distribution of $(\epsilon, X)$ predicted by $p$. An assumption $A$ is popper falsifiable if

$$
\cup_{p \in A} H(p) \neq \cup_{p \in \mathcal{P}} H(p)
$$

Refutability is only defined on the observed distribution, while Popper’s Falsifiability is defined on the joint distribution of latent and observed variable. If an assumption is non-refutable, it does not necessarily mean that it fails popper’s falsifiability criteria. It only says that the current data observation does not allow us to reject such class of theories. The problem of identification is to find a class of theories (structures) that at least the current observations can be justified. If an assumption has empirical content, it means the researcher
will face the possibility that there is no story to tell. Ex ante, before getting the observation, the researcher should have a plan to deal with all possible outcomes that can happen under \( \mathcal{P} \).

**Example 2 (Continued).** We consider the independence assumption

\[
A' = \{ p \mid G^p \text{ s.t. } (Y_i(1), Y_i(0), D_i(1), D_i(0)) \perp Z_i \}
\]

It can be shown that \( \mathcal{H}^{snf}_{\mathcal{P}}(A) = \mathcal{P} \) holds, so \( A' \) is non-refutable. However, \( A' \) is Popper falsifiable since \((Y_i(1), Y_i(0), D_i(1), D_i(0)) \perp Z_i\) can be rejected.

For identification problem, \( \Theta^{ID} \) is the set of parameters that are compatible with data. When \( \Theta^{ID} \) is an empty set, I call it an not well defined identification problem. It can be shown that an identification problem is well defined if and only if \( A \) is strongly non-refutable.

**Definition 11.** An identification problem \((\mathcal{P}, A, \theta, \Theta^{ID})\) is well defined if \( \Theta^{ID}(F) \neq \emptyset \) holds \( \forall F \in \mathcal{F} \).

**Proposition 7.** \( \mathcal{H}^{snf}_{\mathcal{P}}(A) = \mathcal{P} \) if and only if \((\mathcal{P}, A, \theta, \Theta^{ID})\) is well defined for all parameter of interest \( \theta \).

**Proof.** When \( \mathcal{H}^{snf}_{\mathcal{P}}(A) = \mathcal{P}, A \cap \mathcal{P}^{-1}(F) \neq \emptyset \) for all \( F \). So by definition, \( \Theta^{ID}(F) = \{ \theta(p) \mid p \in A \cap \mathcal{P}^{-1}(F) \} \neq \emptyset \), holds for all \( \theta \).

Conversely, if \( \Theta^{ID}(F) = \emptyset \) for some \( \theta \) and \( F \), that means \( A \cap \mathcal{P}^{-1}(F) = \emptyset \), which means \( \mathcal{H}^{snf}_{\mathcal{P}}(A) \neq \mathcal{P} \). \( \square \)

**Example 3.** (Binary Decision As An Identification Problem)

Suppose we want to make the binary decision on assumption \( A \) using the existence rule:

\[
\mathcal{H}_0 : \exists p \in A \cap \mathcal{P}^{-1}(F) \quad \text{vs} \quad \mathcal{H}_1 : \mathcal{H}_0^c
\]

Consider an identification problem \((\mathcal{P}, \bar{A}, \theta, \Theta)\) where \( \bar{A} = \mathcal{P} \), i.e. \( \bar{A} \) does not impose any restriction. Now we consider the following parameter of interest \( \theta \):

\[
\theta(p) = \begin{cases} 
1 & \text{if } p \in A \\
0 & \text{o.w.}
\end{cases}
\]
And we define the identified set $\Theta^{ID}(F)$ as

$$\Theta^{ID}(F) = \{\theta(p) : p \in \mathcal{P}^{-1}(F) \cap \tilde{A}\}$$

Apparently $(\mathcal{P}, \tilde{A}, \theta, \Theta^{ID})$ is a well defined identification problem since $\tilde{A} = \mathcal{P}$. Then we define the decision rule

$$r^1(F) = \begin{cases} 
  \text{don't reject} & \text{if } 1 \in \Theta^{ID}(F) \\
  \text{reject} & \text{o.w.}
\end{cases}$$

Then if $r^1(F) = \text{don't reject}$, then $\mathcal{H}_0$ holds; if $r^1(F) = \text{reject}$, then $\mathcal{H}_1$ holds. Further, if $\mathcal{H}^{snf}_{P}(A) = \mathcal{H}^{wnf}_{P}(A)$ holds, $\Theta^{ID}(F)$ is a singleton for all $F$ by proposition 6, and we do not need a further decision rule $r$.

Similarly, if we want to test $\tilde{\mathcal{H}}_0 : \mathcal{P}^{-1}(F) \subseteq A$ vs $\tilde{\mathcal{H}}_1 : \mathcal{H}^{c}_0$

we can use the same identification problem $(\mathcal{P}, \tilde{A}, \theta^{ID}, \Theta^{ID})$, but we need a different decision rule:

$$r^2(F) = \begin{cases} 
  \text{reject} & \text{if } 0 \in \Theta^{ID}(F) \\
  \text{don't reject} & \text{o.w.}
\end{cases}$$

Binary decision on an assumption $A$ can always be casted into a well defined identification problem, because the story it tries to tell is either yes or no.

In the following sections, I propose a way of transform an ill-defined identification problem $(\mathcal{P}, A, \theta, \Theta^{ID})$ to a well-defined problem and at the same time preserving features of $A$. The key is to extend the existing assumption $A$ to a larger set.

### 3.1 Extended Assumption Approach

For some assumptions $A$, the associated identification problem may not be well defined. When $\Theta^{ID} = \emptyset$, empirical researchers do not have a sound interpretation of the identified set. Since this issue results from $\mathcal{H}^{snf}_{P}(A) \neq \mathcal{P}$, a natural thing is to choose another set of assumption $\tilde{A}$ such that $\mathcal{H}^{snf}_{P}(\tilde{A}) \neq \mathcal{P}$. In this section, I discuss the criteria in choosing a relaxed assumption $\tilde{A}$ and some issues that will happen in incomplete models.
Definition 12. Given a not well defined identification system \((\mathcal{P}, A, \theta, \Theta^{ID})\), we call \(\tilde{A}\) a well defined extension if \(A \subset \tilde{A}\) and \(\mathcal{H}_{\mathcal{P}}^{unf}(\tilde{A}) = \mathcal{P}\). We further call \(\tilde{A}\) a strong consistent extension if \(\tilde{A} \cap \mathcal{H}_{\mathcal{P}}^{unf}(A) = A\).

By proposition 7 above, if \(\tilde{A}\) is a well defined extension, then for all identification problem, \(\Theta^{ID}(F) \neq \emptyset\).

Definition 13. Given a not well defined identification system \((\mathcal{P}, A, \theta, \Theta^{ID})\), suppose \(\tilde{A}\) is a well defined extension of \(A\). Let \(\tilde{\Theta}^{ID}(F)\) be the associated identified set
\[
\tilde{\Theta}^{ID}(F) = \{\theta(p) : p \in \mathcal{P} \cap \tilde{A}\}
\]

We call \(\tilde{A}\) a \(\theta\) - consistent extension if for any \(F \in \mathcal{F}\), if \(\Theta^{ID}(F) \neq \emptyset\), then \(\Theta^{ID}(F) = \tilde{\Theta}^{ID}(F)\)

The following proposition shows that strong consistent extension is a stronger restriction than \(\theta\) - consistent.

Proposition 8. For any parameter \(\theta\), if \((\mathcal{P}, A, \theta, \Theta^{ID})\) is not a well defined identification problem, let \(\tilde{A}\) be a strong consistent extension, then \(\tilde{A}\) is \(\theta\) - consistent.

Proof. Suppose not, for this \(\theta\) there exists \(F\) such that \(\tilde{\Theta}^{ID}(F) \setminus \Theta^{ID}(F) \neq \emptyset\). This means there exists some \(p \notin A\) and \(p \in \tilde{A}\) such that \(F \in M^p(G^p)\). By definition, this shows \(p \in \mathcal{H}_{\mathcal{P}}^{unf}(A)\). This implies \(\mathcal{H}_{\mathcal{P}}^{unf}(A) \cap \tilde{A} \neq A\), so \(\tilde{A}\) is not strong consistent.

Strong consistent extension achieves the goal of preserving features of \(A\). If we view the parameter of interest \(\theta\) as a projection of structure \(p\), strong consistent extension \(\tilde{A}\) gives the same projection for any \(\theta\) when \(A\) holds, and it gives a non-empty set when \(A^c\) holds. Sometimes we are only interested in a particular \(\theta\), it suffices to choose \(\tilde{A}\) to be \(\theta\) - consistent. However, sometimes we may not be able to find a strong extension of \(A\) in the incomplete structure space. For example, suppose the following two conditions hold:

1. \(\mathcal{H}_{\mathcal{P}}^{unf}(A) \setminus \mathcal{H}_{\mathcal{P}}^{unf}(A) \neq \emptyset\)
2. \((\bigcup_{p \in \mathcal{H}_{\mathcal{P}}^{unf}(A)} M^p(G^p)) \cap (\bigcup_{p \in \mathcal{H}_{\mathcal{P}}^{unf}(A)^c} M^p(G^p)) = \emptyset\)
In this case, by the second condition, there exists some $F \in \mathcal{F}$ that can only be generated by structures in $\mathcal{H}_P^{wnf}(A) \setminus \mathcal{H}_P^{snf}(A)$. In this case, if we want to have a well defined extension $\tilde{A}$ such that $\tilde{A} \cap \mathcal{P}^{-1}(F) \neq \emptyset$, $\tilde{A} \cap \left(\mathcal{H}_P^{wnf}(A) \setminus \mathcal{H}_P^{snf}(A)\right) \neq \emptyset$ must hold. This means $\tilde{A} \cap \mathcal{H}_P^{wnf}(A) \neq A$. On the other hand, if our model space is complete, there always exists a strong extension.

**Corollary 3.1.** If $\mathcal{P}$ is a complete structure universe, then $\tilde{A} = A \cup \mathcal{H}_P^{snf}(A)^c$ is a strong extension of $A$.

**Proof.** If $F \notin \cup_{p \in A \cap \mathcal{P}} M_p(G_p)$, then $F \notin \cup_{p \in \mathcal{H}_P^{snf}(A)} M_p(G_p)$. Since $F \in \cup_{p \in \mathcal{P}} M_p(G_p)$, it means $F \in \cup_{p \in \mathcal{H}_P^{snf}(A)^c} M_p(G_p)$. This shows $\mathcal{H}_P^{snf}(A) = \mathcal{P}$, so $\tilde{A}$ is an well defined extension.

Then $\tilde{A} \cap \mathcal{P}^{wnf}(\tilde{A}) = \tilde{A} \cap \mathcal{P}^{snf}(\tilde{A}) = A$ by construction. \hfill $\Box$

While strongly consistent extension may not exist for incomplete structure space, $\theta$ – consistent extension can be found for some cases.

**Example 1** (Continued). By discussion above, for the PSNE assumption $A$, by proposition 3 we have $\mathcal{H}_P^{wnf}(A) = \mathcal{P} \neq \mathcal{H}_P^{snf}(A)$, so both

1. $\mathcal{H}_P^{wnf}(A) \setminus \mathcal{H}_P^{snf}(A) \neq \emptyset$

2. $\left(\cup_{p \in \mathcal{H}_P^{wnf}(A)} M_p(G_p)\right) \cap \left(\cup_{p \in [\mathcal{H}_P^{wnf}(A)]^c} M_p(G_p)\right) = \emptyset$

hold in this example. Therefore, there is no strong consistent extension of assumption $A$. However, if we are only interested in $\beta_1$, the effective direction of $X_{1m}$, we can find a $\beta_1$ – consistent extension. Proposition 3 implies that PSNE assumption cannot generate data distribution such that $\Delta_{11}\Delta_{00} > 0$. We first consider an extension $\tilde{A}$ can generate any $\Delta_{11}\Delta_{00} > 0$, and then justify $\tilde{A}$ is $\beta_1$ – consistent.

I use notation $\Delta_{ii}^F$ to denote the change of conditional probability under $F$ in equation (2.11). First, let’s consider any $F$ such that $\Delta_{11}^F \geq \Delta_{00}^F > 0$. This class of $F$ cannot be generated by PSNE assumption $A$. For this type of $F$, we consider a structure $p$ such that $S^p = L2R$, $\theta^p = (0, 0, 1, -1, -1)$, so

$$A((1, 1)|X_{1m} = 0) \subset A((1, 1)|X_{1m} = 1); \quad A((0, 0)|X_{1m} = 1) \subset A((0, 0)|X_{1m} = 0)$$
and we construct $G^p(\epsilon)$ in the following way:

\[
\int_{A((1,0)|X)} g^p(\epsilon)d\epsilon = \Pr_F(Y_m = 2|X_{1m} = 0)
\]

\[
\int_{A((0,0)|X)} g^p(\epsilon)d\epsilon = \Pr_F(Y_m = 0|X_{1m} = 0)
\]

\[
\int_{A((1,1)|X) \setminus A((1,1)|X_{1m} = 0)} g^p(\epsilon)d\epsilon = \Delta_{11}
\]

\[
\int_{A((0,0)|X) \setminus A((0,0)|X_{1m} = 0)} g^p(\epsilon)d\epsilon = 0
\]

$g^p(\epsilon) = 0$ for $\epsilon \in (A(M|X_{1m} = 0) \setminus A(M|X_{1m} = 1)) \cup (A(M|X_{1m} = 1) \setminus A(M|X_{1m} = 0))$

\[
\int_{A(M|X_{1m} = 1) \cap A(M|X_{1m} = 0)} g^p(\epsilon)d\epsilon = \Delta_{00}
\]

\[
\int_{A((1,0)|X_{1m} = 1) \cup A((0,1)|X_{1m} = 0)} g^p(\epsilon)d\epsilon = \Pr_F(Y_m = 1|X_{1m} = 1)
\]

The support of $G^p$ can be seen in figure 3.

By choosing

\[
c_0^2 = c_0^0 = 0, \quad c_0^1 = \Delta_{00} \times \Pr(X_{1m} = 0)
\]

\[
c_1^2 = c_1^0 = 0, \quad c_1^0 = \Delta_{00} \times \Pr(X_{1m} = 1)
\]
we can show \( F \in M^p(G^p) \).

Under this \( p \), for any other \( \tilde{F} \in M^p(G^p) \), we must have

\[
\Delta_{11}^{\tilde{F}} \equiv Pr_{\tilde{F}}(Y_m = 2|X_{1m} = 1) - Pr_{\tilde{F}}(Y_m = 2|X_{1m} = 0) \in [0, \Delta_{11}^F - \Delta_{00}^F]
\]

Similarly can construct different class of \( p \) such that \( \Delta_{11}^F \Delta_{00}^F > 0 \), as shown in the following table:

| \( F \) | \( \int_{A(M|X=1) \cap A(M|X=0)} g_p(\epsilon) d\epsilon \) | \( \beta_1^p \) | \( M^p(G^p) \) also includes | \( \Theta^{ID}(p) \) under \( A \) |
|---|---|---|---|---|
| \( \Delta_{11} \geq \Delta_{00} > 0 \) | \( \Delta_{00} \) | 1 | \( \Delta_{11} \geq 0 \) | \( \Delta_{00} \leq 0 \) |
| \( \Delta_{00} \geq \Delta_{11} > 0 \) | \( \Delta_{11} \) | -1 | \( \Delta_{11} \leq 0 \) | \( \Delta_{00} \geq 0 \) |
| \( \Delta_{11} \leq \Delta_{00} < 0 \) | \( |\Delta_{00}| \) | -1 | \( \Delta_{11} \leq 0 \) | \( \Delta_{00} \geq 0 \) |
| \( \Delta_{00} \leq \Delta_{11} < 0 \) | \( |\Delta_{11}| \) | 1 | \( \Delta_{11} \geq 0 \) | \( \Delta_{00} \leq 0 \) |

The fourth column lists the class of \( \tilde{F} \) that is also in \( M^p(G^p) \). The last column report the original identified set \( \Theta^{ID}(p) \) under \( A \) if the situation in the fourth column \( M^p(G^p) \) holds.

Now we can show \( \tilde{A} \) is \( \beta_1 \) – consistent. For example, when \( F \) satisfies \( \Delta_{11}^F \geq 0 \) and \( \Delta_{00}^F \leq 0 \), \( \Theta^{ID}(F) = [0, \infty) \) under \( A \). In the construction of \( \tilde{A} \), only \( p \) in the first and fourth row will generates \( \Delta_{11}^F \geq 0 \) and \( \Delta_{11}^F \leq 0 \). In both cases, \( \beta_1^F = 1 \). So the identified set under \( \tilde{A} \) is \( \tilde{\Theta}^{ID}(F) = [0, \infty) \cup \{1\} = \Theta^{ID}(F) \). The identified set also does not change for \( \Delta_{11}^F \leq 0 \) and \( \Delta_{00}^F \geq 0 \).

### 3.2 Structure Completion

We have seen that for an incomplete structure space, binary decision of existence of structures in an assumption \( A \) potentially cannot be rewritten as binary decision of whether the true structure \( p_0 \) is in \( H_{snf}^p(A) \); for an ill defined identification problem \( (P, \tilde{A}, \theta, \Theta^{ID}) \), there may
not exists a strongly consistent extension. This is because in incomplete structures, we are agnostic about how the distribution of outcomes are selected, so we cannot separate a structure $p$ that generates both $F_1$ and $F_2$ into two structures.

**Definition 14.** A completion $C$ of structure $p = (M^p, G^p)$ is a mapping $C : 2^F \setminus \{\emptyset\} \rightarrow F$ and $p^* = (M^*_C, G^p)$ where $M^*_C = C \circ M^p$ is a complete structure. Let $C(p)$ be the collection of completion corresponding to $p$. We call

$$\mathcal{P}^* = \{(M^*_C, G^p) : p \in \mathcal{P} \text{ and } C \in C(p)\}$$

(3.2)

the completion of $\mathcal{P}$.

The definition above consider all possible completion $C$. Intuitively, if $M^p(G^p)$ has two distribution, then the completion procedure simply separate them and put them into two complete structures. The key property is that for any identification problem, the identified set is preserved if parameter of interest in the completed model is properly defined in the following way.

**Proposition 9.** Let $(\mathcal{P}, A, \theta, \Theta^{ID})$ be any identification problem, and let $(\mathcal{P}^*, A^*, \theta^*, \Theta^{*ID})$ be the corresponding completed identification problem such that

$$\mathcal{P}^*$$ is the completion of $\mathcal{P}$

$$A^* = \{(M^*_C, G^p) : p \in \mathcal{P} \cap A \text{ and } C \in C(p)\}$$

(3.3)

Then $\Theta^{ID}(F) = \Theta^{*ID}(F)$ for all $F$.

**Example 1** (Continued). In binary entry game setting, $M^p(G^p)$ is a singleton when $S^p = PSNE$. We only need to find completions for $S^p = L2R$, and the set $M^p(G^p)$ is explicitly defined in (2.2). Therefore, a completion $C$ can be explicitly defined by a $\mathbb{R}^6$ vector: $(c_0^2, c_0^1, c_0^0, c_1^2, c_1^1, c_1^0)$. In the general framework, I do not specify the meaning of a completion, but in this example it can be explained as a random equilibrium selection mechanism that depends on $X_{1m}$. Consider a random variable $\eta_m$ supported on $\{0, 1, 2\}$ such that $Pr(\eta_m = i | X_{1m} = j) = c^i_j / Pr(X_{1m} = j)$. When $L2R$ cannot determine in market $m$ what is
the outcome, nature selects the outcome according to the realization of $\eta$. Then this $\eta$ as the equilibrium selection mechanism generates the same $F$ as our completion $C$.

Under this completion, we can give a strongly consistent extension of $A^*$, where $A^*$ is the assumption of PSNE in the completed universe. Let $\tilde{A}$ be the $\beta_1$-consistent extension of $A$ under $P$, and let $\tilde{A}^*$ be the corresponding assumption under $P^*$. Recall that the construction of $\tilde{A}$ depends on the four mutually exclusive situations of $(\Delta_{11}, \Delta_{00})$, we consider eliminating $p^*$ in the completed structures separately.

