Kitagawa, Toru

Working Paper

The identification region of the potential outcome distributions under instrument independence

cemmap working paper, No. CWP23/20

Provided in Cooperation with:
Institute for Fiscal Studies (IFS), London

Suggested Citation: Kitagawa, Toru (2020) : The identification region of the potential outcome distributions under instrument independence, cemmap working paper, No. CWP23/20, Centre for Microdata Methods and Practice (cemmap), London, https://doi.org/10.1920/wp.cem.2020.2320

This Version is available at:
https://hdl.handle.net/10419/241898

cemmap
centre for microdata methods and practice

The Identification Region of the Potential Outcome Distributions under Instrument Independence

Toru Kitagawa

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP 23/20

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

The Identification Region of the Potential Outcome Distributions under Instrument Independence

Toru Kitagawa*†

20, May, 2020

Abstract

This paper examines the identifying power of instrument exogeneity in the treatment effect model. We derive the identification region of the potential outcome distributions, which are the collection of distributions that are compatible with data and with the restrictions of the model. We consider identification when (i) the instrument is independent of each of the potential outcomes (*marginal independence*), (ii) the instrument is independent of each of the potential outcomes jointly (*joint independence*), and (iii.) the instrument is independent of each of the potential outcomes jointly and is monotonic (the *LATE restriction*). By comparing the size of the identification region under each restriction, we show that joint independence provides more identifying information for the potential outcome distributions than marginal independence, but that the LATE restriction provides no identification gain beyond joint independence. We also, under each restriction, derive sharp bounds for the Average Treatment Effect and sharp testable implication to falsify the restriction. Our analysis covers discrete and continuous outcome cases, and extends the Average Treatment Effect bounds of Balke and Pearl (1997) developed for the dichotomous outcome case to a more general setting.

Keywords: Partial Identification, Treatment Effects, Instrumental Variables

JEL Classification: C14, C21.

*CeMMAP and Department of Economics, University College London. Email: t.kitagawa@ucl.ac.uk

†An earlier version of this paper appears in a chapter of my Ph.D. dissertation Kitagawa (2009b). I thank Guido Imbens, Frank Kleibergen, Charles Manski, and the seminar and conference participants at Harvard Econometrics Lunch, the 2009 SETA/CeMMAP conference in Kyoto, and the 2010 Cowles summer conference for valuable comments. I thank Jeff Rowley for excellent research assistance. Financial support from the Brown University Merit Dissertation Fellowship, the ESRC through the ESRC Centre for Microdata Methods and Practice (CeMMAP) (grant number RES-589-28-0001), and the European Research Council (Starting grant No. 715940) is gratefully acknowledged.

1 Introduction

This paper studies identification of the potential outcome distributions using an instrumental variable in settings where data exhibits imperfect compliance and selection is an issue. A motivating example is a randomized control trial with two treatment arms and where trial-subjects are observed following a different treatment arm to the one that they are allocated by the experimental design (Angrist et al., 1996; Imbens and Angrist, 1994). Of interest throughout this paper is identification of the effect of treatment on the potential outcomes, and the models that we study feature an instrumental variable that facilitates identification. The potential outcomes can be dichotomous, discrete or continuous, and the effect of treatment can be heterogeneous in the population. The models that we study differ in the statistical independence conditions and monotonicity restrictions that they embed, and so in the relationships between the potential outcomes and instrumental variable that they are compatible with.

We make three contributions to the existing literature on partial identification of treatment effects by models featuring selection on unobservables. Firstly, for each model that we study, we derive the identification region of the potential outcome distributions, which we emphasize are counterfactual distributions that describe the outcomes that would be realized if each treatment arm were applied uniformly to the population. Since the statistical independence conditions and monotonicity restrictions that we consider are made successively stronger, we show that these identification regions are nested and provide closed-form expressions for them. Secondly, we similarly derive sharp bounds for the Average Treatment Effect and provide closed-form expressions for these bounds. The expressions that we provide include a novel and non-trivial extension of existing results, with our extension of Balke and Pearl (1997) and its Average Treatment Effect bounds a leading example of this. Thirdly, we show that each model is falsifiable by deriving a sharp testable implication of each set of assumptions that we consider, and that equate to a set of conditions such that the identification region of the potential outcome distributions is empty.

The statistical independence conditions and monotonicity restrictions that we consider in this paper are as follows. Firstly, that the instrumental variable is statistically independent of each potential outcome (marginal statistical independence). Secondly, that the instrumental variable is statistically independent of the potential outcomes and selection heterogeneity jointly (joint statistical independence). Thirdly, that the instrumental variable is statistically independent of the potential outcomes and selection heterogeneity jointly, and that each unit in the population exhibits a weakly

monotonic selection response to the instrumental variable (instrument monotonicity). We note that the third assumption is the so-called LATE restriction (Imbens and Angrist, 1994).

The remainder of the paper is organized as follows. In the remainder of this section, we present a brief review of the existing literature. In Section 2, we introduce the notation that we use and provide a formal definition of the identification region. In Section 3, we derive the identification region of the potential outcome distributions for each model. In Section 4, we compare the size of the obtained identification regions, and present sharp bounds for the Average Treatment Effect. In Section 5, we conclude. Proofs and further discussion are included as appendices.

1.1 Related literature

We identify several papers in the econometrics literature that are related to this paper, which we collect into three broad categories according to their content and their relation to this paper.

Firstly, we recognise the contribution of papers that consider identification of treatment effects under a similar set of assumptions to those that we consider here, and that propose bounds on these treatment effects. Chief amongst these is Manski (1990), which reports sharp bounds on mean outcomes and the Average Treatment Effect when mean independence is assumed to hold. Balke and Pearl (1997) similarly considers identification of treatment effects in settings where data exhibits imperfect compliance and selection is an issue, but restricts attention to the case where the potential outcomes are dichotomous. When outcomes are dichotomous, the mean independence condition that is present in the analysis of Manski (1990) coincides with marginal statistical independence. Balke and Pearl (1997) strengthens marginal statistical independence to full statistical independence, and shows that the Manski bounds are not sharp in this case. Balke and Pearl (1997) provides closed-form expressions for the sharp bounds on mean outcomes and the Average Treatment Effect by solving a linear programme, which is of finite dimension when there are a finite number of treatment arms and both the instrumental variable and outcomes are discrete with finite supports. We extend Balke and Pearl (1997) by allowing for non-scalar and continuous outcomes, providing closed-form expressions for the identified sets of the potential outcome distributions and the Average Treatment Effects. These closed-form expressions complement the general characterizations that are reported in Beresteanu et al. (2012). Gunsilius (2019) also extends Balke and Pearl (1997) but goes further than this paper in allowing for an infinite number of treatment arms and a continuous instrumental variable (in addition to continuous outcomes). To facilitate this extension, Gunsilius (2019) notes

that it is necessary to regularize heterogeneity concerning individual responses to an infinite number of treatment arms. No such regularization is required if the number of treatment arms is finite and, like Balke and Pearl (1997), we allow for unrestricted heterogeneity and rich behavior. Additionally imposing instrument monotonicity, Heckman and Vytlacil (2001) (and Heckman and Vytlacil, 2005, 1999) considers identification of the Average Treatment Effect when outcomes are continuous, and shows that the obtained bounds coincide with the Manski bounds (Manski, 1990, 1994, 2003) under mean independence. If data is not compatible with instrument monotonicity though, then the bounds that are derived in Heckman and Vytlacil (2001) can be wider than the sharp bounds that are derived under marginal or joint statistical independence, which we provide in this paper. Heckman and Vytlacil (2005, 2007) and Mogstad et al. (2018) extend the analysis of Heckman and Vytlacil (2001) to consider identification of the Marginal Treatment Effect and other policy relevant parameters, while Huber et al. (2017) and Huber and Mellace (2015a) focus on partial identification of treatment effects for sub-populations including that of the compliers (Imbens and Angrist, 1994). Chen and Flores (2015) and Cheng and Small (2006) also consider identification of the Average Treatment Effect under instrument monotonicity, but allow for sample selection and three (rather than two) treatment arms, respectively. Bhattacharya et al. (2008), Mourifié (2015), Shaikh and Vytlacil (2011) and Vytlacil and Yildiz (2007) each study a special case where instrumental variables are statistically independent of the potential outcomes, which are dichotomous and monotonic in treatment. Chiburis (2010) also studies the special case of dichotomous outcomes, considering identification of treatment effects under a variety of semiparametric restrictions. Lafférs (2019) adopts a linear programming approach to identification of treatment effects that is similar to the approach taken in Balke and Pearl (1997), adding constraints and restrictions that are not present in that analysis.

Secondly, we recognise the contribution of papers that consider the failure and testing of identifying assumptions. Pearl (1995b) derives a testable implication for instrument independence, which is the so-called (Pearl) Instrument Inequality. Pearl (1995b) shows that this implication is a necessary condition for emptiness of the identification region or, equivalently, for falsification of instrument independence. We show in this paper that this testable implication is, in fact, both necessary and sufficient for emptiness of the identification region when the instrumental variable is dichotomous (the case that Pearl, 1995b and Balke and Pearl, 1997 consider). As such, there does not exist a stronger testable implication than the Instrument Inequality unless further restrictions are maintained. For instance, Kédagni and Mourifie (2020) shows that the Pearl Inequality can be strengthened if the

instrumental variable takes more than two values. Gunsilius (2018) shows that the testability of instrument independence is reliant on there being a finite number of treatment arms, and this assumption is untestable when there are instead an infinite number. Provided that there are a finite number of treatment arms, Heckman and Vytlacil (2005) and Balke and Pearl (1997) provide testable implications for instrument independence and instrument monotonicity jointly. Kitagawa (2015) and Mourifié and Wan (2017) build upon these implications to propose formal tests of these restrictions, which lead to identification of the complier outcome distribution (Imbens and Rubin, 1997) and of the Local Average Treatment Effect (Imbens and Angrist, 1994). Huber and Mellace (2015b) proposes a complementary testing procedure for the weaker statistical independence condition of mean independence. Complementary to this work on falsifiability of a model is de Chaisemartin (2017), Flores and Flores-Lagunes (2013) and Kédagni (2017) that consider identification in instances where various common restrictions of a model are inappropriate. For instance, de Chaisemartin (2017) considers identification in the presence of instrument non-monotonicity, Flores and Flores-Lagunes (2013) considers identification in the absence of exclusion, and Kédagni (2017) considers identification in the absence of instrument independence. Machado et al. (2019) proposes testable implications when outcomes are dichotomous and maintained assumptions can reveal the sign of the Average Treatment Effect.

Thirdly, we recognise the contribution of papers that study the identification of treatment effects by incomplete structural models that do not feature a selection equation. Particular examples include Beresteanu et al. (2012), Chernozhukov and Hansen (2005) and Chesher and Rosen (2017). Chesher and Rosen (2017) is notable since it provides a sharp characterization of the identification region of treatment effects for a broad class of models using tools from random set theory. Chesher and Rosen (2013) illustrates what such incomplete models can deliver in practice by means of a simple application (we refer to Clarke and Windmeijer, 2012 for further evidence of what partially identifying models can deliver in practice in comparison to conventional models). We also recognise the contribution of papers studying complete structural models that impose additional restrictions on their constituent structural equations, including Chesher (2003, 2005, 2010), Chesher et al. (2013), Imbens and Newey (2009) and Vuong and Xu (2017) to list but a few. These additional restrictions constrain the association of the potential outcomes and are a source of additional identifying power in the model.

2 Analytical Framework

2.1 Data Generating Process and the Population

Consider identification of the causal effect of a binary treatment on some outcome of interest. We use $D \in \{1, 0\}$ as an indicator for treatment, where $D = 1$ indicates a treated individual and where $D = 0$ indicates an untreated individual. Following the Neyman-Rubin potential outcome framework, let Y_1 denote the outcome that would be observed if the individual receives treatment and let Y_0 denote the outcome that would be observed if the individual does not receive treatment. The observed outcome in data is then $Y \equiv DY_1 + (1 - D)Y_0$, which need not be scalar. To this end, we let the support of Y_1 and Y_0 be a subset of $\mathcal{Y}$, which we can take to be an arbitrary space equipped with the Borel σ-algebra $B(\mathcal{Y})$ and a probability measure μ. We focus on a situation where treatment status is not randomized and selection is an issue of concern (i.e., treatment status can depend upon the underlying potential outcomes). We suppose that a non-degenerate binary variable $Z \in \{1, 0\}$ is available in data, and that Z qualifies as an instrumental variable (Angrist et al., 1996; Imbens and Angrist, 1994). In particular, we suppose that Z satisfies an exclusion restriction prohibiting it as a (direct) cause of Y, and our notation reflects this. For example, initial assignment to treatment is often used as an instrumental variable in experimental settings with non-compliance.

We denote a conditional distribution of (Y, D) given Z by

$$\begin{aligned}
P_{Y_1}(B) &\equiv \Pr(Y \in B, D = 1|Z = 1) = \Pr(Y_1 \in B, D = 1|Z = 1), \\
P_{Y_0}(B) &\equiv \Pr(Y \in B, D = 0|Z = 1) = \Pr(Y_0 \in B, D = 0|Z = 1), \\
Q_{Y_1}(B) &\equiv \Pr(Y \in B, D = 1|Z = 0) = \Pr(Y_1 \in B, D = 1|Z = 0), \\
Q_{Y_0}(B) &\equiv \Pr(Y \in B, D = 0|Z = 0) = \Pr(Y_0 \in B, D = 0|Z = 0).
\end{aligned} \tag{1}$$

where B is an arbitrary subset of $\mathcal{Y}$. Except for the marginal distribution of Z, $P = (P_{Y_1}(\cdot), P_{Y_0}(\cdot))$ and $Q = (Q_{Y_1}(\cdot), Q_{Y_0}(\cdot))$ uniquely characterize the distribution of data. We represent the *data generating process* by $(P, Q) \in \mathcal{P}$, where $\mathcal{P}$ is the class of data generating processes. Throughout our analysis, we do not restrict the class of data generating processes $\mathcal{P}$ other than to assume the existence of probability density functions with respect to the dominating measure μ, which the researcher has knowledge of. We denote the probability sub-density functions of $P_{Y_j}(\cdot)$ and $Q_{Y_j}(\cdot)$ with respect to μ by $p_{Y_j}(\cdot)$ and $q_{Y_j}(\cdot)$, $j = 1, 0$. That is, for every subset B, we have

$$P_{Y_1}(B) = \int_B p_{Y_1} d\mu, \qquad P_{Y_0}(B) = \int_B p_{Y_0} d\mu,$$

$$Q_{Y_1}(B) = \int_B q_{Y_1} d\mu, \qquad Q_{Y_0}(B) = \int_B q_{Y_0} d\mu.$$

It is important to keep in mind that the integration of the sub-density functions $p_{Y_j}(\cdot)$ and $q_{Y_j}(\cdot)$ over $\mathcal{Y}$ yield the conditional probabilities $\Pr(D = j|Z = 1)$ and $\Pr(D = j|Z = 0)$, which can be less than one. Sub-distribution functions (the integral of a sub-density function over subsets) are common in competing risks analysis, where they are often alternatively referred to as cumulative incidence functions.

Our identification framework features a selection equation with unobserved selection heterogeneity V,

$$D = I\{u(Z, V) \geq 0\}.$$

Here, $u(Z, V)$ is latent utility and rationalizes the individual's choice of treatment status, and V is unobserved heterogeneity that affects the individual's choice and is possibly dependent on the potential outcomes. We interpret this equation as structural in the sense that, with V fixed, $u(z, V)$ yields a counterfactual treatment status for each $z = 1, 0$. Provided that D and Z are binary, there are at most four distinct selection behaviors, which we refer to as *types*. The role of unobserved heterogeneity V is to randomly categorize individuals into one of these four types. A random category variable T is used to indicate type (Angrist et al., 1996), with

$$T = \begin{cases} c: & \text{complier} & \text{if } u(1, V) = 1,\ u(0, V) = 0, \\ n: & \text{never-taker} & \text{if } u(1, V) = u(0, V) = 0, \\ a: & \text{always-taker} & \text{if } u(1, V) = u(0, V) = 1, \\ d: & \text{defier} & \text{if } u(1, V) = 0,\ u(0, V) = 1. \end{cases}$$

If we do not impose any restriction on the distribution of T, then we are also free of any assumption on the functional form of latent utility and on the dimensionality of unobserved heterogeneity V (Pearl, 1995a).

