

# Identifying the Effects of a Program Offer with an Application to Head Start\*

Vishal Kamat

Toulouse School of Economics  
University of Toulouse Capitole  
[vishal.kamat@tse-fr.eu](mailto:vishal.kamat@tse-fr.eu)

December 14, 2024

## Abstract

I develop tools to learn about the average effects of providing an offer to participate in a program given data from an experiment that randomizes offers. I allow for the complication that individuals may not comply with their assigned status in the sense that those who are not provided an offer may receive one from outside the experiment, and that the data may provide only partial information on the receipt of an offer across individuals. To do so, I propose a new nonparametric selection model with unobserved choice sets that provides a conceptual framework to define a range of parameters evaluating the effects of an offer, exploit the partial information available on offer receipt, and also consider an array of identifying assumptions. I illustrate how a computational procedure can be used to sharply learn about the parameters under the various assumptions. Using these tools, I analyze the effects of a policy that provides an offer to participate in the Head Start preschool program given data from the Head Start Impact Study. I find that such a policy affects a large number of children who take up the offer, and that they subsequently have positive effects on test scores. These effects primarily arise from children who do not have any preschool as an outside option. Performing a cost-benefit analysis, I find that the earning benefits associated with the test score gains can outweigh the net costs associated with the take up of the offer.

**KEYWORDS:** Program offer, program evaluation, discrete choice, unobserved choice sets, partial identification, noncompliance, randomized experiments, Head Start Impact Study.

**JEL classification codes:** C14, C25, C31, I21.

---

\*I am extremely grateful to Ivan Canay, Chuck Manski and Alex Torgovitsky for their extensive guidance and feedback. I also thank Eric Auerbach, Gideon Bornstein, Pedro Carneiro, Joel Horowitz, Tom Meling, Magne Mogstad, Francesca Molinari, Sam Norris, Matt Notowidigdo, Nicolas Inostroza, Rob Porter, Pedro Sant'Anna, Lola Segura, Azeem Shaikh, Max Tabord-Meehan and conference and seminar participants at several institutions for useful comments, and Research Connections for providing data on the Head Start Impact Study. Funding from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program) and the Robert Eisner Memorial Fellowship is gratefully acknowledged. First version: arXiv:1711.02048 dated November 6, 2017.

# 1 Introduction

A central task of program evaluation is to evaluate policies relevant to a program. For many programs, the baseline policy corresponds to providing individuals with an offer to participate in the program. For example, in a preschool program, children may be provided with an offer to preschools in the program, while in a microfinance program, individuals may be provided with a loan offer. The evaluation task for such programs then translates to evaluating the effects of providing an offer. Indeed, evaluating the number of individuals who may take up the offer as well as the outcomes they later experience allows one to assess the costs and benefits of providing an offer and, in turn, the returns to the policy.

In this paper, I develop tools to learn about the average effects of providing a program offer given data from an experiment that randomly assigns individuals to either a treatment group with an offer or a control group without one. Several experiments in economics are implemented in such a manner. For example, in the Head Start Impact Study (HSIS), children were randomly provided an offer to specific Head Start preschools. Alternatively, in a microfinance experiment in [Angelucci et al. \(2015\)](#), individuals in certain villages were randomly provided a loan offer.

In such experiments, if no control individual received an offer to the program, one can indeed evaluate the average effects of an offer by estimating standard estimands such as the intention-to-treat (ITT) and instrumental variable (IV) estimands, which compare mean responses of individuals in the treatment and control groups. However, as is the case in the above experiments, some control individuals often do not comply with their assigned status and receive an offer from alternative means. For example, in the HSIS, some control children received an offer by being admitted to Head Start preschools not part of the experiment. In such cases, as illustrated in [Appendix S.1](#), the ITT and IV estimands allow one to only evaluate so-called local average effects of providing an offer local to the subgroup of compliers, i.e. individuals who would not have received an offer had they been assigned to the control group.

