Ernesto San Martín, Marcela Perticará, Inés M. Varas,
Kenzo Asahi, and Jorge González

4 The role of identifiability in empirical research

Abstract: This chapter discusses the general concepts of identification and partial identification of statistical models. We elucidate the identification restrictions to endow with meaning the parameters of interest of the fixed-effects one-parameter logistic model with guessing (1PL-G), a model used in educational measurement. We also review the restrictions for identifying the average treatment effect (ATE) in evaluating a policy or program. To address the fundamental problem of causal inference, we also present a partial identification analysis of the ATE. On the basis of the results, we emphasize the relevance of an identification analysis and the usefulness of considering a partial identification approach in causal inference.

Keywords: Partial identification, Causal inference, Self-selection bias, Finite sample space

> [. . .] it has been considered legitimate to use some of the *tools* developed in statistical theory *without* accepting the very *foundation* upon which statistical theory is built.
> (Haavelmo, 1944)

4.1 Introduction

Manski (2013a) emphasizes combining assumptions and data to draw meaningful conclusions in empirical research and policy analysis. The logic of empirical inference can accordingly be summarized as follows:

Acknowledgments: The authors were partially supported by the Millennium Nucleus on Intergenerational Mobility MOVI NCS2021072. Kenzo Asahi and Jorge González were also partially funded by Chile's Agencia Nacional de Investigación y Desarollo (ANID)'s FONDECYT Regular 1231755 and 1230968 respectively.

Ernesto San Martín, Inés M. Varas, Jorge González, Millennium Nucleus on Intergenerational Mobility: From Modelling to Policy (MOVI), Chile; Interdisciplinary Laboratory of Social Statistics LIES, Pontificia Universidad Católica de Chile; Faculty of Mathematics, Pontificia Universidad Católica de Chile
Marcela Perticará, Millennium Nucleus on Intergenerational Mobility: From Modelling to Policy (MOVI), Chile; Facultad de Administración y Economía, Universidad Diego Portales, Chile
Kenzo Asahi, Millennium Nucleus on Intergenerational Mobility: From Modelling to Policy (MOVI), Chile; Escuela de Gobierno, Pontificia Universidad Católica de Chile; Centre for Sustainable Urban Development (CEDEUS), Chile ORCID ID https://orcid.org/0000-0001-7838-4647

https://doi.org/10.1515/9783111036496-004

$$\text{Assumptions} + \text{Data} \Rightarrow \text{Conclusions}$$

Merely relying on data alone is insufficient for making conclusions; researchers need to make explicit assumptions linking the data to the population of interest.

It is essential to distinguish between these assumptions and the statistical hypotheses that can be falsified with data through the "null ritual" of hypothesis testing (Gigerenzer, 2004). The former are the so-called maintained assumptions, which makes explicit a fundamental difficulty of empirical research: holding fixed the available data, what assumptions can we maintain (Manski et al., 2021)? These considerations allow us to distinguish between data and (scientific) knowledge: "evidence is synonymous with data. Knowledge is the set of conclusions that one draws by combining evidence with assumptions about unobserved quantities" (Manski, 2013b).

Throughout this chapter, we pursue Manski's perspective by relating those assumptions on unobserved quantities with the concept of *parameter of interest*. To make this link explicit, we first establish what we understand by *population* and *structure*, and second, we clarify the importance of distinguishing and defining the identified parameters and the parameters of interest (Section 4.2).

After that, in Section 4.3, we present the identification analysis of models commonly used in educational measurement and psychometrics to represent the underlying response process of test takers to a set of stimuli/items (Lord, 1980; Hambleton & Swaminathan, 1985).

Section 4.4 provides a detailed discussion of the identification problem in causal inference, mainly focusing on the average treatment effect (ATE) in an observational study. We argue that causal inference often relies on strong assumptions and additional information; it is crucial to carefully interpret the estimated causal effects based on the underlying model and data limitations. We also explore the possibility of identifying the parameters of interest using logically weaker identification constraints.

In Section 4.5, we continue the discussion on the partial identification of the parameters of interest using a specific example of self-selection in the context of a leveling program of a Chilean public university. Some conclusions are gathered in Section 4.6.

4.2 The logic of empirical research

4.2.1 Population and structure

Following Koopmans & Reiersol (1950) and Hurwicz (1950), the identification problem arises from distinguishing between a *population* "in the sense of a [probability] distribution of observable variables" and a *structure* referring to the "investigator's ideas regarding the explanation or formation of the phenomena studied." In this way, Koopmans and Reiersol (1950) reformulate the *specification problem* originally made explicit by Fisher (1922). Instead of specifying the "mathematical form of the probability distribution of the population," one specifies a set of structures corresponding to the probability distribution of certain unobserved variables and a specific relationship between observed and unobserved variables.

This framework gives rise to a fundamental problem, specifically, the question of whether a single structure uniquely generates the probability distribution of the observations. It is possible to focus on observations, seeking empirical relationships likely to be caused by the presence and persistence of underlying structural relationships. "However, the direction of this deduction cannot be reversed –from empirical to the structural relationships– except possibly with the help of a theory which specifies the form of the structural relationships, the variables which enter into each, and any further details supported by prior observation or deduction therefrom" (Koopmans, 1949). When multiple structures are compatible with the empirical evidence, it becomes challenging to assess their validity objectively. This situation leads to undesirable implications. At the policy level, different courses of action may be justifiable based on the same empirical evidence. At the scientific level, different conclusions stemming from the same empirical evidence might be considered as equally valid "scientific knowledge." This situation is frustrating: "all you can do is judge the persuasiveness of the arguments offered. If you are persuaded by one social scientist more than by another, it is only because one is a more skilled advocate for his or her position" (Manski, 1995, p. 2).

The previous framework involves the specification of (i) the latent variable model, say $p^{\omega_1}(\theta)$, where ω_1 is the corresponding parameter, and (ii) the conditional model (i.e., the structural relationships between observed and latent variables) of the observed data given the latent ones, say $p^{\omega_2}(y \mid \theta)$, where ω_2 is the corresponding parameter. Both models induce a probability distribution of the observable variables, namely

$$p^{(\omega_1,\omega_2)}(y) = \int p^{\omega_2}(y \mid \theta)p^{\omega_1}(\theta)d\theta \tag{4.1}$$

The parameters ω_1 and ω_2 constitute the essential component of the structure underlying the observed variables. Once these parameters are identified, we obtain a unique structural explanation.

4.2.2 Identified parameters and parameters of interest

Although Koopmans and Reiersol's framework (1950) may seem, at first glance, applicable only to structural models with latent variables, it is possible to highlight from (4.1) the fundamental objective underlying such a framework, namely "to learn what conclusions can and cannot be drawn given specified combinations of assumptions and data" (Manski, 1995, p. 3).

In the first level of analysis, the attention can be focused on a description of a set of data – the population of interest. What we *can learn from the data* corresponds to its probability distribution: its specification “is not arbitrary but requires an understanding of the way in which the data are supposed to, or did in fact, originate” (Fisher, 1973, p. 8). Parameters always index such probability distribution: these parameters exhaustively describe the population of interest “in respect of all qualities under discussion” (Fisher, 1973). This initial level of analysis, accordingly, requires making explicit three components:

1. The set of observations, along with the events of interest. This is the so-called sample space, which corresponds to a measurable space $(M, \mathcal{M})$ corresponding to the statistical units’ labels and constitutes a finite set (Basu, 1977). Consequently, the σ-field $\mathcal{M}$ reduces to the class of all the subsets of M.
2. A probability distribution P^a that is defined on the sample space and indexed by a parameter a. We refer to this probability distribution as *sampling probability* emphasizing its relation to the observed data.
3. The set A of all logical ranges of the parameter a, namely the *parameter space.*

These three components define the *statistical model*, which can be compactly written as

$$\varepsilon = \{(M, \mathcal{M}), P^a : a \in A\} \tag{4.2}$$

and that corresponds to a family of sampling probability distributions defined on the observed data, following Koopmans & Reiersol (1950) and Hurwicz (1950), Gourieroux and Monfort (1995; Chapter 1); McCullagh (2002); Florens et al. (2007, Chapter 1). In ε, the parameter a is said to be identified if the mapping Φ from the parameter space A into the set $P(M, \mathcal{M})$ of sampling probabilities defined on $(M, \mathcal{M})$ is such that $\Phi(a) = P^a$ is injective.