Now we eliminate $p^* \in \tilde{A}^* \cap \{p : \exists F \in M^p(G^p) \text{ s.t. } \Delta_{11}^F \geq \Delta_{00}^F > 0\}^*$ such that

$$S^p = L2R \quad \frac{c^0_0}{P(X_{1m} = 0)} - \frac{c^0_1}{P(X_{1m} = 1)} > 0$$

(3.4)

Call this set $\tilde{A}^*$. Then for all $p^* \in \tilde{A}^* \cap \{p : \Delta_{11}^F \geq \Delta_{00}^F \geq 0\}^*$, it only generates distribution such that $\Delta_{00} > 0$ because $\Delta_{00} = \frac{c^1_0}{P(X_{1m} = 1)} - \frac{c^0_0}{P(X_{1m} = 0)}$ under $p^*$. The eliminations for other cases are

| $p^*$ set | Eliminate | $\beta_1$ | $M^p(G^p)$ only contains |
|-----------|-----------|-----------|-------------------------|
| $\Delta_{11} \geq \Delta_{00} > 0$ | $c^0_0 - c^0_1 \geq 0$ | 1 | $\tilde{\Delta}_{11} > 0$ |
| $\Delta_{00} \geq \Delta_{11} > 0$ | $c^2_0 - c^2_1 \geq 0$ | -1 | $\tilde{\Delta}_{11} < 0$ |
| $\Delta_{11} \leq \Delta_{00} < 0$ | $c^0_0 - c^0_1 \leq 0$ | -1 | $\tilde{\Delta}_{11} < 0$ |
| $\Delta_{00} \leq \Delta_{11} < 0$ | $c^2_0 - c^2_1 \geq 0$ | 1 | $\tilde{\Delta}_{11} > 0$ |

where $c^i_i = c^i_j / P(X_{1m} = i)$. We can give nice interpretation to $\tilde{A}^*$ using the equilibrium selection mechanism. For example the structure $p^*$ such that $c^0_0 - c^0_1 < 0$ implies that $X_{1m}$ has positive effect on selecting no entry outcome, i.e. $Pr(\eta = 0 | X_{1m} = 1) > Pr(\eta = 0 | X_{1m} = 0)$.

As shown in the example, one way to construct a strongly consistent extension $\tilde{A}^*$ is to start with a consistent extension $\tilde{A}$ and then eliminates certain completed structures by
looking at the corresponding completion $C$. However, it can be hard to justify why some particular completions are eliminated. In our example, the eliminated set can be interpreted as $X_{1m}$’s effect on equilibrium selection mechanism, but in other cases the shape of the eliminated set can be hard to justify using economic explanation. So our suggestion to deal with incomplete structure space is to first find a $\theta$ – consistent extension $\tilde{A}$ first, if such set does not exists and then try to find an strongly consistent extension $\tilde{A}^*$ that has clear economic interpretation in the completed space, otherwise, use a consistent extension $\tilde{A}$ on the incomplete space.

### 3.3 Choice of Extended Assumptions

For an ill defined identification problem $(\mathcal{P}, A, \theta, \Theta^{ID})$, there can be multiple strongly consistent extensions $\tilde{A}_1, ... \tilde{A}_l$. Indeed, if $\tilde{A}_1$ is a strongly consistent extension, then $\tilde{A}_2 = \tilde{A}_1 \cup \{p\}$, where $p \notin \mathcal{H}_P^{snf}(A)$, is also a strongly consistent extension. By construction, these extensions agree on set $A$, but differ on $A^c$. In this section, I discuss which extension should we choose, and how should we interpret the extended identification problem.

We should first note that any parameter of interest $\theta$ should be interpreted under the corresponding assumption. In an identification problem, $\mathcal{P}$ indicates the empirical framework we are looking at, $A$ indicates the restrictions that we impose on this particular empirical environment, and the identified set $\Theta^{ID}$ are constructed using $A$. An identified set $\Theta^{ID}$ is the set of $\theta$ under $A$ that can rationalize our data. If $A_1 \subset A_2$, the difference of the identified set can be interpreted as the degree of relaxation in terms of $\Theta^{ID}$. However, if two assumptions $A_1$ and $A_2$ are not nested, comparison of identified sets from these two identification problems is not meaningful, because identified set $\Theta^{ID}$ is assumption specific.

A natural approach is either to seek the maximal or minimal extension of $\tilde{A}$, if such sets exists. Indeed, if we collect all possible strongly consistent extension, it forms a partial ordered set $(\mathcal{A}, \subset)$. We can show that the maximal extension is $A \cup \mathcal{H}_P^{snf}(A)^c$.

**Proposition 10.** When $\mathcal{H}_P^{snf}(A) = \mathcal{H}_P^{wnf}(A)$, the lattice $(\mathcal{A}, \subset)$ exists and is bounded above by $\tilde{A}^{max} = A \cup \mathcal{H}_P^{snf}(A)^c$. If $\mathcal{P}$ is complete, then there exists a minimal extension $\tilde{A}^{min}$ such that no $\tilde{A} \in \mathcal{A}$ such that $\tilde{A} \subset \tilde{A}^{min}$. 

27
If we use the maximal set $\tilde{A}^{\text{max}}$, we put least constraints beyond $A$, and if we use the minimal set $\tilde{A}^{\text{min}}$, we put most constraints beyond $A$. Both sets seem appealing, but the economics behind them can be hard to justify. We not only want the identified set $\Theta^{ID}$ to be consistent with $A$, we also want the economic logic behind $A$ and $\tilde{A}$ to be consistent. For example, in the binary entry game example, we do not impose any sign restriction on $\beta_1$. If one extension $\tilde{A}^*$ entails $\beta^p > 0$ on $p \in \tilde{A} \setminus A$, it would be logically inconsistent. This is because on $A$ we allow any effect of $X_1$ but only allow effect of $X_1$ to be positive. We will explore this criterion in the application to potential outcome framework. On the other hand, using minimal or maximal set may cause discontinuity of identified set. As we will see later, this may cause inconsistency in estimation procedure. The following lemma says that $\theta$ is point identified under $\tilde{A}^{\text{min}}$ if $A$ is rejected.

**Lemma 3.1.** Let $\tilde{A}^{\text{min}}$ be a minimal extension, then for identification problem $(P, \tilde{A}^{\text{min}}, \theta, \tilde{\Theta}^{ID}(F))$:

$$\text{card}(\tilde{\Theta}^{ID}(F)) = 1 \quad \forall F \notin \bigcup_{p \in A} \mathbb{M}^p(G^p)$$

where $\text{card}(\cdot)$ is the cardinality of a set.

**Proof.** Suppose not. Then there exists $F$ such that $\theta_1 \neq \theta_2$ and $\theta_1, \theta_2 \in \tilde{\Theta}^{ID}(F)$. Let $p_1, p_2$ be the structures such that $\theta(p_i) = \theta_i$. Then $\tilde{A} = \tilde{A}^{\text{min}} \setminus p_1$ is a well defined extension since $F \in \bigcup_{p \in \tilde{A}} \mathbb{M}^p(G^p)$ and this violates $\tilde{A}^{\text{min}}$ being minimal.

**Example 2 (Continued).** We define the sup-norm $\| \cdot \|_\infty$ on the space of distribution of $F(Y_i, D_i, Z_i)$, and the induced metric

$$d(F_1, F_2) = \sup_{y, d, z} |Pr_{F_1}(Y_i \leq y, D_i = d, Z_i = z) - Pr_{F_2}(Y_i \leq y, D_i = d, Z_i = z)|$$

We consider the identification problem $(P, \tilde{A}^{\text{min}}, \theta^{ATE}, \tilde{\Theta}^{ID})$, where $\tilde{A}^{\text{min}}$ is any minimal strongly consistent extension, and the average treatment effect $\theta^{ATE}(p) = E[Y_i(1) - Y_i(0)]$. It can be shown that if $F$ satisfies (2.14),

$$\tilde{\Theta}^{ID}(F) = (-\infty, \infty)$$

Now, let’s consider a set $B = [-1, 1]$ and $F^*$ satisfying (2.14) such that the density $p^{F^*} (y, j) = \ldots$
\( \frac{\partial F^* (Y_i \leq y, j)}{\partial y} \) and \( q^F (y, j) \) exists, and \( \forall y \in [-1, 1] \) there exists some \( \delta > 0 \):

\[
\begin{align*}
    p^F (y, 1) &= q^F (y, 1) > \delta \\
    q^F (y, 0) &= p^F (y, 0)
\end{align*}
\]

Heuristically, this \( F^* \) is on the border of \( A \). For \( \epsilon > 0 \) small enough, we construct \( F^\epsilon \) such that

\[
P r_{F^\epsilon} (D_i = j, Z_i = j) = P r_{F^*} (D_i = j, Z_i = j) \quad i, j = 0, 1
\]

otherwise \( F^\epsilon \) agrees with \( F^* \). The construction of \( F^\epsilon \) shifts measures of \( Q^F (y, 1) \) from \([-1, 0] \) to \([0, 1] \), so \( F^\epsilon \) is a well defined probability measure, and \( ||F^* - F^\epsilon|| < \epsilon \). However, by construction:

\[
P^F (B, 1) - Q^F (B, 1) < 0 \quad \forall B \subset [0, 1]
\]

so \( F^\epsilon \) will reject \( A \). By Lemma 3.1, \( \hat{\Theta}^{ID} (F^\epsilon) \) is a singleton holds for all \( \epsilon > 0 \). On the other hand \( \hat{\Theta}^{ID} (F^*) = (-\infty, \infty) \). This implies that \( \hat{\Theta}^{ID} (F) \) fails to be lower hemicontinuous at \( F^* \).

If the identified set is not continuous at \( F \), it means that if we change \( F \) a little bit, the identified set may change abruptly. If we consider the uncertainty from sampling procedure, this will cause problem for estimation and hypothesis testing.

To sum up, when we consider an assumption extension \( \tilde{A} \), the first criterion is to consider the logic consistency of \( A \) and \( \tilde{A} \), then we choose \( \tilde{A} \) such that the identified set \( \hat{\Theta}^{ID} \) is continuous in \( F \). If multiple extensions are available, we can then interpret the identified set differently under different extensions.

4 Estimation and Confidence Region

In empirical studies, researchers only have a sample drown from the distribution of outcomes and need to get an estimator of the identified set \( \hat{\Theta} \) or to do hypothesis testing on \( \Theta^{ID} \). If the identified set can be characterized by moment inequalities, then estimation and hypothesis testing method in [Chernozhukov et al. (2007)] can be applied directly. In this section, I
provide an estimation and hypothesis testing procedure using a modified dilation method from Galichon and Henry (2013). Compared with their method, my estimation and inference procedure can easily deal with \( X \) that is of higher dimension.

The dilation method is closely related to the construction of estimated identified set in CHT. In CHT, each parameter \( \theta \in \Theta^D \) implies \( Q(\theta) = E(m(X_i, \theta))_+ = 0 \), and the estimated identified set is constructed by collecting all \( \theta \) such that \( Q_n(\theta) = (\frac{1}{N} \sum_i m(X_i, \theta))_+ \leq c_n \). Heuristically, they dilate the population criteria function \( Q \) with a number \( c_n \) to accommodate the sampling uncertainty in \( Q_n \), since the identified set is characterized by moment condition \( E[m(X_i, \theta)] \leq 0 \). In my framework, the identified set is characterized by matching the model predicted distribution \( M_p(G_p) \) with the observed distribution \( F \). So, the natural choice is to dilate the model predicted set \( M_p(G_p) \) to accommodate the sampling uncertainty in \( F_n \).

Let \( F^d \) be the collection of all empirical distribution supported on \( supp(X) \). My estimation and inference methods are based on dilating \( M_p(G_p) \) properly and check whether the observed empirical distribution \( F_n \) falls into the dilated set.

**Definition 15.** A dilation is a correspondence \( J : \mathcal{F} \rightharpoonup \mathcal{F} \cup \mathcal{F}^d \). We call \( J_n^\alpha : \mathcal{F} \rightharpoonup \mathcal{F} \cup \mathcal{F}^d \) a sequence of \( \alpha \)-stochastic dilation if conditioned on the sample \( (X_1, ... X_n) \sim F \) with probability \( 1 - \alpha_n \) such that \( \limsup_n \alpha_n \leq \alpha \), the following hold:

\[
Pr(F_n \notin J_n^\alpha(F)) \leq \alpha_n
\]

To slightly abuse notation, we let \( J(M_p(G_p)) = \cup_{F \in M_p(G_p)} J(F) \).

### 4.1 Estimation

The estimated identified set use the following dilation:

\[
J_{c_n}(F) = \{ F^* \in \mathcal{F} \cup \mathcal{F}^d \mid \hat{d}_F(F^*, F) < c_n \}
\]

where \( \hat{d}_F \) is some metric defined on \( \hat{F} \equiv \mathcal{F} \cup \mathcal{F}^d \). For example, when \( \hat{d}_F(F^*, F) = \sup_x ||F^*(x) - F(x)||_\infty \equiv ||F^* - F||_\infty \), the dilation can be written as

\[
J_{c_n}(F) = \{ F^* \in \mathcal{F} \cup \mathcal{F}^d \mid F^*(x) \in [F(x) - c_n/\sqrt{n}, F(x) + c_n/\sqrt{n}] \quad \forall x \in \text{Supp}(X) \}
\]
Estimated identified set is characterized by
\[ \hat{\Theta}^{ID} = \{ \theta(p) | p \in A, \quad \mathbb{P}_n \in J_{c_n}(M^p(G^p)) \} \]

The estimated identified set can also be written as a contour set with respect to \( d_{\tilde{\mathcal{F}}} \):
\[ \hat{\Theta}^{ID} = \{ \theta(p) \mid \inf_{F \in \mathcal{F}^p} d_{\tilde{\mathcal{F}}}(\mathbb{P}_n, F) < c_n/\sqrt{n} \} \]

The idea is very similar to the methods in CHT, where their contour set is formed with respect to the sample criteria function \( Q_n \). The analysis of consistency is based on the following Hausdorff distance between sets:
\[ d_H(A, B) = \max \{ \sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A) \} \]

I assume the following conditions hold:

**Assumption 4.1.** Let \( F \) be the true distribution of outcomes and \( \Theta^{ID}(F) \) is the identified set. Let \( d_{\tilde{\mathcal{F}}} \) be a semi-metric on \( \mathcal{F} \cup \mathcal{F}^d \).

1. The support of \( X \): \( \text{supp}(X) \) is compact in \( \mathbb{R}^d \).
2. \( \Theta^{ID}(F) \) is upper hemi-continuous at \( F \)
3. There exists a sequence of \( a_n \) such that \( c_n = \sqrt{a_n} d_{\tilde{\mathcal{F}}}(\mathbb{P}_n, F) = O_p(1) \). The sequence of constant \( c_n \geq C_n \) holds with probability converging to 1, and \( c_n/\sqrt{a_n} \to 0 \).

For example, when \( d_{\tilde{\mathcal{F}}} = ||\cdot||_\infty \), we can take \( a_n = n, c_n = \log n \). The following proposition states the consistency result.

**Proposition 11.** Under assumption [4.1], \( \Theta^{ID}(F) \subset \hat{\Theta}^{ID} \) with probability approaching 1, and \( d_H(\Theta^{ID}(F), \hat{\Theta}^{ID}) = o_p(1) \)

**Remark 4.1.** In most cases, we can take the metric \( d_{\tilde{\mathcal{F}}} \) to be the sup norm \( ||\cdot||_\infty \). However, if the identified set involves density of \( F \), then we will use other norms sup-norm on the density. However, density of \( \mathbb{P}_n \) is not well defined, in this case we can use the following metric:
\[
\begin{align*}
d_{\tilde{\mathcal{F}}}(F_1, F_2) &= \sup_x |f_1(x) - f_2(x)| \quad \text{if} \quad F_1, F_2 \in \mathcal{F} \\
d_{\tilde{\mathcal{F}}}(\mathbb{P}_n, F) &= \sup_x |f_n(x) - f(x)| \quad \text{if} \quad F \in \mathcal{F}, \quad \mathbb{P}_n \in \mathcal{F}^d
\end{align*}
\]
where \( f_n \) is an estimated density from \( F_n \) using kernel density estimation or sieve regression. For example, when \( \text{dim}(X) = 1 \), and we estimate \( f_n \) using kernel density method with bandwidth \( h_n = n^{-1/5} \), we can take \( a_n = n^{4/5} \), and \( c_n = \log n \).

### 4.2 Confidence Region

In this section, I derive a general method to construct confidence region for the identified set \( \Theta^{ID} \). We need to use a sequence of \( \alpha \)-stochastic dilation to construct confidence region.

**Proposition 12.** Given an identification problem \((\mathcal{P}, A, \theta, \Theta^{ID})\), suppose assumption 4.1 holds, and there exists a sequence of \( \alpha \)-stochastic dilation \( J_{n}^{\alpha} \), then

\[
\Theta_{n}^{\alpha} = \{ \theta(p) \mid p \in A, F_n \in J_{n}^{\alpha}(M^p(G^p)) \}
\]

is a valid \( \alpha \)-confidence region for the identified set \( \Theta^{ID} \).

#### 4.2.1 Construction of \( J_{n}^{\alpha} \) when the metric is induced by \( \| \cdot \|_{\infty} \)

Construction of \( J_{n}^{\alpha} \) can be constructed using bootstrap method. Let \((X_1, \ldots, X_n)\) be the original sample and \((X^b_1, \ldots, X^b_n)\) be the bootstrapped sample. Let \( \bar{F}^b_n \) be the empirical distribution for the bootstrapped sample. For each bootstrapped sample, I calculate \( \eta_n^b = \sup_{x \in \text{supp}(X)} |\sqrt{n}(\bar{F}^b_n(x) - F_n(x))| \), and let \( c^*(\alpha) \) be the \((1 - \alpha)\)-quantile of \( \{\eta_n^b\}_{n=1}^{B} \), where \( B \) is the number of bootstrap simulations.

**Proposition 13.** Under assumption 4.1, the following dilation

\[
J_{n}^{\alpha}(F) = \left\{ F^* \in \hat{\mathcal{F}} \mid F^*(x) \in [F(x) - c^*(\alpha)/\sqrt{n}, F(x) + c^*(\alpha)/\sqrt{n}] \quad \forall x \in \text{supp}(X) \right\}
\]

is a sequence of \( \alpha \)-stochastic dilation.

To implement the method, we first construct \( c^*(\alpha) \) using bootstrap method. To check whether \( F \in J_{n}^{\alpha}(F) \), it suffice to check for each \( x = X_1, \ldots, X_n \)

\[
F_n(x) \in [F(x) - c^*(\alpha)/\sqrt{n}, F(x) + c^*(\alpha)/\sqrt{n}]
\]

The difficulty comes when we need to go through all \( F \in M^p(G^p) \). When \( M^p(G^p) \) is a convex set, it suffice to check all extreme points of \( M^p(G^p) \). This is because, if \( F_n \notin [F_j(x) - c^*(\alpha)/\sqrt{n}, F_j(x) + c^*(\alpha)/\sqrt{n}] \) for \( j = 1, 2 \), and \( F_3 = kF_1 + (1 - k)F_2 \) for some \( k \in [0, 1] \), then \( F_n \notin [F_3(x) - c^*(\alpha)/\sqrt{n}, F_3(x) + c^*(\alpha)/\sqrt{n}] \).
5 Application: Identification of LATE in Potential Outcome Framework

I now revisit the potential outcome framework with binary instrument. The structure space \( \mathcal{P} \) is specified by (2.6)-(2.9), and the Imbens and Angrist monotonicity assumption \( A \) is given by (2.10). Under (2.10), the local average treatment effect (LATE) conditioned on the compliance type is identified as

\[
LATE \equiv E[Y_i(1) - Y_i(0) | D_i(1) = 1, D_i(0) = 0] = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}
\]

Our parameter of interest \( \theta \) is the local average treatment effect, and the identification problem is characterized by \( (\mathcal{P}, A, LATE, LATE^{ID}) \). As shown in proposition 4, the identification problem \( (\mathcal{P}, A, LATE, LATE^{ID}) \) is not well defined since \( \mathcal{H}_{\mathcal{P}}^{\text{nf}}(A) \neq \mathcal{P} \). In most empirical applications, researchers do not test this implication, neither do they specify what should be done when the testable implication is rejected. In this section, I use the extended assumption approach to find relaxed assumptions \( \tilde{A} \) such that \( (\mathcal{P}, \tilde{A}, LATE, LATE^{ID}) \) is well defined, find the identified set under \( \tilde{A} \), and discuss the estimation and inference on \( LATE \) under \( \tilde{A} \).