Every individual in the population of interest possesses a non-random value of (Y_1, Y_0, T, Z) and the parameter of interest is defined on the distribution of (Y_1, Y_0, T, Z). We define *the population* as a joint probability distribution of $(Y_1, Y_0, T, Z) \in \mathcal{Y} \times \mathcal{Y} \times \{c, n, a, d\} \times \{1, 0\}$. Hereafter, f denotes the probability density or sub-density function of population variables, distinguished by subscripts such as f_{Y_1}, $f_{Y_1,T|Z}$, etc. We use $\mathcal{F}$ to denote the class of populations. In the following analysis, equalities or inequalities for density or sub-density functions are interpreted as almost everywhere with respect to the measure μ.

2.2 Defining the Identification Region

Model restrictions take the form of statistical relationships for the population random variables (Y_1, Y_0, T, Z). Let M be the model restriction(s) and let $\mathcal{F}_M \subset \mathcal{F}$ be the sub-class of populations satisfying the imposed restriction M.

For each data generating process $(P, Q) \in \mathcal{P}$, the class of *observationally equivalent* populations $\mathcal{F}^o(P, Q) \subset \mathcal{F}$ is defined as the collection of distributions of (Y_1, Y_0, T, Z) that generate (P, Q). Given a particular data generating process (P, Q), the identification region under restriction M, which we denote by $IR(P, Q|M)$, is defined as *the set of populations that are compatible with* (P, Q) *and restriction* M. That is, $IR(P, Q|M)$ is formulated as the intersection of $\mathcal{F}_M$ and $\mathcal{F}^o(P, Q)$,

$$IR(P, Q|M) \equiv \mathcal{F}_M \cap \mathcal{F}^o(P, Q). \qquad (P, Q) \in \mathcal{P}$$

When $IR(P, Q|M)$ is empty, restriction M is not compatible with observed data and is refutable (Manski, 2003).[1]

If interest instead lies in $\theta : \mathcal{F} \to \Theta$, a feature or parameter of the population, then the identification region of θ under restriction M, which we denote by $IR_\theta(P, Q|M)$, is defined as the range of $\theta(\cdot)$ for the domain $IR(P, Q|M)$. When $IR(P, Q|M)$ is empty, we also define $IR_\theta(P, Q|M)$ as empty so as to reflect the refutability property of the identification region. As such, the identification region of θ under restriction M is defined as

$$IR_\theta(P, Q|M) \equiv \begin{cases} \{\theta(F) : F \in IR(P, Q|M\} \cap \Theta & \text{if } IR(P, Q|M) \neq \emptyset, \\ \emptyset & \text{if } IR(P, Q|M) = \emptyset. \end{cases} \qquad (2)$$

In words, $IR_\theta(P, Q|M)$ is defined as *the set of* θ *such that we can construct a population* F *that is compatible with* (P, Q) *and the imposed restriction* M.

Here, our construction of the identification region does *not* assume that the true population satisfies the imposed restriction M, which matters when M *is observationally restrictive* (Koopmans and Reiersøl, 1950). If we assume that the true population satisfies restriction M and M is observationally restrictive, we *a priori* exclude the possibility of $IR(P, Q|M)$ being empty, even if data provides evidence to refute M. If we then derive sharp bounds on θ under the assumption that the true population satisfies restriction M, the bound formula and its sharpness break-down if $IR(P, Q|M)$ is empty. Moreover, the bound formula does not correspond to an empty set, despite the fact that

[1] Since this rule for refuting restriction M is based on emptiness of $IR(P, Q|M)$, no other testable implication is more powerful in detecting violations of M.

$IR(P, Q|M)$ is empty. This break-down gives rise to a misspecification of the sharp bounds for θ. As we discuss further in Section 4, the bounds on the Average Treatment Effect under instrument independence provide an example of this type of misspecification problem. In order to avoid such a misspecification problem, we do not vary the class of data generating processes $\mathcal{P}$ and construct the bounds for each restriction that we impose by explicitly applying definition (2).

2.3 Instrumental Variable Restrictions

We consider the following three model restrictions in turn.

Restriction MSI:

Marginal Statistical Independence Restriction: Z is marginally independent of each of Y_1 and Y_0.

Restriction RA:

Random Assignment Restriction: Z is jointly independent of (Y_1, Y_0, T).

Restriction LATE:

LATE Restriction: Z is jointly independent of (Y_1, Y_0, T), and $f_T(T = d) = 0$ or $f_T(T = c) = 0$.

The notion of instrument exogeneity is represented in all three model restrictions by statistical independence of the potential outcomes and the instrument. The restrictions are nested and are listed in terms of their strength, from weak to strong. The first restriction, MSI, imposes marginal independence between the instrument and each of the potential outcomes. Since selection heterogeneity T is unaffected by the model restriction, the analysis corresponding to this case is robust to dependence between the instrument and selection heterogeneity.[2] The second restriction, RA, embodies a stronger version of instrument exogeneity such that the instrument is jointly independent of both outcome heterogeneity and selection heterogeneity. RA is justified if the researcher believes that the instrument is generated through some randomization mechanism as in the (quasi-)experimental setting. The final restriction, $LATE$, is due to Imbens and Angrist (1994) and Angrist et al. (1996), and is crucial to identifying the potential outcome distributions for the sub-population of compliers.

We assert that, although MSI is theoretically interesting, it is of limited practical use.[3] It is difficult to think of instances where MSI can be justified but RA cannot. Nonetheless, we study MSI here for its simplicity, as a stepping-stone to analysis under RA and LATE.

[2] Other identification analyses, such as Chernozhukov and Hansen (2005) and Chesher (2010), are also unbound by a selection equation. This analysis differs in that it makes no assumption about the association between Y_1 and Y_0.

[3] We thank an anonymous referee for making this point to us.

Our primary interest lies in identifying f_{Y_1} and f_{Y_0}, which describe the marginal distributions of Y_1 and Y_0. The marginal distributions are of interest if the goal of analysis is to assess the effect of intervention by comparing various features of the marginal distributions of the potential outcomes. For example, the *Average Treatment Effect* is defined as the difference between the mean of f_{Y_1} and of f_{Y_0}. As a further example, we may be interested in the τ-th *quantile differences*, defined as the difference between the τ-th quantiles of the two potential outcome distributions. As a final example, we may be interested in the effect of intervention on the inequality of outcomes, and so in the variances of f_{Y_1} and f_{Y_0} or in some other measure of inequality of outcome such as the Gini index. In all three examples, the parameters of interest are defined in terms of the marginal distributions of Y_1 and Y_0. We focus on constructing the sharp identification region of f_{Y_1} and f_{Y_0}, which we denote by $IR_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$, instead of the identification region for the (full) population distribution. We note that if interest lies instead in a parameter that is defined on the distribution of the individual causal effects $Y_1 - Y_0$, $IR_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$ is less useful, as the distribution of $Y_1 - Y_0$ is sensitive not only to the marginals of Y_1 and Y_0 but also to dependence between Y_1 and Y_0. Identification of the distribution of $Y_1 - Y_0$ is beyond the scope of this paper.[4]

3 Construction of the Identification Region

For the construction of $IR_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$, our first step is to formulate the conditions for $F \in \mathcal{F}^o(P,Q)$ (i.e., for compatibility of a distribution F of (Y_1,Y_0,T,Z) with observed data (P,Q)). These conditions are obtained by rewriting the right-hand side of the identities (1) in terms of the distribution of (Y_1,Y_0,T,Z).

$$\begin{aligned}
p_{Y_1}(y_1) &= f_{Y_1,T|Z}(y_1,T=c|Z=1) + f_{Y_1,T|Z}(y_1,T=a|Z=1),\\
q_{Y_1}(y_1) &= f_{Y_1,T|Z}(y_1,T=d|Z=0) + f_{Y_1,T|Z}(y_1,T=a|Z=0),\\
p_{Y_0}(y_0) &= f_{Y_0,T|Z}(y_0,T=d|Z=1) + f_{Y_0,T|Z}(y_0,T=n|Z=1),\\
q_{Y_0}(y_0) &= f_{Y_0,T|Z}(y_0,T=c|Z=0) + f_{Y_0,T|Z}(y_0,T=n|Z=0).
\end{aligned} \tag{3}$$

The law of total probability implies that $f_{Y_1|Z}(y_1|Z=z) = \sum_{t\in\{c,n,a,d\}} f_{Y_1,T|Z}(y_1,T=t|Z=z)$ and $f_{Y_0|Z}(y_0|Z=z) = \sum_{t\in\{c,n,a,d\}} f_{Y_0,T|Z}(y_0,T=t|Z=z)$, and therefore $f_{Y_j|Z}$ less the observed

[4] In the situation where the marginal distributions of Y_1 and Y_0 are point-identified, Fan and Park (2010, see also Fan et al., 2017), Firpo and Ridder (2019) and Heckman et al. (1997) analyze identification of the distribution of individual causal effects $Y_1 - Y_0$.

sub-densities p_{Y_j} or q_{Y_j} has the mixture form

$$\begin{aligned}
f_{Y_1|Z}(y_1|Z=1) - p_{Y_1}(y_1) &= f_{Y_1,T|Z}(y_1, T=d|Z=1) + f_{Y_1,T|Z}(y_1, T=n|Z=1), \\
f_{Y_1|Z}(y_1|Z=0) - q_{Y_1}(y_1) &= f_{Y_1,T|Z}(y_1, T=c|Z=0) + f_{Y_1,T|Z}(y_1, T=n|Z=0), \\
f_{Y_0|Z}(y_0|Z=1) - p_{Y_0}(y_0) &= f_{Y_0,T|Z}(y_0, T=c|Z=1) + f_{Y_0,T|Z}(y_0, T=a|Z=1), \\
f_{Y_0|Z}(y_0|Z=0) - q_{Y_0}(y_0) &= f_{Y_0,T|Z}(y_0, T=d|Z=0) + f_{Y_0,T|Z}(y_0, T=a|Z=0).
\end{aligned} \tag{4}$$

We use these identities to relate the distribution $f_{Y_j|Z}$ to the distribution $f_{Y_j,T|Z}$.

3.1 Identification Region under Marginal Independence (MSI)

If we impose MSI, $f_{Y_1|Z} = f_{Y_1}$ and $f_{Y_0|Z} = f_{Y_0}$ must hold. Therefore, we substitute f_{Y_1} and f_{Y_0} (the unconditional sub-densities) for $f_{Y_1|Z}$ and $f_{Y_0|Z}$ (the conditional sub-densities) in the left-hand side of (4). We have

$$\begin{aligned}
f_{Y_1}(y_1) - p_{Y_1}(y_1) &= f_{Y_1,T|Z}(y_1, T=d|Z=1) + f_{Y_1,T|Z}(y_1, T=n|Z=1), \\
f_{Y_1}(y_1) - q_{Y_1}(y_1) &= f_{Y_1,T|Z}(y_1, T=c|Z=0) + f_{Y_1,T|Z}(y_1, T=n|Z=0), \\
f_{Y_0}(y_0) - p_{Y_0}(y_0) &= f_{Y_0,T|Z}(y_0, T=c|Z=1) + f_{Y_0,T|Z}(y_0, T=a|Z=1), \\
f_{Y_0}(y_0) - q_{Y_0}(y_0) &= f_{Y_0,T|Z}(y_0, T=d|Z=0) + f_{Y_0,T|Z}(y_0, T=a|Z=0).
\end{aligned} \tag{5}$$

Given $(P, Q) \in \mathcal{P}$, any population contained in $IR(P, Q|MSI)$ satisfy (3) and (5). That is, by noting that the right-hand side of every equation of (5) is non-negative, we find necessary conditions for (f_{Y_1}, f_{Y_0}) to be contained in $IR_{(f_{Y_1}, f_{Y_0})}(P, Q|MSI)$,

$$f_{Y_1}(y_1) \geq \max\{p_{Y_1}(y_1), q_{Y_1}(y_1)\} \quad \text{and} \quad f_{Y_0}(y_0) \geq \max\{p_{Y_0}(y_0), q_{Y_0}(y_0)\}.$$

We hereafter refer to $\max\{p_{Y_j}, q_{Y_j}\}$ as the *density envelope* for Y_j and to $\delta_{Y_j} \equiv \int_{\mathcal{Y}} \max\{p_{Y_j}, q_{Y_j}\} d\mu$ as the *integrated envelope* for Y_j, for each $j = 1, 0$. The next proposition shows that these conditions are sufficient to construct $IR_{(f_{Y_1}, f_{Y_0})}(P, Q|MSI)$ (i.e., that any f_{Y_1} and f_{Y_0} that are above the density envelopes constitute the identification region of (f_{Y_1}, f_{Y_0}) under MSI). This result can be viewed as a straightforward extension to the treatment effect model of Corollary 2.2.1 of Manski (2003) for the missing data model.

Proposition 3.1 (Identification region under marginal independence) *Denote the density envelopes by* $\underline{f_{Y_1}} \equiv \max\{p_{Y_1}, q_{Y_1}\}$ *and* $\underline{f_{Y_0}}(y_0) \equiv \max\{p_{Y_0}, q_{Y_0}\}$*, and the integrated envelopes by* $\delta_{Y_1} \equiv \int_{\mathcal{Y}} \underline{f_{Y_1}} d\mu$ *and* $\delta_{Y_0} \equiv \int_{\mathcal{Y}} \underline{f_{Y_0}} d\mu$*. Define the sets of probability denisty functions that cover* $\underline{f_{Y_1}}$ *and* $\underline{f_{Y_0}}$

respectively, by

$$\mathcal{F}^{env}_{f_{Y_1}}(P,Q) = \left\{ f_{Y_1} : \int_{\mathcal{Y}} f_{Y_1} d\mu = 1,\ f_{Y_1} \geq \underline{f_{Y_1}} \right\},$$
$$\mathcal{F}^{env}_{f_{Y_0}}(P,Q) = \left\{ f_{Y_0} : \int_{\mathcal{Y}} f_{Y_0} d\mu = 1,\ f_{Y_0} \geq \underline{f_{Y_0}} \right\}.$$

The identification region under MSI is non-empty if and only if $\delta_{Y_1} \leq 1$ *and* $\delta_{Y_0} \leq 1$, *and is given by*

$$IR_{(f_{Y_1}, f_{Y_0})}(P,Q|MSI) = \mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q).$$

Proof. See Appendix A. ■

The density envelope $\underline{f_{Y_1}}$ provides the maximal identifying information for the Y_1-distribution. Under MSI, each of the observed sub-densities p_{Y_1} and q_{Y_1} must be a part of the common underlying density of the treated outcome f_{Y_1}. An interpretation is that the density envelope then fills f_{Y_1} as much as is possible with the identified sub-densities p_{Y_1} and q_{Y_1} (and similarly for the untreated outcome). That $IR_{(f_{Y_1}, f_{Y_0})}(P,Q|MSI)$ is the Cartesian product of $\mathcal{F}^{env}_{f_{Y_1}}(P,Q)$ and $\mathcal{F}^{env}_{f_{Y_0}}(P,Q)$ implies that marginal independence does not provide a channel through which p_{Y_1} and q_{Y_1} contribute to identifying f_{Y_0} or through which p_{Y_0} and q_{Y_0} contribute to identifying f_{Y_1}. As such, we can, without loss of identifying information, separate identification analysis of f_{Y_1} from identification analysis of f_{Y_0}.

The refutability condition for marginal independence coincides with the testability result for the instrument exclusion restriction analyzed in Bonet (2001) and Pearl (1995b). Manski (2003) obtained an analogous refutability condition in the context of missing data.[5]

3.2 Identification Region under Random Assignment (RA)

If we strengthen MSI to RA, we replace the conditional distributions that appear on the right-hand side of (3) and (5) with their unconditional equivalents. With this in mind, we claim that[6] that a pair of marginal distributions (f_{Y_1}, f_{Y_0}) belongs to $IR_{(f_{Y_1}, f_{Y_0})}(P,Q|RA)$ if and only if it satisfies the *area constraints*

$$\int_{\mathcal{Y}} f_{Y_1,T}(y_1, T=t) d\mu = \int_{\mathcal{Y}} f_{Y_0,T}(y_0, T=t) d\mu, \quad t = c, n, a, d, \tag{6}$$

[5] Kitagawa (2015) considers estimation and inference for the integrated envelope parameter, so as to develop a specification test for instrument independence.