These local effects, in general, evaluate the effects of a policy providing an offer relative to not through assignment to the treatment or control groups of the experiment. However, while such effects may be of interest, we may also want to learn the policy effects of providing an offer more generally, beyond the experimental assignment, and hence evaluate the effects of an offer unconditional of the compliers. For precisely such reasons, numerous treatment evaluation tools have been developed—see [Mogstad and Torgovitsky \(2018, Section 6\)](#) for a recent overview. But, to apply them to evaluate the effects of a program offer, one would typically require, as highlighted in [Appendix S.1](#), data on the receipt of an offer for every individual. Unfortunately, for many experiments such as those above, such data is often not available. Instead, data on the treatment assignment status and participation decision is available, which may only provide partial information on the offer receipt across individuals. For example, in the HSIS, while data on whether each

child was admitted and received an offer was not available, one can infer if this was the case when they were assigned to the treatment group or when they participated in Head Start.

The main contribution of this paper is to develop a new framework to show how data on only the treatment assignment status, participation decision and outcome in such experiments can be exploited to learn about the average effects of providing a program offer. As discussed below, a key novelty over the existing literature is the introduction of a new nonparametric selection model that provides a natural conceptual framework to straightforwardly analyze a range of parameters capturing the effects of providing an offer. The model is based on the fundamental insight that receiving an offer to an alternative can be framed as receiving that alternative in the choice set from which participation decisions are made. In turn, given one does not directly observe from where an individual receives offers, the model takes the individual’s participation decision to be equal to a utility maximization decision based on their unobserved preferences over the various alternatives and their unobserved choice set of available alternatives.

An important feature of the model is that it introduces variables that capture the heterogeneity in both choice sets and preferences in the participation decision. This novel structure has several benefits towards straightforwardly analyzing the effects of a program offer. First, it allows one to define a range of parameters evaluating these effects by comparing the participation decisions and subsequent outcomes under counterfactual choice sets with and without the program. Second, it allows one to capture the partial information the data provides on from where an offer was received. In particular, the information provided by the participation decision can be captured through the structural relationship between the choice set and decision, and that provided by the treatment assignment status can be captured by imposing restrictions on the feasible values that the choice set can take. Finally, it allows one to similarly impose various other easy-to-interpret restrictions on the variables that can further help learn about the effects of an offer.

The model, moreover, is nonparametric and allows the choices set, preferences and outcomes to be generally related in an unrestricted manner. Given the partial information provided by the data, it follows that the various parameters of interest are generally set identified. Deriving sharp analytical bounds for these parameters, however, can be difficult given the rich structure of the model and geometry of the parameters. In turn, by leveraging results from [Charnes and Cooper \(1962\)](#), I illustrate how a linear programming procedure can be used to sharply characterize the identified sets. The procedure flexibly applies to a range of parameters as well as a range of assumptions that can be potentially imposed on the baseline model.

Using the developed framework, I empirically analyze the average effects of providing an offer to the Head Start program given data from the HSIS. For ease of exposition, I also develop the framework in the context of this experiment.<sup>1</sup> Previous analyses have focused on the ITT and IV estimands that evaluate local average effects of providing an offer for the subgroup of compliers,

---

<sup>1</sup>Appendix [S.6](#) presents the generalized version of the framework and shows how it applies to the microfinance

and variants of these effects that account for where these children would have enrolled in the control group (Feller et al., 2016; Kline and Walters, 2016; Puma et al., 2010). These studies find that the effects on test scores are positive, and that they arise from children who, absent receiving a Head Start offer, do not participate in an alternative (non-Head Start) preschool. However, as noted above, these effects in general will solely capture those of a policy that provides a Head Start offer relative to not through assignment to the treatment or control groups of the experiment. In turn, they leave open the question on what the policy effects of providing an offer beyond the experimental assignment are.

The empirical analysis aims to complement previous findings by analyzing such effects by evaluating those of providing a Head Start offer unconditional of the compliers. The analysis reveals several broad conclusions qualitatively symmetric to those based on the local effects. First, providing an offer to Head Start affects a large percentage of children who take up the offer and participate in Head Start, and can significantly improve their test scores. In particular, 82.5% of children who are provided an offer choose to take it up, and their subsequent tests scores can improve on average between 0.21 and 0.50 standard deviation points. Second, these effects primarily arise from the subgroup of children who do not have an offer to an alternative (non-Head Start) preschool and hence any outside option. Finally, using the estimates to perform a cost-benefit analysis of the policy, the earning benefits associated with the test score gains can outweigh the net costs associated with the take up of the offer.