5.1 Minimal Defier as Strong Extension

Recall \( \mathcal{P} \) is a complete structure space, so finding a strong extension of \( A \) is possible. In this section, I give a strong extension \( \tilde{A} \) of \( A \) that preserve the logic of 'No Defier': there should be as least defier in the data generating process as possible. This restriction can be rationalized by a two step procedures that the researcher first chooses an model through some model selection criteria, and then he estimates the object of interest, which is LATE, under the selected model. I first define the measure of defier in \( G^p \).

**Definition 16.** The measure of defier for any \( p \in \mathcal{P} \) is defined as

\[
m^d(p) = E_{G^p} [ \mathbb{1}\{D_i(1) = 0, D_i(0) = 1\}]
\]

**Assumption 5.1.** (Minimal Defiant with Type Independent Instrument) Let \( \tilde{A} \) be the class of \( p \) such that

\[
33
\]
1. Type Independence: \((Y_i(1), Y_i(0)) \perp Z_i \mid D_i(1), D_i(0)\) holds for \(p \in \tilde{A}\)

2. The measure of always taker and never taker is independent of \(Z_i\) for all \(p \in \tilde{A}\):

\[
E_{G^p}[\mathbb{1}(D_i(1) = D_i(0) = 1) \mid Z_i = 1] = E_{G^p}[\mathbb{1}(D_i(1) = D_i(0) = 1) \mid Z_i = 0]
\]

\[
E_{G^p}[\mathbb{1}(D_i(1) = D_i(0) = 0) \mid Z_i = 1] = E_{G^p}[\mathbb{1}(D_i(1) = D_i(0) = 0) \mid Z_i = 0]
\]

(5.2)

3. Minimal Defier: for any \(p^*\) that satisfies condition 1 and 2 above, and any \(p \in \tilde{A}\), if \(M^p(G^p) = M^{p^*}(G^{p^*})\), then

\[
m^d(p^*) \geq m^d(p)
\]

The first two conditions are relaxations of \((Y_i(1), Y_i(0), D_i(1), D_i(0)) \perp Z_i\), and the minimal defier condition is the relaxation of \(D_i(1) \geq D_i(0)\). Under no defier condition, for all \(p \in A\), \(m^d(p) = 0\) holds. Since measure of defier should be non negative, no defier also implies minimal defier. Our construction of \(\tilde{A}\) relaxes both conditions that defines \(A\), so it is easy to see \(A \subset \tilde{A}\).

**Proposition 14.** The \(\tilde{A}\) defined in assumption 5.1 is a strong extension of \(A\).

A simple corollary is that the local average treatment effect under \(\tilde{A}\) will be the same as before if \(A\) cannot be rejected by the testable implication.

**Corrolary 5.1.** The extended identification problem \((\mathcal{P}, \tilde{A}, LATE, LATE^{ID})\) satisfies

\[
LATE^{ID}(F) = \tilde{LATE}^{ID}(F)
\]

for all distribution \(F(Y, D, Z)\) satisfying (2.14).

### 5.2 Minimal Distance to Independent Instrument as LATE-Consistent Extension

Minimal defier extension relaxes the 'No Defier' assumption but preserves type independent instrument assumption. On the other hand, we can relax the independent instrument assumption but keep the 'No Defier' assumption. Similar to the minimal defier extension, I define the distance to independent instrument as the following quantity:
Definition 17. The marginal integrated distance to \((AT,NT)\)-independent instrument for any \(p \in \mathcal{P}\) is defined as

\[
R(p) = \sum_j \left\{ \int_y \left[ g_p(y_j, D(0) = j, D(1) = j | Z = 1) - g(y_j, D(0) = D(1) = j | Z = 0) \right]^2 dy \right\}
\]

(5.3)

where

\[
g_p(y_j, D(0) = j, D(1) = k | Z) = \frac{\partial Pr_{Gp}(Y(j) \leq y_j, D(0) = j, D(1) = k | Z)}{\partial y_j}
\]

is the conditional density for \(Y(j), D(1), D(0)\) conditioned on \(Z\).

Note that the above definition only measures the deviation from independent instrument using \(D(1) = D(0) = j\), and only measure the deviation corresponding to \(Y(j)\). This is because for always taker \(D(1) = D(0) = 1\), we can never observe the counterfactual outcome \(Y(0)\), so measuring the deviation with respect to \(Y(0)\) is not meaningful. I ignore the complier group and this metric is a measure of the distance to independence instrument \((Y_i(1), Y_i(0), D_i(1), D_i(0)) \perp Z_i\) for the always taker and never taker type. Whenever independent instrument condition holds for \(p\), \(R(p) = 0\). In this sense, I am looking at the extension that relax the instrument independence condition with respect to always taker and never taker.

Assumption 5.2. (Minimal Distance to Independent Instrument with No Defier) Let \(\bar{A}\) be the class of \(p\) such that

1. Instrument Monotonicity holds: \(Pr_{Gp}(D_i(1) \geq D_i(0)) = 1\)

2. Type Independence for complier: \(Y_i(1), Y_i(0) \perp Z_i | (D_i(1) = 1, D_i(0) = 0)\)

3. Minimal Distance to Type Independent Instrument: For any \(p^*\) satisfies condition 1 and 2 above, and any \(p \in \bar{A}\), if \(M^p(Gp) = M^{p^*}(G^{p^*})\), then

\[
R(p^*) \geq R(p)
\]

It is easy to see that \(R(p) \geq 0\) and whenever \(A\) holds, \(R(p) = 0\). So independent instrument achieves the minimal distance. Since the minimal distance \(R(p)\) is only defined for always taker and never taker, the constructed \(\bar{A}\) is not a strong extension. However, as we will see later, if we are interested in LATE, \(\bar{A}\) is LATE-consistent.
5.3 Identified Set of LATE

I focus on the identification of local average treatment effect because it is widely used in literature and it’s policy relevant. Recall that the testable implication of IA monotonicity is characterized by $P(B, i)$ and $Q(B, i)$, our construction of the identified set $\text{LATE}^{ID}$ relies on these two quantities. Since $\mathcal{F}$ contains all $F$ such that $F(y|d, z)$ has a density, then $p(y, i) = \frac{\partial P(Y_i \leq y, i)}{\partial y}$ and $q(y, i) = \frac{\partial Q(Y_i \leq y, i)}{\partial y}$ exist. Define the set

$$\mathcal{Y}_0 = \{ y : q(y, 0) - p(y, 0) \geq 0 \}$$

$$\mathcal{Y}_1 = \{ y : p(y, 1) - q(y, 1) \geq 0 \}$$

(5.4)

Then on $B \subset \mathcal{Y}_0$, the testable implication $Q(B, 0) - P(B, 0) \geq 0$ is not violated. Similarly, on $B \subset \mathcal{Y}_1$, the testable implication $P(B, 1) - Q(B, 1) \geq 0$ is not violated. The identified set of LATE is shown in the following proposition.

**Assumption 5.3.** There exists a constant $c > 0$ such that: (i) $P(\mathcal{Y}_0, 0) - P(\mathcal{Y}_0, 0) > c$ and $P(\mathcal{Y}_1, 1) - Q(\mathcal{Y}_1, 1) > c$.

**Proposition 15.** Let assumption 5.3 holds for $F$ and $\bar{A}$ satisfies assumption 5.1 or assumption 5.2, local average treatment effect $LATE = E[Y_i(1) - Y_i(0)|D_i(1) = 1, D_i(0) = 0]$ is well defined and identified as

$$LATE^{ID}(F) = \frac{\int_{\mathcal{Y}_1} y(p(y, 1) - q(y, 1))dy}{P(\mathcal{Y}_1, 1) - Q(\mathcal{Y}_1, 1)} - \frac{\int_{\mathcal{Y}_0} y(q(y, 0) - p(y, 0))dy}{Q(\mathcal{Y}_0, 0) - P(\mathcal{Y}_0, 0)}$$

(5.5)

Note that $LATE$ is always point identified under $\bar{A}$ that satisfies assumption 5.1 or 5.2.

We can check when $\mathcal{Y}_1 = \mathcal{Y}_0 = \mathcal{Y}$, i.e. assumption $A$ is not violated,

$$P(\mathcal{Y}, 1) - Q(\mathcal{Y}, 1) = E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$$

$$= Q(\mathcal{Y}, 0) - P(\mathcal{Y}, 0)$$

(5.6)

where the second equality holds by $P(\mathcal{Y}, 1) + P(\mathcal{Y}, 0) = Q(\mathcal{Y}, 1) + Q(\mathcal{Y}, 0) = 1$. Also, $\int_{\mathcal{Y}} y(p(y, 1) + p(y, 0))dy = E[Y_i(1)D_i(1) + (1 - D_i(1))Y_i(0)] = E[Y_i|Z_i = 1]$, and $\int_{\mathcal{Y}} y(q(y, 1) + q(y, 0))dy = E[Y_i|Z_i = 0]$, so $LATE^{ID} = LATE^{ID}$ indeed holds.

**Corrollary 5.2.** The extended assumption $\bar{A}$ that satisfies assumption 5.3 is LATE-consistent extension.
If we know \( Y_1 \) and \( Y_0 \), both numerator and denominator can be estimated by sample mean. To estimate \( \hat{Y}_0, \hat{Y}_1 \) we just need to estimate the density \( p(y,j) \) and \( q(y,j) \) using kernel density estimator:

\[
g_h(y, 1) = \frac{\frac{1}{h} \sum_{i=1}^{n} K \left( \frac{Y_i - y}{h} \right) \mathbb{1}(D_j = 1, Z_j = 1)}{\frac{1}{h} \sum_{i=1}^{n} \mathbb{1}(Z_j = 1)} - \frac{\frac{1}{h} \sum_{i=1}^{n} K \left( \frac{Y_i - y}{h} \right) \mathbb{1}(D_j = 1, Z_j = 0)}{\frac{1}{h} \sum_{i=1}^{n} \mathbb{1}(Z_j = 0)}
\]

\[
g_h(y, 0) = \frac{\frac{1}{h} \sum_{i=1}^{n} K \left( \frac{Y_i - y}{h} \right) \mathbb{1}(D_j = 0, Z_j = 0)}{\frac{1}{h} \sum_{i=1}^{n} \mathbb{1}(Z_j = 0)} - \frac{\frac{1}{h} \sum_{i=1}^{n} K \left( \frac{Y_i - y}{h} \right) \mathbb{1}(D_j = 0, Z_j = 1)}{\frac{1}{h} \sum_{i=1}^{n} \mathbb{1}(Z_j = 1)}
\]

Then our LATE estimator can be constructed by

\[
LATE = \frac{\frac{1}{n} \sum_{i=1}^{n} Y_i \left[ \frac{1}{\frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1)} \mathbb{1}(g_h(Y_i, 1) \geq b_n) \right] \left[ \frac{1}{\frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0)} \mathbb{1}(g_h(Y_i, 1) \geq b_n) \right] - \frac{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(D_i = 0, Z_i = 0)}{\frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0)} \mathbb{1}(g_h(Y_i, 0) \geq b_n)}
\]

where \( h \) is the bandwidth for kernel estimator, and \( b_n \) is a sequence of constants that converges to zero. The trimming sequence mimics the selection of \( Y_i \) on the set \( Y_1 \) and \( Y_0 \).

### 5.4 Statistical Property of LATE Estimator

The following assumptions are sufficient to guarantee \( LATE \) will converge to a normal distribution.

**Assumption 5.4.** The kernel function \( K \) satisfies: (i) \( K(u) \) is continuous and supported on \([-A, A]\) and \( \int_{-A}^{A} K(u)du = 1 \); (ii) \( \int_{-A}^{A} uK(u)du = 0 \); (iii) \( \int u^2K(u)du < \infty \).

**Assumption 5.5.** The conditional distribution \( F(y|D_i = k, Z_i = l) \) has a density \( f(y|k, l) \) for all \( k, l \in \{0, 1\} \), and \( f''(y|k, l) \) exists and is uniformly bounded by a constant \( c_f \); (iii) \( E(Y_i^{2+\delta}) < \infty \) for some \( \delta > 0 \).

The above two assumptions are standard in literature and guarantee the uniform convergence rate of deseity estimator to its limit.

**Assumption 5.6.** Let \( g(y, 1) = p(y, 1) - q(y, 1) \) and \( g(y, 0) = q(y, 0) - p(y, 0) \). The following conditions hold for any sequence \( b_n \to 0^+ \): (i) \( \int_{y \in Y} |y|g(y, j)\mathbb{1}(-b_n < g(y, j) \leq b_n)dy = O_p(b_n^2) \); (ii) \( \int_{y \in Y} g(y, j)\mathbb{1}(-b_n < g(y, j) \leq b_n)dy = O_p(b_n) \).
The assumption above controls the bias from trimming \( b_n \).

**Theorem 1.** Under assumption (5.3 - 5.6), let \( b_n = n^{-1/4} / \log n \) and \( h_n = n^{-1/5} \), then

\[
\sqrt{n}(LATE - LATE^{ID}) \rightarrow_d N(0, \Pi' \Sigma D \Gamma \Pi)
\]

where

\[
D = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & -\frac{1}{Pr(Z_i=1)^2} & 0 & 0 & 0 \\
0 & 0 & 0 & -\frac{1}{Pr(Z_i=0)^2} & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]

\[
\Gamma = [\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4].
\]

\[
\Gamma_1 = \begin{pmatrix}
1 \\
0 \\
0 \\
0 \\
E[Y_i \mathbb{1}(D_i = 1, Z_i = 1)(g(Y_i, 1) > 0)] \\
E[Y_i \mathbb{1}(D_i = 1, Z_i = 0)(g(Y_i, 0) > 0)] \\
0 \\
0 \\
E[Y_i \mathbb{1}(D_i = 1, Z_i = 1)(g(Y_i, 1) > 0)] \\
E[Y_i \mathbb{1}(D_i = 1, Z_i = 0)(g(Y_i, 0) > 0)]
\end{pmatrix}
\]

\[
\Gamma_2 = \begin{pmatrix}
0 \\
1 \\
0 \\
0 \\
E[Y_i \mathbb{1}(D_i = 0, Z_i = 0)(g(Y_i, 0) > 0)] \\
E[Y_i \mathbb{1}(D_i = 0, Z_i = 1)(g(Y_i, 1) > 0)]
\end{pmatrix}
\]

\[
\Gamma_3 = \begin{pmatrix}
0 \\
1 \\
0 \\
0 \\
E[Y_i \mathbb{1}(D_i = 0, Z_i = 0)(g(Y_i, 0) > 0)] \\
E[Y_i \mathbb{1}(D_i = 0, Z_i = 1)(g(Y_i, 1) > 0)]
\end{pmatrix}
\]

\[
\Gamma_4 = \begin{pmatrix}
0 \\
0 \\
0 \\
1
\end{pmatrix}
\]

(5.9)

and the matrix

\[
\Pi = \begin{pmatrix}
\frac{1}{\beta_3} \\
-\frac{1}{\beta_1} \\
-\frac{1}{\beta_2} \\
\beta_3 \\
\beta_4
\end{pmatrix}
\]

\[
\beta \equiv \begin{pmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4
\end{pmatrix} = \begin{pmatrix}
\int_{\mathcal{Y}_1} y(p(y, 1) - q(y, 1))dy \\
\int_{\mathcal{Y}_0} y(q(y, 0) - p(y, 0))dy \\
\int_{\mathcal{Y}_1} (p(y, 1) - q(y, 1))dy \\
\int_{\mathcal{Y}_0} (q(y, 0) - p(y, 0))dy
\end{pmatrix}
\]

38
\[ \Sigma = \text{Var} \left\{ \begin{array}{lcl} Y_i \left[ \frac{1(D_i=1, Z_i=1)}{Pr(Z_i=1)} - \frac{1(D_i=1, Z_i=0)}{Pr(Z_i=0)} \right] \mathbb{1}(g(Y_i, 1) \geq 0) \\ Y_i \left[ \frac{1(D_i=0, Z_i=0)}{Pr(Z_i=0)} - \frac{1(D_i=0, Z_i=1)}{Pr(Z_i=1)} \right] \mathbb{1}(g(Y_i, 0) \geq 0) \\ \mathbb{1}(Z_i = 1) \\ \mathbb{1}(Z_i = 0) \\ \frac{1(D_i=1, Z_i=1)}{Pr(Z_i=1)} - \frac{1(D_i=1, Z_i=0)}{Pr(Z_i=0)} \mathbb{1}(g(Y_i, 1) \geq 0) \\ \frac{1(D_i=0, Z_i=0)}{Pr(Z_i=0)} - \frac{1(D_i=0, Z_i=1)}{Pr(Z_i=1)} \mathbb{1}(g(Y_i, 0) \geq 0) \end{array} \right\} \]

Unfortunately, assumption 5.6 can fail for many unbounded distributions. For example, if \( g(y, 1) \) has a Gaussian Tail \( g(y, 1) = C_{g,1} e^{-y^2/2} \), where \( C_{g,1} \) is a constant. Then trimming region equals

\[ \{ y : |g(y, 1)| \leq b_n \} = \left( -\infty, -\sqrt{-2 \log(b_n/C_{g,1})} \right] \cup \left[ \sqrt{-2 \log(b_n/C_{g,1})}, \infty \right) \]

and the integral

\[ \int_{y \in Y} |y| \mathbb{1}(g(y, 1) \geq b_n) dy \]

\[ = 2 \int_{y = \sqrt{-2 \log(b_n/C_{g,1})}}^{\infty} y C_{g,1} e^{-y^2/2} dy \]

Using L'Hospital rule, we can show

\[ \lim_{b_n \to 0} \frac{2 \int_{y = \sqrt{-2 \log(b_n/C_{g,1})}}^{\infty} y C_{g,1} e^{-y^2/2} dy}{b_n} \]

\[ = \lim_{b_n \to 0} -2C_{g,1} \sqrt{-2 \log(b_n/C_{g,1}) b_n} \frac{\partial \sqrt{-2 \log(b_n/C_{g,1})}}{\partial b_n} \]

\[ = \lim_{b_n \to 0} -2C_{g,1} \sqrt{-2 \log(b_n/C_{g,1}) b_n} \frac{-2/(b_n C_{g,1})}{2 \sqrt{-2 \log(b_n/C_{g,1})}} \]

\[ = 2 \]

so the trimming bias is of order \( O_p(b_n) \). This is because on the tail of the distribution, the density \( p(y, 1) \) and \( q(y, 1) \) will both converge to zero, and the trimming \( b_n \) that takes care of sampling error of \( g_h(y, 1) \) will result in large bias in the tail. Since the density estimator is a first step estimator, the sampling error will not be killed in the second step when we calculate the LATE estimator.
5.4.1 Compact Support Assumption

Since the problem of assumption 5.6 comes from the tail, we can instead assume the support of $Y_i$ is bounded. The following is a condition such that assumption 5.6 will hold.

**Assumption 5.7.** Suppose $Y = [Y_{\min}, Y_{\max}]$, the set $C_1 = \{y : g(y, 1) = 0\}$ has $M < \infty$ points, and $\sup_{y \in B(C_1, \delta)} |\frac{d(g(y, 1))}{dy}| > 1/C$ for some $\delta > 0$, and $C_0 = \{y : g(y, 0) = 0\}$ has $M < \infty$ points, and $\sup_{y \in B(C_0, \delta)} |\frac{d(g(y, 0))}{dy}| > 1/C$, where $B(C_j, \delta) = \cup_y \in C_j B(y, \delta)$ for $j = 0, 1$.