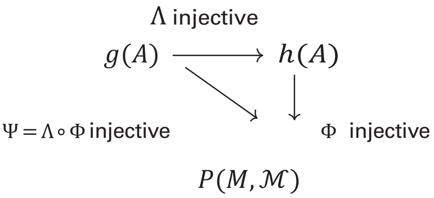
Remark 4.2.2 It should be noted that (4.2) involves the so-called parametric, nonparametric, and semi-parametric models. The parametric models are characterized by the fact that the parameter space A is (a subset of) a finite-dimensional vector space. For the nonparametric models, A corresponds to (a subset of) an infinite-dimensional vector space, for instance, a functional space or a space of probability distributions. Finally, for the semi-parametric model, the set A corresponds to a Cartesian product between a finite and infinite vector space.

It can be shown that a statistical model always involves an identified parameter; see Florens et al. (1985) and Florens et al. (1990, chapter 4). This suggests specifying the statistical model (4.2) in an identified way, namely, to index the sampling probabilities by the identified parameter, say $h(a)$, where h is a function defined from A into $h(A)$. Therefore, this identified parameterization fully captures the characteristics of the population of interest that can be gathered at this first level of analysis.

However, empirical research usually develops a second level of analysis. Faced with a set of observations, a researcher asks substantive questions. For example, in the second round of a presidential election, we only observe the proportion of votes in favor of one candidate or the other. If the difference in votes is minimal, a substantive question is whether there was electoral fraud. A political scientist would *like to learn from the data* whether there was electoral fraud. However, considering the observed data, such a political phenomenon is unobserved. The modeling challenge is, therefore, to express the electoral fraud as a parameter, which we will call a *parameter of interest*. The identification problem arises when the parameter of interest cannot be expressed as an injective function of the identified parameter. That is, when what we can learn from the data does not match what we want to know.

Technically speaking, a parameter of interest is a function of a, namely $g(a)$, where g is a function defined from A into $g(A)$; see Engle et al. (1983). Therefore, $g(a)$ is identified if the mapping Ψ from $g(A)$ into $P(M, \mathcal{M})$ such that $\Psi(g(a)) = P^a$ is injective; see LeCam & Schwartz (1960) and Mouchart and Oulhaj (2006).

A strategy of identification analysis consists of establishing in (4.2) an injection Λ between the parameter of interest $g(a)$ and the identified parameter $h(a)$. Since only the identified parameters capture properties of the population under study, this strategy shows what needs to be established so that *what is to be learned from the data* matches *what can be learned from it*: that the mapping Λ is injective. Technically speaking, this strategy is based on the following commutative diagram:



This strategy will be pursued in the examples discussed in the following sections.

4.3 Item response theory models: What do difficulty and guessing parameters mean?

Item response theory (IRT) models are typically used to analyze a set of binary data related to the reaction/response of test takers to a set of stimuli/items. More precisely, the investigator is confronted with a matrix of dimension $I \times J$, whose entries are 0's and 1's: I corresponds to the number of persons, whereas J corresponds to the number of stimuli or items; an entry 0 corresponds either to a negative reaction of a person to a stimulus or to an incorrect response of a person to an item; an entry 1 corresponds to a positive reaction of a person to a stimulus or to a correct response of a person to an item. The $I \times J$ matrix can accordingly be represented as

$$\begin{pmatrix} y_{11} & \cdots & y_{1J} \\ \vdots & \vdots & \vdots \\ y_{I1} & \cdots & y_{IJ} \end{pmatrix}, \tag{4.3}$$

where $y_{mj} \in \{0, 1\}$ for each $(m, j) \in \{1, \ldots, I\} \times \{1, \ldots, J\}$; here y_{mj} corresponds to the reaction or response of a person m to stimuli/item j. Therefore, the available information consists not only of the response patterns but also of a set of labels representing those people and a set of labels representing the stimuli/items. Thus, the available data are assembled into a set whose elements are ordered triplets (label of a person, label of a stimulus/item, response pattern):

$$\{(1, 1, y_{11}), (1, 2, y_{12}), \ldots, (1, J, y_{1J}), \ldots, (I, 1, y_{I1}), (I, 2, y_{I2}), \ldots, (I, J, y_{IJ})\}$$

for a similar way of describing the available data, see Bahadur et al. (2002).

To formalize the components of the statistical model, note that once the set of stimuli/items is defined, the observed data comes only from the people exposed to such a set of stimuli/items. Thus, the sample space is defined as the set of labels associated with each person, that is, $M = \{1, \ldots, I\}$. Taking into account that a function is fully characterized by its image space, it is possible to define a random variable (or a function) Y as follows:

$$Y: M \longrightarrow \{0, 1\}^J$$

such that for all $m \in M$,

$$Y(m) = (y_{m1}, \ldots, y_{mJ})$$

represents the person's response pattern. Note that the function Y can also be written as

$$Y(m) = (X(m, 1), \ldots, X(m, J))$$

where $X: M \longrightarrow \{0, 1\}$ and, consequently, $X(m, j) = (\pi_j \circ Y)(m, (1, \ldots, J))$, with π_j defined as $\pi_j(1, \ldots, j, \ldots, J) = j$ (Itô, 1984). It is assumed that the random variables $\{X(m, j): m \in M, j \in \{1, \ldots, J\}\}$ are mutually independent, and that each $X(m, j)$ is distributed according to a Bernoulli distribution with parameter p_{mj}. To simplify notation, we will denote by $X_j(m)$ the random variable $X(m, j)$ for all $(m, j) \in M \times \{1, \ldots, J\}$.

The reader may wonder what is the advantage in specifying the response variable Y in this way compared to the standard form in the IRT model literature (Lord & Novick, 1968; De Boeck & Wilson, 2004). There are at least three advantages:

1. The sample space is made explicit considering no elements related to the items, thus emphasizing the uniqueness of the available data (Fisher, 1955). This establishes the limits of statistical inference, which, in principle, must be reduced to the sample space.

2. The behavior of each person exposed to each stimulus/item is described by *one and only one* random variable. This contrasts with Lord and Novick's representation (1968, Chapter 2), in which one random variable is defined for each pair person-stimulus/item.
3. The definition of a unique random variable for a person's response pattern clarifies that people are distinguished whenever they have different response patterns, that is, when people belong to different elements on the equivalence class defined by the random variable Y on M.

4.3.1 A fixed-effects logistic model

To demonstrate the significance of the question that introduces this section, let us briefly examine the problem of parameter interpretability in a fixed-effects model.

Following Rasch (1960), the substantive problem is comparing persons using their responses to stimuli/elements. This problem led Rasch to introduce two parameters, one representing a specific characteristic of a person, say $\epsilon_m \in \mathbb{R}_+$, and the other representing a particular characteristic of the stimulus/item, say $\eta_j \in \mathbb{R}_+$, such that

$$p_{mj} = P(X_j(m) = 1) = F\left(\frac{\epsilon_m}{\eta_j}\right), \qquad m \in M, j = 1, \ldots, J \tag{4.4}$$

where $F(x) = \mathrm{x}/[1 + \mathrm{x}]$ for $x \in \mathbb{R}_+$. It should be remarked that the parameters of interest are $\{(\epsilon_m, \eta_j) : (m,j) \in M \times \{1, \ldots, J\}\}$, whereas the identified parameters are $\{p_{mj} : (m,j) \in M \times \{1, \ldots, J\}\}$.