[6] See Lemma A.1 in Appendix for a formal justification of this claim.

and the *compatibility constraints*

$$\begin{aligned}
p_{Y_1}(y_1) &= f_{Y_1,T}(y_1,c) + f_{Y_1,T}(y_1,a),\\
q_{Y_1}(y_1) &= f_{Y_1,T}(y_1,d) + f_{Y_1,T}(y_1,a),\\
p_{Y_0}(y_0) &= f_{Y_0,T}(y_0,d) + f_{Y_0,T}(y_0,n),\\
q_{Y_0}(y_0) &= f_{Y_0,T}(y_0,c) + f_{Y_0,T}(y_0,n),\\
f_{Y_1}(y_1) - p_{Y_1}(y_1) &= f_{Y_1,T}(y_1,d) + f_{Y_1,T}(y_1,n),\\
f_{Y_1}(y_1) - q_{Y_1}(y_1) &= f_{Y_1,T}(y_1,c) + f_{Y_1,T}(y_1,n),\\
f_{Y_0}(y_0) - p_{Y_0}(y_0) &= f_{Y_0,T}(y_0,c) + f_{Y_0,T}(y_0,a),\\
f_{Y_0}(y_0) - q_{Y_0}(y_0) &= f_{Y_0,T}(y_0,d) + f_{Y_0,T}(y_0,a).
\end{aligned} \tag{7}$$

Subject to (6) and (7), we propose a compatible population as,[7] for $t = c, n, a, d$,

$$\begin{aligned}
f_{Y_1,Y_0,T}(y_1,y_0,T=t) &= f_{Y_1,Y_0,T|Z}(y_1,y_0,T=t|Z=1) = f_{Y_1,Y_0,T|Z}(y_1,y_0,T=t|Z=0)\\
&= \begin{cases} \left[\int_{\mathcal{Y}} f_{Y_1,T}(y_1,t)d\mu\right]^{-1} f_{Y_1,T}(y_1,t) f_{Y_0,T}(y_0,t) & \text{if } \int_{\mathcal{Y}} f_{Y_1,T}(y_1,t)d\mu > 0,\\ 0 & \text{if } \int_{\mathcal{Y}} f_{Y_1,T}(y_1,t)d\mu = 0. \end{cases}
\end{aligned}$$

By construction, the proposed population satisfies RA, and is compatible with the data generating process as it satisfies (3). Accordingly, $IR_{(f_{Y_1},f_{Y_0})}(P,Q|RA)$ is obtained by *characterizing the conditions for* (f_{Y_1}, f_{Y_0}) *such that we can find feasible* $(f_{Y_1,T}(y_1,t), f_{Y_0,T}(y_0,t))$, $t = c, n, a, d$.

The next proposition provides a closed-form expression of $IR_{(f_{Y_1},f_{Y_0})}(P,Q|RA)$.

Proposition 3.2 (Identification region under random assignment) *Let* λ_{Y_1} *be the inner integrated envelope of* p_{Y_1} *and* q_{Y_1}*, defined as* $\lambda_{Y_1} = \int_{\mathcal{Y}} \min\{p_{Y_1}, q_{Y_1}\}d\mu$.
(i) The identification region of (f_{Y_1}, f_{Y_0}) *under RA is*

$$IR_{(f_{Y_1},f_{Y_0})}(P,Q|RA) = \begin{cases} \mathcal{F}^*_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q) & \text{if } 1-\delta_{Y_0} < \lambda_{Y_1}\\ \mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q) & \text{if } 1-\delta_{Y_0} = \lambda_{Y_1}\\ \mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^*_{f_{Y_0}}(P,Q) & \text{if } 1-\delta_{Y_0} > \lambda_{Y_1} \end{cases}$$

where $\mathcal{F}^*_{f_{Y_1}}(P,Q)$ *and* $\mathcal{F}^*_{f_{Y_0}}(P,Q)$ *are proper subsets of* $\mathcal{F}^{env}_{f_{Y_1}}(P,Q)$ *and* $\mathcal{F}^{env}_{f_{Y_0}}(P,Q)$ *respectively, defined as*

$$\mathcal{F}^*_{f_{Y_1}}(P,Q) = \left\{ f_{Y_1} : f_{Y_1} \in \mathcal{F}^{env}_{f_{Y_1}}(P,Q),\ \int_{\mathcal{Y}} \min\left\{ f_{Y_1} - \underline{f_{Y_1}}, \min\{p_{Y_1}, q_{Y_1}\} \right\} d\mu \geq \lambda_{Y_1} + \delta_{Y_0} - 1 \right\},$$

[7] There are many ways to combine the density of (Y_1,T) and (Y_0,T) to obtain the joint density of (Y_1,Y_0,T). The one employed here is called the conditional independence coupling: the association of Y_1 and Y_0 satisfies $Y_1 \perp Y_0|T$.

$$\mathcal{F}^*_{f_{Y_0}}(P,Q) = \left\{ f_{Y_0} : f_{Y_0} \in \mathcal{F}^{env}_{f_{Y_0}}(P,Q),\ \int_{\mathcal{Y}} \min\left\{ f_{Y_0} - \underline{f_{Y_0}}, \min\{p_{Y_0}, q_{Y_0}\} \right\} d\mu \geq 1 - \delta_{Y_0} - \lambda_{Y_1} \right\}.$$

(ii) $IR_{(f_{Y_1}, f_{Y_0})}(P,Q|RA)$ *is non-empty if and only if* $\delta_{Y_1} \leq 1$ *and* $\delta_{Y_0} \leq 1$.

Proof. See Appendix A. ■

The proof of this proposition, which is provided in Appendix A, proceeds by the method of "guess and verify," and so the reader might think that the origins of the inequalities that appear in the definitions of $\mathcal{F}^*_{f_{Y_1}}(P,Q)$ and $\mathcal{F}^*_{f_{Y_0}}(P,Q)$ are rather obscure. In Appendix B, with the intent of providing intuition for this result, we present a geometric illustration of the additional identification gain of RA relative to MSI.

The above proposition makes clear that the identification region under RA can be strictly smaller than the identification region under MSI. In particular, this identification gain arises if the data reveals that $1 - \delta_{Y_0} \neq \lambda_{Y_1}$, as $\mathcal{F}^*_{f_{Y_1}}(P,Q)$ and $\mathcal{F}^*_{f_{Y_0}}(P,Q)$ are strictly smaller than $\mathcal{F}^{env}_{f_{Y_1}}(P,Q)$ and $\mathcal{F}^{env}_{f_{Y_0}}(P,Q)$ respectively, due to the inequality constraints appearing in their definitions. For the case of $1 - \delta_{Y_0} < \lambda_{Y_1}$, the fact that the inequality in the definition of $\mathcal{F}^*_{f_{Y_1}}(P,Q)$ involves δ_{Y_0} implies that p_{Y_0} and q_{Y_0} can contribute to identifying f_{Y_1} despite RA not explicitly constraining the association between Y_1 and Y_0. Symmetrically, for the case of $1 - \delta_{Y_0} > \lambda_{Y_1}$, p_{Y_1} and q_{Y_1} can contribute to identifying f_{Y_0} through the parameter λ_{Y_1}.

Figure 1 illustrates the intuition behind this identification gain. We draw a data generating process corresponding to the case of $1-\delta_{Y_0} < \lambda_{Y_1}$, which is equivalent to $1-\delta_{Y_1} > \lambda_{Y_0} \equiv \int_{\mathcal{Y}} \min\{p_{Y_0}, q_{Y_0}\} d\mu$.[8] We also draw marginal distributions of the potential outcomes (f_{Y_1}, f_{Y_0}) that belong to $\mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$. There, the subgraphs of f_{Y_1} and f_{Y_0} are partitioned into $(c(1), n(1), a(1), d(1))$ and $(c(0), n(0), a(0), d(0))$ respectively. If the identification region of f_{Y_1} were $\mathcal{F}^{env}_{f_{Y_1}}(P,Q)$, then the area of $n(1)$, which equals $1 - \delta_{Y_1}$, would coincide with the fraction of never-takers. If not, then $\mathcal{F}^{env}_{f_{Y_1}}(P,Q)$ cannot be spanned by the Y_1-distribution of never-takers $f_{Y_1,T}(\cdot, n)$, the shape of which is not constrained by data. However, this violates the third and fourth equations of (7) since the fraction of never-takers cannot be greater than the area of $n(0)$, which is smaller than $1 - \delta_{Y_1}$ for the drawn data generating process. Hence, we claim that the identification region for f_{Y_1} must be strictly smaller than $\mathcal{F}^{env}_{f_{Y_1}}(P,Q)$. To summarize, the source of the identification gain of RA relative to MSI is that RA allows us to learn the feasible type distributions from the observed sub-densities of Y_0 and that further constrain the feasible marginal distribution of Y_1.

[8] See Lemma A.2 in Appendix A.

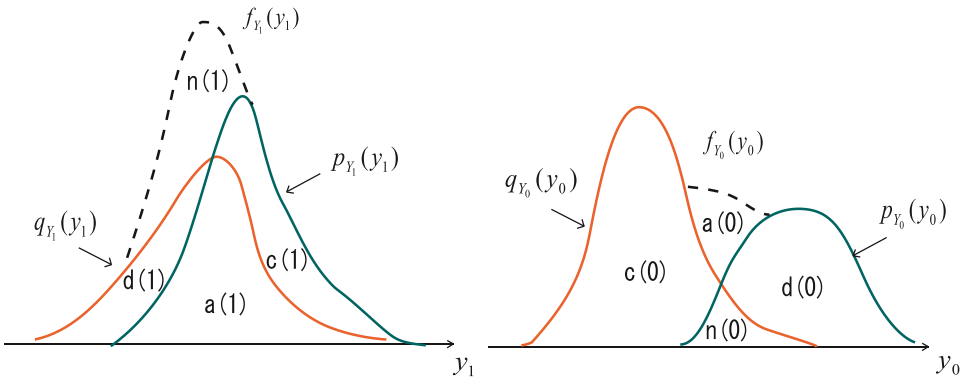

Figure 1: The drawn data generating process satisfies $1-\delta_{Y_0} < \lambda_{Y_1}$ (the area of $a(0)$ is strictly smaller than the area of $a(1)$).

3.3 Identification Region under the LATE restriction

Proposition 3.2 makes clear that if the observed data satisfies $1 - \delta_{Y_0} = \lambda_{Y_1}$, then the difference between MSI and RA does not matter for identification of f_{Y_1} and f_{Y_0}. This condition is satisfied when the data generating process reveals *nested sub-densities*

$$\begin{aligned} &p_{Y_1}(y_1) \geq q_{Y_1}(y_1) \text{ and } q_{Y_0}(y_0) \geq p_{Y_0}(y_0), \text{ or} \\ &p_{Y_1}(y_1) \leq q_{Y_1}(y_1) \text{ and } q_{Y_0}(y_0) \leq p_{Y_0}(y_0). \end{aligned} \tag{8}$$

Nested sub-densities come into play once we consider imposing the LATE restriction.

The LATE restriction further constrains the population by eliminating one of the selection types from the population. Specifically, in the case of $\Pr(D = 1|Z = 1) \geq \Pr(D = 1|Z = 0)$, the LATE restriction implies the no-defier condition $f_T(T = d) = 0$. Since analysis of the no-compliers case and the no-defiers case is symmetric, we consider the case of $\Pr(D = 1|Z = 1) \geq \Pr(D = 1|Z = 0)$ without loss of generality.

Under the LATE restriction (equivalent to RA plus the no-defier condition), (7) simplify to

$$\begin{aligned} p_{Y_1}(y_1) &= f_{Y_1,T}(y_1, T = c) + f_{Y_1,T}(y_1, T = a), \\ q_{Y_1}(y_1) &= f_{Y_1,T}(y_1, T = a), \\ p_{Y_0}(y_0) &= f_{Y_0,T}(y_0, T = n), \\ q_{Y_0}(y_0) &= f_{Y_0,T}(y_0, T = c) + f_{Y_0,T}(y_0, T = n), \\ f_{Y_1}(y_1) - p_{Y_1}(y_1) &= f_{Y_1,T}(y_1, T = n), \\ f_{Y_1}(y_1) - q_{Y_1}(y_1) &= f_{Y_1,T}(y_1, T = c) + f_{Y_1,T}(y_1, T = n), \end{aligned}$$

$$f_{Y_0}(y_0) - p_{Y_0}(y_0) = f_{Y_0,T}(y_0, T = c) + f_{Y_0,T}(y_0, T = a),$$
$$f_{Y_0}(y_0) - q_{Y_0}(y_0) = f_{Y_0,T}(y_0, T = a).$$

The first four of the above constraints imply that when the population satisfies the LATE restriction, the data generating process must reveal nested sub-densities since $p_{Y_1}(y_1) - q_{Y_1}(y_1) = f_{Y_1,T}(y_1, T = c) \geq 0$ and $q_{Y_0}(y_0) - p_{Y_0}(y_0) = f_{Y_0,T}(y_0, T = c) \geq 0$. This is equivalent to saying that observing non-nested sub-densities must yield an empty identification region under the LATE restriction. On the other hand, when data reveals nested sub-densities then, for every $(f_{Y_1}, f_{Y_0}) \in \mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$, we can uniquely solve the above constraints to obtain the (non-negative) probability density functions of (Y_1, T) and (Y_0, T), and these can be combined to obtain the probability density function of (Y_1, Y_0, T) independent of Z. Accordingly, it can be seen that any $(f_{Y_1}, f_{Y_0}) \in \mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$ belongs to $IR_{(f_{Y_1}, f_{Y_0})}(P, Q|LATE)$.

Proposition 3.3 (Identification region under the LATE restriction) *The identification region of* (f_{Y_1}, f_{Y_0}) *under the LATE restriction is*

$$IR_{(f_{Y_1}, f_{Y_0})}(P, Q|LATE) = \begin{cases} \mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q) & \text{for nested sub-densities (8),} \\ \emptyset & \text{otherwise.} \end{cases}$$

Proof. A proof is given in the preceding paragraphs of this section. ■

If the data generating process reveal nested sub-densities then the the identification region under the LATE restriction coincides with the identification region under MSI. Moreover, the fact that nested sub-densities satisfy $1 - \delta_{Y_0} = \lambda_{Y_1}$ implies that the identification region under the LATE restriction also coincides with the identification region under RA. If nested sub-densities are not observed, then the LATE restriction is refuted but the identification region under RA or MSI can be non-empty. In other words, as far as the distributions of the potential outcomes are concerned, adding instrument monotonicity[9] to the instrument independence restriction *only constrains the data generating process without helping us to learn more about* (f_{Y_1}, f_{Y_0}) *than under MSI or RA*. In this sense, we can safely drop instrument monotonicity from the LATE restriction and still acquire the maximal identifying information for the potential outcome distributions. Note that the refutability result of the LATE restriction is not new in the literature. Heckman and Vytlacil (2005) demonstrate

[9]Equivalently, imposing weak-separability of unobserved selection heterogeneity V in the selection equation (Vytlacil, 2002).

a testable implication for the LATE restriction, which is equivalent to the nested sub-density condition given here.

4 Bounding Causal Parameters

Since the analysis of the previous section does not rely on the choice of dominating measure μ, the constructed identification regions are applicable for discrete, continuous, unbounded or multi-dimensional outcomes. Moreover, for a parameter (vector) θ that maps (f_{Y_1}, f_{Y_0}) to Θ, we can make a comparison of the size of the sharp bounds of θ under the different model restrictions without explicitly computing them.

Theorem 1 *Let θ be a parameter (vector) that maps (f_{Y_1}, f_{Y_0}) to Θ. Then, for each layer of the data generating process (see Figure 2), the sharp bounds of θ under MSI, RA, and the LATE restriction have the following properties.*

(A) If $\delta_{Y_1} > 1$ or $\delta_{Y_0} > 1$, then

$$IR_\theta(P, Q|\cdot) = \emptyset \quad \text{for all of MSI, RA, and the LATE restriction.}$$

(B)-(i) If $\delta_{Y_1} \leq 1$ and $\delta_{Y_0} \leq 1$, and $1 - \delta_{Y_0} \neq \lambda_{Y_1}$, then,

$$IR_\theta(P, Q|MSI) \supset IR_\theta(P, Q|RA) \neq \emptyset, \qquad IR_\theta(P, Q|LATE) = \emptyset.$$

(B)-(ii) If $\delta_{Y_1} \leq 1$ and $\delta_{Y_0} \leq 1$, $1 - \delta_{Y_0} = \lambda_{Y_1}$, and the data generating process does not reveal nested sub-densities, then

$$IR_\theta(P, Q|MSI) = IR_\theta(P, Q|RA) \neq \emptyset, \qquad IR_\theta(P, Q|LATE) = \emptyset.$$

(B)-(iii) If the data generating process reveals nested sub-densities, then

$$IR_\theta(P, Q|MSI) = IR_\theta(P, Q|RA) = IR_\theta(P, Q|LATE) \neq \emptyset.$$

Proof. By the definition of $IR_\theta(P, Q|\cdot)$ given in (2), Proposition 3.1, 3.2, and 3.3 directly imply the results. ■

Provided that the outcome is scalar with compact support $\mathcal{Y} = [y_l, y_u]$, this theorem clearly applies to the sharp bounds of the Average Treatment Effect (ATE) $\theta = E(Y_1) - E(Y_0)$.