However, kernel density estimator on bounded support will have boundary bias and will not be consistent. The remedy is to use boundary kernel method as in Karunamuni and Alberts (2005) to kill the bias. The kernel density estimator for $y \in [Y_{\min} + h, Y_{\max} - h]$ can still be estimated by equation (5.7). When $y \in [Y_{\min}, Y_{\min} + h)$, we use the following boundary kernel estimator:

\[
g_h(y, 1) = \frac{1}{h_n} \sum_{i=1}^{n} \left\{ K \left( y - Y_{\min} - \frac{\phi_c'}{h}(y - Y_{\min}) \right) + K \left( y - Y_{\min} + \frac{\phi_c'}{h}(y - Y_{\min}) \right) \right\} \mathbb{1}(D_j = 1, Z_j = 1) \\
- \frac{1}{h_n} \sum_{i=1}^{n} \left\{ K \left( y - Y_{\min} - \frac{\phi_c'}{h}(y - Y_{\min}) \right) + K \left( y - Y_{\min} + \frac{\phi_c'}{h}(y - Y_{\min}) \right) \right\} \mathbb{1}(D_j = 1, Z_j = 0) \\
\sum_{i=1}^{n} \mathbb{1}(Z_j = 1) \\
\sum_{i=1}^{n} \mathbb{1}(Z_j = 0) \\
\right(5.10)\]

where the constant $c = (y - Y_{\min})/h$ measures the localization to boundary point, and function $\phi_c$ is given by

\[
\frac{c}{2} \hat{d}_j k_c' y^2 + \lambda_0 (\hat{d}_j k_c')^2 y^3
\]

\[
k_c' = 2 \int_c^1 (t - c)K(t)dt \left[ \left( c + 2 \int_c^1 (t - c)K(t)dt \right) \right]
\]

\[
\hat{d}_j = \frac{1}{h_1} \log \left( \frac{\frac{1}{nh} \sum_i K(h - (y - Y_{\min})) \mathbb{1}(D_i = 1, Z_i = j) + \frac{1}{n^2}}{\frac{1}{n} \sum_i \mathbb{1}(Z_i = j)} \right) + \max \log \left\{ \frac{\frac{1}{nh_0} \sum_i K(0) \left( \frac{-(y - Y_{\min})}{h_0} \right) \mathbb{1}(D_i = 1, Z_i = j) + \frac{1}{n^2}}{\mathbb{1}(Z_i = j)} \right\}
\]

where $K(0)$ is an end point order two kernel satisfying

\[
\int_{-1}^{0} K(0)(t)dt = 1, \quad \int_{-1}^{0} tK(0)(t)dt = 0, \quad 0 < \int_{-1}^{0} t^2 K(0)(t)dt < \infty
\]
And $h_0 = b_0 h_1$ where

$$b_0 = \left\{ \left( \int_{-1}^{1} t^{2} K(t) dt \right)^{2} \left( \int_{-1}^{0} K^{2}(t) dt \right)^{1/5} \right\}$$

Boundary kernel estimator for When $y \in (\mathcal{Y}_{\max} - h, \mathcal{Y}_{\max}]$ can be estimated similarly by replacing $y - \mathcal{Y}_{\min}$ by $\mathcal{Y}_{\max} - y$ and replace the indicator function $1(D_i = 1, Z_i = j)$ accordingly.

**Theorem 2.** Suppose assumption 5.3 - 5.5 and 5.7 holds, and $g_h(y, j)$ is defined by (5.10). Also, we suppose the $f^m(y|k, l)$ exists and is uniformly bounded. Let $h = n^{-1/5}$, $h_1 = n^{-1/7}$, and $b_n = n^{-1/4}/\log n$. Then

$$\sqrt{n}(\hat{LATE} - \tilde{LATE}^{ID}) \to_d N(0, \Pi' \Sigma_2 D \Gamma \Pi)$$

### 5.4.2 LATE Conditioned on Bounded Interval

Sometimes researchers may only want to study the treatment for compliers when the potential outcome is not in the tail of distribution. In this case, there is a pre-specified interval $[\underline{y}, \bar{y}]$ and the LATE conditioned on this interval is defined as

$$LATE = E[Y(1) - Y(0)|D(1) - D(0) = 1, Y(1), Y(0) \in [\underline{y}, \bar{y}]]$$

The identified value of $LATE$ is given by

$$\tilde{LATE}^{ID}_{[\underline{y}, \bar{y}]}(F) = \frac{\int_{\mathcal{Y}_1 \cap [\underline{y}, \bar{y}]} y(p(y, 1) - q(y, 1))dy}{P(\mathcal{Y}_1 \cap [\underline{y}, \bar{y}], 1) - Q(\mathcal{Y}_1 \cap [\underline{y}, \bar{y}], 1)} - \frac{\int_{\mathcal{Y}_0 \cap [\underline{y}, \bar{y}]} y(q(y, 0) - p(y, 0))dy}{Q(\mathcal{Y}_0 \cap [\underline{y}, \bar{y}], 0) - P(\mathcal{Y}_0 \cap [\underline{y}, \bar{y}], 0)} \tag{5.11}$$

The estimator is slightly adapted to be

$$\hat{LATE}^{ID}_{[\underline{y}, \bar{y}]} = \frac{1}{n} \sum_{i=1}^{n} Y_i \left[ \frac{1(D_i = 1, Z_i = 1)}{\frac{1}{n} \sum_{j=1}^{Z_i = 1} 1(Z_j = 1)} - \frac{1(D_i = 1, Z_i = 0)}{\frac{1}{n} \sum_{j=1}^{Z_i = 0} 1(Z_j = 0)} \right] \mathbb{I}(g_h(Y_i, 1) \geq b_n, Y_i \in [\underline{y}, \bar{y}])$$

$$- \frac{1}{n} \sum_{i=1}^{n} Y_i \left[ \frac{1(D_i = 0, Z_i = 0)}{\frac{1}{n} \sum_{j=1}^{Z_i = 0} 1(Z_j = 0)} - \frac{1(D_i = 0, Z_i = 1)}{\frac{1}{n} \sum_{j=1}^{Z_i = 1} 1(Z_j = 1)} \right] \mathbb{I}(g_h(Y_i, 0) \geq b_n, Y_i \in [\underline{y}, \bar{y}])$$

$$- \frac{1}{n} \sum_{i=1}^{n} Y_i \left[ \frac{1(D_i = 0, Z_i = 0)}{\frac{1}{n} \sum_{j=1}^{Z_i = 0} 1(Z_j = 0)} - \frac{1(D_i = 0, Z_i = 1)}{\frac{1}{n} \sum_{j=1}^{Z_i = 1} 1(Z_j = 1)} \right] \mathbb{I}(g_h(Y_i, 0) \geq b_n, Y_i \in [\underline{y}, \bar{y}]) \tag{5.12}$$

where $g_h$ is given by (5.7). Since we focus on $Y_i \in [\underline{y}, \bar{y}]$, assumption 5.6 only needs to hold for $Y_i \in [\underline{y}, \bar{y}]$. 

41
Assumption 5.8. Let \( g(y, 1) = p(y, 1) - q(y, 1) \) and \( g(y, 0) = q(y, 0) - p(y, 0) \). The following conditions hold for any sequence \( b_n \to 0_+ \): (i) \( \int y \cdot y \cdot g(y, j) \cdot \mathbb{1}(-b_n < g(y, j) \leq b_n) \, dy = O_p(b_n^2) \); (ii) \( \int y \cdot g(y, j) \cdot \mathbb{1}(-b_n < g(y, j) \leq b_n) \, dy = O_p(b_n) \).

5.5 Minimal Defier Assumption \( \tilde{A} \) as Bayesian Model Selection

In this section, I discuss the economic reasoning behind the minimal defier restrictions. Consider a policy maker who has a prior distribution \( H(m^d) \) on the measure of defier. This prior distribution should be understood in the following way: there are many LATE studies, each LATE study is associated with a measure of defiers, and \( H(m^d) \) is the distribution of measure of defiers across these LATE studies. In this setting, it is natural to set the belief to have a decreasing density function \( h(m^d) \).

Assumption 5.9. The prior distribution \( H(m^d) \) has a decreasing density function \( h(m^d) \) on \( m^d \in [0, 1] \).

The assumption says that the policy maker believes that defier is something unnatural and should only happen with decreasing probabilities. Now, suppose \( F(Y, D, Z) \) implies that there must be at least \( m \) measure of defiers, his updated belief is \( H(m^d|m^d \geq m) \), and the likelihood is maximized at \( m^d = m \).

Proposition 16. If assumption 5.9 holds, and \( F \) implies there is at least \( m \) measure of defier under \( (Y_i(1), Y_i(0)) \perp \perp Z_i|D_i(1), D_i(0) \) and (5.2), then \( \tilde{A} \) maximize the posterior belief density of \( m^d \).

Under assumption 5.9, \( \tilde{A} \) is the class of structures that maximize the likelihood of defier. Therefore, assumption 5.1 can be viewed as the result of the first step model selection via maximal likelihood.

Remark 5.1. Another natural way is to use the posterior belief \( H \) as a weighting function. However, this pose significant challenge to the interpretation of the final LATE object. Given a measure of defier \( m^d \), LATE is potentially set identified. Then we have a distribution of identified set of LATE, which requires random set interpretation. One reason that the policy analyzer chooses this two step procedure is to balance the complexity and model consistency.
5.5.1 An Example of Minimal Defier

In this section, I present a model where the Bayesian posterior selected minimal defier model has economic interpretation. Consider the two-instrument case in Mogstad et al. (2019). Suppose there are two binary instruments $Z = (Z_1, Z_2) \in \{0, 1\}^2$. Decision maker $i$ will take treatment when the following utility $V_i(Z_i)$ is positive:

$$V_i(Z_i) = B_{i,0} + B_{i,1}Z_{i,1} + \beta Z_{i,2}$$

where $(B_{i,0}, B_{i,1}) \in \mathbb{R} \times \mathbb{R}_+$ are latent variables that vary across individuals, $\beta$ is some positive parameter. Using the terminology in Mogstad et al. (2019), I partition the space of $(B_{i,0}, B_{i,1})$ into six regions Without loss of generality, we normalize $(B_{i,0}, B_{i,1}) \in [-2, 1] \times [0, 3]$. The space of $(B_{i,0}, B_{i,1})$ is partitioned into 6 regions:

1. $\tilde{AT}$: Always takers, who take treatment
2. $\tilde{NT}$: Never takers, who never take treatment
3. $\tilde{EC}$: Eager takers, who take treatment when $\max\{Z_{i,1}, Z_{i,2}\} > 0$
4. $\tilde{RC}$: Reluctant takers, who take treatment when $\min\{Z_{i,1}, Z_{i,2}\} > 0$
5. $Z_1C$: $Z_1$ compliers, who take treatment when $Z_{i,1} = 1$
6. $Z_2C$: $Z_2$ compliers, who take treatment when $Z_{i,2} = 1$

Now suppose an econometrician observes only $Z_1$ but not $Z_2$. Now, consider for possible types of $(Z_1, Z_2)$ pairs in table 2. For example, type 1 is defined as the individuals who get $Z_2 = 1$ regardless of the value of $Z_1$.

| $(Z_1, Z_2)$ Types | $Z_1$ = 1 | $Z_2$ = 1 | $Z_2$ = 1 | $Z_2$ = 0 | $Z_2$ = 0 |
|------------------|-----------|-----------|-----------|-----------|-----------|
| $Z_1 = 1$        | $Z_2$ = 1 | $Z_2$ = 1 | $Z_2$ = 0 | $Z_2$ = 0 |
| $Z_1 = 0$        | $Z_2$ = 1 | $Z_2$ = 0 | $Z_2$ = 1 | $Z_2$ = 0 |
For an econometrician who use only $Z_1$ to define compliance type, the latent variable in the $V_i(Z)$ is $(B_{i,0}, B_{i,1}, \text{Type}_j)$. We now define four compliance types

\[
\{AT, NT, \text{Compliance}(C), \text{Defier}(D) \}
\]

with respect to the observed instrument $Z_1$. The following table shows the connection to the $(B_{i,0}, B_{i,1}, \text{Type}_j)$ space. For example, $(B_{i,0}, B_{i,1}, \text{Type}_j) \in \tilde{Z}_2C \times \{\text{Type I}\}$ looks as if they are always taker under $Z_1$, since they comply with $Z_2$ and $Z_2 = 1$ always hold for Type I. Defier only happens when an individual is $Z_2$ complier in the full model, and his instrument of $Z_1$ and $Z_2$ is always different, i.e. type III.

Table 3: Projection of Compliance Type into 1 Instrument

| Type I | Type II | Type III | Type IV |
|--------|---------|----------|---------|
| $AT$   | AT      | AT       | AT      | AT      |
| $ET$   | AT      | C        | AT      | C       |
| $\tilde{Z}_1C$ | C | C        | C       | C       |
| $\tilde{Z}_2C$ | AT | C        | D       | NT      |
| $\tilde{R}T$ | C | C        | NT      | NT      |
| $\tilde{N}T$ | NT | NT       | NT      | NT      |
Suppose \((B_{i,0}, B_{i,1})\) is independent of \(Type\ j\), the measure of defier is

\[ m^d = Pr((B_{i,0}, B_{i,1}) \in Z^C) \times Pr(Type = III) \]

Now there are two ways to justify the minimal defier assumption using the Bayesian model selection. One is to fix the distribution of \((B_{i,0}, B_{i,1})\) and parameter \(\beta\), have a prior belief on Type III: the possibility of always having opposite \(Z_1\) and \(Z_2\) should be small. The second is to fix the distribution of \((B_{i,0}, B_{i,1}, Type\ j)\) but have a prior belief on the parameter \(\beta\): \(\beta\) should be as small as possible, i.e. the omitted instrument should have little effect on treatment decision, and smaller value of \(\beta\) implies smaller measure of defier.

### 6 Conclusion

In this paper, I study incomplete economic models with testable assumptions. I generalize the definition of refutability and confirmability to incomplete economic structures and study the identification problem when an assumption \(A\) is testable. I provide a general method for estimation and inference. In the main application to Potential Outcome framework, I give a set of minimal defier condition and characterize the identified set of local average treatment effect under the relaxed assumption.

One of the main limitation of the estimation and inference method in Section 4 is the computational issue. When the latent variable falls into a parametric family \(G^p(\theta)\), the method can be done by partition the parameter space using grids. However, when \(G^p\) is fully non-parametric, the method cannot be implemented effectively. In most case when \(G^p\) is non-parametric, the identified set can be characterized by moment inequalities, and moments based method can out-perform my procedure.

For future researches, one should consider more application to economic structures where existing assumptions have testable implications. For example, in the \(2 \times 2\) entry game framework, one can similarly think about a 'maximal Nash' condition that generalize PSNE assumption.
References

Aradillas-Lopez, Andres and Elie Tamer (2008), “The identification power of equilibrium in simple games.” *Journal of Business & Economic Statistics*, 26, 261–283.

Belloni, Alexandre, Federico Bugni, and Victor Chernozhukov (2018), “Subvector inference in partially identified models with many moment inequalities.” *arXiv preprint arXiv:1806.11466*.

Bresnahan, Timothy F and Peter C Reiss (1991), “Empirical models of discrete games.” *Journal of Econometrics*, 48, 57–81.

Breusch, Trevor S (1986), “Hypothesis testing in unidentified models.” *The Review of Economic Studies*, 53, 635–651.

Chernozhukov, Victor, Han Hong, and Elie Tamer (2007), “Estimation and confidence regions for parameter sets in econometric models 1.” *Econometrica*, 75, 1243–1284.

Ciliberto, Federico and Elie Tamer (2009), “Market structure and multiple equilibria in airline markets.” *Econometrica*, 77, 1791–1828.

Galichon, Alfred and Marc Henry (2013), “Dilation bootstrap.” *Journal of Econometrics*, 177, 109–115.

Imbens, Guido and Joshua Angrist (1994), “Identification and estimation of local average treatment effects.” *Econometrica*, 62.

Jovanovic, Boyan (1989), “Observable implications of models with multiple equilibria.” *Econometrica: Journal of the Econometric Society*, 1431–1437.

Karunamuni, Rhoana and Tom Alberts (2005), “On boundary correction in kernel density estimation.” *Statistical Methodology*, 2, 191–212.

Kitagawa, Toru (2015), “A test for instrument validity.” *Econometrica*, 83, 2043–2063.

Masten, Matthew and Alexandre Poirier (2018), “Salvaging falsified instrumental variable models.” *Working Paper*. 

46
Mogstad, Magne, Alexander Torgovitsky, and Christopher R Walters (2019), “Identification of causal effects with multiple instruments: Problems and some solutions.” Working Paper.

Mourifie, Ismael, Marc Henry, and Romuald Méango (2018), “Sharp bounds and testability of a roy model of stem major choices.” Available at SSRN 2043117.

Mourifié, Ismael and Yuanyuan Wan (2017), “Testing local average treatment effect assumptions.” Review of Economics and Statistics, 99, 305–313.

Roy, Andrew Donald (1951), “Some thoughts on the distribution of earnings.” Oxford economic papers, 3, 135–146.

A Proofs in section 2 and 3

A.1 Proof of proposition 1

Proof. Since $G^p$ is a proper probability measure, this implies $M^p(G^p) \subset \mathcal{F}$.

On the other hand, for any $F$ we consider a model with $\beta_1^p = 0$, and $S^p = L2R$ so $A(y|X_{1m} = 0) = A(y|X_{1m} = 1)$ for all $y \in \{(1, 1), (1, 0), (0, 1), (0, 0), M\}$. Define set $Y_{-M} = \{(1, 1), (1, 0), (0, 1), (0, 0)\}$ Now we choose the density $g^p$ such that

$$\int_{A(y|X_{1m})} g^p(\epsilon)d\epsilon = \min_{i=0, 1} Pr_F(Y_m = \sum_{j=1}^2 y_j, X_{1m} = i)$$

for $y \in Y_{-M}$, and choose density such that

$$\int_{A(M|X_{1m})} g^p(\epsilon)d\epsilon = 1 - \sum_{y \in Y_{-M}} \int_{A(y|X_{1m})} g^p(\epsilon)d\epsilon$$

Then $g^p$ is a well defined density supported on $\mathbb{R}^2$. Then it can be shown that $F \in M^p(G^p)$ by choosing the constant

$$c_0^i = Pr_F(Y_m = 0, X_{1m} = i) - \int_{A((0, 0)|X_{1m})} g^p(\epsilon)d\epsilon$$

$$c_1^i = Pr_F(Y_m = 1, X_{1m} = i) - \int_{A((1, 0)\cup(0,1)\cup M|X_{1m})} g^p(\epsilon)d\epsilon$$

$$c_2^i = Pr_F(Y_m = 2, X_{1m} = i) - \int_{A((1,1)|X_{1m})} g^p(\epsilon)d\epsilon$$
in equation (2.2)

A.2 Proof of proposition 2

Proof. 1. If $p \in H_{P}^{snf}(A)^c$, by definition it means $M^p(G^p) \cap \left[ \bigcup_{p^* \in A} M^{p^*}(G^{p^*}) \right]^c \neq \emptyset$. Since $(A^c)^c = A$ and

$$\left[ \bigcup_{p^* \in A} M^{p^*}(G^{p^*}) \right]^c = \bigcap_{p^* \in (A^c)^c} M^{p^*}(G^{p^*})^c$$

this shows $H_{P}^{snf}(A)^c \subset H_{P}^{wcon}(A^c)$. Use the same logic to find $H_{P}^{snf}(A)^c \supset H_{P}^{wcon}(A^c)

2. If $p \in H_{P}^{wnf}(A)^c$, by definition it means $\forall p^* \in A$, $M^p(G^p) \cap M^{p^*}(G^{p^*}) = \emptyset$. Since $(A^c)^c = A$ and

$$M^p(G^p) \subset \bigcap_{p^* \in (A^c)^c} (M^{p^*}(G^{p^*})^c)$$

By definition it means $p \in H_{P}^{wcon}(A)^c$, use the same logic to find the reversed inclusion.

3. Suppose not, we can find $p^* \in H_{P}^{wcon}(A)^c$ such that

$$M^p(G^p) \cap M^{p^*}(G^{p^*})^c = \emptyset \iff M^p(G^p) \subset M^{p^*}(G^{p^*})$$

Now, since $p \in H_{P}^{wcon}(A)$, it means

$$M^p(G^p) \cap \left[ \bigcap_{\tilde{p} \in A^c} M^{\tilde{p}}(G^{\tilde{p}})^c \right] \neq \emptyset$$

Since $M^p(G^p) \subset M^{p^*}(G^{p^*})$, it means

$$M^{p^*}(G^{p^*}) \cap \left[ \bigcap_{\tilde{p} \in A^c} M^{\tilde{p}}(G^{\tilde{p}})^c \right] \neq \emptyset$$

which by definition means $p^* \in H_{P}^{wcon}(A)$, which is a contradiction.