The substantive problem can be solved once the meaning of ϵ_m and η_j is made explicit. Since Lord and Novick (1968, chapter 17) (see also De Boeck and Wilson, 2004; Van der linden and Hambleton, 1997; Baker and Kim, 2004), the standard psychometric literature has made explicit the meaning of the parameters of interest using the item characteristic curves, which correspond to the conditional probability of the observable variable $X_j(m)$ given a latent variable (in this case, what they call the person's ability). However, because the parameters correspond to the characteristics of the population under study, their meaning must be established concerning the statistical model. In passing, this confusion in the psychometric literature is (almost) the same as the one seen in the econometric literature, where models with fixed and random effects are confused and even compared; for some examples, see Longford (2012); Castellano et al. (2014); Clarke et al. (2015); Bell et al. (2019).

As discussed in the previous section, the strategy to identify the parameters of interest consists of establishing a one-to-one relationship between them and the identified parameters. Such a relationship follows noticing in (4.4) that, for every pair (m, j),

$$\frac{\epsilon_m}{\eta_j} = F^{-1}(p_{mj})$$

If there are at least two items and one person, it follows that

$$\epsilon_m = \eta_1 \cdot F^{-1}(p_{m1}) \text{ for all } m \in M, \quad \eta_j = \eta_1 \cdot \frac{F^{-1}(p_{m1})}{F^{-1}(p_{mj})} \text{ for all } j = 2, \ldots, J \tag{4.5}$$

Thus, ϵ_m's and η_j's depend on η_1 and, consequently, a necessary and sufficient condition for identifying the parameters of interest is to fix η_1. If $\eta_1 = 1$, then the characteristic ϵ_m of a person m corresponds to the betting odd of a correct answer to the standard item 1, and the characteristic η_j of item j could be interpreted as an odd ratio between item 1 and item j for each person m; for details, see Rasch (1966) and San Martín et al. (2009). It is important to emphasize that these interpretations of the parameters of interest are not based on psychological or educational considerations.

4.3.2 A model with a guessing parameter

Considering the same data (4.3), researchers have wondered whether a person can answer an item or a stimulus by "guessing." This question is more pressing in educational measurement, particularly when a standardized test has no consequences for individuals. In Chapter 17 of Lord and Novick (1968), Birnbaum introduced a latent trait model that allows random guessing so that "subjects of very low ability will sometimes give correct responses to multiple-choice items, just by chance." Birnbaum emphasizes the substantive side of the problem:

> A highly schematised psychological hypothesis has suggested one model for such items. This model assumes that if an examinee has ability θ, then the probability that he will know the correct answer is given by a normal ogive function $\Phi\left[a_g(\theta - b_g)\right]$ [here, Φ is the cumulative distribution function of a standard normal distribution, whereas a_g and b_g are item parameters] [. . .]; it further assumes that if he does not know it he will guess, and, with probability c_g, will guess correctly. It follows from these assumptions that [. . .] the probability of a correct response is the item characteristic curve
>
> $$Q_g(\theta) = c_g + (1 - c_g)\Phi\left[a_g(\theta - b_g)\right]$$
>
> The psychological hypothesis implicit here has been mentioned primarily to point up a mathematical feature of this form; the empirical validity of this form is not dependent on this psychological hypothesis. (p. 404)

For Birnbaum, "answering an item correctly by chance" is formulated using a probability that depends only on the item and not on the person's characteristics. However, when interpreting this probability, Birnbaum does not do so concerning the statistical model but only based on a conditional, unobservable model.

To understand the meaning of a guessing parameter, we will focus on a slightly simplified model called the 1PL-G model (Weitzman, 1996; San Martin et al., 2006). The 1PL-G fixed-effects model is specified as follows:

$$P\big(X_j(m)=1\big)=c_j+\big(1-c_j\big)F(\theta_m-\beta_j)\quad \text{for all } (m,j)\in M\times\{1,\ldots,J\} \tag{4.6}$$

where θ_m is a person parameter, β_j and c_j are known as the difficulty and guessing parameter related to the item, respectively; $F(x)=\exp(x)/[1+\exp(x)]$ with $x\in\mathbb{R}$ and $(c_j,\beta_j,\theta_m)\in[0,1]\times\mathbb{R}\times\mathbb{R}$. It is also assumed that $\{X_j(m):(m,j)\in M\times\{1,\ldots,J\}\}$ are mutually independent.

Note that (4.6) can be rewritten as

$$q_{mj}=P\big(X_j(m)=0\big)=\delta_j G(\theta-\beta_j)\quad \text{for all } (m,j)\in M\times\{1,\ldots,J\} \tag{4.7}$$

where $\delta_j=1-c_j\in[0,1]$ and G is a function such that $G(x)+F(x)=1$ for all $x\in\mathbb{R}$.

Assuming that there are at least two persons and two items, (4.7) implies the following equations:

$$\theta_m=G^{-1}\left(\frac{q_{m1}}{\delta_1}\right)+\beta_1 \tag{4.8}$$

$$\beta_j=G^{-1}\left(\frac{q_{m1}}{\delta_1}\right)-G^{-1}\left(\frac{q_{mj}}{\delta_j}\right)+\beta_1 \tag{4.9}$$

$$G^{-1}\left(\frac{q_{1j}}{\delta_{\mathrm{j}}}\right)-G^{-1}\left(\frac{q_{2j}}{\delta_j}\right)=G^{-1}\left(\frac{q_{11}}{\delta_1}\right)-G^{-1}\left(\frac{q_{2j}}{\delta_j}\right) \tag{4.10}$$

Thus, $\theta_m=\theta_m(\beta_1,\delta_1)$, $\beta_j=(\beta_1,\delta_1)$, and $\delta_j=\delta_j(\delta_1)$. Therefore, a necessary and sufficient condition for identifying the parameters of interest is to fix the item parameters of an item, namely $(\beta_1,\delta_1)=(0,1)$ or, equivalently, $(\beta_1,c_1)=(0,0)$. This restriction reveals that there is no other way to know about the guessing parameter of the items than when there is at least one item with a guessing parameter equal to zero. Details about these results can be found in Appendix A.

These equations allow us to interpret the parameters of interest of the 1PL-G fixed-effect model:

1. Regarding the person parameter θ_m, it can be verified that its meaning is the same as that in an identified Rasch fixed-effects model. Moreover,

   $$\theta_m>\theta_l\Leftrightarrow P(X_1(l)=1)<P(X_1(m)=1),$$

   which provides empirical insight.
2. The item parameter β_j does not have the same meaning as in the identified Rasch fixed-effect model, which implies that comparing these parameters from one model to another is incorrect. Moreover,

$$\beta_j > \beta_k \Leftrightarrow \frac{q_{mj}}{\delta_j} > \frac{q_{mk}}{\delta_k}.$$

Thus, the sentence *item j* is *more difficult than item k* needs to be understood in the following terms: the probability of answering item j incorrectly is greater than the probability of answering item k incorrectly once both probabilities are normalized by δ_j and δ_k, respectively.

3. The previous inequality shows the role of the so-called nonguessing parameter δ_j: a normalization factor to ensure correct comparisons between items and persons. As a matter of fact, equality (4.10) provides us with an interpretation for the parameter δ_j: the difference in answering incorrectly the item standard 1 for two persons $(m = 1, 2)$ must be the same as the difference of these two persons in answering incorrectly any other item provided that the probabilities of answering incorrectly are normalized by the parameter δ_j. In other words, even if an item "invites" to be answered by chance, the differences between persons' characteristics will always be based on their performance in an item that "does not invite" to be answered by chance.

This identification analysis limits the empirical applicability of the 1PL-G model as one can compare the characteristics of stimulus/items and persons only after arbitrarily deciding which stimuli/item will be assumed to have a parameter $c_j = 0$. Assuming that the conclusions will change dramatically if that item is changed seems plausible.

4.4 Identification problems in causal inference

The evaluation of the impact of policy interventions, the effect of a leveling program on students' performance, and the effectiveness of a disease drug are common topics of interest in economic, education, and health-related fields, respectively. In all these contexts, the interest is to recover a treatment effect by comparing the mean outcome difference between sample units under treatment and sample units under the status quo: this corresponds to the so-called ATE; see, for example, Rubin (1974, 1978). Despite the field of application, there is an inherent missing data problem in all treatment effect analyses: each unit in our sample experiences only one of the statuses (treatment/status quo). Thus, "It is a fundamental problem of empirical inference that can be addressed only by making assumptions that relate observed and counterfactual outcomes" (Manski, 2013a, p. 53).