The developed framework contributes to several literatures. It is related to a large literature that studies partial identification of various average treatment effects—see Ho and Rosen (2017, Section 5) for a recent survey. Papers in this literature typically use the observed distribution of the response and treatment along with additional data and assumptions to partially identify these effects. In contrast, as argued in Appendix S.1, the treatment of interest in the setting studied in this paper, corresponding to the receipt of a program offer, is not directly observed for any individual. The main novelty here is to show how a selection model provides a natural conceptual framework to exploit the partial information available on the treatment of an offer and to define and learn about effects related to this treatment.

A number of studies have similarly used different types of selection models to define and analyze parameters capturing various treatment effects—see Mogstad and Torgovitsky (2018) for an overview of such studies in the binary treatment setup and Heckman et al. (2006, 2008), Heckman and Pinto (2018), Kirkeboen et al. (2016), Kline and Walters (2016), Lee and Salanié (2018) and Pinto (2019) for examples in the case of multiple treatments. However, as elaborated in Section 3.2, the structures in these models do not introduce variables that capture the unobserved heterogeneity potentially experiment from Angelucci et al. (2015). It also shows how the framework applies to the Oregon Health Insurance experiment, where not only the control individual may non comply but also the treated individuals may do so in the sense that they may not receive an offer to the program.

present in both preferences and choice sets. This feature, a novelty of the model proposed in this paper, plays an essential role to straightforwardly define and analyze various parameters capturing the effects of providing a program offer. Moreover, the arguments behind the identification analysis also differ from those typically employed in these studies. Specifically, while they usually focus on cases where the parameters are point identified, the analysis here allows for the case of partial identification.

To develop the model, I build on elements from the literature on nonparametric discrete choice analysis. As elaborated in Section 3.2, Manski (2007) proposed a choice model with heterogeneity in both preferences and choice sets where one only assumed individuals to have a strict preference relationship over the various alternatives. However, while this model allowed for unobserved preferences, it presumed that the choice sets were observed and statistically independent of preferences. Indeed, as from where an individual receives offers is unobserved and may depend on their preferences, exogenous variation in observed choice sets is inherently absent in the setting studied in this paper. I illustrate how the structure of this model can be extended to allow for the choice sets to be both unobserved and correlated with preferences in an unrestricted manner.

More broadly, since the model takes choice sets to be unobserved, the analysis also contributes to the growing literature on the analysis of preferences in discrete choice settings with unobserved heterogeneity in choice sets—see Abaluck and Adams (2018), Barseghyan et al. (2018), Cattaneo et al. (2020) and Crawford et al. (2020) for some recent examples. These papers provide tools for researchers to analyze preferences under various models that impose different types of assumptions on the relation between the unobserved choice sets and preferences. The model proposed in this paper is distinct with a different structure and imposed assumptions. The resulting analysis therefore complements these papers by providing tools that are applicable to alternative settings and hence by expanding the sets of tools available to researchers.

The remainder of the paper is organized as follows. Section 2 briefly describes the setting of the HSIS and expands on the motivation for the developed analysis. Section 3 develops the framework to learn about the average effects of providing an offer to Head Start. Section 4 applies the framework to the HSIS data. Section 5 concludes. Proofs and auxiliary results complementing the main analysis are presented in the Supplementary Appendix.

## 2 The Head Start Impact Study

Head Start is the largest early childhood education program in the United States that provides free preschool education to three- and four-year-old children from disadvantaged households. Program eligibility for households is primarily determined by the federal poverty line, yet certain exceptions qualify additional low-income-households. The Head Start Impact Study (HSIS) was a publicly-

funded randomized experiment implemented in Fall 2002 to evaluate Head Start. [Puma et al. \(2010\)](#) provides further details on the program and experiment. Below, I briefly summarize the details of the experiment that are relevant for the analysis in this paper.

The experiment acquired a sample of Head Start preschools and participating children using a multistage stratified sampling scheme. Preschools were sampled from the subpopulation of so-called non-saturated Head Start preschools, which referred to preschools where the number of available slots was strictly smaller than the number of applicants. From each sampled preschool, children were then sampled separately from the subpopulation of three- and four-year-old applicants in Fall 2002, who were not previously enrolled in any Head Start services.