4. The last statement follows from 3 and 1 by set operation.

A.3 Proof of proposition 3

For any structure $p \in A$, we associate another structure $p^* \in P \setminus A$ such that: $G^p = G^{p^*}$ and $\theta^p = \theta^{p^*}$. The only difference is that $S^p = PSNE$ and $S^{p^*} = L2R$. Then $M^p(G^p) \subset M^{p^*}(G^{p^*})$, and the inclusion is strict if $\delta_1 \delta_2 > 0$, since $G^p$ has full support over $\mathbb{R}^2$. Then 1 and 3 follows directly.
To show the second claim, we need to show two things: any \( p \in A \) generates (2.12), and any \( F \) satisfying (2.12) can be generated by some \( p \in A \).

First, we show \( p \in A \) implies (2.12). First consider \( \beta_1 \geq 0 \). In this case

\[
A((1,1)|X_{1m} = 0) \subset A((1,1)|X_{1m} = 1) \text{ and } A((0,0)|X_{1m} = 1) \subset A((0,0)|X_{1m} = 0)
\]

Since \( Pr_F(Y_m = \sum y_j, X_{1m} = i) = \int_{A(y|X_{1m} = i)} dG^p(\epsilon)G^p(X_{1m} = i) \). Since \( \epsilon \) has a density

\[
\Pr_F(\sum y_j, X_{1m} = i) = \int_{A(y|X_{1m} = i)} dG^p(\epsilon)G^p(X_{1m} = i) \geq 0
\]

Repeat this for \( Y_m = 0 \) to get when \( \beta_1 \geq 0 \), (2.12) holds. The same reasoning holds for \( \beta_1 \leq 0 \).

Next we show that any \( F \) satisfying (2.12) can be generated by some \( p \in A \). Without loss of generality, we assume

\[
\Delta_{11} \equiv Pr_F(Y_m = 2|X_{1m} = 1) - Pr_F(Y_m = 2|X_{1m} = 0) \geq 0
\]

\[
\Delta_{00} \equiv Pr_F(Y_m = 0|X_{1m} = 1) - Pr_F(Y_m = 0|X_{1m} = 0) \leq 0
\]

So in this case, we simply take \( \theta^p = (0, 0, 1, 1, 1) \), \( S^p = PSNE \) and choose choose \( g^p \) such that

\[
\int_{A((1,1)|X_{1m} = 0)} dG(\epsilon) = Pr_F(Y_m = 2|X_{1m} = 1)
\]

\[
\int_{A((0,0)|X_{1m} = 1)} dG(\epsilon) = Pr(F(Y_m = 0|X_{1m} = 1) \, (A.3)
\]

\[
\int_{A((1,1)|X_{1m} = 1)\backslash A((1,1)|X_{1m} = 0)} dG(\epsilon) = \Delta_{11}
\]

\[
\int_{A((0,0)|X_{1m} = 0)\backslash A((0,0)|X_{1m} = 1)} dG(\epsilon) = \Delta_{00}
\]

and let \( g \) to have arbitrary density on \( \mathbb{R}^2 \backslash \bigcup_{y \in \{(0,0),(1,1)\}, i = 0,1} A(y|X_{1m} = i) \). Then \( F \in M^p(G^p) \). The case with \( \Delta_{11} \leq 0 \), \( \Delta_{00} \geq 0 \) can be constructed similarly with \( \theta^p = (0, 0, -1, 1, 1) \).

**A.4 Proof of Lemma 2.1**

**Proof.** Recall that the following inclusion holds

\[
\mathcal{H}_p^{scn}(A) \subset \mathcal{H}_p^{wcon}(A) \subset A \subset \mathcal{H}_p^{snf}(A) \subset \mathcal{H}_p^{unf}(A)
\]
$1 \Rightarrow 2$ It suffices to show $\mathcal{H}_P^{snf}(A) \subset A$. Let $p \in \mathcal{H}_P^{snf}(A)$, by definition of $\mathcal{H}_P^{snf}(A)$, there exists $p^* \in A$ such that $M^p(G^p) \cap M^{p^*}(G^{p^*}) \neq \emptyset$. Since $A$ is a hypothesis, $p^* \in A = \mathcal{H}_P^{scon}(A)$. Now suppose $p \notin A$, by definition of $\mathcal{H}_P^{scon}(A)$, $M^p(G^p) \subset M^p(G^p)^c$ which is a contradiction.

$2 \Rightarrow 3$ It suffices to show that $A \subset \mathcal{H}_P^{scon}(A)$. Suppose not, so there exists a $p \in A \setminus \mathcal{H}_P^{scon}(A)$. Since $p \notin \mathcal{H}_P^{scon}(A)$, by definition, there exists $p^* \in A^c$ such that

$$M^p(G^p) \cap M^{p^*}(G^{p^*}) \neq \emptyset$$

Now since, $p \in A$, this shows that $p^* \in \mathcal{H}_P^{scon}(A)$. However, by 2, $\mathcal{H}_P^{scon}(A) = A$, so this yields the contradiction.

$3 \Rightarrow 1$ It suffices to show that $\mathcal{H}_P^{snf}(A) \subset A$. Suppose not, there exists $p \in \mathcal{H}_P^{snf} \setminus A$. By definition of $\mathcal{H}_P^{snf}$

$$M^p(G^p) \subset \cup_{p^* \in A} M^{p^*}(G^{p^*}) = \cup_{p^* \in \mathcal{H}_P^{scon}(A)} M^{p^*}(G^{p^*})$$

where the equality holds by 3. Now, since $p^* \in \mathcal{H}_P^{scon}(A)$, by definition of $\mathcal{H}_P^{scon}(A)$ and $p \in A^c$, we have

$$M^{p^*}(G^{p^*}) \cap M^p(G^p) = \emptyset \quad p^* \in A$$

This generates contradiction.

\[ \square \]

A.5 Proof of proposition 5

Proof. $\Rightarrow$: Suppose $\mathcal{H}_P^{scon}(A) \neq \mathcal{H}_P^{snf}(A)$, by Lemma 2.1 it means $\exists p \in \mathcal{H}_P^{snf}(A)$ but $p \notin A$. By definition, $M^p(G^p) \cap (\cup_{p^* \in A} M^{p^*}(G^{p^*})) \neq \emptyset$. By definition of $p \in \mathcal{H}_P^{snf}(A)$, $\exists F \in M^p(G^p) \cap (\cup_{p^* \in A} M^{p^*}(G^{p^*}))$, it means $A$ cannot be detected by $F$, because $F \in \cup_{p^* \in A^c} M^{p^*}(G^{p^*})$ and there exists some $p^* \in A$ such that $F \in M^{p^*}(G^{p^*})$. So $A \setminus \mathcal{P}^{-1}(F) \neq \emptyset$.

$\Leftarrow$: Suppose there exists $F$ such that $A$ cannot be detected by $F$, then $F \in \cup_{p^* \in A^c} M^{p^*}(G^{p^*})$ and $A \setminus \mathcal{P}^{-1}(F) \neq \emptyset$. This means we can find $p \in A^c$, such that $F \in M^p(G^p)$, and find $\tilde{p} \in A$ such that $F \in M^{\tilde{p}}(G^{\tilde{p}})$. So $p \in \mathcal{H}_P^{snf}(A)$. This shows $A \neq \mathcal{H}_P^{snf}(A)$. By Lemma 2.1 and $\mathcal{H}_P^{scon}(A) \subset A$, we have $\mathcal{H}_P^{scon}(A) \neq \mathcal{H}_P^{snf}(A)$.

\[ \square \]

A.6 Proof of proposition 6

Proof. We prove the first part of the proposition.
We now show the second part of the proposition.

"1 ⇒ 2": If \( \mathcal{P}^{-1}(F) \subseteq A \), and since \( F \) is what we observe, then \( F \in M^\rho(G^\rho) \). By definition \( p^0 \in \mathcal{H}^\text{snf}_\mathcal{P}(A) \). By assumption \( (\mathcal{H}^\text{snf}_\mathcal{P}(A) \setminus \mathcal{H}^\text{snf}_\mathcal{P}(A)) \cap \mathcal{P}^{-1}(F) = \emptyset \), it implies that \( p^0 \in \mathcal{H}^\text{snf}_\mathcal{P}(A) \).

"1 ⇐ 2": If \( p^0 \in \mathcal{H}^\text{snf}_\mathcal{P}(A) \) and \( F \in M^\rho(G^\rho) \), by definition of \( \mathcal{H}^\text{snf}_\mathcal{P}(A) \), there exists \( p \in A \) such that \( F \in M^p(G^p) \).

We now show the second part of the proposition.

"2 ⇒ 1": Suppose \( 1 \) does not hold. So we can find a structure \( p^* \in A^c \) such that \( F \in M^{p^*}(G^{p^*}) \). Since \( p_0 \) will generate \( F \), \( F \in (\bigcup_{p \in A^c}M^p(G^p)) \cap M^{p_0}(G^{p_0}) \). As a result \( M^{p_0}(G^{p_0}) \not\subseteq (\bigcap_{p \in A^c}M^p(G^p)^c) \), \( 2 \) does not hold.

### A.7 Proof of proposition 10

Proof. To show \( \tilde{A}^{\text{max}} \) is the maximal element, consider any \( A^* \) such that \( \tilde{A}^{\text{max}} \subseteq A^* \). Let \( p \in A^* \setminus \tilde{A}^{\text{max}} \), then \( p \in \mathcal{H}^\text{snf}_\mathcal{P}(A) \setminus A \), so \( A^* \cap \mathcal{H}^\text{snf}_\mathcal{P}(A) \neq \emptyset \), and this implies \( A^* \not\subseteq A \).

I give an explicit construction of minimal element of \( \tilde{A}^{\text{min}} \) when \( \mathcal{P} \) is complete. For any \( F \not\in \bigcup_{p \in A} M^p(G^p) \), by construction \( \mathcal{P}^{-1}(F) \cap \mathcal{H}^\text{snf}_\mathcal{P}(A) = \emptyset \). By axiom of choice, we can define a choice function \( h : \mathcal{P}^{-1}(F) \rightarrow \mathcal{P} \). Let \( \tilde{A}^{\text{min}} = A \cup (\bigcup_{F \not\in \bigcup_{p \in A} M^p(G^p)} h(\mathcal{P}^{-1}(F))) \), then \( \tilde{A}^{\text{min}} \cap \mathcal{H}^\text{snf}_\mathcal{P}(A) = A \), so \( \tilde{A}^{\text{min}} \in \mathcal{A} \).

Now, for any \( A^* \subseteq \tilde{A}^{\text{min}} \), let \( p \in \tilde{A}^{\text{min}} \setminus A^* \). If \( p \in A^* \), then \( A^* \cap \mathcal{H}^\text{snf}_\mathcal{P}(A) \neq \emptyset \). If \( p \in (\bigcup_{F \not\in \bigcup_{p \in A} M^p(G^p)} h(\mathcal{P}^{-1}(F))) \), and \( F \in M^p(G^p) \), then \( F \not\in \mathcal{H}^\text{snf}_\mathcal{P}(A^*) \). In both cases, \( A^* \not\subseteq A \).

### B Proofs of General Statistical Properties

**Lemma B.1.** Let \( \Theta^{ID,\epsilon}(F) = \{ \theta \in \Theta | d(\theta, \Theta^{ID,\epsilon}(F)) < \epsilon \} \). Let assumption 4.1.2 holds, then \( \forall \epsilon > 0 \), there exists \( \delta(\epsilon) > 0 \) such that for all \( \tilde{\theta} \in \Theta \setminus \Theta^{ID,\epsilon}(F) \) and \( \forall p \) such that \( \theta(p) = \tilde{\theta} \),
the following holds:
\[ \inf_{F^* \in M^p(G^p)} d_{\tilde{F}}(F^*, F) \geq \delta(\epsilon) \]

Proof. Since \( \Theta^{ID}(F) \) is upper hemi-continuous at \( F \), and \( \Theta^{ID,\epsilon}(F) \) is an open neighborhood of \( \Theta^{ID}(F) \), there exists an open neighborhood \( U \) of \( F \) such that \( \Theta^{ID}(F^*) \subset \Theta^{ID,\epsilon}(F) \) for all \( F^* \in U \). Let \( \delta(\epsilon) = \sup_{F_1, F_2 \in U} d_{\tilde{F}}(F_1, F_2)/2 \), then for any \( \tilde{\theta} \in \Theta \setminus \Theta^{ID,\epsilon}(F) \) and \( p \) such that \( \theta(p) = \tilde{\theta} \),
\[ \inf_{F^* \in M^p(G^p)} d_{\tilde{F}}(F^*, F) \geq \delta(\epsilon) \]
holds. \( \square \)

B.1 Proof of proposition 11

Proof. If \( \tilde{\theta} \in \Theta^{ID}(F) \), there exists \( p \) such that \( \theta(p) = \tilde{\theta} \), and \( F \in M^p(G^p) \), this implies
\[ \inf_{F^* \in M^p(G^p)} \sqrt{a_n} d_{\tilde{F}}(F_n, F^*) \leq \sqrt{a_n} d_{\tilde{F}}(F_n, F) \leq c_n \]
with probability approaching 1. This the right hand side does not depend on \( \tilde{\theta} \), it implies \( \Theta^{ID}(F) \subset \hat{\Theta}^{ID} \) with probability approaching 1. This proves the first claim.

Next, we show \( d_H(\hat{\Theta}^{ID}, \Theta^{ID}(F)) \rightarrow_p 0 \) by showing \( \hat{\Theta}^{ID} \) does not intersect \( \Theta \setminus \Theta^{ID,\epsilon} \) with probability approaching 1 for all \( \epsilon > 0 \). It suffices to show that
\[ \inf_{F \in \Theta \setminus \Theta^{ID,\epsilon}(F)} \inf_{F^* \in M^p(G^p)} d_{\tilde{F}}(F_n, F^*) > c_n/\sqrt{a_n} \]
holds with probability approaching 1.

\[ \inf_{F \in \Theta \setminus \Theta^{ID,\epsilon}(F)} \inf_{F^* \in M^p(G^p)} d_{\tilde{F}}(F_n, F^*) \geq \inf_{F \in \Theta \setminus \Theta^{ID,\epsilon}(F)} \inf_{F^* \in M^p(G^p)} d_{\tilde{F}}(F, F^*) - d_{\tilde{F}}(F_n, F) \]
\[ \geq \delta(\epsilon) + O_p(1/\sqrt{a_n}) \]
where the last inequality follows from Lemma B.1. Since \( c_n/\sqrt{a_n} \rightarrow 1 \), \( \Theta^{ID} \subset \Theta^{ID,\epsilon}(F) \) with probability approaching 1. \( \square \)
### B.2 Proof of Proposition 12

**Proof.** Let $F$ be the distribution of $X$ and $\tilde{\theta} \in \Theta^D(F)$, then there exists a $p \in A$ such that $\theta(p) = \tilde{\theta}$ and $F \in M^p(G^p)$. If $F_n \in J^n_{\alpha_n}(F)$ holds, then $\tilde{\theta} \in \Theta_n^\alpha$ holds by construction of $\Theta_n^\alpha$. Since this holds for all $\tilde{\theta} \in \Theta^D(F)$, $\Theta^D(F) \in \Theta_n^\alpha$ holds whenever $F_n \in J^n_{\alpha_n}(F)$ holds. By assumption of $J^n_{\alpha_n}$,

$$Pr(\Theta^D(F) \subset \Theta_n^\alpha) \geq Pr(F_n \notin J^n_{\alpha_n}(F)) \leq \limsup_n 1 - \alpha_n = 1 - \alpha$$

\[ \square \]

### C Proofs in Potential Outcome Application

#### C.1 Proofs of Minimal Defier Assumption in section 5.1

**Lemma C.1.** Let $F$ be the distribution of observed variables, and let $P, Q$ be defined in (2.13). Let $p(y, i) = \frac{\partial P(Y_i \leq y | i)}{\partial y}$ and $q(y, i) = \frac{\partial Q(Y_i \leq y | i)}{\partial y}$, and define the set

$$Y_0 = \{y : q(y, 0) - p(y, 0) \geq 0\}$$

$$Y_1 = \{y : p(y, 1) - q(y, 1) \geq 0\}$$

(C.1)

For any $F \in \mathcal{F}$, consider the following $G^p$:

**For always taker:**

$$Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0, D(1) = 1, D(0) = 1 | Z_i = 1) = Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0, D(1) = 1, D(0) = 1 | Z_i = 0)$$

$$= [Q(B_1 \cap Y_i^c, 1) + P(B_1 \cap \mathcal{Y}_1^c, 1)] \times F\alpha_0^n(B_0)$$

(C.2)

where $F\alpha_0^n$ is any probability measure.

**For never taker:**

$$Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0, D(1) = 0, D(0) = 0 | Z_i = 1) = Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0, D(1) = 0, D(0) = 0 | Z_i = 0)$$

$$= [Q(B_0 \cap \mathcal{Y}_0^c, 0) + P(B_0 \cap \mathcal{Y}_0, 0)] \times F\alpha_1^n(B_1)$$

(C.3)

where $F\alpha_1^n$ is any probability measure.
Then

\[ \text{For defier:} \]

\[ \Pr_{G^p}(Y_i(1) \in B, Y_i(0) \in B_0, D(1) = 0, D(0) = 1|Z_i = 0) = \frac{[Q(B_1 \cap \mathcal{Y}_i^c, 1) - P(B_1 \cap \mathcal{Y}_i^c, 1)] \times [P((B_0 \cap \mathcal{Y}_0^c, 0) - Q(B_0 \cap \mathcal{Y}_0^c, 0)]}{Q(\mathcal{Y}_0^c, 0) - P(\mathcal{Y}_0^c, 0)} \]  \tag{C.4} 

\[ \Pr_{G^p}(Y_i(1) \in B, Y_i(0) \in B_0, D(1) = 0, D(0) = 1|Z_i = 1) = \frac{[Q(B_1 \cap \mathcal{Y}_i^c, 1) - P(B_1 \cap \mathcal{Y}_i^c, 1)] \times [P((B_0 \cap \mathcal{Y}_0^c, 0) - Q(B_0 \cap \mathcal{Y}_0^c, 0)]}{P(\mathcal{Y}_i^c, 1) - Q(\mathcal{Y}_i^c, 1)} \]  \tag{C.5} 

\[ \text{For complier:} \]

\[ \Pr_{G^p}(Y_i(1) \in B, Y_i(0) \in B_0, D(1) = 1, D(0) = 0|Z_i = 0) = \frac{[P(B_1 \cap \mathcal{Y}_1, 1) - Q(B_1 \cap \mathcal{Y}_1, 1)] \times [Q(B_0 \cap \mathcal{Y}_0, 0) - P(B_0 \cap \mathcal{Y}_0, 0)]}{P(\mathcal{Y}_1, 1) - Q(\mathcal{Y}_1, 1)} \]  \tag{C.6} 

\[ \Pr_{G^p}(Y_i(1) \in B, Y_i(0) \in B_0, D(1) = 1, D(0) = 0|Z_i = 1) = \frac{[P(B_1 \cap \mathcal{Y}_1, 1) - Q(B_1 \cap \mathcal{Y}_1, 1)] \times [Q(B_0 \cap \mathcal{Y}_0, 0) - P(B_0 \cap \mathcal{Y}_0, 0)]}{Q(\mathcal{Y}_0^c, 0) - P(\mathcal{Y}_0^c, 0)} \]  \tag{C.7} 

Then \( F = M^p(G^p) \) and \( p \in \tilde{A} \) where \( \tilde{A} \) is defined by assumption \[5.1\]

**Proof.** First we show \( F = M^p(G^p) \). Since \( Z \) is in both observed and latent variable, it suffices to show the conditional distribution \( F(Y_i, D_i|Z_i) \) can be matched. Note that

\[ \Pr_{M^p(G^p)}(Y_i \in B, D_i = 1|Z_i = 1) = \sum_{j=0}^{1} \Pr(Y_i(1) \in B, Y_i(0) \in \mathcal{Y}, D_i(1) = 1, D_i(0) = j|Z_i = 1) \]

\[ = [Q(B \cap \mathcal{Y}_1, 1) + P(B \cap \mathcal{Y}_1^c, 1)] + [P(B \cap \mathcal{Y}_1, 1) - Q(B \cap \mathcal{Y}_1, 1)] \]

\[ = P(B, 1) \]

\[ = \Pr_F(Y_i \in B, D_i = 1|Z_i = 1) \]  \tag{C.8} 

where the first equality holds by potential outcome model \[2.6\], the second follows from our construction and the last follows by definition of \( P \). We can check the same reasoning holds for all \( P(B, i) \) and \( Q(B, i) \), so \( F = M^p(G^p) \). The construction of \( G^p \) also directly satisfies conditions 1 and 2 in \( \tilde{A} \). 