Different approaches have been developed to overcome the unobservability of counterfactual outcomes. In what follows, we revisit the identification of the ATE in the context of an observational study. After that, we will compare this analysis with a partial identification approach.

4.4.1 Point identification of the ATE

Let us begin by explicitly defining the sample space M: it consists of all labels of the sample units (typically persons), and it is, therefore, a finite set. Consequently, the class $\mathcal{M}$ consists of all the subsets of M. Let $\mathcal{T}$ be the set of mutually exclusive treatment indexes. Under these considerations, we define the outcome as follows:

$$Y: M \times \mathcal{T} \to \{0,1\}$$

$$(m,t) \to Y(m,t) \in \{0,1\}$$

where $Y(m,t)$ is the outcome experienced by person m when she/he is exposed to treatment t. Thus, the event $\{m \in M: Y(m,t) = 1\}$ includes *all the persons* in M who have experienced a "positive" outcome when they are exposed to treatment t. The complement of this event represents *all the persons* in M exposed to treatment t and who have experienced a "negative" outcome.

Additionally, we define a random variable (or a function) Z as follows:

$$Z: M \to \mathcal{T}$$

$$m \to Z(m) \in \mathcal{T}$$

where $Z(m)$ indicates the treatment received by person m. Thus, the event $\{m \in M: Z(m) = t\}$ represents all the persons in M exposed to treatment t.

In an observational study, the identified parameters of the statistical model are the following:

(i) The proportion of persons who experienced a "positive" outcome when exposed to treatment t, that is,

$$P\left(\{m \in M: Y(m,t_k) = 1\} \mid \{m \in M: Z(m) = t_j\}\right) \text{ if } t_j = t_k$$

Note that if $t_j \neq t_k$, this probability is not identified because no comparable persons are exposed to treatments t_k and t_j. Therefore, it is impossible to characterize what would have been the outcome of persons exposed to a treatment different from the one they received. This is known in the econometric literature as the common support problem (or assumption) (Lechner, 2008; Blundell & Costa Dias, 2009).

(ii) The proportion of people who received treatment t, namely

$$P(\{m \in M: Z(m) = t\}) \quad \text{for each } t \in \mathcal{T}$$

In what follows, we will analyze the identification problem for two exclusive treatments: innovation, labeled by 1, and status quo, labeled by 0; in this case, $\mathcal{T} = \{0,1\}$. Let us also simplify the notation as follows:

$$\{Y(t) = y\} \doteq \{m \in M: Y(m,t) = y\} \qquad \text{for all } y \in \{0,1\},\ t \in \mathcal{T}$$

$$\{Z = t\} \doteq \{m \in M: Z(m) = t\} \qquad \text{for all } t \in \mathcal{T}$$

From a policymaker's perspective, the interest relies on comparing a "positive" outcome when all persons are exposed to the innovation and the "positive" outcomes when all persons are exposed to the status quo. This is precisely the ATE, namely

$$\mathrm{ATE} = P(Y(1)=1) - P(Y(0)=1) \tag{4.11}$$

To relate the parameters of interest $P(Y(1)=1)$ and $P(Y(0)=1)$ with the identified parameters $P(Y(1)=1 \mid Z=1)$, $P(Y(0)=1 \mid Z=0)$, and $P(Z=1)$, we use the law of total probability (Kolmogorov, 1950):

$$P(Y(1)=1) = P(Y(1)=1 \mid Z=1)P(Z=1) + P(Y(1)=1 \mid Z=0)P(Z=0) \tag{4.12}$$

$$P(Y(0)=1) = P(Y(0)=1 \mid Z=1)P(Z=1) + P(Y(0)=1 \mid Z=0)P(Z=0) \tag{4.13}$$

As we noticed before, $P(Y(1)=1 \mid Z=0)$ and $P(Y(0)=1 \mid Z=1)$ are not identified. Therefore, it is impossible to establish an injection between the parameters of interest $P(Y(1)=1)$ and $P(Y(0)=1)$ and the identified parameters $P(Y(1)=1 \mid Z=1)$, $P(Y(0)=1 \mid Z=0)$, and $P(Z=1)$. In the parlance of causal inference, this is due to the fundamental problem of causal inference:

> It is impossible *to observe* the value of $Y_t(u)$ and $Y_c(u)$ on the same unit and, therefore, it is impossible *to observe* the effect of t on $uY_t(u)tY_c(u)c$. (Holland, 1986, p. 947)

It is typically argued that those parameters of interest can be identified if additional information is collected. Such information is contained in a (vector of) covariate(s) X, namely a random variable (or a function) $X\colon M \to \mathcal{X}$, where $\mathcal{X}$ the image space of X, such that $X(m) \in \mathcal{X}$ is associated with each person $m \in M$. Once X is fixed, the parameters of interest are $P(Y(1)=1 \mid X)$, $P(Y(0)=1 \mid X)$, whereas the identified parameters are $P(Y(1)=1 \mid X, Z=1)$, $P(Y(0)=1 \mid X, Z=0)$, and $P(Z=1 \mid X)$. These parameters are related through the law of total probability, namely

$$\begin{aligned} P(Y(1)=1 \mid X) &= P(Y(1)=1 \mid X, Z=1)P(Z=1 \mid X) \\ &\quad + P(Y(1)=1 \mid X, Z=0)P(Z=0 \mid X) \end{aligned} \tag{4.14}$$

$$\begin{aligned} P(Y(0)=1 \mid X) &= P(Y(0)=1 \mid X, Z=1)P(Z=1 \mid X) \\ &\quad + P(Y(0)=1 \mid X, Z=0)P(Z=0 \mid X). \end{aligned} \tag{4.15}$$

The problem of nonidentifiability is analogous to the one mentioned above: $P(Y(1)= 1 \mid X, Z=0)$ and $P(Y(0)=1 \mid X, Z=1)$ are not identified, and consequently, neither are the parameters of interest. The additional information captured in X is used to get identifiability through the so-called strong ignorability conditions (Rosenmbaum & Rubin, 1983), namely

$$P(Y(1)=1 \mid X, Z=0) = P(Y(1)=1 \mid X, Z=1) \tag{4.16}$$

$$P(Y(0)=1 \mid X, Z=1) = P(Y(0)=1 \mid X, Z=0). \tag{4.17}$$

By replacing (4.16)–(4.17) in (4.14)–(4.15), respectively, we get the identifiability of the parameters of interest, namely

$$P(Y(1)=1 \mid X)=P(Y(1)=1 \mid X,Z=1)$$

$$P(Y(0)=1 \mid X)=P(Y(0)=1 \mid X,Z=0)$$

and consequently the ATE as a function of X, namely

$$\text{ATE}(X) \doteq P(Y(1)=1 \mid X,Z=1) - P(Y(0)=1 \mid X)$$

represents a parameter of interest relative to all persons in M with characteristics $X=x$. Once we consider the identification restrictions (4.16) and (4.17), we can point-identify the ATE:

$$ATE(X)=P(Y(1)=1 \mid X,Z=1) - P(Y(0)=1 \mid X,Z=0). \tag{4.18}$$

Note at this point that the restrictions allow us to identify the parameter in those elements in M with characteristics $\{\mathrm{X}=\mathrm{x}\}$ belonging to two mutually exclusive groups, namely $Z^{-1}\{0\}$ and $Z^{-1}\{1\}$, which is a partition of M (here Z^{-1} denotes the preimage of the function Z). It should be remarked that this ATE is a function of the (vector of) covariate(s) X. Using the law of total probability, it is possible to obtain a *marginal* ATE:

$$\text{ATE}(X) = \sum_{x\in\mathcal{X}} \text{ATE}(x)P(X=x)$$

Note that, here, we are restricting all covariates to be discrete.[1]

These results deserve some comments:

1. The strong ignorability conditions (4.16) and (4.17) are *identification restrictions* rather than a component of the model specification. Under this constraint, what a researcher *wants to learn from the data* coincides, for those persons with characteristics $\{m \in M: X(m)=x\}$, with *what can be learned from such data.*
2. Moreover, the ignorability conditions are equivalent to the following conditions:

$$Y(0) \perp Z \mid X, \quad Y(1) \perp Z \mid X;$$

for a discussion, see San Martín and González (2022). Using the symmetry property of conditional independence (Florens et al., 1990, chapter 2), it follows that, for $z \in \{0,1\}$,

$$P(Z=z \mid X, Y(1)) = P(Z=z \mid X), \quad P(Z=z \mid X, Y(0)) = P(Z=z \mid X)$$

As the reader can recognize, these probabilities are the ones used to estimate the propensity scores and perform the matching procedure. It should be noted that

---

1 It is relevant to recall rigorous definition of an absolutely continuous random variable: X is an absolutely continuous random variable if and only if $P(X=x)=0$ for all $x \in \mathcal{X}$. It is (almost) hard to find this type of random variable in concrete applications.

the strong ignorability conditions are well defined once a (vector of) covariate(s) X has been chosen. Consequently, the propensity score procedure is not a procedure to select covariates based on goodness-of-fit indices such as the ones used in confirmatory factor analysis –a widely used method in psychometrics (Brown, 2015): it is only a procedure to perform the matching exercise.

3. The strong ignorability conditions (4.16) and (4.17) do not solve the fundamental problem of causal inference because, according to Holland (1986), such a problem is due to the impossibility to observe *the same statistical unit* exposed to both the innovation $(t=1)$ and the status quo $(t=0)$. For instance, condition (4.16) is an equality between two different mutually exclusive groups of statistical units: those exposed to the innovation, namely $\{m \in M: Z(m)=1\}$ and those exposed to the status quo, namely $\{m \in M: Z(m)=0\}$.

4.4.2 Partial identification of the ATE

Admitting the challenging nature of justifying the strong ignorability condition and the interpretation of the ATE, one might explore the possibility of identifying the parameters of interest using logically weaker identification constraints. Answering this question means moving from point identification of the parameters of interest to partial identification:

> A parameter in a probabilistic model is partially identified if the sampling process and maintained assumptions reveal that the object lies in a set, called the identification region or identified set, that is smaller than the logical range of possibilities but larger than a single point. Sample estimates of partially identified objects generically are set-valued. (Manski, Sanstad, DeCanio, 2021)

In what follows we develop a partial identification approach to partially identify both $P(Y(1)=1 \mid X)$ and $P(Y(0)=1 \mid X)$. We accordingly introduce an identification restriction leading to solving the fundamental problem of causal inference: the conditional probabilities on which these restrictions are based will be conditional on the same statistical units.

Let us explore the impact of the following identification restriction on the identifiability of the parameters of interest:

$$P\ (Y(1)=1 \mid X,\ Z=1) \geq P\ (Y(0)=1 \mid X,\ Z=1) \tag{4.19}$$

$$P\ (Y(1)=1 \mid X,\ Z=0) \geq P\ (Y(0)=1 \mid X,\ Z=0) \tag{4.20}$$

These conditions intend to represent an *optimistic policymaker*. Condition (4.19) implies that individuals exposed to the innovation are less likely to encounter a "positive" outcome under the status quo than when they are genuinely subjected to the intervention. Meanwhile, condition (4.20) means that for those persons exposed to the status quo, it is more probable to experience a "positive" outcome if exposed to the innovation than when exposed to the status quo. In other words, the treatment is better than the status quo for both persons under the innovation and the status quo.

Combining (4.20) and (4.14), we get an interval for all possible values of $P(Y(1)=1 \mid X)$, namely

$$\begin{aligned} &P(Y(1)=1, Z=1 \mid X) + P(Y(0)=1, Z=0 \mid X) \leq \\ &\quad \leq P(Y(1)=1 \mid X) \leq \\ &\quad \leq P(Y(1)=1, Z=1 \mid X) + P(Y(0)=1, Z=0 \mid X) + P(Y(0)=0, Z=0 \mid X) \end{aligned} \tag{4.21}$$

Similarly, combining (4.19) and (4.15), we get an interval for all possible values of $P(Y(0)=1 \mid X)$, namely

$$\begin{aligned} &P(Y(0)=1, Z=0 \mid X) \leq P(Y(0)=1 \mid X) \leq \\ &P(Y(1)=1, Z=1 \mid X) + P(Y(0)=1, Z=0 \mid X) \end{aligned} \tag{4.22}$$

A proof of this result can be found in Appendix B.

To interpret these partial identification intervals, note that

$$\begin{aligned} &P(Y(1), Z=1 \mid X) + P(Y(0), Z=0 \mid X) = \\ &\qquad P(\{m \in M: Y(m,1)=1, Z(m)=1\} \cup \{m \in M: Y(m,0)=1, Z(m)=0\} \mid X) \end{aligned}$$

which corresponds to the proportion of persons who experience a "positive" outcome regardless of whether they are exposed to the intervention or to the status quo. Therefore,

4. Under the optimistic policymaker assumption, the proportion of persons in M who would experience a "positive" outcome if all of them were exposed to the innovation would improve the current proportion of "positive" outcomes. Conversely, the proportion of persons in M who would experience a "positive" outcome if all of them were exposed to the status quo would decrease such proportion.
5. Moreover, the partial identification interval of the ATE(X) is given by

$$\text{ATE}(X) \in [0, P(Y(1)=1, Z=1 \mid X) + P(Y(0)=1, Z=0 \mid X)] \tag{4.23}$$

   that is, it is always positive, and its upper bound corresponds to the current proportion of persons experiencing a "positive" outcome regardless of whether they are exposed to the intervention or the status quo!
6. It should be noted that the point identified $\text{ATE}(X)$ under ignorability conditions (see equality (4.18)) not necessarily is a plausible value of a partially identified $\text{ATE}(X)$ under the optimistic policymaker assumption. As a matter of fact, the point identified $\text{ATE}(X)$ belongs to the partial identification interval (4.23) if and only if

$$\frac{P(Y(1)=1, Z=1 \mid X)}{P(Y(0)=1, Z=0 \mid X)} < \frac{P(Z=1 \mid X)}{P(Z=0 \mid X)} \left(\frac{1}{P(Z=0)} + 1 \right)$$

Consequently, the conditions of strong ignorability cannot necessarily be interpreted in line with the optimistic policymaker assumption.

7. Indeed, the above conclusions are tautological with the optimistic hypothesis of policymakers. Why is, then, this type of analysis relevant? To answer this, it is necessary to provide an interpretation of the parameters of interest, which requires making explicit the role of the sample space M. Thus, the partial identification (4.21) should be interpreted in the following terms:

   > If all the persons in M (with characteristics $\{X = x\}$) had been exposed to the innovation, then the proportion of those who experienced a "positive" outcome would have been at least equal to the actual proportion of "positive" outcomes, regardless of whether the persons were under the innovation or the status quo. Moreover, this proportion could have increased by a proportion equivalent to the proportion of persons in M (with characteristics $\{X = x\}$) who are under the status quo and who have a "negative" outcome (i.e., $P(Y(0) = 0, Z = 0 \mid X)$ in (4.21)).

   A similar interpretation can be made for the partial identification interval (4.22).

8. This interpretation emphasizes that the evaluation of the policy or program only concerns the population in M, so a policy evaluation should not be confused with a prediction of what might happen if the innovation is implemented. If we want to forecast outcomes for another population $\widetilde{M} \neq M$, we would be facing a new identification problem, which is beyond the scope of this chapter.