Once these samples were acquired, the randomization was performed. To better understand certain details, it is useful to describe the remainder of the experiment in stages:

**Stage 1:** At each sampled Head Start preschool, the experiment randomly assigned sampled children to either a treatment group where they were provided an offer to that preschool or a control group where they were not. The experiment, however, did not control the child’s chances of receiving an offer from Head Start preschools from which they were not sampled. In particular, they could receive an offer to Head Start if they could be admitted to these other Head Start preschools. In addition, in a similar manner, children in both groups could also receive an offer to alternative (non-Head Start) preschools.

**Stage 2:** The experiment collected data on the type of the care setting where the parents enrolled the child, which consisted of Head Start and alternative preschools and a home care or non-preschool setting where the child was taken care of either by the parents, a relative or some known individual.

**Stage 3:** The experiment collected data on a number of outcomes, such as various test scores, at the end of the school year in Spring 2003.

The objective of this paper is to use data from the above experiment to evaluate the average effects of providing an offer to Head Start. To this end, it is useful to highlight two features in Stage 1 of the experiment and their consequences towards this objective—see Appendix [S.1](#) for a formal illustration. First, note that some children assigned to the control group may not have complied with their assigned status and received a Head Start offer from preschools outside the experiment. As highlighted in Section [4.1](#), the data reveals that this is indeed the case. In turn, due to this noncompliance, standard intention-to-treat and instrumental variable estimands based on comparing mean responses in the treatment and control groups generally do not evaluate the average effects of a Head Start offer. In particular, they can be shown to only evaluate so-called local average effects of an offer local to the subgroup of compliers, i.e. children who do not receive an offer when assigned to the control group. These local effects in general solely capture the effects

of a policy that provides a Head Start offer relative to not through assignment to the treatment and control groups of the experiment. As a result, they leave open the question on what the policy effects of providing an offer beyond the experimental assignment are.

For this question, we generally need to evaluate the effects of an offer unconditionally of the complier subgroup, which is the focus of the analysis in this paper. To this end, note secondly that the experiment did not collect data on the preschools from where any child received offers. This implies that we do not directly observe whether each child received a Head Start offer or not. As a result, many of the treatment effect tools developed in the literature to move beyond local treatment effects cannot be directly applied to learn about those of providing a Head Start offer as these tools would typically require data on the receipt of a Head Start offer for each child. In addition, if we also wanted to apply these tools to better understand these effects and evaluate their heterogeneity based on whether an offer to an alternative preschool was received, as I do in Section 4.3, we would similarly require data on the offer receipt to an alternative preschool for each child, which again is not directly observed given the data.

The experiment nonetheless collected data on the treatment assignment status and the participation decision that provides some partial information across children on whether an offer to Head Start or an alternative preschool was received. In particular, if a child was assigned to the treatment group or if they participated in Head Start then we can infer that the child must have received an offer from Head Start. Similarly, if a child participated in an alternative preschool then we can infer that the child must have received an offer to that preschool. The objective of the analysis presented below is to develop tools to show how we can exploit this partial information to learn about the average effects of providing a Head Start offer.

### 3 Identification Framework

In this section, I develop a framework to show how to evaluate the average effects of providing an offer to Head Start using only data on the treatment assignment status, participation decision and outcome, taken to be some test score of interest.

#### 3.1 Model with Unobserved Choice Sets

For each child, we observe variables  $Z$ ,  $D$  and  $Y$ , where  $Z$  denotes an indicator for whether they were assigned to the treatment group or not,  $D$  denotes the type of care setting taking values in the set  $\mathcal{D}$  where they participated, and  $Y$  denotes their observed test score outcome. As further discussed below, I take the types of care settings where the child can participate to be parsimoniously given by  $\mathcal{D} = \{n, a, h\}$ , where  $h$  denotes a Head Start preschool,  $a$  denotes an alternative (non-Head Start) preschool, and  $n$  denotes a no preschool or equivalently a home care setting.