54
It suffice to show that any $p^*$ satisfies condition 1 and 2, has larger defier of measure than $p$. We first note that

$$P(B_1, 1) = Pr_{G^p}(Y_1(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 1, D_i(0) = 1|Z_i = 1)$$

$$+ Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 1, D_i(0) = 0|Z_i = 1)$$

$$Q(B_1, 1) = Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 1, D_i(0) = 1|Z_i = 0)$$

$$+ Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 0, D_i(0) = 1|Z_i = 0)$$

(C.9)

Since $p^*$ satisfies condition 1 (type independence) and 2 (same always taker and never taker),

$$Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 1, D_i(0) = 1|Z_i = 1) = Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 1, D_i(0) = 0|Z_i = 1) = Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 1, D_i(0) = 1|Z_i = 0)$$

Take the difference to get

$$Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 0, D_i(0) = 1|Z_i = 0)$$

$$= Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 1, D_i(0) = 0|Z_i = 1) + Q(B_1, 1) - P(B_1, 1)$$

$$\geq \max\{0, Q(B_1, 1) - P(B_1, 1)\}$$

holds for all $B_1$. Then

$$Pr_{G^p}(Y_i(1) \in \mathcal{Y}, Y_i(0) \in \mathcal{Y}, D_i(1) = 0, D_i(0) = 1|Z_i = 0) \geq Q(\mathcal{Y}_i^c, 1) - P(\mathcal{Y}_i^c, 1)$$

holds by definition of $\mathcal{Y}_i$. Similar we can show

$$Pr_{G^p}(Y_i(1) \in \mathcal{Y}, Y_i(0) \in \mathcal{Y}, D_i(1) = 0, D_i(0) = 1|Z_i = 1) \geq P(\mathcal{Y}_0^c, 0) - Q(\mathcal{Y}_0^c, 0)$$

The measure of defier under $p^*$ is

$$m^d(p^*) = Pr_{G^p}(D_i(1) = 0, D_i(0) = 1|Z_i = 0) \times Pr_{G^p}(Z_i = 0)$$

$$+ Pr_{G^p}(D_i(1) = 0, D_i(0) = 1|Z_i = 1) \times Pr_{G^p}(Z_i = 1)$$

$$\geq [Q(\mathcal{Y}_i^c, 1) - P(\mathcal{Y}_i^c, 1)] \times Pr_{G^p}(Z_i = 0) + [P(\mathcal{Y}_0^c, 0) - Q(\mathcal{Y}_0^c, 0)] \times Pr_{G^p}(Z_i = 1)$$

$$= Pr_{G^p}(D_i(1) = 0, D_i(0) = 1|Z_i = 0) \times Pr_{G^p}(Z_i = 0)$$

$$+ Pr_{G^p}(D_i(1) = 0, D_i(0) = 1|Z_i = 1) \times Pr_{G^p}(Z_i = 1)$$

where the last equality holds by construction of $G^p$ and the fact that $M^p(G^p) = M^p(G^{p^*})$ implies $Pr_{G^p}(Z_i = j) = Pr_{G^{p^*}}(Z_i = j)$ for $j = 0, 1$.

□

55
C.1.1 Proof of proposition \[14\]

**Proof.** By construction of $\tilde{A}$, $A \subset \tilde{A}$ holds. Also, by Lemma \[C.1\] any $F$ can be rationalize by a structure $p \in \tilde{A}$, so $\mathcal{H}_P^{snf}(\tilde{A}) = \mathcal{P}$.

It remains to show $\mathcal{H}_P^{snf}(A) \cap \tilde{A} = A$. Since $A \subset \mathcal{H}_P^{snf}(A)$ and $A \subset \tilde{A}$, it suffices to show $\mathcal{H}_P^{snf}(A) \cap \tilde{A} \subset A$. Let $p \in \mathcal{H}_P^{snf}(A) \cap \tilde{A}$, then $P(B, 1) - Q(B, 1) \geq 0$ and $Q(B, 0) - P(B, 0) \geq 0$ holds for all $B$. The construction in Lemma \[C.1\] gives zero measure of defier. Since minimal measure of defier is unique, and $p \in \tilde{A}$, $m^d(p) = 0$ must hold, this implies $D_i(1) \geq D_i(0)$ almost surely under $G^p$. Then

\[
\begin{align*}
Pr_{G^p}(Y_i(1) &\in B_1, Y_i(0) \in B_0, D_i(1) = 1, D_i(0) = 0|Z_i = 1) \\
= &\ (1) \ Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0|D_i(1) = 1, D_i(0) = 0, Z_i = 1) \times Pr_{G^p}(D_i(1) = 1, D_i(0) = 0|Z_i = 1) \\
= &\ (2) \ Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0|D_i(1) = 1, D_i(0) = 0, Z_i = 1) \\
&\times (1 - Pr_{G^p}(D_i(1) = 1, D_i(0) = 1|Z_i = 1) - Pr_{G^p}(D_i(1) = 0, D_i(0) = 0|Z_i = 1)) \\
= &\ (3) \ Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0|D_i(1) = 1, D_i(0) = 0, Z_i = 0) \\
&\times (1 - Pr_{G^p}(D_i(1) = 1, D_i(0) = 1|Z_i = 1) - Pr_{G^p}(D_i(1) = 0, D_i(0) = 0|Z_i = 1)) \\
= &\ (4) \ Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0|D_i(1) = 1, D_i(0) = 0, Z_i = 0) \\
&\times (1 - Pr_{G^p}(D_i(1) = 1, D_i(0) = 1|Z_i = 0) - Pr_{G^p}(D_i(1) = 0, D_i(0) = 0|Z_i = 0)) \\
= &\ (5) \ Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0, D_i(1) = 1, D_i(0) = 0|Z_i = 0)
\end{align*}
\]

(C.10)

where (1) and (5) follows by conditional probability, (2) follows by $D_i(1) \geq D_i(0)$ almost surely, (3) follows by type independence condition 1, (4) follows by measure of always taker and never taker is independent of $Z_i$. We can similarly show

\[
Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0, D_i(1) = D_i(0) = j|Z_i = 1)
\]

(C.11)

for $j = 0, 1$. Then equations (C.10) and (C.11) implies $(Y_i(1), Y_i(0), D_i(1), D_i(0)) \perp Z_i$ holds for $p$. 

\[\square\]
C.2 Proof of Minimal Distance to Independent Instrument Assumption in section \[5.2\]

**Lemma C.2.** For any \( F \in \mathcal{F} \), consider the following \( G^p \):

For always taker:

\[
\Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0, D(1) = 1, D(0) = 1 \mid Z_i = 1) = \left[ Q(B_1 \cap Y_1, 1) + P(B_1 \cap Y_1^c, 1) \right] \times F^0_a(B_0)
\]

\[
\Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0, D(1) = 1, D(0) = 1 \mid Z_i = 0) = Q(B_1, 1) \times F^0_a(B_0)
\]  \hspace{1cm} (C.12)

where \( F^0_a \) is any probability measure.

For never taker:

\[
\Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0, D(1) = 0, D(0) = 0 \mid Z_i = 1) = \left[ Q(B_0 \cap Y_0^c, 0) + P(B_0 \cap Y_0, 0) \right] \times F^1_n(B_1)
\]

\[
\Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0, D(1) = 0, D(0) = 0 \mid Z_i = 0) = P(B_0, 0) \times F^1_n(B_1)
\]  \hspace{1cm} (C.13)

where \( F^1_n \) is any probability measure.

For complier:

\[
\Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0, D(1) = 1, D(0) = 0 \mid Z_i = 0) = \frac{[P(B_1 \cap Y_1, 1) - Q(B_1 \cap Y_1, 1)] \times \left[ Q(B_0 \cap Y_0, 0) - P(B_0 \cap Y_0, 0) \right]}{P(Y_1, 1) - Q(Y_1, 1)}
\]

\[
\Pr_{G^p}(Y_i(1) \in B_1, Y_i(0) \in B_0, D(1) = 1, D(0) = 0 \mid Z_i = 1) = \frac{[P(B_1 \cap Y_1, 1) - Q(B_1 \cap Y_1, 1)] \times \left[ Q(B_0 \cap Y_0, 0) - P(B_0 \cap Y_0, 0) \right]}{Q(Y_0, 0) - P(Y_0, 0)}
\]  \hspace{1cm} (C.14)

For defier, the probability measure is always zero: \( \Pr_{G^p}(D(1) = 0, D(0) = 1) = 0 \)

Then \( F = M^p(G^p) \) and \( p \in \tilde{A} \) where \( \tilde{A} \) is defined by assumption \[5.2\]

**Proof.** First we show \( F = M^p(G^p) \). Since \( Z \) is in both observed and latent variable, it suffices
to show the conditional distribution $F(Y_i, D_i | Z_i)$ can be matched. Note that

$$Pr_{MP(G^p)}(Y_i \in B, D_i = 1 | Z_i = 1) = \sum_{j=0}^{1} Pr(Y_i(1) \in B, Y_i(0) \in \mathcal{Y}, D_i(1) = j, D_i(0) = j | Z_i = 1)$$

$$= [Q(B \cap \mathcal{Y}_i, 1) + P(B \cap \mathcal{Y}_c, 1)] + [P(B \cap \mathcal{Y}_i, 1) - Q(B \cap \mathcal{Y}_1, 1)]$$

$$= P(B, 1)$$

$$= Pr_F(Y_i \in B, D_i = 1 | Z_i = 1)$$

(C.16)

where the first equality holds by potential outcome model (2.6), the second follows from our construction and the last follows by definition of $P$. Similarly,

$$Pr_{MP(G^p)}(Y_i \in B, D_i = 1 | Z_i = 0) = \sum_{j=0}^{1} Pr(Y_i(1) \in B, Y_i(0) \in \mathcal{Y}, D_i(1) = j, D_i(0) = 1 | Z_i = 0)$$

$$= 0 + Q(B, 1)$$

$$= Pr_F(Y_i \in B, D_i = 1 | Z_i = 0)$$

(C.17)

We can check the condition holds for $P(B, 0)$ and $Q(B, 0)$ similarly. So $F = MP(G^p)$.

The constructed measure $G^p$ has is the same for compliance group as in Lemma [C.1], so type independence condition for the complier holds as proved in Lemma [C.1].

Now I show $p \in \bar{A}$. For any $p^*$ that satisfies condition 1 and 2 of $\bar{A}$, since there is no defier,

$$P(B_0, 0) = Pr_{Gp^*}(Y_i(0) \in B_0, D(1) = 0, D(0) = 0 | Z_i = 0)$$

$$P(B_1, 1) = Pr_{Gp^*}(Y_i(1) \in B, Y_i(0) \in \mathcal{Y}, D_i(1) = 1, D_i(0) = 1 | Z_i = 1)$$

$$+ Pr_{Gp^*}(Y_i(1) \in B, Y_i(0) \in \mathcal{Y}, D_i(1) = 1, D_i(0) = 0 | Z_i = 1)$$

$$Q(B_1, 1) = Pr_{Gp^*}(Y_i(1) \in B_1, D(1) = 1, D(0) = 1 | Z_i = 0)$$

$$Q(B_1, 1) = Pr_{Gp^*}(Y_i(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 1, D_i(0) = 1 | Z_i = 0)$$

$$+ Pr_{Gp^*}(Y_i(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 0, D_i(0) = 1 | Z_i = 0)$$

(C.18)
must hold. Use the definition of $R(p)$ and the expression above, we have

$$R(p^*) = \int_{y_1} [g_{p^*}(y_1, D(1) = D(0) = 1|Z = 1) - g_{p^*}(y_1, D(1) = D(0) = 0|Z = 0)]^2 dy_1$$

$$+ \int_{y_0} [g_{p^*}(y_0, D(1) = D(0) = 0|Z = 1) - g_{p^*}(y_0, D(1) = D(0) = 0|Z = 0)]^2 dy_0$$

$$= \int_{y_1} [p(y_1, 1) - q(y_1, 1) - g_{p^*}(y_1, D(1) = 1, D(0) = 0|Z = 1)]^2 dy_1$$

$$+ \int_{y_0} [q(y_0, 0) - p(y_0, 0) - g_{p^*}(y_0, D(1) = 1, D(0) = 0|Z = 0)]^2 dy_0$$

(C.19)

Where the last equality holds by (C.18), since (C.18) implies

$$g_{p^*}(y_1, D(1) = D(0) = 1|Z = 0) = q(y_1, 1)$$

$$g_{p^*}(y_1, D(1) = D(0) = 1|Z = 1) = p(y_1, 1) - g_{p^*}(y_1, D(1) = 1, D(0) = 0|Z = 1)$$

$$g_{p^*}(y_0, D(1) = D(0) = 0|Z = 1) = p(y_0, 0)$$

$$g_{p^*}(y_0, D(1) = D(0) = 0|Z = 0) = q(y_0, 0) - g_{p^*}(y_0, D(1) = 1, D(0) = 0|Z = 0)$$

(C.20)

To minimize $R(p^*)$, we can consider an optimization problem

$$\min_{t \geq 0} (x - t)^2$$

(C.21)

When $x \geq 0$, the minimizer is $t^* = x$, and the minimized value is 0; When $x < 0$, the minimizer is $t^* = 0$, and the minimized value is $x^2$.

Recall the definition of $\mathcal{Y}_1$ and $\mathcal{Y}_0$, and density $g_{p^*}(y_1, D(1) = 1, D(0) = 0|Z = 1) \geq 0$ and $g_{p^*}(y_0, D(1) = 1, D(0) = 0|Z = 0) \geq 0$, we have

$$R(p^*) \geq \int_{y_1 \in \mathcal{Y}_1} [p(y_1, 1) - q(y_1, 1)]^2 dy_1 + \int_{y_0 \in \mathcal{Y}_0} [q(y_0, 0) - p(y_0, 0)]^2 dy_0$$

$$= R(p)$$

(C.22)

where the last equality holds by construction of structure $p$. The inequality holds with equality when $p^*$ satisfies

$$g_{p^*}(y_1, D(1) = 1, D(0) = 0|Z = 1) = \max\{p(y_1, 1) - q(y_1, 1), 0\}$$

$$g_{p^*}(y_0, D(1) = 1, D(0) = 0|Z = 0) = \max\{q(y_0, 0) - p(y_0, 0), 0\}$$

(C.23)

This proves the $p$ constructed is $R(p)$ minimal. □
C.3 Proof of Proposition 15

I first prove the following Lemma.

Lemma C.3. Let \( p \) be constructed in Lemma C.1, for any \( p^* \in \tilde{A} \) such that \( F = M^{p^*}(G^{p^*}) \) and

\[
Pr_{G^{p^*}}(D_i(1) = 0, D_i(0) = 1) = P_F(Z_i = 1)(Q(\mathcal{Y}_i^c, 1) - P(\mathcal{Y}_i^c, 1)) + P_F(Z_i = 0)(P(\mathcal{Y}_0^c, 0) - Q(\mathcal{Y}_0^c, 0))
\]

it must satisfies:

\[
Pr_{G^{p^*}}(Y_i(1) \in B_1, D_i(1) = 1, D_i(0) = 1 | Z_i = 1) = Pr_{G^{p^*}}(Y_i(1) \in B_1, D_i(1) = 1, D_i(0) = 1 | Z_i = 1)
\]
\[
Pr_{G^{p^*}}(Y_i(0) \in B_0, D_i(1) = 1, D_i(0) = 1 | Z_i = 1) = Pr_{G^{p^*}}(Y_i(0) \in B_0, D_i(1) = 1, D_i(0) = 1 | Z_i = 1)
\]

Proof. By proof of Lemma C.1 since \( p^* \) is in \( \tilde{A} \),

\[
Pr_{G^{p^*}}(Y_i(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 0, D_i(0) = 1 | Z_i = 0) = \max\{0, Q(B_1, 1) - P(B_1, 1)\}
\]

holds for all set \( B_1 \), where the equality holds because \( p^* \) also achieves the minimal defier measure. Note that

\[
Pr_{G^{p^*}}(Y_i(1) \in B_1, Y_i(0) \in \mathcal{Y}, D_i(1) = 0, D_i(0) = 1 | Z_i = 0) = \max\{0, Q(B_1, 1) - P(B_1, 1)\}
\]

holds by construction of \( p \), so the result follows.

C.3.1 Main Proof

Proof. If first show that the displayed identified LATE holds for \( \tilde{A} \) that satisfies C.1. Since local average treatment effect is defined as \( E[Y_i(1) - Y_i(0) | D_i(1) - D_i(0) = 1] \), we look at the term \( E[Y_i(1) | D_i(1) - D_i(0)] = 1 \) first. We define the following quantity

\[
g_{p^*}^1(y) = \frac{\partial Pr_p(Y_i(1) \leq y | D_i(1) = 1, D_i(0) = 1)}{\partial y}
\]
\[
g_{p^*}^0(y) = \frac{\partial Pr_p(Y_i(0) \leq y | D_i(1) = 1, D_i(0) = 1)}{\partial y}
\] (C.24)
\[ E[Y_i(1)|D_i(1) - D_i(0) = 1] = \int_y y \cdot g_{p^*}(y) dy \]

where (1) follows by definition of conditional expectation, (2) follows by \((Y_i(1) \perp D_i(1), D_i(0))|Z_i\), (3) follows by Bayes’ rule, (4) follows by Lemma C.3, (5) holds by construction of \(p\). We can similarly construct the quantity for \(E[Y_i(0)|D_i(1) - D_i(0) = 1]\), the result follows.

Now I show that the displayed identified LATE holds for \(\bar{A}\) that satisfies assumption 5.2. For any \(\bar{p} \in \bar{A}\), it achieves the minimal \(R(p^\ast)\) over all \(p^\ast\) that satisfies no defier and type independence of complier. So \(\bar{p}\) must satisfy density constraint in (C.23):

\[
\begin{align*}
  g_{\bar{p}}(y_1, D(1) = 1, D(0) = 0|Z = 1) &= \max\{p(y_1, 1) - q(y_1, 1), 0\} \\
  g_{\bar{p}}(y_0, D(1) = 1, D(0) = 0|Z = 0) &= \max\{q(y_0, 0) - p(y_0, 0), 0\}
\end{align*}
\]

The identified value of LATE is

\[
\text{LATE}(\bar{p}) = \frac{\int y_1 g_{\bar{p}}(y_1, D(1) = 1, D(0) = 0|Z = 1) dy_1}{P(Y_1, 1) - Q(Y_1, 1)} - \frac{\int y_0 g_{\bar{p}}(y_0, D(1) = 1, D(0) = 0|Z = 0) dy_0}{P(Y_0, 1) - Q(Y_0, 1)}
\]

\[
= \frac{\int y_1 (p(y_1) - q(y_1)) dy}{P(Y_1, 1) - Q(Y_1, 1)} - \frac{\int y_0 (q(y_0) - p(y_0)) dy}{P(Y_0, 1) - Q(Y_0, 1)}
\]

\[\square\]

### C.4 LATE statistical properties

#### C.4.1 Useful Lemmas

**Lemma C.4.** Let \(h_n = n^{-\gamma}\) for some \(\gamma \in (0, 1)\), such that \(\frac{\sqrt{n}}{\log h_n} \rightarrow \infty\), define \(a_n = \min\{\sqrt{\frac{\sqrt{n}}{\log h_n}}, h_n^{-2}\}\). Suppose assumptions 5.4 and 5.5 hold, then there exists a constant \(C\)
such that

\[
\limsup_{n \to \infty} a_n \sup_y \left| \frac{1}{nh_n} \sum_{i=1}^{n} K \left( \frac{Y_i - y}{h_n} \right) \mathbb{1}(D_i = 1, Z_i = 1) - p(y, 1) \Pr(Z_i = 1) \right| \leq C \quad \text{a.s.}
\]

\[
\limsup_{n \to \infty} a_n \sup_y \left| \frac{1}{nh_n} \sum_{i=1}^{n} K \left( \frac{Y_i - y}{h_n} \right) \mathbb{1}(D_i = 0, Z_i = 1) - p(y, 0) \Pr(Z_i = 1) \right| \leq C \quad \text{a.s.}
\]

\[
\limsup_{n \to \infty} a_n \sup_y \left| \frac{1}{nh_n} \sum_{i=1}^{n} K \left( \frac{Y_i - y}{h_n} \right) \mathbb{1}(D_i = 1, Z_i = 0) - q(y, 1) \Pr(Z_i = 1) \right| \leq C \quad \text{a.s.}
\]

\[
\limsup_{n \to \infty} a_n \sup_y \left| \frac{1}{nh_n} \sum_{i=1}^{n} K \left( \frac{Y_i - y}{h_n} \right) \mathbb{1}(D_i = 0, Z_i = 0) - q(y, 1) \Pr(Z_i = 0) \right| \leq C \quad \text{a.s.}
\]

(C.27)

Proof. We prove the first inequality and the rest follows similarly. By triangular inequality

\[
a_n \sup_y |f_n(y) - p(y, 1) \Pr(Z_i = 1)| \leq a_n \sup_y |\bar{f}_n(y) - p(y, 1) \Pr(Z_i = 1)| + a_n |f_n(y) - \bar{f}_n(y)|
\]

\[
\leq a_n \sup_y |\bar{f}_n(y) - p(y, 1) \Pr(Z_i = 1)| + \sqrt{\frac{nh_n}{\log h_n}} |f_n(y) - \bar{f}_n(y)|
\]

The first term \(a_n \sup_y |\bar{f}_n(y) - p(y, 1) \Pr(Z_i = 1)|\) is the bias term and can be calculated as

\[
a_n \sup_y |\bar{f}_n(y) - p(y, 1) \Pr(Z_i = 1)|
\]

\[
= a_n \left| \frac{1}{h_n} \int K \left( \frac{t - y}{h_n} \right) f(t|D_i = 1, Z_i = 1) - f(y|D_i = 1, Z_i = 1) \right| \Pr(D_i = 1, Z_i = 1)
\]

\[
= a_n \left| \int K(u)[f(y + uh_n|D_i = 1, Z_i = 1) - f(y|D_i = 1, Z_i = 1)] \right| \Pr(D_i = 1, Z_i = 1)
\]

\[
= a_n \left| \int K(u)uh_n f'(y|D_i = 1, Z_i = 1) + K(u)u^2 h_n^2 f''(y|D_i = 1, Z_i = 1) + o(h_n^2) \right| \Pr(D_i = 1, Z_i = 1)
\]

\[
\leq h_n^{-2} \left| \int K(u)u^2 h_n^2 f''(y|D_i = 1, Z_i = 1) + o(h_n^2) \right| \Pr(D_i = 1, Z_i = 1)
\]

\[
\leq \left| \int K(u)u^2 f''(y|D_i = 1, Z_i = 1) \right| \Pr(D_i = 1, Z_i = 1) + o(1)
\]

(C.28)

Let \(C\) be the constant in Lemma [D.3], and set set \(C = \bar{C} + \left| \int_u K(u)u^2 f''(y|D_i = 1, Z_i = 1) \right|\).