9. Despite this, it is essential to think about the usefulness of a policy evaluation such as the one above. One possible answer is to consider the concept of *inductive behavior* introduced by Neyman (1938) and developed in Neyman (1950):

   > With many phenomena certain permanencies appear quite stable. This created the habit of regulating our actions in regard to some observed events by referring to the permanencies which at the particular moment seem to be established. This is what we call inductive behaviour. (Neyman, 1950, p. 1)

The interpretation of the partial identification intervals shows what the situation of population M would have been if all of them had been exposed to the innovation or the status quo. But such a situation is a logical consequence of the optimistic policymaker assumption. Consequently, the evaluation of a policy or program aims to persuade the policymaker to *act following this optimistic view*. This is in line with Neyman's inductive behavior concept:

> Nous pouvons *savoir* que la loi mathématique des grands nombres subsiste dans les cas précisés par les conditions des théorèmes qu'on a démontrés. Nous pouvons aussi *savoir* que la loi empirique des grands nombre s'était réaisée dans telles expériences déjà effectueées. Mais nous ne pouvons que *croire* qu'elle continuera à être réalisée dans les expériences futures.
> [. . .]
> Mais se décider à 'affirmer' ne veut pas dire 'savoir' ni même 'croire'. C'est un acte de voloté précédé par quelques expériences et quelqes raisonnements déductifs, tout à fait comme de s'assurer sur la vie, que l'on fait, même si l'on espère vivre longetemps. (Neyman, 1938, p. 352)

In other words, evaluating a policy or a program is intended to modify the policymaker's willingness to act. But let's be clear: it is an evaluation that assumes certain invariance once the population under study is changed.

4.5 How to model self-selection?

Let us continue with the previous discussion on the partial identification of the parameters of interest $P(Y(0) = 1 \mid X)$ and $P(Y(1) = 1 \mid X)$ in the context of a leveling program of a Chilean public university.

4.5.1 Context

In this Chilean public university, students are selected either by a national admission process (considering high school background and scores from standardized tests) or through an inclusive access program. The Inclusive Access, Equity and Permanence Program (PAIEP, by its Spanish name) is a program developed by the university to support students during their first year at the university. Among other activities, the program considers tutorial classes in both academic and socio-educational topics. Although all students enrolled in the university are invited to participate in this program, the targeted group is the one enrolled through the inclusive access program. Program activities take place throughout the year, although students can stop participating at any time during the year. Furthermore, once the students know their grades for the first semester, the students decide whether to continue in the program during the second semester.

One of the response variables of interest to the program is the grade point average (GPA) at the end of each semester of the first year of university. In addition, the program considers that a student has been intervened if he/she attends at least 10 tutorial sessions per semester.

Once the first semester has ended, students who have participated in the leveling program may choose whether to continue in the program during the second semester. We will now outline how to evaluate the students' decision to continue or not in the program during the second semester. More specifically, for illustrative purposes, we will focus our attention on those students who attended the leveling program during the first semester and obtained a GPA score at least equal to 4.0 (which in Chile is the minimum score to pass a course): we will observe their decision to continue or not in the program during the second semester. Thus, using the notation of the previous section, the labels of these students are gathered in the set M.

4.5.2 Parameters of the problem

For the students in M, we consider the following random variables (always using the notation previously introduced):

– $Z(m) = 1$ if the student $m \in M$ participated in the leveling program during the second semester, and $Z(m) = 1$ if not.
– $Y(m,1) = 1$ if a student $m \in M$ who decides to continue in the leveling program obtains a GPA at least equal to 4.0, and $Y(m,1) = 0$ if he/she obtains a GPA smaller than 4.0.
– $Y(m,0) = 1$ if a student $m \in M$ who decides not to continue in the leveling program obtains a GPA at least equal to 4.0, and $Y(m,0) = 0$ if he/she obtains a GPA smaller than 4.0.
– The covariates X include eventual additional information at the student level.

Thus, the identified parameters are the following:
(a) The proportion of students (with characteristics $\{X = x\}$) who participated in the leveling program on the second semester, namely $P(Z = 1 \mid X)$.
(b) The proportion of students who actually decided to continue in the leveling program and obtained a second semester GPA at least equal to 4.0, namely $P(Y(1) = 1 \mid X, Z = 1)$.
(c) The proportion of students who actually decided not to continue in the leveling program and obtained a second semester GPA at least equal to 4.0, namely $P(Y(0) = 1 \mid X, Z = 0)$.

The parameters of interest are $P(Y(0) = 1 \mid X)$ and $P(Y(1) = 1 \mid X)$, which are unidentified because, as discussed above, $P(Y(1) = 1 \mid X, Z = 0)$ and $P(Y(0) = 1 \mid X, Z = 1)$ are unidentified.

4.5.3 Partial identification analysis

Let us assume that the students' decision to continue or not to continue in the program is "rational" in the sense that if a student decides to continue (relative to the decision to not continue), he/she does so because he/she believes that if he/she does not continue (resp. continues) he/she will have a worse outcome than if he/she continues (resp. does not continue). This assumption can be expressed as follows:

$$P(Y(1) = 1 \mid X, Z = 0) \leq P(Y(0) = 1 \mid X, Z = 0) \tag{4.24}$$

$$P(Y(0) = 1 \mid X, Z = 1) \leq P(Y(1) = 1 \mid X, Z = 1) \tag{4.25}$$

Condition (4.24) implies that among students who actually opt not to continue in the leveling program, if they had chosen to continue, they would have had a lower likelihood of obtaining a GPA $\geq$ 4.0 compared to not continuing in the program. Similarly, condition (4.25) suggests that among students who actually decide to continue in the leveling program, if they had chosen not to continue, they would have had a lower likelihood of achieving a GPA $\geq$ 4.0 compared to continuing in the program.

These identification restrictions imply that

$$P(Y(1)=1,Z=1 \mid X) \leq P(Y(1)=1 \mid X) \leq P(\{m \in M: Y(m,1)=1, Z(m)=1\} \cup \{m \in M: Y(m,0)=1, Z(m)=0\} \mid X) \tag{4.26}$$

$$P(Y(0)=1,Z=0 \mid X) \leq P(Y(0)=1 \mid X) \leq P(\{m \in M: Y(m,1)=1, Z(m)=1\} \cup \{m \in M: Y(m,0)=1, Z(m)=0\} \mid X) \tag{4.27}$$

These partial identification intervals deserve the following comments:

- If we are willing to assume that student behavior is "rational" according to assumptions (4.24) and (4.25), then the actual proportion of students obtaining a GPA $\geq$ 4.0 at the end of the second semester will deteriorate if either *all students* in M decide to continue in the program or if they all decide not to continue.
- Considering that the choice is voluntary, what is less bad, to continue or not in the program? One way to respond is to choose the least adverse scenario, which means comparing the lower bounds of the partial identification intervals. That is, we have to compare $P(Y(0)=1,Z=0 \mid X)$ and $P(Y(1)=1,Z=1 \mid X)$. Assessing whether it is better for students to decide to continue in the program, thus, boils down to comparing the current proportion of students who do not decide to continue and have a GPA $\geq$ 4.0, with the current proportion of students who decide to continue and have a GPA $\geq$ 4.0.

These conclusions show, on the one hand, that believing in a "rational" behavior does not ensure an improvement of the current situation and, on the other hand, they show that the conclusion depends on a given population M and therefore cannot be automatically extrapolated to a different population.

4.5.4 Illustration

We illustrate the previous results with available information regarding the participation of students in the leveling program implemented by a Chilean public university. There are a total of 214 students who attended the leveling program during the first semester and obtained a GPA score at least equal to 4.0. This number represents the cardinality of the sample space M described before.