The observed variables are assumed to be generated by a selection model. For a clearer exposition, it is useful to describe the model in terms of the various stages of the experiment described in Section 2. In Stage 1, the child receives offers to Head Start and alternative preschools. From where they receive offers depends on whether they are assigned to the treatment or control group, and their application and admission decisions at preschools not part of the experiment. I model receiving an offer to a preschool as obtaining that preschool in their choice set of care settings from which the participation decision is made. Let  $C$  denote the value of this choice set. Based on the preschools from where an offer is received, this choice set can take values in  $\mathcal{C} = \{\{n\}, \{n, a\}, \{n, h\}, \{n, a, h\}\}$ , i.e. the set of possible choice sets containing home care. To account for the fact that the experiment altered the choice set by providing a Head Start offer to the treatment group, let  $C(1)$  and  $C(0)$  respectively denote the potential choice sets had the child been assigned to the treatment and control group, which are related to the obtained choice set by

$$C = \begin{cases} C(1) & \text{if } Z = 1, \\ C(0) & \text{if } Z = 0. \end{cases} \quad (1)$$

In Stage 2, the parents decide where to enroll their child given their obtained choice set from Stage 1. I assume this participation decision is a product of a utility maximizing decision. Let  $U(d)$  denote the parents' indirect utility had their child been enrolled in  $d \in \mathcal{D}$ . The observed enrollment decision is assumed to be given by the following utility maximization relationship

$$D = \arg \max_{d \in C} U(d). \quad (2)$$

The utilities here can be viewed as capturing the perceived benefits and costs of participating in a given care setting and include, for example, costs such as tuition and transport costs. However, it is useful to emphasize that they do not include the costs endured to receive an offer to a preschool such as, for example, those related to making application decisions. As discussed further in Section 3.2, these costs are separately captured and subsumed in the value of the choice set  $C$  that the individual faces.

Since the utility under each care setting does not possess any cardinal value, different monotonic transformations of these utilities will generate observationally equivalent choices. For the purposes of the analysis, it is therefore more useful to directly refer to the underlying preference type that these utilities represent. To this end, assuming that the parents have a strict preference relation over the set of care settings, let  $U$  denote the parents' preference type corresponding to a strict preference relationship over the set of care settings  $\mathcal{D}$ . As there are three possible care settings, note that there are six possible preference types that can be denoted by the following set  $\mathcal{U} = \{(h \succ a \succ n), (h \succ n \succ a), (a \succ h \succ n), (a \succ n \succ h), (n \succ h \succ a), (n \succ a \succ h)\}$ . In addition, let  $d(u, c)$  denote the known choice function that corresponds to what preference type  $u \in \mathcal{U}$  would choose under a non-empty subset  $c \subseteq \mathcal{D}$ , i.e. under a given feasible choice set. For example,



$d((h \succ a \succ n), \{n, a\}) = a$  as preference type  $(h \succ a \succ n)$  prefers  $a$  to  $n$  and hence would choose  $a$  when faced with the choice set containing  $n$  and  $a$ . Using this notation, the utility maximization relationship in (2) can be alternatively re-written in terms of the preference type and the obtained choice set through the following relationship

$$D = \sum_{u \in \mathcal{U}, c \in \mathcal{C}} d(u, c) I\{U = u, C = c\} = d(U, C) . \quad (3)$$

In Stage 3, the child's test score is observed under the enrolled care setting from Stage 2. I model this using the usual potential outcome setup. In particular, denoting by  $Y(d)$  the potential test score had the child been enrolled in care setting  $d \in \mathcal{D}$ , the observed test score is assumed to be given by the following relationship

$$Y = \sum_{d \in \mathcal{D}} Y(d) I\{D = d\} = Y(D) . \quad (4)$$

The model captures the fact that we do not observe where an individual receives an offer from by taking the value of the obtained choice set  $C$  to be unobserved. Nonetheless, as noted earlier, the participation decision and treatment assignment status provide some partial information on where an offer was received and in turn on the value of the choice set. The information provided by the participation decision is automatically captured by the structure of the above model. In particular, the structure of the choice equation in (3) reveals that the observed choice must be in the choice set, i.e. if  $D = d$  then it must be the case that  $C \in \{c \in \mathcal{C} : d \in c\}$ . The information provided by the treatment assignment status, on the other hand, can be captured by imposing a restriction on the possible values the underlying variables of the model can take. I state this restriction formally in the following assumption.

**Assumption HSIS.**

(i)  $(Y(n), Y(a), Y(h), U, C(0), C(1)) \perp Z$  .

(ii)  $h \in C(1)$  .

Assumption **HSIS**(i) states that children were randomly assigned to either the treatment or control group. Assumptions **HSIS**(ii) states that being assigned to the treatment group guaranteed a Head Start offer and captures the information that the experiment provides on the values that the choice set can take. In particular, it reveals that Head Start must be in the choice set when assigned to the treatment group, i.e. if  $Z = 1$  then it must be the case that  $C \in \{c \in \mathcal{C} : h \in c\} = \{\{n, h\}, \{n, a, h\}\}$  .