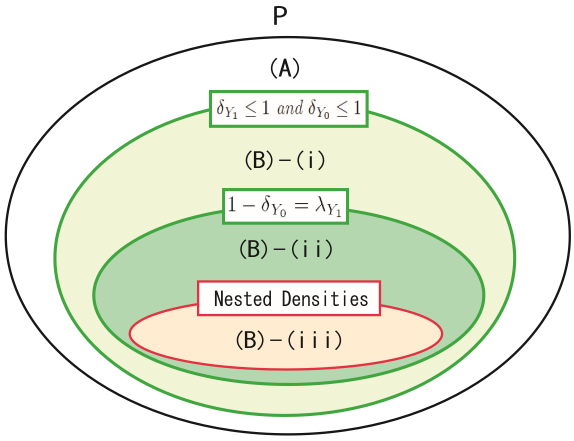

Figure 2: The classification of the data generating processes in Theorem 1.

In order to present a closed-form expression of the sharp ATE bounds, we define the α-th left- or right-trimming of a non-negative integrable function $g : \mathcal{Y} \to \mathbb{R}$. For $\alpha < \int_{\mathcal{Y}} g d\mu$, we let $q_\alpha^{left} \equiv \inf\left\{t : \int_{(-\infty,t]} g d\mu \geq \alpha\right\}$ and we define the *α-th left-trimming* of g as

$$[g]_\alpha^{ltrim}(y) \equiv g(y)1\left\{y > q_\alpha^{left}\right\} + \left(\int_{(-\infty, q_\alpha^{left}]} g(y)d\mu - \alpha\right) 1\left\{y = q_\alpha^{left}\right\}.$$

Similarly, we let $q_\alpha^{right} \equiv \sup\left\{t : \int_{[t,\infty)} g d\mu \geq \alpha\right\}$ and we define the *α-th right-trimming* of g as

$$[g]_\alpha^{rtrim}(y) \equiv g(y)1\left\{y < q_\alpha^{right}\right\} + \left(\int_{[q_\alpha^{right}, \infty)} g(y)d\mu - \alpha\right) 1\left\{y = q_\alpha^{right}\right\}.$$

The α-th (right-) left-trimming is obtained by trimming the (right-) left-tail part of the function g so that the trimmed mass is exactly equal to α. Note that if the underlying measure is atomic then the second terms on the right-hand sides of the above definitions can be non-zero, and these adjustment terms are needed to make the trimmed area exactly equal to α.

Proposition 4.1 (The sharp ATE bounds) *Assume that Y_1 and Y_0 have compact support $\mathcal{Y} = [y_l, y_u]$ and that their marginal distributions are absolutely continuous with respect to the measure μ that allows point mass at y_l and y_u. Further assume that the data generating process satisfies $\delta_{Y_1} \leq 1$ and $\delta_{Y_0} \leq 1$ so as to exclude Case (A) of Theorem 1.*

(i) The sharp ATE bounds under MSI are

$$IR_{ATE}(P, Q|MSI) = \Bigg[(1 - \delta_{Y_1})y_l + \int_{\mathcal{Y}} y_1 \underline{f_{Y_1}} d\mu - \int_{\mathcal{Y}} y_0 \underline{f_{Y_0}} d\mu - (1 - \delta_{Y_0})y_u,$$
$$\int_{\mathcal{Y}} y_1 \underline{f_{Y_1}} d\mu + (1 - \delta_{Y_1})y_u - (1 - \delta_{Y_0})y_l - \int_{\mathcal{Y}} y_0 \underline{f_{Y_0}} d\mu\Bigg]. \quad (9)$$

(ii) The sharp ATE bounds under RA are, for $1-\delta_{Y_0}=\lambda_{Y_1}$,

$$IR_{ATE}(P,Q|RA)=IR_{ATE}(P,Q|MSI),$$

for $1-\delta_{Y_0}<\lambda_{Y_1}$,

$$\begin{aligned}&IR_{ATE}(P,Q|RA)\\&=\Bigg[\int_{\mathcal{Y}} y_1\left(\underline{f_{Y_1}}+[\min\{p_{Y_1},q_{Y_1}\}]^{rtrim}_{1-\delta_{Y_0}}\right)d\mu+\lambda_{Y_0}y_l-\int_{\mathcal{Y}} y_0\underline{f_{Y_0}}d\mu-(1-\delta_{Y_0})y_u,\\&\quad\int_{\mathcal{Y}} y_1\left(\underline{f_{Y_1}}+[\min\{p_{Y_1},q_{Y_1}\}]^{ltrim}_{1-\delta_{Y_0}}\right)d\mu+\lambda_{Y_0}y_u-\int_{\mathcal{Y}} y_0\underline{f_{Y_0}}(y_0)d\mu-(1-\delta_{Y_0})y_l\Bigg],\end{aligned}\tag{10}$$

and, for $1-\delta_{Y_0}>\lambda_{Y_1}$,

$$\begin{aligned}&IR_{ATE}(P,Q|RA)\\&=\Bigg[\int_{\mathcal{Y}} y_1\underline{f_{Y_1}}d\mu+(1-\delta_{Y_1})y_l-\int_{\mathcal{Y}} y_0\left(\underline{f_{Y_0}}+[\min\{p_{Y_0},q_{Y_0}\}]^{ltrim}_{1-\delta_{Y_1}}\right)d\mu-\lambda_{Y_1}y_u,\\&\quad\int_{\mathcal{Y}} y_1\underline{f_{Y_1}}d\mu+(1-\delta_{Y_1})y_u-\int_{\mathcal{Y}} y_0\left(\underline{f_{Y_0}}+[\min\{p_{Y_0},q_{Y_0}\}]^{rtrim}_{1-\delta_{Y_1}}\right)d\mu-\lambda_{Y_1}y_l\Bigg].\end{aligned}\tag{11}$$

(iii) The sharp ATE bounds under the LATE restriction are

$$IR_{ATE}(P,Q|LATE)$$
$$=\begin{cases}\left[\begin{array}{l}\max_z\{E(Y|D=1,Z=z)\Pr(D=1|Z=z)+y_l\Pr(D=0|Z=z)\}\\-\min_z\{E(Y|D=0,Z=z)\Pr(D=0|Z=z)+y_u\Pr(D=1|Z=z)\},\\\quad\min_z\{E(Y|D=1,Z=z)\Pr(D=1|Z=z)+y_u\Pr(D=0|Z=z)\}\\-\max_z\{E(Y|D=0,Z=z)\Pr(D=0|Z=z)+y_l\Pr(D=1|Z=z)\}\end{array}\right] & \textit{for nested sub-densities,}\\ \emptyset & \textit{otherwise.}\end{cases}$$

Proof. See Appendix A. ■

The identification region for (f_{Y_1},f_{Y_0}) under MSI or RA collapses to a singleton if and only if $\delta_{Y_1}=1$ and $\delta_{Y_0}=1$, and determines whether ATE is non-parametrically (point-)identified or not. We emphasize that this condition is weaker than the well-known argument of identification at infinity (Chamberlain, 1986; Heckman, 1990). Whereas identification at infinity requires that the propensity score is zero or one at some instrument values, the above condition on the integrated envelopes can

be satisfied even when the propensity score is away from zero and one for all instrument values. However, when (P, Q) reveals nested sub-densities, the integrated envelopes equal the maximum propensity score (or one minus it) and so identification is attained only at infinity.

When the data generating process reveals $1 - \delta_{Y_0} \neq \lambda_{Y_1}$, the ATE bounds under RA are strictly narrower than the bounds under MSI. In the case of $1 - \delta_{Y_0} < \lambda_{Y_1}$, comparison of the lower bound of (10) and the lower bound of (9) shows that the former is greater by

$$[\lambda_{Y_1} - (1 - \delta_{Y_0})] \times \int_{y_1} (y_1 - y_l) \frac{[\min\{p_{Y_1}, q_{Y_1}\}]_{1-\delta_{Y_0}}^{rtrim}}{[\lambda_{Y_1} - (1 - \delta_{Y_0})]} d\mu.$$

By noting that $[\lambda_{Y_1} - (1 - \delta_{Y_0})]^{-1} [\min\{p_{Y_1}, q_{Y_1}\}]_{1-\delta_{Y_0}}^{rtrim}$ is a probability measure, we see that this identification gain for ATE becomes greater as $[\lambda_{Y_1} - (1 - \delta_{Y_0})]$ increases or as $[\min\{p_{Y_1}, q_{Y_1}\}]_{1-\delta_{Y_0}}^{rtrim}$ departs from degeneracy at y_l.

When (P, Q) reveals nested sub-densities, the sharp ATE bounds are given by (9), irrespective of the imposed restrictions, as claimed in Theorem 1. Moreover, with nested sub-densities, (9) reduces to the expression in (iii) of Proposition (4.1). This expression is identical to the ATE bounds of Manski (1994) under the mean independence restriction, $E(Y_1|Z) = E(Y_1)$ and $E(Y_0|Z) = E(Y_0)$. This observation supports the result of Heckman and Vytlacil (1999, 2001, 2007), which says that the sharp ATE bounds under the LATE restriction coincide with Manski's mean independence bounds. However, this statement is no longer valid if the data reveals non-nested sub-densities. Furthermore, a naïve implementation of the expression of the ATE bounds under the LATE restriction does not necessarily yield the emptyset even if $IR_{(f_{Y_1}, f_{Y_0})}(P, Q|LATE)$ is empty. As such, there is arguably some advantage to explicit statement of the model and its associated identification region, rather than working solely with the expression for the ATE due to this misspecification problem.

In the special case where the outcome variables are binary, the sharp ATE bounds under RA that are presented above coincide with the treatment effect bounds of Balke and Pearl (1997) (see Appendix C of Kitagawa, 2009a for a proof of this claim). Since the analysis of Balke and Pearl (1997) relies on a linear optimization procedure with a finite number of choice variables, such an approach cannot be directly applied to the case in which the outcome variable has continuous variation. Thus, the bound formula obtained here can be seen as a non-trivial generalization of the Balke and Pearl bounds to a more general case (see Gunsilius, 2019 for more recent advances).

As is discussed elsewhere, the potential outcomes have a structural equation analog (see Athey and Imbens, 2006; Chernozhukov and Hansen, 2005; Pearl, 2009), and requiring that this equation

is monotonic can lead to substantial identification gains. Like Balke and Pearl (1997), we do not rely on any type of assumption on the functional form of the structural equation analog, nor on the dimension or distribution of the unobserved heterogeneity that it features. In contrast, the analyses of Chesher (2003, 2005) and Chernozhukov and Hansen (2005) impose what is referred to as *outcome monotonicity in unobservable* or *rank invariance*, which necessarily restrict the structural equation and unobserved heterogeneity.[10] In the special case where the outcome variables are binary, Chesher (2010) obtains bounds on the Average Treatment Effect that are substantially narrower than the ones that are presented in this paper (Hahn, 2010). Moreover, in the continuous outcome case, Chernozhukov and Hansen (2005) shows that rank invariance and random assignment leads to (point-)identification of the potential outcome distributions. In each case, the imposed assumption limits individual behavior through association of the potential outcomes and requires justification that it is appropriate for the studied economic envrionment.

5 Concluding Remarks

With partial-identification in mind, this paper clarifies the identifying power of instrument independence assumptions in the heterogeneous treatment effect model. We derive the identification regions of the marginal distributions of the potential outcomes under each restriction that we consider, and compare their size. For some data generating processes, strengthening instrument independence from marginal statistical independence to joint statistical independence results in a tightening of the identification region. We clarify which data generating processes exhibit this property and which processes do not. We find that instrument monotonicity is redundant for identification of the potential outcome distributions when assumed in conjunction with instrument independence since monotonicity constrains the data generating process without further identifying the potential outcome distributions (see also Heckman and Vytlacil, 1999, 2001, 2007). We also present sharp bounds for the Average Treatment Effect under each restruction that we consider. Our analysis covers binary, discrete, and continuous support of an outcome of interest, and our bounds under joint independence amount to a generalization of the bounds of Balke and Pearl (1997) from the binary outcome case to the continuous outcome case.

[10] Chernozhukov and Hansen (2005) also consider a weaker condition, *rank similarity*, and establish similar results for this.

References

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association 91*(434), 444–455.

Athey, S. and G. W. Imbens (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica 74*(2), 431–497.

Balke, A. and J. Pearl (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association 92*(439), 1171–1176.

Beresteanu, A., I. Molchanov, and F. Molinari (2012). Partial identification using random set theory. *Journal of Econometrics 166*(1), 17–32.

Bhattacharya, J., A. M. Shaikh, and E. Vytlacil (2008). Treatment effect bounds under monotonicity assumptions: an application to Swan-Ganz catheterization. *American Economic Review 98*(2), 351–56.

Bonet, B. (2001). Instrumentality tests revisited. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pp. 48–55.

Chamberlain, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *journal of Econometrics 32*(2), 189–218.

Chen, X. and C. A. Flores (2015). Bounds on treatment effects in the presence of sample selection and noncompliance: the wage effects of Job Corps. *Journal of Business & Economic Statistics 33*(4), 523–540.

Cheng, J. and D. S. Small (2006). Bounds on causal effects in three-arm trials with non-compliance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(5), 815–836.

Chernozhukov, V. and C. Hansen (2005). An IV model of quantile treatment effects. *Econometrica 73*(1), 245–261.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica 71*(5), 1405–1441.

Chesher, A. (2005). Nonparametric identification under discrete variation. *Econometrica 73*(5), 1525–1550.

Chesher, A. (2010). Instrumental variable models for discrete outcomes. *Econometrica 78*(2), 575–601.

Chesher, A. and A. M. Rosen (2013). What do instrumental variable models deliver with discrete dependent variables? *American Economic Review 103*(3), 557–62.

Chesher, A. and A. M. Rosen (2017). Generalized instrumental variable models. *Econometrica 85*(3), 959–989.

Chesher, A., A. M. Rosen, and K. Smolinski (2013). An instrumental variable model of multiple discrete choice. *Quantitative Economics 4*(2), 157–196.

Chiburis, R. C. (2010). Semiparametric bounds on treatment effects. *Journal of Econometrics 159*(2), 267–275.

Clarke, P. S. and F. Windmeijer (2012). Instrumental variable estimators for binary outcomes. *Journal of the American Statistical Association 107*(500), 1638–1652.

de Chaisemartin, C. (2017). Tolerating defiance? Local average treatment effects without monotonicity. *Quantitative Economics 8*(2), 367–396.

Fan, Y., E. Guerre, and D. Zhu (2017). Partial identification of functionals of the joint distribution of "potential outcomes". *Journal of econometrics 197*(1), 42–59.

Fan, Y. and S. S. Park (2010). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory 26*(3), 931–951.

Firpo, S. and G. Ridder (2019). Partial identification of the treatment effect distribution and its functionals. *Journal of Econometrics 213*(1), 210–234.

Flores, C. A. and A. Flores-Lagunes (2013). Partial identification of local average treatment effects with an invalid instrument. *Journal of Business & Economic Statistics 31*(4), 534–545.

Gunsilius, F. (2018). Testability of the exclusion restriction in continuous instrumental variable models. *arXiv preprint arXiv:1806.09517*.

Gunsilius, F. (2019). Bounds in continuous instrumental variable models. *arXiv preprint arXiv:1910.09502*.

Hahn, J. (2010). Bounds on ATE with discrete outcomes. *Economics Letters 109*(1), 24–27.

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review 80*(2), 313–318.

Heckman, J. J., J. Smith, and N. Clements (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies 64*(4), 487–535.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the national Academy of Sciences 96*(8), 4730–4734.

Heckman, J. J. and E. J. Vytlacil (2001). Instrumental variables, selection models, and tight bounds on the average treatment effect. In *Econometric Evaluation of Labour Market Policies*, pp. 1–15. Springer.

Heckman, J. J. and E. J. Vytlacil (2007). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics 6*, 4875–5143.

Huber, M., L. Laffers, and G. Mellace (2017). Sharp IV bounds on average treatment effects on the treated and other populations under endogeneity and noncompliance. *Journal of Applied Econometrics 32*(1), 56–79.

Huber, M. and G. Mellace (2015a). Sharp bounds on causal effects under sample selection. *Oxford bulletin of economics and statistics 77*(1), 129–151.

Huber, M. and G. Mellace (2015b). Testing instrument validity for LATE identification based on inequality moment constraints. *Review of Economics and Statistics 97*(2), 398–411.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of Local Average Treatment Effects. *Econometrica: Journal of the Econometric Society*, 467–475.

Imbens, G. W. and W. K. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica 77*(5), 1481–1512.

Imbens, G. W. and D. B. Rubin (1997). Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies 64*(4), 555–574.