Because the focus of the PAIEP program is on students selected by the inclusive access program, as an additional characteristic of interest for the students, we define the random variable X – using the notation previously introduced – by $X(m)=1$ if student $m \in M$ was selected by the inclusive access program; $X(m) = 0$ if the student was selected by the national admission process. Thus, for the students selected by the inclusive program, the identified parameters from the available data are:

(i) The proportion of students who participated in the leveling program in the second semester, that is,

$$P(Z = 1 \mid X = 1) = \frac{153}{201} = 0.7612$$

(ii) The proportion of students who continue in the leveling program and obtained a GPA at least to 4.0 in the second semester, that is,

$$P(Y = 1 \mid X = 1, Z = 1) = \frac{120}{153} = 0.7853$$

(iii) The proportion of students who decided not to continue in the leveling program and obtained a GPA at least to 4.0 in the second semester, that is,

$$P(Y(0) = 1 \mid X = 1, Z = 0) = \frac{33}{48} = 0.6875$$

However, the parameters of interest are $P(Y(1) = 1 \mid X = 1)$ and $P(Y(0) = 1 \mid X = 0)$, which represent the proportion of students in M selected by the inclusive program who obtain a GPA at least 4.0 at the end of the second semester, when they decide to continue and not continue in the leveling program, respectively. Under the identified restrictions (4.24) and (4.25), the evaluation of the partial identification intervals (4.26) and (4.27) are as follows:

– $0.5970 \leq P(Y(1) = 1 \mid X = 1) \leq 0.7612$
– $0.1642 \leq P(Y(0) = 1 \mid X = 1) \leq 0.7612$

Thus, under a rationality decision assumption, when all students choose to continue or not continue in the leveling program, the proportion of students who obtain a GPA at least 4.0 on the second semester will never be greater than the observed proportion (0.7612). Moreover, considering what was discussed about these intervals in the previous section, and given that the lower bound of $P(Y(1) = 1 \mid X = 1)$ is greater than the same quantity for $P(Y(0) = 1 \mid X = 0)$, it would be better for students to continue in the program. We highlight at this point that these results are valid only for the observed students belonging to M and that were selected by the inclusive access program.

4.6 Conclusions and discussion

In this chapter, we have delved into the critical concept of *identifiability* across two fields, namely, econometrics and psychometrics. We have highlighted the significance of identifiability analysis not only in the specification of statistical models but also in conferring statistical meaning upon parameters of interest. A key aspect underscored throughout the discussion is the distinction between "identified parameterization" and "parameters of interest." While identified parameterization pertains to population characteristics and corresponds to functionals of the data-generating process, parameters of interest are substantive issues specific to the data under examination. The crux of the problem lies in establishing a functional injective relationship between identified parameters and parameters of interest. This pursuit of identification is crucial for drawing meaningful and reliable inferences from data. Understanding identifiability is pivotal in ensuring that the parameters we estimate have a valid statistical interpretation and can inform us about the real-world phenomena we seek to study.

As one of the illustrative examples shown in this chapter, the identifiability analysis of the 1PL-G model brings significant clarity not only by resolving the identification issue but also by facilitating the interpretation of crucial parameters of interest, particularly the meaning of guessing, a concept not easy to define. We have seen that the empirical applicability of the 1PL-G model is subject to constraints imposed by the identifiability analysis. The requirement to make arbitrary decisions regarding which stimuli or items will be assigned a parameter $c_j = 0$ presents challenges in comparing the characteristics of stimuli and persons effectively. Consequently, the interpretation and generalizability of the model's outcomes become limited, and we must recognize that alterations to the designated items could significantly influence the resulting conclusions.

As another example, this chapter presents a comprehensive review of the identification conditions for estimating the ATE in observational studies. We offered a formal presentation of the problem, defined the sample space, and identified parameters and parameters of interest. A key point discussed is that the parameters of interest are not identified, leading to limitations in drawing definitive causal inferences. We have argued that the ignorability condition can serve as an identification restriction, enabling the expression of parameters of interest as functions of the identified parameters.

While the ignorability assumption helps to solve the identification problem, it does not fully address the fundamental issue of causal inference, as proposed by Holland in 1986. In response to this challenge, we have presented the concept of partial identification and offered four distinct solutions for causal inference. These solutions draw inspiration from Manski's empirical research approach and Neyman's concept of "behavioral inference", offering promising avenues to overcome the limitations of point identification.

By combining rigorous theoretical analysis with practical approaches, this example contributes valuable insights to the field of observational studies and causal inference. The exploration of partial identification and its integration with established methodologies paves the way for a more nuanced understanding of the ATE estimation in observational settings. Our exposition provides a significant step forward in addressing the identification challenges surrounding the ATE estimation and highlights the importance of considering partial identification methods for advancing causal inference research.

In summary, in this chapter we have shown that the concept of identifiability serves as a cornerstone in statistical analysis, providing a framework for establishing connections between theoretical models and empirical data. By comprehending and addressing the challenges of identification, we can enhance the rigor and validity of our research, ultimately advancing knowledge and understanding across a wide array of disciplines.

Appendix A

The identification analysis of the 1PL-G fixed-effects model with a guessing parameter follows from (4.7). In fact, note that:

$$G\left(\theta_i - \beta_j\right) = \frac{q_{ij}}{\delta_j} \Leftrightarrow \theta_i - \beta_j = G^{-1}\left(\frac{q_{ij}}{\delta_j}\right) \tag{4.28}$$

where $G^{-1}(\cdot)$ represents the inverse function of G. By considering $j = 1$ be the standard item (Rasch, 1960), then:

$$\theta_m = G^{-1}\left(\frac{q_{m1}}{\delta_1}\right) + \beta_1$$

which is precisely (4.8). Thus, the meaning of the ability parameter is given by when two different persons indexed by m and l are compared. In fact,

$$\begin{aligned} \theta_m > \theta_l &\Leftrightarrow G^{-1}\left(\frac{q_{m1}}{\delta_1}\right) + \beta_1 > G^{-1}\left(\frac{q_{l1}}{\delta_1}\right) + \beta_1 \\ &\Leftrightarrow q_{m1} < q_{l1} \\ &\Leftrightarrow P(X_1(l) = 1) < P(X_1(m) = 1) \end{aligned}$$

where the second inequality comes from the fact that G is a nonincreasing function and the last one by the definition of q_{mj}. Thus, more ability means greater probability to correctly answer the standard item.

Regarding the parameter β_j, by replacing (4.8) in (4.28), it satisfies:

$$\beta_j = G^{-1}\left(\frac{q_{m1}}{\delta_1}\right) - G^{-1}\left(\frac{q_{m1}}{\delta_1}\right) + \beta_1$$

obtaining equality (4.9). The interpretation of this parameter is obtained by comparing two different items. In fact,

$$\beta_j > \beta_k \Leftrightarrow G^{-1}\left(\frac{q_{m1}}{\delta_1}\right) - G^{-1}\left(\frac{q_{mj}}{\delta_j}\right) > G^{-1}\left(\frac{q_{m1}}{\delta_1}\right) - G^{-1}\left(\frac{q_{mk}}{\delta_k}\right)$$
$$\Leftrightarrow \frac{q_{mj}}{\delta_j} > \frac{q_{mk}}{\delta_k}$$

Regarding to the nonguessing parameter δ_j, when comparing two persons $(m = 1, 2)$ from (4.9) it follows that:

$$\beta_j - \beta_1 = G^{-1}\left(\frac{q_{11}}{\delta_1}\right) - G^{-1}\left(\frac{q_{1j}}{\delta_j}\right)$$
$$\beta_j - \beta_1 = G^{-1}\left(\frac{q_{21}}{\delta_1}\right) - G^{-1}\left(\frac{q_{2j}}{\delta_j}\right)$$

Given that these results are equal, rearranging terms it holds that:

$$G^{-1}\left(\frac{q_{1j}}{\delta_j}\right) - G^{-1}\left(\frac{q_{2j}}{\delta_j}\right) = G^{-1}\left(\frac{q_{11}}{\delta_1}\right) - G^{-1}\left(\frac{q_{2j}}{\delta_j}\right)$$

recovering equality (4.10). It is important to emphasize that all these results are independent from the function G, the item standard j, and the persons compared.

Appendix B

The identification bounds for the parameters of interest $P(Y(1) = 1 \mid X)$ and $P(Y(0) = 1 \mid X)$ obtained by using the law of total probability, *Optimistic policymaker* perspective restrictions for the nonidentified probabilities and recognizing that they range in the interval $[0, 1]$.