The result follows.
Lemma C.5. Let $h_n = n^{-\gamma}$ for some $\gamma \in (0,1)$, such that $\frac{nh_n}{\log h_n} \to \infty$ and let $a_n = \min\{\sqrt{\frac{nh_n}{\log h_n}}, h_n^{-2}\}$. If there exists a constant $c > 0$ such that $Pr(Z_i = 1) \in [c, 1 - c]$, then there exists an $\epsilon > 0$ such that

$$n^{-\epsilon}a_n \sup_{y} |g_h(y, 1) - (p(y, 1) - q(y, 1))| = o_p(1)$$

$$n^{-\epsilon}a_n \sup_{y} |g_h(y, 0) - (q(y, 0) - q(y, 0))| = o_p(1) \tag{C.29}$$

Proof. Note that

$$|g_h(y, 1) - (p(y, 1) - q(y, 1))| \leq \sup_{y} \left| \frac{f_n^{1,1}(y)}{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(Z_i = 1)} - \frac{p(y, 1)Pr(Z_i = 1)}{Pr(Z_i = 1)} \right| + \sup_{y} \left| \frac{f_n^{1,0}(y)}{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(Z_i = 1)} - \frac{q(y, 1)Pr(Z_i = 1)}{Pr(Z_i = 1)} \right| \tag{C.30}$$

For the first term, we have

$$n^{-\epsilon}a_n \sup_{y} \left| \frac{f_n^{1,1}(y)}{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(Z_i = 1)} - \frac{p(y, 1)Pr(Z_i = 1)}{Pr(Z_i = 1)} \right| \leq n^{-\epsilon}a_n \left| \frac{f_n^{1,1}(y)}{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(Z_i = 1)} - \frac{f_n^{1,1}(y)}{Pr(Z_i = 1)} \right| + n^{-\epsilon}a_n \left| \frac{f_n^{1,1}(y)}{Pr(Z_i = 1)} - \frac{p(y, 1)}{Pr(Z_i = 1)} \right|$$

$$= \sup_{y} |p(y, 1) + o(1)|n^{-\epsilon}a_n O_p\left(\frac{1}{\sqrt{n}}\right) + \frac{1}{Pr(Z_i = 1)} n^{-\epsilon}a_n \sup_{y} |f_n^{1,1}(y) - p(y, 1)|$$

$$\leq \sup_{y} p(y, 1)O_p\left(\sqrt{\frac{n^{-2\epsilon}h_n}{\log h_n^{-1}}} \right) + o(n^{-\epsilon})$$

$$= o_p(1)$$

where equality $\ast$ follows by $|\sum_{i=1}^{n} \mathbb{1}(Z_i = 1) - Pr(Z_i = 1)| = O_p(1/\sqrt{n})$ and continuous mapping holds when $Pr(Z_i = 1) > 0$. By the same argument, the second term in (C.30) is also $o_p(1)$. The result follows. \qed

Lemma C.6. (Limit Distribution of Infeasible Components) Define $g(y, 1) = p(y, 1) - q(y, 1)$
and $g(y, 0) = q(y, 0) - p(y, 0)$, then for any $b_n \to 0$,

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( Y_i \left[ \frac{1(D_i=1, Z_i=1)}{Pr(Z_i=1)} - \frac{1(D_i=1, Z_i=0)}{Pr(Z_i=0)} \right] \mathbb{1}(g(Y_i, 1) \geq b_n) - \int_{y \in \mathcal{Y}} yg(y, 1) \mathbb{1}(g(y, 1) \geq b_n) \\
- Y_i \left[ \frac{1(D_i=0, Z_i=0)}{Pr(Z_i=0)} - \frac{1(D_i=0, Z_i=1)}{Pr(Z_i=1)} \right] \mathbb{1}(g(Y_i, 0) \geq b_n) - \int_{y \in \mathcal{Y}} yg(y, 0) \mathbb{1}(g(y, 0) \geq b_n) \right)
\to_d N(0, \Sigma)
$$

where

$$
\Sigma = \text{Var} \left( \begin{array}{c}
Y_i \left[ \frac{1(D_i=1, Z_i=1)}{Pr(Z_i=1)} - \frac{1(D_i=1, Z_i=0)}{Pr(Z_i=0)} \right] \mathbb{1}(g(Y_i, 1) \geq 0) \\
Y_i \left[ \frac{1(D_i=0, Z_i=0)}{Pr(Z_i=0)} - \frac{1(D_i=0, Z_i=1)}{Pr(Z_i=1)} \right] \mathbb{1}(g(Y_i, 0) \geq 0) \\
\mathbb{1}(Z_i = 1) \\
\mathbb{1}(Z_i = 0) \\
\mathbb{1}(Z_i = 1) \\
\mathbb{1}(Z_i = 0) \\
\left[ \frac{1(D_i=1, Z_i=1)}{Pr(Z_i=1)} - \frac{1(D_i=1, Z_i=0)}{Pr(Z_i=0)} \right] \mathbb{1}(g(Y_i, 1) \geq 0) \\
\left[ \frac{1(D_i=0, Z_i=0)}{Pr(Z_i=0)} - \frac{1(D_i=0, Z_i=1)}{Pr(Z_i=1)} \right] \mathbb{1}(g(Y_i, 0) \geq 0)
\end{array} \right)
$$

Proof. First note that

$$
E \left( \begin{array}{c}
Y_i \left[ \frac{1(D_i=1, Z_i=1)}{Pr(Z_i=1)} - \frac{1(D_i=1, Z_i=0)}{Pr(Z_i=0)} \right] \mathbb{1}(g(Y_i, 1) \geq b_n) - \int_{y \in \mathcal{Y}} yg(y, 1) \mathbb{1}(g(y, 1) \geq b_n) \\
Y_i \left[ \frac{1(D_i=0, Z_i=0)}{Pr(Z_i=0)} - \frac{1(D_i=0, Z_i=1)}{Pr(Z_i=1)} \right] \mathbb{1}(g(Y_i, 0) \geq b_n) - \int_{y \in \mathcal{Y}} yg(y, 0) \mathbb{1}(g(y, 0) \geq b_n) \\
\mathbb{1}(Z_i = 1) - Pr(Z_i = 1) \\
\mathbb{1}(Z_i = 0) - Pr(Z_i = 0) \\
\left[ \frac{1(D_i=1, Z_i=1)}{Pr(Z_i=1)} - \frac{1(D_i=1, Z_i=0)}{Pr(Z_i=0)} \right] \mathbb{1}(g(Y_i, 1) \geq b_n) - \int_{y \in \mathcal{Y}} g(y, 1) \mathbb{1}(g(y, 1) \geq b_n) \\
\left[ \frac{1(D_i=0, Z_i=0)}{Pr(Z_i=0)} - \frac{1(D_i=0, Z_i=1)}{Pr(Z_i=1)} \right] \mathbb{1}(g(Y_i, 0) \geq b_n) - \int_{y \in \mathcal{Y}} g(y, 0) \mathbb{1}(g(y, 0) \geq b_n)
\end{array} \right) = 0
$$

(C.32)
and there exists $\delta > 0$ such that $E[Y_i^{2+\delta}] < \infty$. Let

$$X_{ni} = \left\{ \begin{array}{l}
Y_i\left[\frac{1(D_i=1,Z_i=1)}{Pr(Z_i=1)} - \frac{1(D_i=1,Z_i=0)}{Pr(Z_i=0)}\right] \mathbb{1}(g(Y_i, 1) \geq b_n) - \int_{y \in \mathcal{Y}} yg(y, 1) \mathbb{1}(g(y, 1) \geq b_n)dy
\end{array} \right. - \int_{y \in \mathcal{Y}} yg(y, 0) \mathbb{1}(g(y, 0) \geq b_n)dy - \int_{y \in \mathcal{Y}} yg(y, 1) \mathbb{1}(g(y, 1) \geq b_n)dy
\]$$

we have $E[||X_{ni}||_2^{2+\delta}] \leq \max\{E[Y_i^{2+\delta}], 1\} < \infty$, and

$$Var(X_{ni}) = Var\left( \begin{array}{l}
Y_i\left[\frac{1(D_i=1,Z_i=1)}{Pr(Z_i=1)} - \frac{1(D_i=1,Z_i=0)}{Pr(Z_i=0)}\right] \mathbb{1}(g(Y_i, 1) \geq b_n) - \int_{y \in \mathcal{Y}} yg(y, 1) \mathbb{1}(g(y, 1) \geq b_n)dy
\end{array} \right. - \int_{y \in \mathcal{Y}} yg(y, 0) \mathbb{1}(g(y, 0) \geq b_n)dy - \int_{y \in \mathcal{Y}} yg(y, 1) \mathbb{1}(g(y, 1) \geq b_n)dy
\]$$

by dominated convergence theorem, since $\mathbb{1}(g(y, 1) \geq b_n) \to \mathbb{1}(g(y, 1) \geq 0)$ and $E[Y_i^{2+\delta}] < \infty$. So by Lyapounov CLT, the triangular array converges to $N(0, \Sigma)$.

**Lemma C.7.** Suppose assumption 5.7 holds, then 5.6 holds. In particular

$$\int_{y \in \mathcal{Y}} yg(y, 1) \mathbb{1}(g(y, 1) \geq b_n)dy - \int_{y \in \mathcal{Y}} yg(y, 1) \mathbb{1}(g(y, 1) \geq -b_n)dy \leq MC \sup\{|y| : y \in \mathcal{C}_1\} b_n^2$$

$$\int_{y \in \mathcal{Y}} yg(y, 0) \mathbb{1}(g(y, 0) \geq b_n)dy - \int_{y \in \mathcal{Y}} yg(y, 1) \mathbb{1}(g(y, 1) \geq -b_n)dy \leq MC \sup\{|y| : y \in \mathcal{C}_0\} b_n^2$$

$$\int_{y \in \mathcal{Y}} g(y, 1) \mathbb{1}(g(y, 1) \geq b_n)dy - \int_{y \in \mathcal{Y}} g(y, 1) \mathbb{1}(g(y, 1) \geq -b_n)dy \leq MCb_n^2$$

$$\int_{y \in \mathcal{Y}} g(y, 0) \mathbb{1}(g(y, 0) \geq b_n)dy - \int_{y \in \mathcal{Y}} g(y, 1) \mathbb{1}(g(y, 1) \geq -b_n)dy \leq MCb_n^2$$

(C.33)
Proof. It is easy to see that
\[
\Bigg| \int_{y \in \mathcal{Y}} yg(y, 1) \mathbb{I}(g(y, 1) \geq b_n) dy - \int_{y \in \mathcal{Y}} yg(y, 1) \mathbb{I}(g(y, 1) \geq -b_n) dy \Bigg|
\]
\[
= \Bigg| \int_{y \in \mathcal{Y}} yg(y, 1) \mathbb{I}(-b_n \leq g(y, 1) \leq b_n) dy \Bigg|
\]
\[
\leq \sup\{ |y| : y \in \mathcal{C}_i \} \int_{y \in \mathcal{Y}} g(y, 1) \mathbb{I}(-b_n \leq g(y, 1) \leq b_n) dy
\]
\[
= \sup\{ |y| : y \in \mathcal{C}_1 \} \int_{y \in \mathcal{Y}} g(y, 1) \mathbb{I}(-b_n \leq g(y, 1) \leq b_n) dy
\]
\[
\leq MC \sup\{ |y| : y \in \mathcal{C}_1 \} b_n^2
\]
The same argument holds the rest of the inequalities. □

Lemma C.8. Let \( h_n = n^{-1/5} \) and \( b_n = n^{-1/4} / \log n \), and assumption [5.3]-[5.6] holds, then
\[
\frac{1}{n} \sum_{i=1}^{n} Y_i \left[ \frac{\mathbb{I}(D_i = 1, Z_i = 1)}{Pr(Z_i = 1)} - \frac{\mathbb{I}(D_i = 1, Z_i = 0)}{Pr(Z_i = 0)} \right] \left( \mathbb{I}(g_h(Y_i, 1) > b_n) - \mathbb{I}(g(Y_i, 1) > b_n + cn^{-2/5+\epsilon}) \right)
\]
\[
= o_p(1/\sqrt{n})
\]
\[
\frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{I}(D_i = 1, Z_i = 1) \left( \mathbb{I}(g_h(Y_i, 1) > b_n) - \mathbb{I}(g(Y_i, 1) > b_n + cn^{-2/5+\epsilon}) \right)
\]
\[
= o_p(1)
\]
\[
(C.36)
\]
Proof. For any \( \epsilon > 0 \),
\[
Pr \left( \frac{\sqrt{n}}{n} \sum_{i=1}^{n} Y_i \left[ \frac{\mathbb{I}(D_i = 1, Z_i = 1)}{Pr(Z_i = 1)} - \frac{\mathbb{I}(D_i = 1, Z_i = 0)}{Pr(Z_i = 0)} \right] \times \left( \mathbb{I}(g_h(Y_i, 1) > b_n) - \mathbb{I}(g(Y_i, 1) > b_n + cn^{-2/5+\epsilon}) \right) > \epsilon \right)
\]
\[
\leq Pr(\sup_y |g_h(y, 1) - g(y, 1)| \geq cn^{-2/5+\epsilon})
\]
\[
+ Pr \left( \frac{\sqrt{n}}{n} \sum_{i=1}^{n} Y_i \left[ \frac{\mathbb{I}(D_i = 1, Z_i = 1)}{Pr(Z_i = 1)} - \frac{\mathbb{I}(D_i = 1, Z_i = 0)}{Pr(Z_i = 0)} \right] \mathbb{I}(0 \leq g(Y_i, 1) < b_n + cn^{-2/5+\epsilon}) > \epsilon \right)
\]
\[
(C.37)
\]
where the inequality hold because on the event $\sup_y |g_h(y, 1) - g(y, 1)| < cn^{-2/5+\epsilon}$,

$$|\mathbb{1}(g_h(Y_i, 1) > b_n) - \mathbb{1}(g(Y_i, 1) > b_n + cn^{-2/5+\epsilon})| \leq |\mathbb{1}(-b_n - cn^{-2/5+\epsilon} \leq g(Y_i, 1) < b_n + cn^{-2/5+\epsilon})|
$$

Note that

$$\sqrt{n} \sum_{i=1}^{n} Y_i \left[ \frac{\mathbb{1}(D_i = 1, Z_i = 1)}{Pr(Z_i = 1)} - \frac{\mathbb{1}(D_i = 1, Z_i = 0)}{Pr(Z_i = 0)} \right] (\mathbb{1}(|g(Y_i, 1)| < b_n + cn^{-2/5+\epsilon}))$$

$$\leq E \left[ \left( \sqrt{n} \sum_{i=1}^{n} Y_i \left[ \frac{\mathbb{1}(D_i = 1, Z_i = 1)}{Pr(Z_i = 1)} - \frac{\mathbb{1}(D_i = 1, Z_i = 0)}{Pr(Z_i = 0)} \right] \mathbb{1}(|g(Y_i, 1)| < b_n + cn^{-2/5+\epsilon}) \right)^2 \right]$$

$$\leq E \left[ Y_i^2 \left[ \frac{\mathbb{1}(D_i = 1, Z_i = 1)}{Pr(Z_i = 1)} - \frac{\mathbb{1}(D_i = 1, Z_i = 0)}{Pr(Z_i = 0)} \right]^2 \mathbb{1}(|g(Y_i, 1)| < b_n + cn^{-2/5+\epsilon}) \right]$$

$$+ (n - 1)E[Y_i \left[ \frac{\mathbb{1}(D_i = 1, Z_i = 1)}{Pr(Z_i = 1)} - \frac{\mathbb{1}(D_i = 1, Z_i = 0)}{Pr(Z_i = 0)} \right] \mathbb{1}(|g(Y_i, 1)| < b_n + cn^{-2/5+\epsilon})]$$

Term $A = o(1)$ by dominated convergence theorem since $\mathbb{1}(0 \leq g(Y_i, 1) < b_n + cn^{-2/5+\epsilon}) \to 0$.

Term $B$, by assumption 5.6

$$E[Y_i \left[ \frac{\mathbb{1}(D_i = 1, Z_i = 1)}{Pr(Z_i = 1)} - \frac{\mathbb{1}(D_i = 1, Z_i = 0)}{Pr(Z_i = 0)} \right] \mathbb{1}(0 \leq g(Y_i, 1) < b_n + cn^{-2/5+\epsilon})]$$

$$= O((b_n + cn^{-2/5+\epsilon})^2) = O\left(\frac{1}{\sqrt{n} \log^2 n}\right)$$

therefore $B = (n - 1)O\left(\frac{1}{n \log^4 n}\right) = o(1)$. Therefore,

$$Pr \left( \left| \frac{\sqrt{N}}{N} \sum_{i=1}^{N} Y_i \left[ \frac{\mathbb{1}(D_i = 1, Z_i = 1)}{Pr(Z_i = 1)} - \frac{\mathbb{1}(D_i = 1, Z_i = 0)}{Pr(Z_i = 0)} \right] \mathbb{1}(0 \leq g(Y_i, 1) < b_n + cn^{-2/5+\epsilon}) \right| > \epsilon \right) \to 0$$

by mean squared error convergence. The first result follows since $Pr(\sup_y |g_h(y, 1) - g(y, 1)| \geq cn^{-2/5+\epsilon}) \to 0$ by Lemma C.5.