Kédagni, D. (2017). Identifying treatment effects in the presence of confounded types. *Available at SSRN 3064230*.

Kédagni, D. and I. Mourifie (2020). Generalized instrumental inequalities: Testing the IV independence assumption. *Biometrika*.

Kitagawa, T. (2009a). The identification region of the potential outcome distributions. *Cemmap Working Paper Series 30*(9).

Kitagawa, T. (2009b). *Three Essays on Instrumental Variables*. Ph. D. thesis, Brown University.

Kitagawa, T. (2015). A test for instrument validity. *Econometrica 83*(5), 2043–2063.

Koopmans, T. C. and O. Reiersøl (1950). The identification of structural characteristics. *The Annals of Mathematical Statistics 21*(2), 165–181.

Lafférs, L. (2019). Bounding average treatment effects using linear programming. *Empirical Economics 57*(3), 727–767.

Machado, C., A. M. Shaikh, and E. J. Vytlacil (2019). Instrumental variables and the sign of the average treatment effect. *Journal of Econometrics 212*(2), 522–555.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review 80*(2), 319–323.

Manski, C. F. (1994). The selection problem. In *Advances in Econometrics, Sixth World Congress*, Volume 1, pp. 143–70.

Manski, C. F. (2003). *Partial identification of probability distributions*. Springer Science & Business Media.

Mogstad, M., A. Santos, and A. Torgovitsky (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica 86*(5), 1589–1619.

Mourifié, I. (2015). Sharp bounds on treatment effects in a binary triangular system. *Journal of Econometrics 187*(1), 74–81.

Mourifié, I. and Y. Wan (2017). Testing local average treatment effect assumptions. *Review of Economics and Statistics 99*(2), 305–313.

Pearl, J. (1995a). From Bayesian networks to causal networks. In *Mathematical models for handling partial knowledge in artificial intelligence*, pp. 157–182. Springer.

Pearl, J. (1995b). On the testability of causal models with latent and instrumental variables. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pp. 435–443.

Pearl, J. (2009). *Causality*. Cambridge university press.

Shaikh, A. M. and E. J. Vytlacil (2011). Partial identification in triangular systems of equations with binary dependent variables. *Econometrica 79*(3), 949–955.

Vuong, Q. and H. Xu (2017). Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity. *Quantitative Economics 8*(2), 589–610.

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica 70*(1), 331–341.

Vytlacil, E. and N. Yildiz (2007). Dummy endogenous variables in weakly separable models. *Econometrica 75*(3), 757–779.

Appendix A: Proofs

Proofs for constructing $IR_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$ proceed in the manner of "guess and verify." We first propose $IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$ as a guess for $IR_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$. In order to verify that the guess $IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$ is correct, we need to show the two things. First, for an arbitrary $(f_{Y_1},f_{Y_0}) \in IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$, we shall show that there exists a distribution of (Y_1,Y_0,T,Z) that is compatible with (P,Q) and the imposed model restrictions. This first step proves $IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot) \subset IR_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$. Next, in order to prove $IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot) \supset IR_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$, it suffices to show that a necessary condition for $(f_{Y_1},f_{Y_0}) \in IR_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$ is satisfied for every $(f_{Y_1},f_{Y_0}) \in IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$ (e.g., the proof of Proposition 3.1). Alternatively, we may demonstrate that any

$(f_{Y_1}, f_{Y_0}) \notin IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$ delivers a contradiction of some of the imposed restrictions (e.g., the proof of Proposition 3.2.). In either fashion, we can conclude that $IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot) \supset IR_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$. By combining them, we conclude that the guess is correct, $IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot) = IR_{(f_{Y_1},f_{Y_0})}(P,Q|\cdot)$.

Throughout the proof, we do not explicitly state μ-a.e but any equalities or inequalities between the density functions should be interpreted in the sense of almost everywhere with respect to the measure μ.

Proof of Proposition 3.1. Fix $(P,Q) \in \mathcal{P}$, and guess the identification region under MSI to be $IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|MSI) = \mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$. Clearly, $IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|MSI)$ is non-empty if and only if $\delta_{Y_1} \leq 1$ and $\delta_{Y_0} \leq 1$, as otherwise no probability density functions can cover the entire density envelopes. Let us pick an arbitrary $(f_{Y_1}, f_{Y_0}) \in \mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$. Consider the distribution of (Y_1, Y_0, T) given Z as follows.

$$
\begin{aligned}
&f_{Y_1,Y_0,T|Z}(y_1,y_0,T=a|Z=1) = \frac{1}{\Pr(D=1|Z=1)} p_{Y_1}(y_1)[f_{Y_0}(y_0) - p_{Y_0}(y_0)],\\
&f_{Y_1,Y_0,T|Z}(y_1,y_0,T=a|Z=0) = \frac{1}{\Pr(D=1|Z=0)} q_{Y_1}(y_1)[f_{Y_0}(y_0) - q_{Y_0}(y_0)],\\
&f_{Y_1,Y_0,T|Z}(y_1,y_0,T=n|Z=1) = \frac{1}{\Pr(D=0|Z=1)} [f_{Y_1}(y_1) - p_{Y_1}(y_1)]p_{Y_0}(y_0),\\
&f_{Y_1,Y_0,T|Z}(y_1,y_0,T=n|Z=0) = \frac{1}{\Pr(D=0|Z=0)} [f_{Y_1}(y_1) - q_{Y_1}(y_1)]q_{Y_0}(y_0),\\
&f_{Y_1,Y_0,T|Z}(y_1,y_0,T=c|Z=z) = 0 \text{ for } z=1,0,\\
&f_{Y_1,Y_0,T|Z}(y_1,y_0,T=d|Z=z) = 0 \text{ for } z=1,0.
\end{aligned}
$$

By integrating out y_1 or y_0 from these densities, we can see that the constructed population meets the constraints (3). Furthermore, by plugging the constructed population densities into the identities, $f_{Y_1|Z} = \sum_{t\in\{c,n,a,d\}} \int_{y_0} f_{Y_1,Y_0,T|Z} d\mu$ and $f_{Y_0|Z} = \sum_{t\in\{c,n,a,d\}} \int_{y_1} f_{Y_1,Y_0,T|Z} d\mu$, we can confirm that $f_{Y_1|Z}$ and $f_{Y_0|Z}$ do not depend on Z, so the constructed population meets MSI. Therefore, $IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|MSI) \subset IR_{(f_{Y_1},f_{Y_0})}(P,Q|MSI)$. On the other hand, if $(f_{Y_1}, f_{Y_0}) \in IR_{(f_{Y_1},f_{Y_0})}(P,Q|MSI)$, $f_{Y_1} \geq \underline{f_{Y_1}}$ and $f_{Y_0} \geq \underline{f_{Y_0}}$ must hold because the right-hand side of (5) is always non-negative. Hence, $IR_{(f_{Y_1},f_{Y_0})}(P,Q|MSI) \subset IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|MSI)$, and this completes the proof. ■

The following lemmata are used for the proof of Propositions 3.2 and 4.1

Lemma A.1. *Let the data generating process $(P,Q) \in \mathcal{P}$ be given. Consider a pair of marginal probability density functions $(f^*_{Y_1}, f^*_{Y_0})$. $(f^*_{Y_1}, f^*_{Y_0}) \in IR_{(f_{Y_1},f_{Y_0})}(P,Q|RA)$ if and only if there exist*

non-negative functions $\{(h_{Y_1,t}, h_{Y_0,t}), t = c, n, a, d\}$ *that satisfy the following constraints,*

$$p_{Y_1}(y_1) = h_{Y_1,c}(y_1) + h_{Y_1,a}(y_1), \tag{12}$$

$$q_{Y_1}(y_1) = h_{Y_1,d}(y_1) + h_{Y_1,a}(y_1), \tag{13}$$

$$p_{Y_0}(y_0) = h_{Y_0,d}(y_0) + h_{Y_0,n}(y_0), \tag{14}$$

$$q_{Y_0}(y_0) = h_{Y_0,c}(y_0) + h_{Y_0,n}(y_0), \tag{15}$$

$$f^*_{Y_1}(y_1) - p_{Y_1}(y_1) = h_{Y_1,d}(y_1) + h_{Y_1,n}(y_1), \tag{16}$$

$$f^*_{Y_1}(y_1) - q_{Y_1}(y_1) = h_{Y_1,c}(y_1) + h_{Y_1,n}(y_1), \tag{17}$$

$$f^*_{Y_0}(y_0) - p_{Y_0}(y_0) = h_{Y_0,c}(y_0) + h_{Y_0,a}(y_0), \tag{18}$$

$$f^*_{Y_0}(y_0) - q_{Y_0}(y_0) = h_{Y_0,d}(y_0) + h_{Y_0,a}(y_0), \tag{19}$$

$$\int_{\mathcal{Y}} h_{Y_1,c}(y_1)d\mu = \int_{\mathcal{Y}} h_{Y_0,c}(y_0)d\mu, \tag{20}$$

$$\int_{\mathcal{Y}} h_{Y_1,n}(y_1)d\mu = \int_{\mathcal{Y}} h_{Y_0,n}(y_0)d\mu, \tag{21}$$

$$\int_{\mathcal{Y}} h_{Y_1,a}(y_1)d\mu = \int_{\mathcal{Y}} h_{Y_0,a}(y_0)d\mu, \tag{22}$$

$$\int_{\mathcal{Y}} h_{Y_1,d}(y_1)d\mu = \int_{\mathcal{Y}} h_{Y_0,d}(y_0)d\mu. \tag{23}$$

Proof of Lemma A.1. The "only if" part is implied by (7) in the main text, by substituting h for f. So, we focus on proving the "if" part of the lemma. Given the non-negative functions $\{(h_{Y_1,t}, h_{Y_0,t}), t = c, n, a, d\}$ satisfying the above constraints, let $\pi_t = \int_{\mathcal{Y}} h_{Y_1,t}d\mu = \int_{\mathcal{Y}} h_{Y_0,t}d\mu \geq 0$ for $t \in \{c, n, a, d\}$. Consider the conditional densities of (Y_1, Y_0, T) given Z constructed as

$$\begin{aligned} f_{Y_1,Y_0,T|Z}(y_1, y_0, T = t|Z = 1) &= f_{Y_1,Y_0,T|Z}(y_1, y_0, T = t|Z = 0) \\ &= \begin{cases} \pi_t^{-1} h_{Y_1,t}(y_1)h_{Y_0,t}(y_0) & \text{if } \pi_t > 0, \\ 0 & \text{if } \pi_t = 0. \end{cases} \end{aligned}$$

By construction the constructed population satisfies RA. Also, the constraint (12) and the construction of the population implies that

$$\begin{aligned} p_{Y_1}(y_1) &= h_{Y_1,c}(y_1) + h_{Y_1,a}(y_1) \\ &= f_{Y_1,T|Z}(y_1, T = c|Z = 1) + f_{Y_1,T|Z}(y_1, T = a|Z = 1). \end{aligned}$$

A similar result holds for p_{Y_0}, q_{Y_1}, and q_{Y_0}. Hence, the constructed population is compatible with the data generating process. Lastly, this way of constructing the population distribution yields the

provided $f^*_{Y_1}$ as the population marginal distribution of Y_1 since $\sum_{t=c,n,a,d}\int_{y_0\in\mathcal{Y}} f_{Y_1,Y_0,T}(y_1,y_0,t)d\mu = \sum_{t=c,n,a,d} h_{Y_1,t}(y_1) = f^*_{Y_1}$, as implied by the constraints (12) and (16). This is also the case for $f^*_{Y_0}$ and the population marginal distribution of Y_0, as implied by the constraints (14) and (18). Thus, the given $(f^*_{Y_1}, f^*_{Y_0})$ belongs to $IR_{(f_{Y_1},f_{Y_0})}(P,Q|RA)$. This completes the proof. ■

Lemma A.2. *Let δ_{Y_1}, δ_{Y_0}, λ_{Y_1}, be the parameters defined in the statement of Propositions 3.1 and 3.2. In addition, define $\lambda_{Y_0} \equiv \int_{\mathcal{Y}} \min\{p_{Y_0}, q_{Y_0}\}d\mu$.*

$$\delta_{Y_1} + \delta_{Y_0} + \lambda_{Y_1} + \lambda_{Y_0} = 2.$$

Proof of Lemma A.2.

$$\begin{aligned}\Pr(D=1|Z=1) + \Pr(D=1|Z=0) &= \int_{\mathcal{Y}} [p_{Y_1} + q_{Y_1}]d\mu \\ &= \int_{\mathcal{Y}} [\max\{p_{Y_1}, q_{Y_1}\} + \min\{p_{Y_1}, q_{Y_1}\}]d\mu \\ &= \delta_{Y_1} + \lambda_{Y_1}.\end{aligned}$$

On the other hand,

$$\begin{aligned}\Pr(D=1|Z=1) + \Pr(D=1|Z=0) &= 2 - \Pr(D=0|Z=1) + \Pr(D=0|Z=0) \\ &= 2 - \int_{\mathcal{Y}} [p_{Y_0} + q_{Y_0}]d\mu \\ &= 2 - \int_{\mathcal{Y}} [\max\{p_{Y_0}, q_{Y_0}\} + \min\{p_{Y_0}, q_{Y_0}\}]d\mu \\ &= 2 - \delta_{Y_0} - \lambda_{Y_0}.\end{aligned}$$

Hence, $\delta_{Y_1} + \lambda_{Y_1} = 2 - \delta_{Y_0} - \lambda_{Y_0}$ holds. ■

Proof of Proposition 3.2. As shown in Proposition 3.1, if the data generating process reveals $\delta_{Y_1} > 1$ or $\delta_{Y_0} > 1$, no population is compatible with MSI, and this clearly implies that $IR_{(f_{Y_1},f_{Y_0})}(P,Q|RA)$ is empty. So, we preclude this trivial case from the proof and focus on a data generating process with $\delta_{Y_1} \leq 1$ and $\delta_{Y_0} \leq 1$.

First, consider a data generating process with $1 - \delta_{Y_0} < \lambda_{Y_1}$, and guess the identification region to be $IR^{guess}_{(f_{Y_1},f_{Y_0})}(P,Q|RA) = \mathcal{F}^*_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$. Note that $\mathcal{F}^*_{f_{Y_1}}(P,Q)$ is non-empty since it always contains $f_{Y_1} = \underline{f_{Y_1}} + \frac{1-\delta_{Y_1}}{\lambda_{Y_1}} \min\{p_{Y_1}, q_{Y_1}\}$.

Pick an arbitrary f_{Y_1} from $\mathcal{F}^*_{f_{Y_1}}(P,Q)$ and an arbitrary f_{Y_0} from $F^{env}_{f_{Y_0}}(P,Q)$. Define a non-negative function

$$g_{Y_1} = \frac{\lambda_{Y_1} + \delta_{Y_0} - 1}{\int_{\mathcal{Y}} \min\left\{f_{Y_1} - \underline{f_{Y_1}}, \min\{p_{Y_1}, q_{Y_1}\}\right\} d\mu} \min\left\{f_{Y_1} - \underline{f_{Y_1}}, \min\{p_{Y_1}, q_{Y_1}\}\right\}, \tag{24}$$

and consider the following choice of $\{(h_{Y_1,t}, h_{Y_0,t}), t = c, n, a, d\}$,

$$\begin{aligned}
h_{Y_1,c} &= p_{Y_1} - \min\{p_{Y_1}, q_{Y_1}\} + g_{Y_1}, \\
h_{Y_1,n} &= f_{Y_1} - \underline{f_{Y_1}} - g_{Y_1}, \\
h_{Y_1,a} &= \min\{p_{Y_1}, q_{Y_1}\} - g_{Y_1}, \\
h_{Y_1,d} &= q_{Y_1} - \min\{p_{Y_1}, q_{Y_1}\} + g_{Y_1}, \\
h_{Y_0,c} &= q_{Y_0} - \min\{p_{Y_0}, q_{Y_0}\}, \\
h_{Y_0,n} &= \min\{p_{Y_0}, q_{Y_0}\}, \\
h_{Y_0,a} &= f_{Y_0} - \underline{f_{Y_0}}, \\
h_{Y_0,d} &= p_{Y_0} - \min\{p_{Y_0}, q_{Y_0}\}.
\end{aligned} \tag{25}$$

Since the first multiplicative term on the right-hand side of (24) is less than or equal to one, $g_{Y_1} \leq \min\left\{f_{Y_1} - \underline{f_{Y_1}}, \min\{p_{Y_1}, q_{Y_1}\}\right\} \leq \min\{p_{Y_1}, q_{Y_1}\}$ and $g_{Y_1} \leq f_{Y_1} - \underline{f_{Y_1}}$. Hence, $\{h_{Y_1,t}(y_1), t = c, n, a, d\}$ constructed above are all non-negative functions. It can be seen that the constraints (12) through (19) are all satisfied. Also, by utilizing Lemma A.2, we can confirm that the area constraints (20) through (23) are satisfied. By Lemma A.1, we conclude that the proposed (f_{Y_1}, f_{Y_0}) belongs to $IR_{(f_{Y_1}, f_{Y_0})}(P, Q|RA)$, and hence $IR^{guess}_{(f_{Y_1}, f_{Y_0})}(P, Q|RA) \subset IR_{(f_{Y_1}, f_{Y_0})}(P, Q|RA)$.