In particular, considering restriction (4.20) in (4.14), the lower bound for the parameter $P(Y(1) = 1 \mid X)$ is given by

$$\begin{aligned} P(Y(1) = 1 \mid X) &= P(Y(1) = 1 \mid X, Z = 1)P(Z = 1 \mid X) + P(Y(1) = 1 \mid X, Z = 0)P(Z = 0 \mid X) \\ &\geq P(Y(1) = 1 \mid X, Z = 1)P(Z = 1 \mid X) + P(Y(0) = 1 \mid X, Z = 0)P(Z = 0 \mid X) \\ &= P(Y(1) = 1, Z = 1 \mid X) + P(Y(0) = 1 \mid X, Z = 1) \end{aligned}$$

The upper is obtained by taking into account that the nonidentified related parameter is a probability, that is, $P(Y(1) = 1 \mid X, Z = 1) \leq 1$. Then,

$$\begin{aligned}P(Y(1)=1 \mid X) &\leq P(Y(1)=1 \mid X, Z=1)P(Z=1 \mid X) + P(Z=0 \mid X)\\ &= P(Y(1)=1, Z=1 \mid X) + P(Z=0 \mid X)\\ &= P(Y(1)=1, Z=1 \mid X) + P(Y(0)=1, Z=0 \mid X) + P(Y(0)=0, Z=0 \mid X)\end{aligned}$$

Thus, the lower and upper bound obtained here are precisely the ones shown in (4.21). We emphasize at this point that this interval contains all the possible values for the proportion of person in M who would experience a positive outcome if all of them were exposed to the innovation.

In a similar way, the identification bound for the parameter of interest $P(Y(0)=1 \mid X)$ is derived. As a matter of fact, by replacing restriction (4.19) in (4.15) it holds that:

$$\begin{aligned}P(Y(0)=1 \mid X) &= P(Y(0)=1 \mid X, Z=1)P(Z=1 \mid X) + P(Y(0)=1 \mid X, Z=0)P(Z=0 \mid X)\\ &\geq P(Y(1)=1 \mid X, Z=1)P(Z=1 \mid X) + P(Y(0)=1 \mid X, Z=0)P(Z=0 \mid X)\\ &= P(Y(1)=1, Z=1 \mid X) + P(Y(0)=1, Z=0 \mid X)\end{aligned}$$

The lower bound for the parameter is attained when the maximum value for the non-identified probability is considered, that is, $P(Y(0)=1, Z=1 \mid X) \geq 0$. Then,

$$\begin{aligned}P(Y(0)=1 \mid X) &\leq P(Y(0)=1 \mid X, Z=0)P(Z=0 \mid X)\\ &= P(Y(0)=1, Z=0 \mid X)\end{aligned}$$

Thus, under the last two restrictions mentioned before the identification bound (4.22) is recovered. This interval represents all the possible values compatible with the observables for the proportion of person in M who would experience a positive outcome if all of them were exposed to the status quo.

References

Bahadur, R. R., Stigler, S. M., Wong, W. H. and Xu, D. (2002). R.R. *Bahadur's Lectures on the Theory of Estimation*. Institute of Mathematical Statistics Lecture Notes-Monograph series. Institute of Mathematical Statistics.

Baker, F., & Kim, S. (2004). *Item response theory: Parameter estimating techniques*. New York: Marcel Dekker.

Basu, D. (1977). On the elimination of nuisance parameters. *Journal of the American Statistical Association*, *72*, 355–366.

Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: Making an informed choice. *Quality & Quantity*, *53*, 1051–1074.

Blundell, R., & Costa Dias, M. (2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources*, *44*, 565–640.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Castellano, K. E., Rabe-Hesketh, S., & Skrondal, A. (2014). Composition, context, and endogeneity in school and teacher comparisons. *Journal of Educational and Behavioural Statistics*, *39*, 333–367.

Clarke, P., Crawford, C., Steele, F., & Vignoles, A. (2015). Revisiting fixed-and random-effects models: Some considerations for policy-relevant education research. *Education Economics*, *23*, 259–277.

De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.

Engle, R. F., Hendry, D. F., & Richard, J. F. (1983). Exogeneity. *Econometrica*, *51*, 277–304.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A*, *22*, 309–368.

Fisher, R. A. (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *17*, 69–78.

Fisher, R. A. (1973). *Statistical methods for research workers*. Hafner Publishing.

Florens, J. P., Marimoutou, V., & Péguin-Feissolle, A. (2007). *Econometric modeling and inference*. Cambridge University Press.

Florens, J. P., Mouchart, M., & Rolin, J. M. (1985). On two definitions of identification. *Statistics*, *16*, 213–218.

Florens, J. P., Mouchart, M., & Rolin, J. M. (1990). *Elements of Bayesian statistics*. Marcel Dekker, Inc.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, *33*, 587–606.

Gourieroux, C., & Monfort, A. (1995). *Statistics and econometric models*. Vol. 1, Cambridge University Press.

Haavelmo, T. (1944). The Probability Approach in Econometrics. *Econometrica*, *12*, iii–115.

Hambleton, R., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Dordrecht: Kluwer Nijhoff Publishing.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*, 945–960.

Hurwicz, L. (1950). Generalization of the concept of identification. *Statistical Inference in Dynamic Economic Models*, *10*, 245–257.

Itô, K. (1984). *An introduction to probability theory*. Cambridge University Press.

Kolmogorov, A. N. (1950). *Foundations of the theory of probability*. New York: Chelsea Pub. Co.

Koopmans, T. (1949). Identification problems in economic model construction. *Econometrica*, *17*, 125–144.

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

LeCam, L., & Schwartz, L. (1960). A necessary and sufficient condition for the existence of consistent estimates. *The Annals of Mathematical Statistics*, *31*, 140–150.

Lechner, M. (2008). A note on the common support problem in applied evaluation studies. *Annales D'économie Et de Statistique*, *91/92*, 217–235.

Longford, N. T. (2012). A revision of school effectiveness analysis. *Journal of Educational and Behavioral Statistics*, *37*, 157–179.

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison Wesley.

Manski, C. (2013a). *Public policy in an uncertain world: Analysis and decisions*. Harvard University Press.

Manski, C. (2013b). Diagnostic testing and treatment under ambiguity: Using decision analysis to inform clinical practice. Proceedings of the National Academy of Sciences, 110:2064–2069.

Manski, C., Sanstad, A., & DeCanio, S. (2021). Addressing partial identification in climate modeling and policy analysis. *Proceedings of the National Academy of Sciences*, *11*, e2022886118.

Manski, C. F. (1995). *Identification problems in the social sciences*. Harvard University Press.

McCullagh, P. (2002). What is a statistical model? *The Annals of Statistics*, *30*, 1225–1310.

Mouchart, M., & Oulhaj, A. (2006). The role of the exogenous randomness in the identification of conditional models. *Metron*, *64*, 253–271.

Neyman, J. (1938). L'estimation statistique traitée comme un problème classique de probabilité'. *Actualités Scientifiques Et Industrielles*, *739*, 25–57.

Neyman, J. (1950). *First course in probability and statistics*. Rinehart and Winston: Holt, Inc.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. The Danish Institute for Educational Research.

Rasch, G. (1966). An individualistic approach to item analysis. In P. F. Lazarsfeld & N. W. Henry (Eds.). *Readings in mathematical social sciences* (pp. 89–107). MIT Press.
Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.
Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701.
Rubin, D. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, *6*, 34–58.
San Martín, E., del Pino, G., & De Boeck, P. (2006). IRT models for ability-based guessing. *Applied Psychological Measurement*, *30*, 183–203.
San Martín, E., & González, J. (2022). A critical view on the NEAT equating design: Statistical modeling and identifiability problems. *Journal of Educational and Behavioural Statistics*, *47*, 406–437.
San Martín, E., González, J., & Tuerlinckx, F. (2009). Identified parameters, parameters of interest and their relationships. *Measurement: Interdisciplinary Research & Perspective*, *7*, 97–105.
Van der Linden, W., & Hambleton, R. (1997). *Handbook of modern item response theory*. New York: Springer.
Weitzman, R. A. (1996). The Rasch model plus guessing. *Educational and Psychological Measurement*, *56*, 779–90.