It is worth highlighting that the model is entirely nonparametric and places no restrictions on the dependence between the various underlying variables. The potential test scores, the preference type and the choice sets are allowed to be correlated in an unrestricted manner. As a result, the model

allows for various selection patterns where parents make participation decisions in Stage 2 based on unobservables correlated to potential test scores in Stage 3, and from where children receive offers in Stage 1 based on unobservables correlated to preferences in Stage 2 and potential test scores in Stage 3. As noted below, the baseline model, however, may not be sufficiently restrictive and, in turn, may not generally allow us to reach informative conclusions. To this end, with the objective of reaching stronger conclusions, I also illustrate in Section 3.5 several additional nonparametric assumptions that could be imposed, which restrict how the various underlying variables may be related to each other.

It is, however, also worth highlighting that the model imposes some implicit assumptions through its structure. In particular, by taking the set of care settings to be categorized into only three values, the model assumes in some sense that for each child all Head Start preschools and all alternative preschools are homogeneous in terms of both preferences and test scores. In turn, this rules out scenarios, for example, where assignment to the treatment group may change preferences for Head Start as the Head Start preschools in the treatment group may potentially be in a more convenient location, or may change test scores as these preschools may be of different quality. The main reason why I impose this implicit assumption, similar to previous analyses of the HSIS (Feller et al., 2016; Kline and Walters, 2016), is because the HSIS data encodes the enrollment decision only under any Head Start and any alternative preschool and does not allow distinguishing the preschools in each of these categories.

Nonetheless, had the appropriate data been available, note that the developed framework can be extended to allow for a general number of discrete care settings as described in Appendix S.6. In particular, in Appendix S.6, I present the generalized version of the model that the framework permits. This generalized model allows for a general number of discrete feasible choices. In addition, it also allows for alternative experimental setups that may provide information from where an offer is received not satisfying the form of Assumption HSIS(ii).<sup>2</sup> I also present there examples of two alternative experiments and illustrate how these extensions to the model allow it to accommodate certain details specific to the setting of these experiments.

### 3.2 Comparison to a Standard Choice Model

A standard baseline choice model in the literature on treatment effect analysis is to take the observed choice to be given by  $D = D(1)Z + D(0)(1 - Z)$ , where

$$D(z) = \arg \max_{d \in \mathcal{D}} \bar{U}(d, z) \quad (5)$$

---

<sup>2</sup>For example, it allows the setup where  $h \notin C(1)$ , i.e. a two-sided noncompliance setup where even the individuals in the treatment group may not comply and not receive an offer to the program, as well as a setup where individuals can simultaneously participate in multiple alternatives.

for  $z \in \{0, 1\}$  with  $\bar{U}(d, z)$  corresponding to the indirect utility under  $d \in \mathcal{D}$  and  $z \in \{0, 1\}$ —see Section 1 for references of various papers employing such a model and Kline and Walters (2016) who employ such a model in the context of the HSIS. Given this choice model, the analysis in these papers is based on taking either  $(D(0), D(1))$  or  $(\bar{U}(d, z) : d \in \mathcal{D}, z \in \mathcal{Z})$  to be the model primitives, and then defining and studying parameter based on these primitives. Observe that this choice model is indeed more agnostic about the choice process than that in (2) and can nest it by taking  $\bar{U}(d, z) = U(d)$  if  $d \in C(z)$  and  $\bar{U}(d, z) = -\infty$  if  $d \notin C(z)$ , for each  $d \in \mathcal{D}$  and  $z \in \{0, 1\}$ , i.e. by taking utility to be equal to that of enrolling in that alternative if it is in the choice set and to be equal to negative infinity if not as that alternative cannot be feasibly chosen. In other words, the utility in this model can be interpreted as combining the utility of choosing that alternative and whether that alternative is in the choice set into a single dimension.