Next, consider f_{Y_1} that does *not* satisfy $\int_{\mathcal{Y}} \min\left\{f_{Y_1} - \underline{f_{Y_1}}, \min\{p_{Y_1}, q_{Y_1}\}\right\} d\mu \geq \lambda_{Y_1} + \delta_{Y_0} - 1$ and $f_{Y_0} \in \mathcal{F}^{env}_{f_{Y_0}}(P, Q)$. In order to find a contradiction of RA, suppose that the non-negative functions $\{(h_{Y_1,t}, h_{Y_0,t}), t = c, n, a, d\}$ satisfying the constraints (12) through (19) exist. Then, the constraints (18) and (19) imply that $\int_{\mathcal{Y}} h_{Y_0,a} d\mu \leq 1 - \delta_{Y_0}$. Moreover,

$$\begin{aligned}
f_{Y_1} &= \sum_{t=c,n,a,d} h_{Y_1,t} \\
&\geq p_{Y_1} + q_{Y_1} - h_{Y_1,a} \\
&= \underline{f_{Y_1}} + \min\{p_{Y_1}, q_{Y_1}\} - h_{Y_1,a},
\end{aligned}$$

implies that

$$f_{Y_1} - \underline{f_{Y_1}} \geq \min\{p_{Y_1}, q_{Y_1}\} - h_{Y_1,a}. \tag{26}$$

Now, since $f_{Y_1} \notin \mathcal{F}^*_{f_{Y_1}}(P,Q)$, it follows that

$$\begin{aligned}
\lambda_{Y_1} + \delta_{Y_0} - 1 &> \int_{\mathcal{Y}} \min\left\{ f_{Y_1} - \underline{f_{Y_1}}, \min\{p_{Y_1}, q_{Y_1}\} \right\} d\mu \\
&\geq \int_{\mathcal{Y}} \min\left\{ \min\{p_{Y_1}, q_{Y_1}\} - h_{Y_1,a}, \min\{p_{Y_1}, q_{Y_1}\} \right\} d\mu \\
&= \int_{\mathcal{Y}} [\min\{p_{Y_1}, q_{Y_1}\} - h_{Y_1,a}] d\mu \\
&= \lambda_{Y_1} - \int h_{Y_1,a} d\mu,
\end{aligned}$$

where the second line follows by inequality (26). Hence, $\int h_{Y_1,a} d\mu > 1 - \delta_{Y_0}$. This and $\int_{\mathcal{Y}} h_{Y_0,a} d\mu \leq 1 - \delta_{Y_0}$ contradict the area constraint for $t = a$. So, we conclude that there are no feasible $\{(h_{Y_1,t}, h_{Y_0,t}), t = c, n, a, d\}$ that meet all the constraints of Lemma A.1, implying that such a f_{Y_1} is not contained in the identification region under RA. Note $f_{Y_0} \notin \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$ immediately implies violation of $Y_0 \perp Z$. Therefore, any $(f_{Y_1}, f_{Y_0}) \notin IR^{guess}_{(f_{Y_1}, f_{Y_0})}(P,Q|RA)$ do not belong to $IR_{(f_{Y_1}, f_{Y_0})}(P,Q|RA)$, implying $IR^{guess}_{(f_{Y_1}, f_{Y_0})}(P,Q|RA) \supset IR_{(f_{Y_1}, f_{Y_0})}(P,Q|RA)$.

By combining these results, we conclude that, if the data generating process satisfies $1 - \delta_{Y_0} < \lambda_{Y_1}$, $\mathcal{F}^*_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$ is the identification region of (f_{Y_1}, f_{Y_0}) under RA.

For the case of $1 - \delta_{Y_0} > \lambda_{Y_1}$, we can construct the identification region by a similar argument to that for the case of $1 - \delta_{Y_0} < \lambda_{Y_1}$. Guess the identification region to be $IR^{guess}_{(f_{Y_1}, f_{Y_0})}(P,Q|RA) = \mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^*_{f_{Y_0}}(P,Q)$. Note that $\mathcal{F}^*_{f_{Y_0}}(P,Q)$ is non-empty since it always contains $f_{Y_0} = \underline{f_{Y_0}} + \frac{1-\delta_{Y_0}}{\lambda_{Y_0}} \min\{p_{Y_0}, q_{Y_0}\}$. Pick an arbitrary f_{Y_0} from $\mathcal{F}^*_{f_{Y_0}}(P,Q)$ and an arbitrary f_{Y_1} from $F^{env}_{f_{Y_1}}(P,Q)$. Similar to (24), define the non-negative function

$$g_{Y_0} = \frac{1 - \delta_{Y_0} - \lambda_{Y_1}}{\int_{\mathcal{Y}} \min\left\{ f_{Y_0} - \underline{f_{Y_0}}, \min\{p_{Y_0}, q_{Y_0}\} \right\} d\mu} \min\left\{ f_{Y_0} - \underline{f_{Y_0}}, \min\{p_{Y_0}, q_{Y_0}\} \right\},$$

and consider the following choice of $\{(h_{Y_1,t}, h_{Y_0,t}), t = c, n, a, d\}$,

$$\begin{aligned}
h_{Y_1,c} &= p_{Y_1} - \min\{p_{Y_1}, q_{Y_1}\}, \\
h_{Y_1,n} &= f_{Y_1} - \underline{f_{Y_1}}, \\
h_{Y_1,a} &= \min\{p_{Y_1}, q_{Y_1}\}, \\
h_{Y_1,d} &= q_{Y_1} - \min\{p_{Y_1}, q_{Y_1}\}, \\
h_{Y_0,c} &= q_{Y_0} - \min\{p_{Y_0}, q_{Y_0}\} + g_{Y_0}, \\
h_{Y_0,n} &= \min\{p_{Y_0}, q_{Y_0}\} - g_{Y_0}, \\
h_{Y_0,a} &= f_{Y_0} - \underline{f_{Y_0}} - g_{Y_0}, \\
h_{Y_0,d} &= p_{Y_0} - \min\{p_{Y_0}, q_{Y_0}\} + g_{Y_0}.
\end{aligned}$$

Note that these $\{(h_{Y_1,t}, h_{Y_0,t}), t = c, n, a, d\}$ are non-negative functions and that they meet the constraints (12) through (23). Again, by Lemma A.1, we conclude that the proposed (f_{Y_1}, f_{Y_0}) belongs to $IR_{(f_{Y_1},f_{Y_0})}(P, Q|RA)$, so $IR^{guess}_{(f_{Y_1},f_{Y_0})}(P, Q|RA) \subset IR_{(f_{Y_1},f_{Y_0})}(P, Q|RA)$.

Next, consider $f_{Y_0} \notin \mathcal{F}^*_{f_{Y_0}}(P, Q)$ and $f_{Y_1} \in \mathcal{F}^{env}_{f_{Y_1}}(P, Q)$. Similar to the previous case, we suppose that non-negative functions $\{(h_{Y_1,t}, h_{Y_0,t}), t = c, n, a, d\}$ satisfying the constraints (12) through (19) exist. Then, the constraints (16) and (17) imply that $\int_{\mathcal{Y}} h_{Y_1,n} d\mu \leq 1 - \delta_{Y_1}$. Moreover,

$$\begin{aligned} f_{Y_0} &= \sum_{t=c,n,a,d} h_{Y_0,t} \\ &\geq p_{Y_0} + q_{Y_0} - h_{Y_0,n} \\ &= \underline{f_{Y_0}} + \min\{p_{Y_0}, q_{Y_0}\} - h_{Y_0,n}, \end{aligned}$$

implies that

$$f_{Y_0} - \underline{f_{Y_0}} \geq \min\{p_{Y_0}, q_{Y_0}\} - h_{Y_0,n}. \tag{27}$$

Now, since $f_{Y_0} \notin \mathcal{F}^*_{f_{Y_0}}(P, Q)$, it follows that

$$\begin{aligned} 1 - \delta_{Y_0} - \lambda_{Y_1} &> \int_{\mathcal{Y}} \min\left\{f_{Y_0} - \underline{f_{Y_0}}, \min\{p_{Y_0}, q_{Y_0}\}\right\} d\mu \\ &\geq \int_{\mathcal{Y}} \min\left\{\min\{p_{Y_0}, q_{Y_0}\} - h_{Y_0,n}, \min\{p_{Y_0}, q_{Y_0}\}\right\} d\mu \\ &= \int_{\mathcal{Y}} [\min\{p_{Y_0}, q_{Y_0}\} - h_{Y_0,n}] d\mu \\ &= \lambda_{Y_0} - \int h_{Y_0,n} d\mu, \end{aligned}$$

where $\lambda_{Y_0} \equiv \int_{\mathcal{Y}} \min\{p_{Y_0}, q_{Y_0}\} d\mu$. By Lemma A.2, $1 - \delta_{Y_0} - \lambda_{Y_1} - \lambda_{Y_0} = \delta_{Y_1} - 1$, so we have $\int h_{Y_0,n} d\mu > 1 - \delta_{Y_1}$. This and $\int_{\mathcal{Y}} h_{Y_1,n} d\mu \leq 1 - \delta_{Y_1}$ violate the area constraint for $t = n$. So, we conclude that there are no feasible $\{(h_{Y_1,t}, h_{Y_0,t}), t = c, n, a, d\}$ that meet all of the constraints of Lemma A.1, implying that such a f_{Y_0} is not contained in the identification region under RA. Since $f_{Y_1} \notin \mathcal{F}^{env}_{f_{Y_0}}(P, Q)$ implies violation of $Y_1 \perp Z$, any $(f_{Y_1}, f_{Y_0}) \notin IR^{guess}_{(f_{Y_1},f_{Y_0})}(P, Q|RA)$ do not belong to $IR_{(f_{Y_1},f_{Y_0})}(P, Q|RA)$, implying $IR^{guess}_{(f_{Y_1},f_{Y_0})}(P, Q|RA) \supset IR_{(f_{Y_1},f_{Y_0})}(P, Q|RA)$. Thus, if the data generating process satisfies $1 - \delta_{Y_0} > \lambda_{Y_1}$, $\mathcal{F}^{env}_{f_{Y_1}}(P, Q) \times \mathcal{F}^*_{f_{Y_0}}(P, Q)$ is the identification region of (f_{Y_1}, f_{Y_0}) under RA.

Lastly, consider the case of $1 - \delta_{Y_0} = \lambda_{Y_1}$. In this case, for every $f_{y_1} \in F^{env}_{f_{Y_1}}(P, Q)$ and $f_{Y_0} \in F^{env}_{f_{Y_0}}(P, Q)$, we consider the following choice of $\{(h_{Y_1,t}, h_{Y_0,t}), t = c, n, a, d\}$,

$$h_{Y_1,c} = p_{Y_1} - \min\{p_{Y_1}, q_{Y_1}\},$$

$$
\begin{aligned}
h_{Y_1,n} &= f_{Y_1} - \underline{f_{Y_1}}, \\
h_{Y_1,a} &= \min\{p_{Y_1}, q_{Y_1}\}, \\
h_{Y_1,d} &= q_{Y_1} - \min\{p_{Y_1}, q_{Y_1}\}, \\
h_{Y_0,c} &= q_{Y_0} - \min\{p_{Y_0}, q_{Y_0}\}, \\
h_{Y_0,n} &= \min\{p_{Y_0}, q_{Y_0}\}, \\
h_{Y_0,a} &= f_{Y_0} - \underline{f_{Y_0}}, \\
h_{Y_0,d} &= p_{Y_0} - \min\{p_{Y_0}, q_{Y_0}\}.
\end{aligned}
$$

This choice satisfies all the constraints of Lemma A.1, including the area constraints. Since $f_{y_1} \notin F^{env}_{f_{Y_1}}(P,Q)$ or $f_{Y_0} \notin F^{env}_{f_{Y_0}}(P,Q)$ leads to violation of MSI, $\mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$ is the identification region of (f_{Y_1}, f_{Y_0}) under RA.

As we have discussed, $\mathcal{F}^{*}_{f_{Y_1}}(P,Q)$ and $\mathcal{F}^{*}_{f_{Y_1}}(P,Q)$ are not empty whenever $\delta_{Y_1} \leq 1$ and $\delta_{Y_0} \leq 1$ so (ii) of the proposition is proved. ■

The next lemma is used for the proof of Proposition 4.1.

Lemma A.3. *$F^{env}_{f_{Y_j}}(P,Q)$ and $F^{*}_{f_{Y_j}}(P,Q)$, $j = 1.0$, are convex sets.*

Proof of Lemma A.3. Convexity of $F^{env}_{f_{Y_j}}(P,Q)$ is trivial. Consider $k, l \in F^{*}_{f_{Y_1}}(P,Q)$. Note that $\min\{x - c_1, c_2\}$ is a convex function for arbitrary constants c_1 and c_2. Hence, for μ-almost every $y_1 \in \mathcal{Y}$, and any $\xi \in [0,1]$,

$$
\begin{aligned}
& \min\{\xi k(y_1) + (1-\xi) l(y_1) - \underline{f_{Y_1}}(y_1), \min\{p_{Y_1}(y_1), q_{Y_1}(y_1)\}\}. \\
\geq\ & \xi \min\left\{k(y_1) - \underline{f_{Y_1}}(y_1), \min\{p_{Y_1}(y_1), q_{Y_1}(y_1)\}\right\} \\
& + (1-\xi) \min\left\{l(y_1) - \underline{f_{Y_1}}(y_1), \min\{p_{Y_1}(y_1), q_{Y_1}(y_1)\}\right\}.
\end{aligned}
$$

By integrating this inequality, we obtain

$$
\int_{\mathcal{Y}} \min\{\xi k + (1-\xi) l - \underline{f_{Y_1}}, \min\{p_{Y_1}, q_{Y_1}\}\} d\mu \geq \lambda_{Y_1} + \delta_{Y_0} - 1.
$$

Hence, $\xi k + (1-\xi) l \in F^{*}_{f_{Y_1}}(P,Q)$. A similar result holds for $F^{*}_{f_{Y_0}}(P,Q)$. ■

Proof of Proposition 4.1. The mean parameter respects stochastic dominance (Manski, 2003). So, the sharp lower bound of $E(Y_1)$ is obtained by finding $f^{lower}_{Y_1}$ within the identification region such that $f^{lower}_{Y_1}$ is first-order stochastically dominated by all other probability density functions

contained in the identification region. Similarly, the sharp upper bound of $E(Y_1)$ is obtained by finding $f_{Y_1}^{upper}$ within the identification region such that $f_{Y_1}^{upper}$ first-order stochastically dominates all other probability density functions in the identification region. By Lemma A.3, the identification regions to be considered are always convex, so we can span any intermediate values between the lower and upper bound of $E(Y_1)$ by a mixture of $f_{Y_1}^{lower}$ and $f_{Y_1}^{upper}$. Hence, for the construction of the sharp ATE bounds, it suffices to find such $f_{Y_1}^{upper}$ and $f_{Y_1}^{lower}$.

We first consider bounding the mean of Y_1 when the density f_{Y_1} belongs to the class of densities $\mathcal{F}_{f_{Y_1}}^{env}(P,Q)$ and $\mathcal{F}_{f_{Y_1}}^{*}(P,Q)$ respectively. For the former, it is known that the bound of $E(Y_1)$ is given by

$$(1-\delta_{Y_1})y_l + \int_{\mathcal{Y}} y_1 \underline{f_{Y_1}} d\mu \leq E(Y_1) \leq (1-\delta_{Y_1})y_u + \int_{\mathcal{Y}} y_1 \underline{f_{Y_1}} d\mu.$$

See Corollary 2.2.2 in Manski (2003) for the discrete outcome case and Kitagawa (2009b) for the continuous outcome case.