The second result holds similarly, since

$$Pr \left( \left| \frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{1}(D_i = 1, Z_i = 1)(\mathbb{1}(g_h(Y_i, 1) > b_n) - \mathbb{1}(g(Y_i, 1) > b_n + cn^{-2/5+\epsilon})) \right| > \epsilon \right)$$

$$\leq Pr(\sup_y |g_h(y, 1) - g(y, 1)| \geq cn^{-2/5+\epsilon})$$

$$+ Pr \left( \left| \frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{1}(D_i = 1, Z_i = 1)(\mathbb{1}(|g(Y_i, 1)| < b_n + cn^{-2/5+\epsilon})) \right| > \epsilon \right)$$

67
Note that

\[
\text{Var}\left(\frac{1}{N} \sum_{i=1}^{N} Y_i \mathbb{1}(D_i = 1, Z_i = 1)\left(1(0 \leq g(Y_i, 1) < b_n + cn^{-2/5+\epsilon})\right)\right) 
\]

\[
\leq E\left(\left|\frac{1}{N} \sum_{i=1}^{N} Y_i \mathbb{1}(D_i = 1, Z_i = 1)\left(1(0 \leq g(Y_i, 1) < b_n + cn^{-2/5+\epsilon})\right)\right|^2\right) 
\]

\[
\leq \frac{1}{N} E\left(\left|Y_i^2 \mathbb{1}(D_i = 1, Z_i = 1)(1(0 \leq g(Y_i, 1) < b_n + cn^{-2/5+\epsilon}))\right|^2\right) + \frac{N-1}{N} \left(E\left|Y_i \mathbb{1}(D_i = 1, Z_i = 1)(1(0 \leq g(Y_i, 1) < b_n + cn^{-2/5+\epsilon}))\right|^2\right) = o(1) 
\]

where the last equality holds by dominated convergence theorem and assumption 5.6. Then by mean squared error convergence the second statement also holds.

\[\square\]

**Lemma C.9. (Asymptotic Linear Expansion of numerator and denominator)** Let \( b_n = n^{-1/5}, b_n = n^{-1/4}/\log n\) and \( c_n = cn^{-2/5+\epsilon}\) as in Lemma C.8, and assumption 5.3 - 5.6 holds, then

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) \right] \mathbb{1}(g(Y_i, 1) \geq b_n) - \int_{Y_i} y(p(y, 1) - q(y, 1)) dy 
\]

\[
= o_p(1) + E[Y_i \mathbb{1}(D_i = 1, Z_i = 1) \mathbb{1}(g(Y_i, 1) \geq 0)] \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{Pr(Z_j = 1)} \right] 
\]

\[
- E[Y_i \mathbb{1}(D_i = 1, Z_i = 0) \mathbb{1}(g(Y_i, 0) > 0)] \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) - \frac{1}{Pr(Z_j = 0)} \right] 
\]

\[
+ \frac{1}{n} \sum_{i=1}^{n} Y_i \left[ \frac{1}{Pr(Z_i = 1)} - \frac{1}{Pr(Z_i = 0)} \right] \mathbb{1}(g(Y_i, 1) > b_n + c_n) - \int_{Y_i} y(p(y, 1) - q(y, 1)) \mathbb{1}(g(Y_i, 1) > b_n + c_n) dy 
\]

(C.38)

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) - \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) \right] \mathbb{1}(g(Y_i, 0) \geq b_n) - \int_{Y_i} y(q(y, 0) - p(y, 0)) dy 
\]

\[
= o_p(1) + E[Y_i \mathbb{1}(D_i = 0, Z_i = 0) \mathbb{1}(g(Y_i, 0) \geq 0)] \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) - \frac{1}{Pr(Z_j = 0)} \right] 
\]

\[
- E[Y_i \mathbb{1}(D_i = 0, Z_i = 1) \mathbb{1}(g(Y_i, 1) > 0)] \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{Pr(Z_j = 1)} \right] 
\]

\[
+ \frac{1}{n} \sum_{i=1}^{n} Y_i \left[ \frac{1}{Pr(Z_i = 0)} - \frac{1}{Pr(Z_i = 1)} \right] \mathbb{1}(g(Y_i, 0) > b_n + c_n) - \int_{Y_i} y(q(y, 0) - p(y, 0)) \mathbb{1}(g(Y_i, 0) > b_n + c_n) dy 
\]

(C.39)
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) \right] \mathbb{1}(g(Y_i, 1) \geq b_n) - \int_{Y_i} (p(y, 1) - q(y, 1)) dy \\
= o_p(1) + E[\mathbb{1}(D_i = 1, Z_i = 1) \mathbb{1}(g(Y_i, 1) \geq 0)] \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{Pr(Z_j = 1)} \right] \\
- E[\mathbb{1}(D_i = 1, Z_i = 0) \mathbb{1}(g(Y_i, 0) > 0)] \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) - \frac{1}{Pr(Z_j = 0)} \right] \\
+ \frac{1}{n} \sum_{i=1}^{n} \left[ \mathbb{1}(D_i = 1, Z_i = 1) - \mathbb{1}(D_i = 1, Z_i = 0) \right] \mathbb{1}(g(Y_i, 1) > b_n + c_n) - \int_{Y_i} (p(y, 1) - q(y, 1)) \mathbb{1}(g(Y_i, 1) > b_n + c_n) dy \\
\text{(C.40)}
\]

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) - \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) \right] \mathbb{1}(g(Y_i, 0) \geq b_n) - \int_{Y_i} (q(y, 0) - p(y, 0)) dy \\
= o_p(1) + E[\mathbb{1}(D_i = 0, Z_i = 0) \mathbb{1}(g(Y_i, 0) \geq 0)] \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) - \frac{1}{Pr(Z_j = 0)} \right] \\
- E[\mathbb{1}(D_i = 0, Z_i = 1) \mathbb{1}(g(Y_i, 1) > 0)] \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{Pr(Z_j = 1)} \right] \\
+ \frac{1}{n} \sum_{i=1}^{n} \left[ \mathbb{1}(D_i = 0, Z_i = 0) - \mathbb{1}(D_i = 0, Z_i = 1) \right] \mathbb{1}(g(Y_i, 0) > b_n + c_n) - \int_{Y_i} (q(y, 0) - p(y, 0)) \mathbb{1}(g(Y_i, 0) > b_n + c_n) dy \\
\text{(C.41)}
\]

**Proof.** I prove the first equality, and the rest holds similarly.
We look at the expansion

\[
\frac{1}{n} \sum_{i=1}^{n} Y_i \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) \right] \mathbb{1}(g_h(Y_i, 1) \geq b_n) - \int_{\mathcal{Y}_1} y(p(y, 1) - q(y, 1)) dy
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} Y_i \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) \right] \mathbb{1}(g_h(Y_i, 1) \geq b_n)
\]

\[
- \frac{1}{n} \sum_{i=1}^{n} Y_i \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) \right] \mathbb{1}(g_h(Y_i, 1) \geq b_n)
\]

\[
+ \frac{1}{n} \sum_{i=1}^{n} Y_i \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) \right] \mathbb{1}(g(Y_i, 1) > b_n + c_n) - \int_{\mathcal{Y}_1} y(p(y, 1) - q(y, 1)) \mathbb{1}(g(Y_i, 1) > b_n + c_n) dy
\]

\[
+ \int_{\mathcal{Y}_1} y(p(y, 1) - q(y, 1)) (\mathbb{1}(g(y, 1) > b_n + c_n) - \mathbb{1}(g(y, 1) > 0)) dy
\]

(\text{C.42})

For term $A_1$, we can write it as

\[
A_1 = \frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{1}(D_i = 1, Z_i = 1) \mathbb{1}(g_h(Y_i, 1) \geq b_n) \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) \right] \mathbb{1}(g(Y_i, 1) \geq b_n + c_n) + o_p(1)
\]

\[
= \left[ \frac{1}{n} \sum_{i=1}^{n} Y_i \mathbb{1}(D_i = 1, Z_i = 1) \mathbb{1}(g(Y_i, 1) \geq b_n + c_n) + o_p(1) \right] \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) \right]
\]

\[
= \left[ \mathbb{E}[Y_i \mathbb{1}(D_i = 1, Z_i = 1) \mathbb{1}(g(Y_i, 1) \geq 0)] + o_p(1) \right] \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 1) - \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) \right]
\]

where the second equality holds by Lemma \textbf{C.8}, the third holds by dominated convergence theorem, and the last holds by $\frac{1}{n} \sum_{j=1}^{n} \frac{1}{1} \frac{1}{P(1/\sqrt{n})} = O_p(1/\sqrt{n})$ by delta method. Similarly,

\[
A_2 = \mathbb{E}[Y_i \mathbb{1}(D_i = 1, Z_i = 0) \mathbb{1}(g(Y_i, 0) \geq 0)] \left[ \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) - \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}(Z_j = 0) \right] + o_p(1/\sqrt{n})
\]

By Lemma \textbf{C.8}, $B = o_p(1/\sqrt{n})$, by Lemma \textbf{C.8} $D = O(b_n^2) = O(\frac{1}{\sqrt{n} \log^4 n}) = o(\frac{1}{\sqrt{n}})$
C.5 Proof of Theorem 1

Proof. Note that by Lemma C.7 assumption 5.6 can be replace by 5.7 Now, let

\[
\hat{\beta} = \left( \begin{array}{c}
\frac{1}{n} \sum_{i=1}^{N} Y_i \left[ \frac{1(D_i=1,Z_i=1)}{Pr(Z_i=1)} - \frac{1(D_i=0,Z_i=0)}{Pr(Z_i=0)} \right] \mathbb{1}(g(Y_i, 1) \geq b_n) \\
\frac{1}{n} \sum_{i=1}^{N} Y_i \left[ \frac{1(D_i=1,Z_i=0)}{Pr(Z_i=1)} - \frac{1(D_i=0,Z_i=0)}{Pr(Z_i=0)} \right] \mathbb{1}(g(Y_i, 0) \geq b_n)
\end{array} \right) \mathbb{1}(\hat{\theta})
\]

and let

\[
\beta = \left( \begin{array}{c}
\int_{y \in Y} yg(y, 1) \mathbb{1}(g(y, 1) \geq b_n) \int_{y \in Y} yg(y, 0) \mathbb{1}(g(y, 0) \geq b_n)
\end{array} \right)
\]

By delta method and Lemma C.6, \(\sqrt{n}(\hat{\theta} - \theta) \rightarrow_d N(0, D^\prime \Sigma D)\). Now let

\[
\hat{\beta} = \left( \begin{array}{c}
\frac{1}{n} \sum_{i=1}^{N} Y_i \left[ \frac{1(D_i=1,Z_i=1)}{n \sum_{j=1}^{N} 1(Z_j=1)} - \frac{1(D_i=0,Z_i=0)}{n \sum_{j=1}^{N} 1(Z_j=0)} \right] \mathbb{1}(g(Y_i, 1) \geq b_n) \\
\frac{1}{n} \sum_{i=1}^{N} Y_i \left[ \frac{1(D_i=1,Z_i=0)}{n \sum_{j=1}^{N} 1(Z_j=1)} - \frac{1(D_i=0,Z_i=0)}{n \sum_{j=1}^{N} 1(Z_j=0)} \right] \mathbb{1}(g(Y_i, 0) \geq b_n)
\end{array} \right) \mathbb{1}(\hat{\theta})
\]

By Lemma C.9

\[
\sqrt{n}(\hat{\beta} - \beta) = o_p(1) + \Gamma \sqrt{n}(\hat{\theta} - \theta)
\]

And we notice that \(LAT_E = \frac{\beta_1}{\beta_3} - \frac{\beta_2}{\beta_4}\) and \(LAT_E^{ID} = \frac{\beta_1}{\beta_3} - \frac{\beta_2}{\beta_4}\), and \(\Pi\) is the Jacobian matrix of function \(f(\beta) = \frac{\beta_1}{\beta_3} - \frac{\beta_2}{\beta_4}\). The result follows by delta method.

\[\square\]

D Auxiliary Lemmas

Lemma D.1. (Gine and Guillou, 2002) Let \(G\) be a measurable uniformly bounded VC class of functions, such that

\[
N(G, L_2(P), \tau ||G||_{L_2(P)}) \leq \left( \frac{A^n}{\tau} \right)
\]

and let \(\sigma\) and \(U\) be the number such that \(\sigma^2 \geq \sup_{g \in G} \text{Var}_P g\) and \(U \geq \sup_{g \in G} ||g||_{\infty}\), and \(0 < \sigma < U/2\), \(\sqrt{n} \sigma \geq U \sqrt{\frac{U}{\log \sigma}}\). Then there exist constant \(L, C\) that depends on \(A\) and \(\nu\)
only such that

\[
Pr\left(\sup_{g \in G} \left| \sum_{i=1}^{n} g(x_i) - Eg(x_i) \right| > C\sigma \sqrt{n \log \frac{U}{\sigma}} \right)
\leq L \exp\left\{ - \frac{C\log(1 + C/(4L))}{L} \log \frac{U}{\sigma} \right\}
\] (D.1)

**Lemma D.2. (Montgomery-Smith’s Maximal Inequality)**

\[
Pr\left(\max_{\frac{k-1}{2^k} \leq n \leq 2^k} \left| \sum_{i=1}^{k} g(x_i) - Eg(x_i) \right| > t \right) \leq 9 Pr\left(\sup_{g \in G} \left| \sum_{i=1}^{n} g(x_i) - Eg(x_i) \right| > t/30 \right)
\] (D.2)

**Lemma D.3.** Let \( h_n = n^{-\gamma} \) for some \( \gamma \in (0, 1) \), such that \( \frac{nh_n}{\log h_n} \to \infty \). Denote the estimator and its expectation of the estimator as

\[
f_{l,m}^n(y) \equiv \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{Y_i - y}{h_n}\right) \mathbb{1}(D_i = l, Z_i = m)
\]

\[
\bar{f}_{l,m}^n(y) \equiv \frac{1}{h_n} \mathbb{E}\left[ K\left(\frac{Y_i - y}{h_n}\right) \mathbb{1}(D_i = l, Z_i = m) \right]
\] (D.3)

Then, the following uniform bounds holds for some constant \( \bar{C} \):

\[
\limsup_{n \to \infty} \left( \sup_y \sqrt{\frac{nh_n}{\log h_n^{-1}}} |f_{l,m}^n(y) - \bar{f}_{l,m}^n(y)| \right) \leq \bar{C} \quad \text{a.s.} \quad (D.4)
\]

**Proof.** We prove the a.s. convergence result for \( l = m = 1 \) and omit the superscript \( l,m \) in \( f_n \) and \( \bar{f}_n \), and the rest inequalities hold similarly. Use Montgomery-Smith’s Maximal inequality, we have

\[
Pr\left(\max_{\frac{2^{k-1}h_{2^{k-1}}} \leq h \leq \frac{2^k}{h}} \left| \sum_{i=1}^{2^k} K\left(\frac{Y_i - y}{h}\right) \mathbb{1}(D_i = 1, Z_i = 1) \right| > t\sqrt{\frac{nh_n \log h_n^{-1}}{2^{k-1}h_{2^k} \log h_{2^k}^{-1}/30}} \right)
\leq 9 Pr\left(\sup_{y,h_{2^{k-1}}} \left| \sum_{i=1}^{2^k} K\left(\frac{Y_i - y}{h}\right) \mathbb{1}(D_i = 1, Z_i = 1) \right| > t\sqrt{\frac{nh_n \log h_n^{-1}}{2^{k-1}h_{2^k} \log h_{2^k}^{-1}/30}} \right)
\]

"
By Gine and Guillou (2002), the class of function $\mathcal{K}_k = \{K(\frac{t}{n}) \mid t \in R, h_{2k} \leq h \leq h_{2k-1}\}$ is a VC class, and we multiply it by a fixed function $\mathbb{1}(d = 1, z = 1)$, the class of function $\tilde{\mathcal{K}}_k = \{K(\frac{t}{n})\mathbb{1}(d = 1, z = 1) \mid t \in R, h_{2k} \leq h \leq h_{2k-1}\}$ is still a VC class. So we take $U_k = ||K(y)||_\infty$, $\sigma_k^2 = h_{2k-1} \sup_y |f(y|D_i = 1, Z_i = 1)| \int t K^2(t) dt$, then we have

$$
\sup_{y, h_{2k} \leq h \leq h_{2k-1}} \text{Var} \left[ K\left(\frac{Y_i - y}{h}\right) \mathbb{1}(D_i = 1, Z_i = 1) \right] \\
\leq \sup_{y, h_{2k} \leq h \leq h_{2k-1}} \mathbb{E} \left[ K^2\left(\frac{Y_i - y}{h}\right) \mathbb{1}(D_i = 1, Z_i = 1) \right] \\
\leq \sup_{y, h_{2k} \leq h \leq h_{2k-1}} h \int K^2(t) f(y - th|D_i = 1, Z_i = 1) dt \times Pr(D_i = 1, Z_i = 1) \\
\leq h ||f||_\infty \int t K^2(t) dt \leq \sigma_k^2
$$

and $\sup_{y, h_{2k} \leq h \leq h_{2k-1}} |K(\frac{Y_i - y}{h}) \mathbb{1}(D_i = 1, Z_i = 1)| \leq U_k$.

Then since $h_{2k-1} \rightarrow 0$ and $\frac{2^k h_{2k}}{\log h_{2k}} \rightarrow \infty$ as $k \rightarrow \infty$, we can find $k_0$ such that $\sigma_k^2 \leq U_k/2$ and $\sqrt{2^k} \sigma_k \geq U_k \frac{t}{\log \sigma_k}$ for all $k \geq k_0$. So, we can apply (D.1). Take $t = 30C \sqrt{2 \times 2^{-\gamma} \sup_y |f(y|D_i = 1, Z_i = 1)| \int t K^2(t) dt}$, then

$$
t \sqrt{2^{k-1} h_{2k} \log h_{2k}^{-1}} / 30 = C \sqrt{2^k (ch_{2k}) \log h_{2k}^{-1} \sup_y |f(y|D_i = 1, Z_i = 1)| \int t K^2(t) dt} \\
\geq C \sqrt{2^k h_{2k-1} \log h_{2k}^{-1} \sup_y |f(y|D_i = 1, Z_i = 1)| \int t K^2(t) dt} \\
= C \sqrt{2^k \sigma_k \log h_{2k}^{-1}}
$$

Since

$$
\frac{h_{2k}^{-1}}{U_k / \sigma_k} \geq h_{2k}^{-0.5} \sqrt{\sup_y |f(y|D_i = 1, Z_i = 1)| \int t K^2(t) dt ||K(y)||_\infty} \rightarrow \infty
$$

where we use the construction of $\sigma_k$ and $h_{2k}^{-1} \geq h_{2k-1}^{-1}$, and $h_n \rightarrow 0$. So we can find $k_1$ such
that $C \sqrt{2^k \sigma_k \log h_2^1} > C \sqrt{2^k \sigma_k \sqrt{\frac{U_k}{\sigma_k}}}$ holds for all $k > k_1$. Then for $k > \max\{k_0, k_1\}$,

$$9 \Pr \left( \sup_{y, h_{2^k-1} \leq h \leq h_{2^k}} \left| \sum_{i=1}^{2^k} K \left( \frac{Y_i - y}{h} \right) 1(D_i = 1, Z_i = 1) \right| > t \sqrt{2^{k-1} h_{2^k} \log h_{2^k}^{-1} / 30} \right)$$

$$\leq 9 \Pr \left( \sup_{y, h_{2^k-1} \leq h \leq h_{2^k}} \left| \sum_{i=1}^{2^k} K \left( \frac{Y_i - y}{h} \right) 1(D_i = 1, Z_i = 1) \right| > C \sqrt{2^k \sigma_k \sqrt{\frac{U_k}{\sigma_k}}} \right)$$

$$\leq L \exp \left\{ - \frac{C \log(1 + C/(4L))}{L} \log \frac{U_k}{\sigma_k} \right\}$$

$$\leq L \exp \left\{ - \frac{C \log(1 + C/(4L))}{L} \log \frac{\|K(y)\|_{\infty}}{\sup_y |f(y|D_i = 1, Z_i = 1)| \int_t K^2(t) dt} \right\} h_{2^{k-1}}$$

$$\leq \text{Constant} \times \left( \frac{1}{2^\gamma} \right)^{k-1}$$

Note that $\sum_{k=\max\{k_0, k_1\}+1}^{\infty} \left( \frac{1}{2^\gamma} \right)^{k-1} < \infty$ holds, so by Borel-Cantelli lemma,

$$\Pr \left( \lim_{n \to \infty} \sup_{y} \sqrt{\frac{n h_n}{\log h_n^{-1}}} \sup_{y} |f_n(y) - \tilde{f}_n(y)| > 30C \sqrt{2 \times 2^{-\gamma} \sup_{y} |f(y|D_i = 1, Z_i = 1)| \int_t K^2(t) dt} \right) = 0$$

(D.7)