However, by not separating the heterogeneity in preferences and choice sets, this model does not straightforwardly allow analyzing the effects of providing an offer. The richer structure of the choice model in (2) aims to do so and provide a natural conceptual framework for the analysis of such effects. In particular, as we observed above, the structure allowed us to introduce in the model the concept of whether an individual received an offer to an alternative—namely, by taking it to be whether that alternative was present in their choice set. Relative to the model in (5), this feature in turn has several benefits towards straightforwardly analyzing the effects of an offer. First, as we observed above, it allows us to capture how the data on observed choices can reveal information on from where an offer was received. Second, it allows imposing an assumption to exactly capture how the treatment group altered the receipt of an offer. As we will observe in Section 3.5, it also allows similarly introducing additional easy-to-interpret identifying assumptions on how the various variables may be related. Finally, as we will observe below, it allows us to straightforwardly define a range of effects of providing an offer by comparing responses under counterfactual choice sets with and without the program.

Explicitly modeling the role of choice sets is indeed a common practice in the literature on discrete choice analysis when one wants to analyze choice under counterfactual choice sets. The above proposed model draws on this conceptual connection and, in particular, to the model in Manski (2007) with some novel modifications—see also the analysis of Marschak (1960) on which Manski (2007) builds. The model in Manski (2007) similarly introduced for each individual an unobserved preference type and a choice set which then together determine the observed choice as in (3). However, the model assumed that the choice set  $C$  is observed and further that it is statistically independent of preferences, i.e.  $C \perp U$ . In the context studied in this paper, this amounts to assuming that from where an offer is received is observed and that it is independent of the individual’s preferences, both of which are inherently not valid. As a result, the proposed model shows how to relax these requirements in the choice equation. In addition, it also shows how to instead exploit additional information provided by an alternative variable, namely the treatment

assignment status  $Z$ , that may provide partial information on the value of the choice set.

### 3.3 Parameters Capturing the Average Effects of a Program Offer

In the context of the proposed model, I next define the main parameters of interest that aim to measure the average effects of providing an offer to Head Start and, in turn, those of a policy providing an offer to the program. To this end, observe that providing a Head Start offer corresponds to the child receiving Head Start in their choice set from which their parents make participation decisions. As a result, the effects of an offer can be defined by comparing potential responses under counterfactual choice sets with and without Head Start. In order to define these effects on participation decisions and test scores, it is therefore useful to introduce additional notation to refer to the potential participation decision and test score under a given choice set. Had the child's choice set been pre-specified to a non-empty subset  $c \subseteq D$  in Stage 1, let  $D_c = d(U, c)$  denote the care setting in which their parent would then enroll them in Stage 2, and let  $Y_c = Y(D_c)$  denote the test score that they would then earn in Stage 3.

For each child, if an offer to Head Start was provided then the child's resulting choice set would be given by  $C_+ = C \cup \{h\}$ , while if it was not then their resulting choice set would be given by  $C_- = C \setminus \{h\}$ . The average (treatment) effect (ATE) of providing a Head Start offer on participation in Head Start and subsequent test scores can then be defined by taking the average value of the difference in the respective potential responses under these two choice sets given by

$$\text{ATE}^D(C_+, C_-) = E [1\{D_{C_+} = h\} - 1\{D_{C_-} = h\}] , \quad (6)$$

$$\text{ATE}^Y(C_+, C_-) = E [Y_{C_+} - Y_{C_-}] . \quad (7)$$

Note that by construction we have  $1\{D_{C_-} = h\} = 0$ , which captures the fact that if the child is not provided a Head Start offer then they naturally cannot participate in Head Start. As a result, observe that the former parameter can more simply be rewritten as

$$\text{ATE}^D(C_+, C_-) = \text{Prob}\{D_{C_+} = h\} \quad (8)$$

i.e. the proportion of children who participate in Head Start when provided an offer. Observe that providing an offer does not affect the subgroup of children who do not take up the offer and participate in Head Start, i.e. those with  $D_{C_+} \neq h$ , as for these children we have  $Y_{C_+} = Y_{C_-}$ . It is therefore useful to also define the average (treatment) effect of providing an offer conditional on participation (ATOP), i.e. for those take it up and are hence affected, by

$$\text{ATOP}(C_+, C_-) = E [Y_{C_+} - Y_{C_-} | D_{C_+} = h] . \quad (9)$$

In a similar manner, we can also consider the average effects of an offer conditional on other subgroups of interest. Of particular interest in the empirical analysis is conditioning on children