To derive the bounds of $E(Y_1)$ for the latter case, consider the probability density function

$$f_{Y_1}^{lower}(y_1) = \lambda_{Y_0} 1\{y_1 = y_l\} + \underline{f_{Y_1}}(y_1) + [\min\{p_{Y_1}, q_{Y_1}\}]_{1-\delta_{Y_0}}^{rtrim}(y_1).$$

Note that $f_{Y_1}^{lower} \geq \underline{f_{Y_1}}$ and

$$\int_{\mathcal{Y}} \min\left\{ f_{Y_1}^{lower} - \underline{f_{Y_1}}, \min\{p_{Y_1}, q_{Y_1}\} \right\} d\mu = \lambda_{Y_1} - (1-\delta_{Y_0}),$$

so $f_{Y_1}^{lower} \in \mathcal{F}_{f_{Y_1}}^{*}(P,Q)$. By applying the decomposition (25) proposed in the proof of Proposition 3.2, we can decompose $f_{Y_1}^{lower}$ into non-negative functions $\{h_{Y_1,t}^{lower},\ t = c,n,a,d\}$. Specifically, for $t = a$ and $t = n$, we obtain

$$\begin{aligned} h_{Y_1,a}^{lower} &= \min\{p_{Y_1}, q_{Y_1}\} - [\min\{p_{Y_1}, q_{Y_1}\}]_{1-\delta_{Y_0}}^{rtrim}, \\ h_{Y_1,n}^{lower} &= \lambda_{Y_0} 1\{y_1 = y_l\}. \end{aligned}$$

Note that by using $h_{Y_1,t}^{lower}$, $t = c,n,a,d$, we can express $f_{Y_1}^{lower}$ as

$$\begin{aligned} f_{Y_1}^{lower} &= \sum_t h_{Y_1,t}^{lower} \\ &= p_{Y_1} + q_{Y_1} - h_{Y_1,a}^{lower} + h_{Y_1,n}^{lower}, \end{aligned} \tag{28}$$

where in the second line we use the constraints (12) and (13). Let $\tilde{f}_{Y_1}$ be an arbitrary element of $\mathcal{F}_{f_{Y_1}}^{*}(P,Q)$. By Lemma A.1 and Proposition 3.2, there exist non-negative functions $\{\tilde{h}_{Y_1,t}, t =$

$c, n, a, d\}$ such that $\tilde{f}_{Y_1}$ can be represented as

$$\begin{aligned}\tilde{f}_{Y_1} &= \sum_t \tilde{h}_{Y_1,t} \\ &= p_{Y_1} + q_{Y_1} - \tilde{h}_{Y_1,a} + \tilde{h}_{Y_1,n}, \end{aligned} \tag{29}$$

and, again, by applying decomposition (25) of the proof of Proposition 3.2, $\tilde{h}_{Y_1,a}$ and $\tilde{h}_{Y_1,n}$ can be expressed as

$$\begin{aligned}\tilde{h}_{Y_1,a} &= \min\{p_{Y_1}, q_{Y_1}\} - \tilde{g}_{Y_1}, \\ \tilde{h}_{Y_1,n} &= \tilde{f}_{Y_1} - \underline{f_{Y_1}} - \tilde{g}_{Y_1},\end{aligned}$$

where $\tilde{g}_{Y_1}$ is obtained by plugging $\tilde{f}_{Y_1}$ into (24). From (28) and (29), for $t \in [y_l, y_u]$, the difference between $\int_{[y_l,t]} f_{Y_1}^{lower} d\mu$ and $\int_{[y_l,t]} \tilde{f}_{Y_1} d\mu$ is written as

$$\begin{aligned}&\int_{[y_l,t]} f_{Y_1}^{lower} d\mu - \int_{[y_l,t]} \tilde{f}_{Y_1} d\mu \\ &= \int_{[y_l,t]} [h_{Y_1,n}^{lower} - \tilde{h}_{Y_1,n}] d\mu + \int_{[y_l,t]} [\tilde{h}_{Y_1,a} - h_{Y_1,a}^{lower}] d\mu \\ &= \lambda_{Y_0} - \int_{[y_l,t]} \left(\tilde{f}_{Y_1} - \underline{f_{Y_1}} - \tilde{g}_{Y_1}\right) d\mu + \int_{[y_l,t]} \left([\min\{p_{Y_1}, q_{Y_1}\}]_{1-\delta_{Y_0}}^{rtrim} - \tilde{g}_{Y_1}\right) d\mu. \end{aligned} \tag{30}$$

Regarding the second term of (30), as $\tilde{f}_{Y_1} - \underline{f_{Y_1}} - \tilde{g}_{Y_1} \geq 0$, it can be bounded from above by

$$\begin{aligned}\int_{[y_l,t]} \left(\tilde{f}_{Y_1} - \underline{f_{Y_1}} - \tilde{g}_{Y_1}\right) d\mu &\leq \int_{\mathcal{Y}} \left(\tilde{f}_{Y_1} - \underline{f_{Y_1}} - \tilde{g}_{Y_1}\right) d\mu \\ &= 1 - \delta_{Y_1} - \lambda_{Y_1} - 1 + \delta_{Y_0}.\end{aligned}$$

Regarding the third term of (30), if t is strictly less than the $(1-\delta_{Y_0})$-th right-trimming point $q_{1-\delta_{Y_0}}^{right} = \sup\left\{s : \int_{[s,y_u]} \min\{p_{Y_1}, q_{Y_1}\} d\mu \geq 1 - \delta_{Y_0}\right\}$, then $[\min\{p_{Y_1}, q_{Y_1}\}]_{1-\delta_{Y_0}}^{rtrim} = \min\{p_{Y_1}, q_{Y_1}\} \geq \tilde{g}_{Y_1}$ holds on $y_1 \in [y_l, t]$. So the integral is non-negative. On the other hand, if $t \geq q_{1-\delta_{Y_0}}^{right}$,

$$\begin{aligned}&\int_{[y_l,t]} \left([\min\{p_{Y_1}, q_{Y_1}\}]_{1-\delta_{Y_0}}^{rtrim} - \tilde{g}_{Y_1}\right) d\mu \\ &= \lambda_{Y_1} - (1 - \delta_{Y_0}) - \int_{[y_l,t]} \tilde{g}_{Y_1} d\mu \\ &\geq \lambda_{Y_1} - (1 - \delta_{Y_0}) - \int_{\mathcal{Y}} \tilde{g}_{Y_1} d\mu \\ &= \lambda_{Y_1} - (1 - \delta_{Y_0}) - [\lambda_{Y_1} - (1 - \delta_{Y_0})] = 0.\end{aligned}$$

By combining them, for each $t \in [y_l, y_u]$, (30) is bounded from below by $\lambda_{Y_0} + \lambda_{Y_1} + \delta_{Y_1} + \delta_{Y_0} - 2$, and this is zero by Lemma A.2. Therefore, we conclude that $f_{Y_1}^{lower}$ first order stochastically dominates $\tilde{f}_{Y_1}$, and the mean of Y_1 with respect to $f_{Y_1}^{lower}$ minimizes $E(Y_1)$ over $f_{Y_1} \in \mathcal{F}_{f_{Y_1}}^*(P,Q)$.

Next, we shall find the upper bound of $E(Y_1)$ by essentially repeating the same procedure as above. Define

$$f_{Y_1}^{upper}(y_1) = \underline{f_{Y_1}}(y_1) + [\min\{p_{Y_1}, q_{Y_1}\}]_{1-\delta_{Y_0}}^{ltrim}(y_1) + \lambda_{Y_0} 1\{y_1 = y_u\},$$

which is shown to belong to $\mathcal{F}_{f_{Y_1}}^*(P,Q)$. Similar to the lower bound case (29), represent $f_{Y_1}^{upper}$ by

$$\begin{aligned} f_{Y_1}^{upper} &= \sum_t h_{Y_1,t}^{upper} \\ &= p_{Y_1} + q_{Y_1} - h_{Y_1,a}^{upper} + h_{Y_1,n}^{upper}, \\ h_{Y_1,a}^{upper} &= \min\{p_{Y_1}, q_{Y_1}\} - [\min\{p_{Y_1}, q_{Y_1}\}]_{1-\delta_{Y_0}}^{ltrim}, \\ h_{Y_1,n}^{upper} &= \lambda_{Y_0} 1\{y_1 = y_u\}. \end{aligned}$$

For an arbitrary $\tilde{f}_{Y_1} \in \mathcal{F}_{f_{Y_1}}^*(P,Q)$, consider the difference between $\int_{(t,y_u]} f_{Y_1}^{upper} d\mu$ and $\int_{(t,y_u]} \tilde{f}_{Y_1} d\mu$. Analogous to (30), we obtain

$$\begin{aligned} &\int_{(t,y_u]} f_{Y_1}^{upper} d\mu - \int_{(t,y_u]} \tilde{f}_{Y_1} d\mu \\ &\qquad = \lambda_{Y_0} - \int_{(t,y_u]} \left(\tilde{f}_{Y_1} - \underline{f_{Y_1}} - \tilde{g}_{Y_1}\right) d\mu + \int_{(t,y_u]} \left([\min\{p_{Y_1}, q_{Y_1}\}]_{1-\delta_{Y_0}}^{ltrim} - \tilde{g}_{Y_1}\right) d\mu. \end{aligned}$$

Now, by repeating the same procedure as above, the right-hand side is bounded from below by $\lambda_{Y_0} + \lambda_{Y_1} + \delta_{Y_1} + \delta_{Y_0} - 2 = 0$. Hence, we conclude that $f_{Y_1}^{upper}$ is first order stochastically dominated by $\tilde{f}_{Y_1}$, and the mean of Y_1 with respect to $f_{Y_1}^{upper}$ maximizes $E(Y_1)$ over $f_{Y_1} \in \mathcal{F}_{f_{Y_1}}^*(P,Q)$.

The bounds for $E(Y_0)$ when the density f_{Y_0} belongs to the class of densities $\mathcal{F}_{f_{Y_0}}^{env}(P,Q)$ and $\mathcal{F}_{f_{Y_0}}^*(P,Q)$ are derived by a symmetric argument to that for the case of $E(Y_1)$, so we do not duplicate the proof here.

In order to combine the bounds of $E(Y_1)$ and $E(Y_0)$, we note that the identification region of (f_{Y_1}, f_{Y_0}) takes the form of the Cartesian product of $\mathcal{F}_{f_{Y_1}}^{env}(P,Q)$ or $\mathcal{F}_{f_{Y_1}}^*(P,Q)$ and $\mathcal{F}_{f_{Y_0}}^{env}(P,Q)$ or $\mathcal{F}_{f_{Y_0}}^*(P,Q)$. Hence, by applying the argument of the outer bounds of Manski (2003), it is valid to bound $E(Y_1) - E(Y_0)$ by subtracting the upper (lower) bound of $E(Y_0)$ from the lower (upper) bound of $E(Y_1)$ for each corresponding underlying identification region of f_{Y_1} and f_{Y_0} . This completes the proof of the sharp bounds under MSI and RA.

As for the bounds under LATE, the sharp bounds become empty when the nested densities are not observed because the identification region of (f_{Y_1}, f_{Y_0}) in this case is empty (see Proposition 3.3 and (2)). On the other hand, when the data generating process exhibits nested densities, the formula of the sharp ATE bounds corresponding to $\mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$ is reduced to the presented formula since, for $j = 1, 0$, we have $\delta_{Y_j} = \max_z\{\Pr(D = j|Z = z)\}$ and $\int_{\mathcal{Y}} y_j \underline{f_{Y_j}} d\mu = \max_z \{E(Y|D = j, Z = z) \Pr(D = j|Z = z)\}$. ■

Appendix B: A Geometric Illustration for Proposition 3.2

In this appendix, we provide a geometric illustration of how the inequalities in the definition of $\mathcal{F}^*_{f_{Y_1}}(P,Q)$ and $\mathcal{F}^*_{f_{Y_0}}(P,Q)$ emerge in constructing the identification region under RA. For ease of exposition, we first consider the case of $1-\delta_{Y_0} = \lambda_{Y_1}$ where Proposition 3.2 says RA does *not* provide further identification gain beyond MSI. Figure 3 draws the data generating process and an arbitrary $(f_{Y_1}, f_{Y_0}) \in \mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$ for this case. There, we partition the subgraph of f_{Y_1} into four, $c(1)$, $a(1)$, $n(1)$, and $d(1)$, and similarly partition the subgraph of f_{Y_0} into $c(0)$, $a(0)$, $n(0)$, and $d(0)$. The condition $1 - \delta_{Y_0} = \lambda_{Y_1}$ means that the area of the partition outlined between f_{Y_0} and $\underline{f_{Y_0}}$ is equal to the area of the subgraph of $\min\{p_{Y_1}, q_{Y_1}\}$ (i.e., the area of $a(1)$ is equal to the area of $a(0)$). Moreover, it can be shown that, $1 - \delta_{Y_0} = \lambda_{Y_1}$ implies not only that $a(1)$ and $a(0)$ but also that $c(1)$ and $c(0)$, $n(1)$ and $n(0)$, and $d(1)$ and $d(0)$ each have the same area. This enables us to pin down $h_{Y_1,t}(y_1)$ and $h_{Y_0,t}(y_0)$ to the height of the partitions $t(1)$ and $t(0)$ for each $t = c, n, a, d$, without violating the area constraints (6). Moreover, this way of pinning down $(h_{Y_1,t}(y_1), h_{Y_0,t}(y_0))$ is compatible with the constraints (7) (see also Figure 4). Thus, we can successfully find feasible non-negative functions $(h_{Y_1,t}, h_{Y_0,t})$ that allow us to construct a population that is compatible with RA and (P,Q) (see Lemma A.1 in Appendix A). Hence, we conclude that the drawn (f_{Y_1}, f_{Y_0}) belongs to $IR_{(f_{Y_1}, f_{Y_0})}(P,Q|RA)$. Note that this way of imputing $h_{Y_1,t}(y_1)$ and $h_{Y_0,t}(y_0)$ works for arbitrary $(f_{Y_1}, f_{Y_0}) \in \mathcal{F}^{env}_{f_{Y_1}}(P,Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$, so the identification region of (f_{Y_1}, f_{Y_0}) under RA is obtained as the Cartesian product of $\mathcal{F}^{env}_{f_{Y_1}}(P,Q)$ and $\mathcal{F}^{env}_{f_{Y_0}}(P,Q)$.

Next, let us consider the case of $1-\delta_{Y_0} < \lambda_{Y_1}$ as drawn in Figure 1 in the main text (i.e., the area of $a(0)$ is smaller than the area of $a(1)$). The preceding way of pinning down $h_{Y_1,t}(y_1)$ and $h_{Y_0,t}(y_0)$ to $t(1)$ and $t(0)$ will now violate the area constraints, so we need to come up with a different way of finding feasible $h_{Y_1,t}(y_1)$ and $h_{Y_0,t}(y_0)$. The following algorithm, graphically illustrated as Figure 5 through Figure 8, presents a way of proposing feasible $h_{Y_1,t}(y_1)$ and $h_{Y_0,t}(y_0)$ in this case.

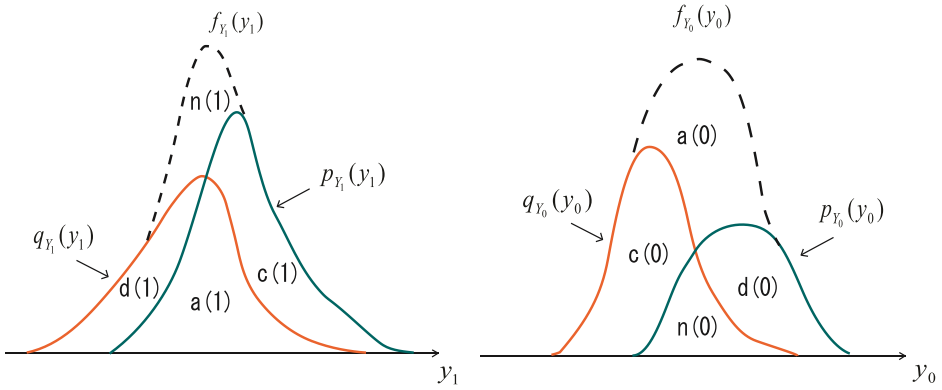

Figure 3: This figure depicts a data gerating process with $1 - \delta_{Y_0} = \lambda_{Y_1}$ (the area of $a(0)$ is equal to the area of $a(1)$). For each $t = c, n, a, d$, $t(1)$ and $t(0)$ have the same area.

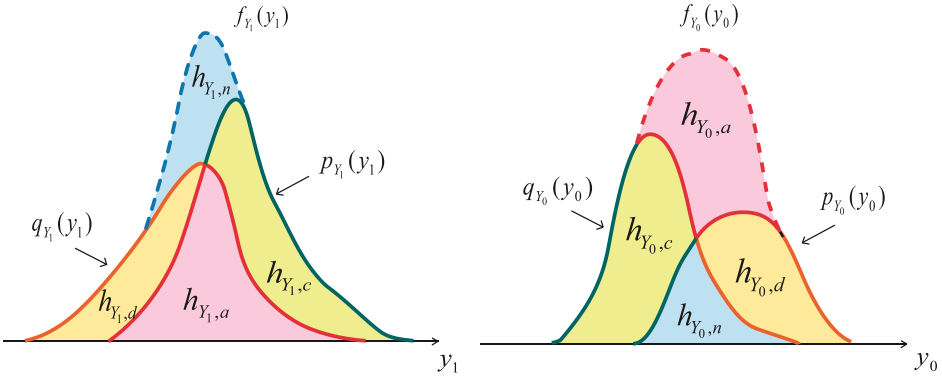

Figure 4: If the data generating process satisfies $1 - \delta_{Y_0} = \lambda_{Y_1}$, we can set $h_{Y_1,t}$ and $h_{Y_0,t}$ to the partitions $t(1)$ and $t(0)$ of Figure 3 without contradicting the scale and compatibility constraints.

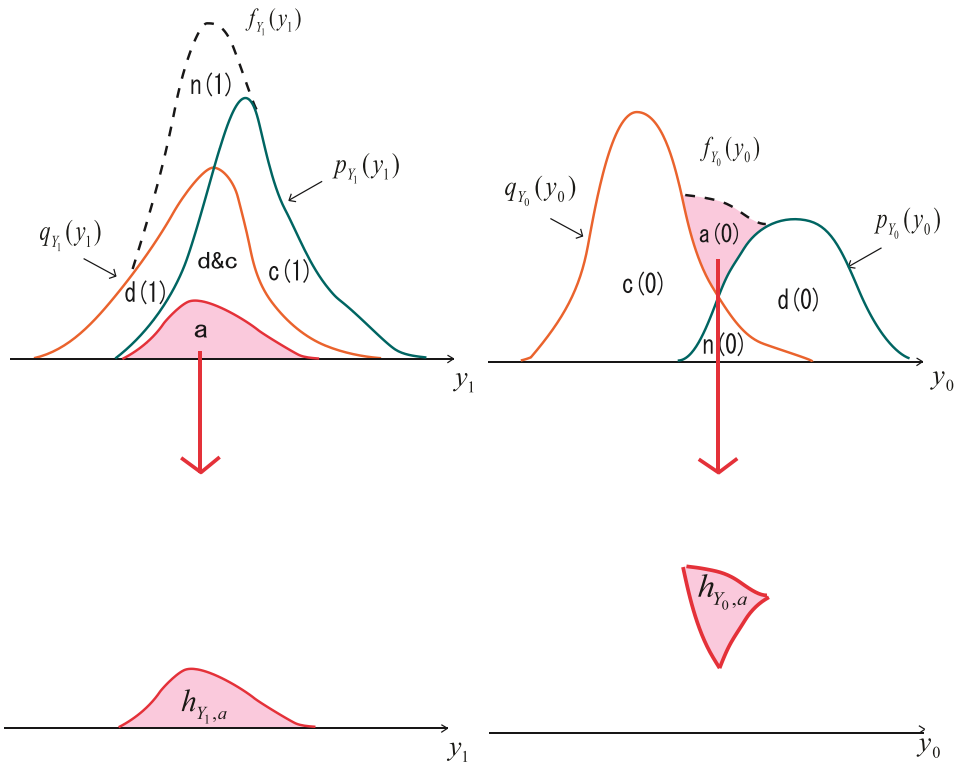

Figure 5: Step 1 – Imputation of $h_{Y_1,a}$ and $h_{Y_0,a}$.

Algorithm to impute $(h_{Y_1,t}, h_{Y_0,t})$, $t = c, n, a, d$.

Step 1: (Figure 5) Draw an arbitrary $f_{Y_1} \in \mathcal{F}^{env}_{f_{Y_1}}(P,Q)$ and $f_{Y_0} \in \mathcal{F}^{env}_{f_{Y_0}}(P,Q)$. We first set $h_{Y_0,a}$ to the height of the partition $a(0)$ and set $h_{Y_1,a}$ to the height of some subset within $\min\{p_{Y_1}, q_{Y_1}\}$ such that its area is equal to the area of $a(0)$. Note that this equal area requirement is due to the area constraint $\int h_{Y_1,a} d\mu = \int h_{Y_0,a} d\mu$. In the top figure, the subset imputed for $h_{Y_1,a}$ is labeled as a. As we pin down $h_{Y_0,a}$ and $h_{Y_1,a}$, we put their copies in the bottom figure for convenience in later steps. How to choose the subset a turns out to be a key for this algorithm and it will be further discussed in Step 4. For now, let us proceed to Step 2 with the drawn subset a.

Step 2: (Figure 6) Impute $h_{Y_1,c}$ and $h_{Y_0,c}$ through the first and seventh constraints of (7). That is, we impute $h_{Y_1,c}$ to the height of subset $c(1) \cup (d\&c)$ and $h_{Y_0,c}$ to the height of subset $c(0)$ as drawn in the top figure. The equal area restriction $\int h_{Y_1,c} d\mu = \int h_{Y_0,c} d\mu$ is automatically satisfied.

Step 3: (Figure 7) Impute $h_{Y_1,d}$ and $h_{Y_0,d}$ via the second and eighth constraints of (7). That is, we impute $h_{Y_1,d}$ to the height of subset $d(1) \cup (d\&c)$ and $h_{Y_0,d}$ to the height of subset $d(0)$ as drawn

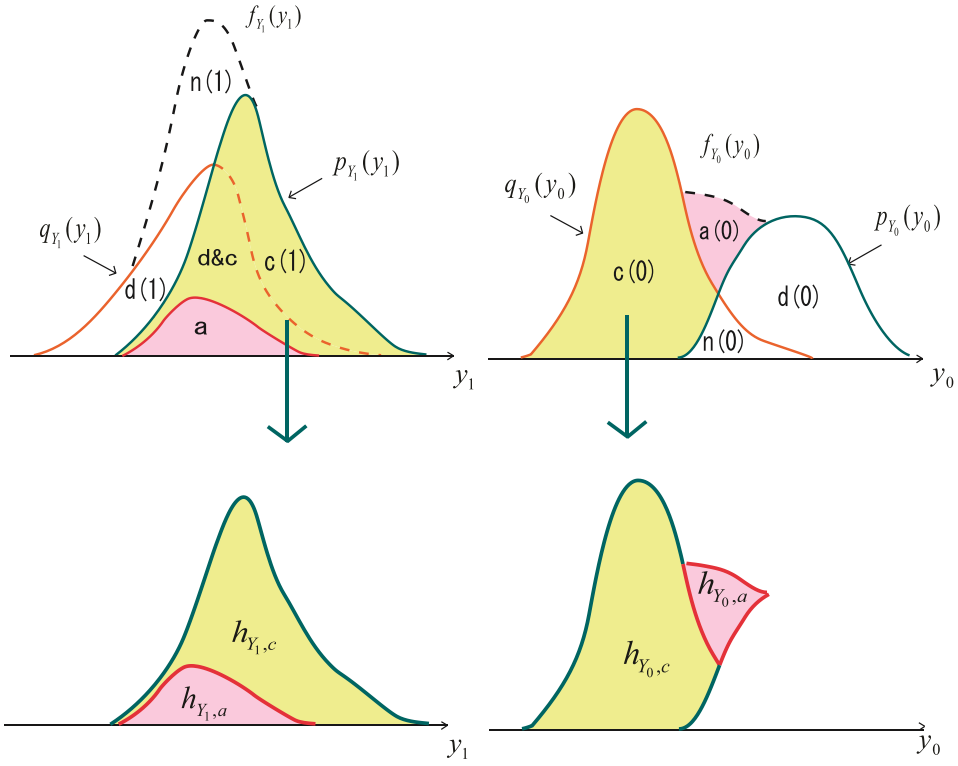

Figure 6: Step 2 – Imputation of $h_{Y_1,c}$ and $h_{Y_0,c}$.

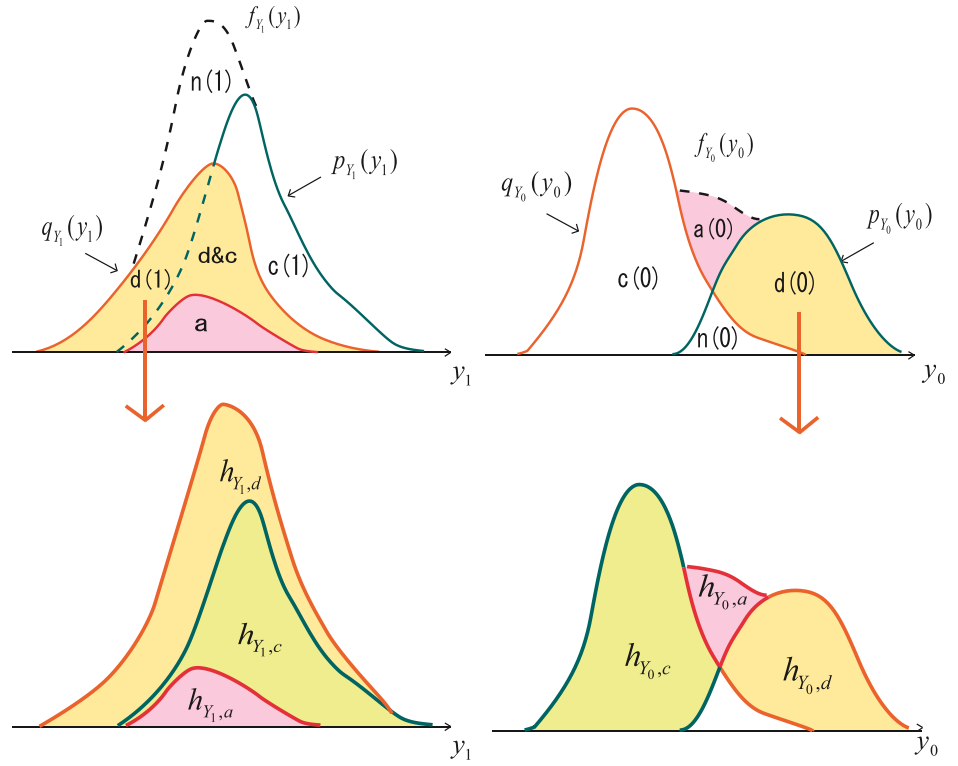

Figure 7: Step 3 – Imputation of $h_{Y_1,d}$ and $h_{Y_0,d}$.

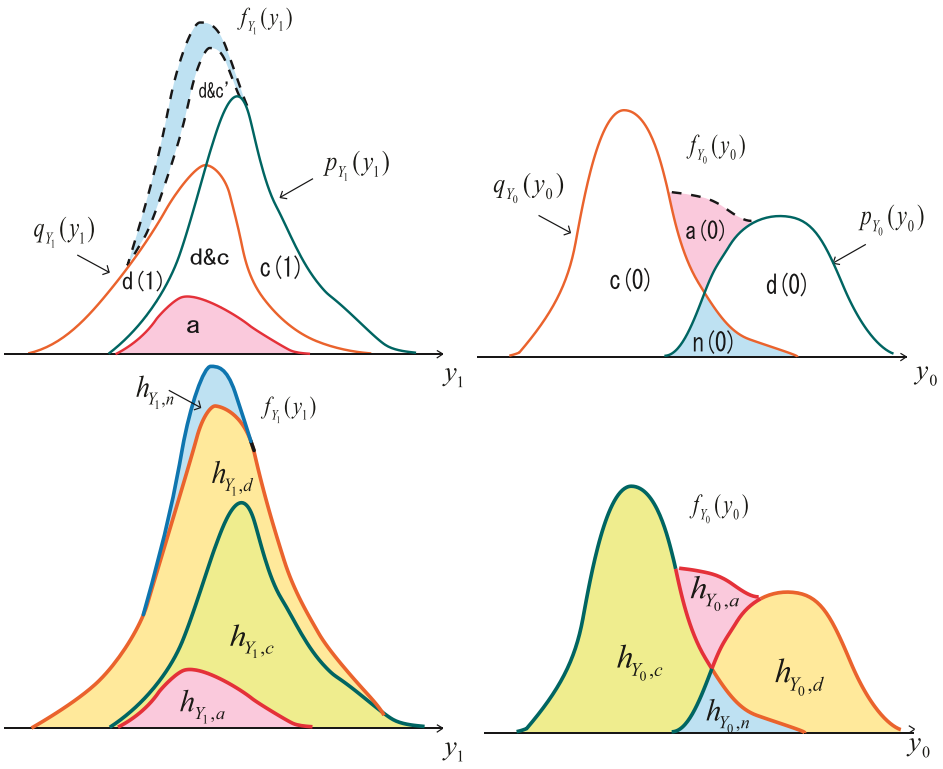

Figure 8: Step 4 – Imputation of $h_{Y_1,n}$ and $h_{Y_0,n}$.

in the top figure. Note that the equal area restriction $\int h_{Y_1,d}d\mu = \int h_{Y_0,d}d\mu$ is again automatically satisfied. In the bottom figure, the imputed $h_{Y_1,d}$ is piled up on the top of $h_{Y_1,a}$ and $h_{Y_1,c}$.

Step 4: (Figure 8) Since the densities of the other three types have been already imputed, $h_{Y_1,n}$ and $h_{Y_0,n}$ must be set at the parts of f_{Y_1} and f_{Y_0} that were left out from the other imputed densities. The imputed $h_{Y_1,n}$ and $h_{Y_0,n}$ are drawn as the shadow areas in the top figure. Algebraically, the imputed $h_{Y_1,n}$ and $h_{Y_0,n}$ are expressed as

$$h_{Y_1,n} = f_{Y_1} - \sum_{t=a,c,n} h_{Y_1,t} = f_{Y_1} - \underline{f_{Y_1}} - [\min\{p_{Y_1}, q_{Y_1}\} - h_{Y_1,a}],$$

$$h_{Y_0,n} = \min\{p_{Y_0}, q_{Y_0}\}.$$

Since $h_{Y_1,n}$ must be non-negative, $h_{Y_1,n} \geq 0$ yields the inequality constraint for the possible choices of $h_{Y_1,a}$ (given the proposed f_{Y_1}) that has not been considered in Step 1,

$$h_{Y_1,a} \geq \max\left\{\underline{f_{Y_1}} + \min\{p_{Y_1}, q_{Y_1}\} - f_{Y_1}, 0\right\}, \tag{31}$$

where the maximum operator is needed on the right-hand side since $h_{Y_1,a}$ must be non-negative.

Step 5: As seen in Step 1, the integration of $h_{Y_1,a}$ has been constrained to be equal to $\int h_{Y_0,a} d\mu = 1 - \delta_{Y_0}$. So, the integration of (31) gives

$$1 - \delta_{Y_0} \geq \int \max\left\{\underline{f_{Y_1}} + \min\{p_{Y_1}, q_{Y_1}\} - f_{Y_1}, 0\right\} d\mu,$$

and this can be rewritten as

$$\begin{aligned} & 1 - \delta_{Y_0} \geq -\int \min\left\{f_{Y_1} - \underline{f_{Y_1}}, \min\{p_{Y_1}, q_{Y_1}\}\right\} d\mu + \lambda_{Y_1} \\ \iff \quad & \int \min\left\{f_{Y_1} - \underline{f_{Y_1}}, \min\{p_{Y_1}, q_{Y_1}\}\right\} d\mu \geq \underbrace{\lambda_{Y_1} - [1 - \delta_{Y_0}]}_{\text{the area of } d\&c}. \end{aligned} \tag{32}$$

This inequality is exactly the one appearing in the definition of $\mathcal{F}^*_{f_{Y_1}}(P, Q)$. If f_{Y_1} proposed in Step 1 meets this inequality, it implies that there exists a choice of $h_{Y_1,a} \geq 0$ based on which Step 2 through Step 4 guarantee the existence of feasible $(h_{Y_1,t}, h_{Y_0,t})$, $t = c, n, d$.

By the implication obtained in Step 5 of the above algorithm, we claim that $IR_{(f_{Y_1}, f_{Y_0})}(P, Q|RA) \supset \mathcal{F}^*_{f_{Y_1}}(P, Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P, Q)$. In fact, it is also possible to show that $IR_{(f_{Y_1}, f_{Y_0})}(P, Q|RA) \subset \mathcal{F}^*_{f_{Y_1}}(P, Q) \times \mathcal{F}^{env}_{f_{Y_0}}(P, Q)$ (see the proof of Proposition 3.2 in Appendix A). A symmetric argument works for the case of $1 - \delta_{Y_0} > \lambda_{Y_1}$. In this case, the identification region for f_{Y_0} becomes smaller than $\mathcal{F}^{env}_{f_{Y_0}}(P, Q)$.