based on whether they have or do not have an offer to an alternative preschool. This allows us to better understand the heterogeneity in the effects based on availability of an alternative preschool option and hence what might be driving the effects. The empirical results from [Feller et al. \(2016\)](#) and [Kline and Walters \(2016\)](#) suggest that such heterogeneity may indeed be present, as they find that the local effects are driven from compliers who absent a Head Start offer did not participate in an alternative preschool. To this end, we can define

$$\text{ATE}_{+a}^D(C_+, C_-) = \text{Prob}\{D_{C_+} = h | a \in C\} , \quad (10)$$

$$\text{ATOP}_{+a}(C_+, C_-) = E[Y_{C_+} - Y_{C_-} | D_{C_+} = h, a \in C] \quad (11)$$

to capture the proportion of children who take up the offer and their subsequent outcomes conditional on the subgroup who have an offer to an alternative preschool, and also

$$\text{ATE}_{-a}^D(C_+, C_-) = \text{Prob}\{D_{C_+} = h | a \notin C\} , \quad (12)$$

$$\text{ATOP}_{-a}(C_+, C_-) = E[Y_{C_+} - Y_{C_-} | D_{C_+} = h, a \notin C] \quad (13)$$

to capture the analogous effects conditional on the subgroup who do not have an alternative preschool offer.

It is useful to make several remarks on the above set of parameters. First, these parameters can be viewed to be conceptually related to the class of so-called policy relevant treatment effects introduced in [Heckman and Vytlacil \(2001, 2005, 2007\)](#). These treatment effects are based on comparing responses between well-defined policies as well as these effects conditional on the subgroup of individuals affected between these policies. Indeed, the above parameters similarly do so where one policy corresponds to that of providing children with an offer to Head Start while the other corresponds to not. These parameters can in turn be viewed as the policy relevant treatment effects of providing an offer versus not.

Second, it is worth providing a more concrete interpretation of the exact policy of providing an offer the above parameters capture. As noted in [Kline and Walters \(2016\)](#), providing an offer potentially opens slots at alternative preschools and hence also changes the distribution of who has offers to these preschools. The parameters do not capture these general equilibrium type effects as the proposed model does not model the supply of all preschool slots and how from where each child receives an offer may be interdependent. In turn, they may more appropriately be viewed as capturing the policy effects of providing a Head Start offer taking the slots at other schools fixed. Alternatively, they may be viewed as the effects of a marginal policy providing an offer to a single child as such a policy is more likely to keep the slots fixed.

Finally, note that one can define additional policy parameters based on comparing alternative choice sets. For example, instead of comparing choice sets  $C_+$  and  $C_-$  as the parameters above do, we can also compare these choice sets to the child's obtained choice  $C$ . Since  $C$  corresponds to the child's choice set under the status quo regime present in the data, these comparisons can

be viewed as capturing the policy effects of providing a Head Start offer or not relative to the status quo regime. Alternatively, we can compare these choice sets with and without an alternative preschool in them to compare the policy effects of a Head Start offer in the case where an offer to an alternative preschool is available or not is also a policy variable. Appendix S.2 discusses such parameters in more detail.

### 3.4 Identification Analysis

For a given parameter of interest, I next show how to characterize what we can learn about it given the distribution of the data on the treatment assignment status, participation decision and outcome, and the assumptions imposed on the model. For the purposes of this identification analysis, I assume that the chosen test score of interest takes values in a known discrete set  $\mathcal{Y} = \{y_1, \dots, y_M\}$ . As we will observe below, the use of a discretized outcome allows us to characterize what we can learn about each of the parameters using a general finite dimensional computational program.

In order to formally describe the analysis, it is useful to first introduce some additional notation. Let  $W = (Y(n), Y(a), Y(h), U, C(0), C(1))$  denote the random variable that summarizes the underlying latent variables for each child in the model. The analysis is based on the distribution of this variable. Given the discrete outcome, the distribution of this random variable can be characterized by a probability mass function  $Q$  with support contained in  $\mathcal{W}$ , i.e.  $Q : \mathcal{W} \rightarrow [0, 1]$  and  $\sum_{w \in \mathcal{W}} Q(w) = 1$ . Similarly, let  $Q_z$  denote the probability mass function of the summary random variable conditional on the treatment group assignment indicator  $Z$  equal to  $z \in \mathcal{Z} \equiv \{0, 1\}$ . Let  $w$  denote  $(y(n), y(a), y(h), u, c(0), c(1)) \in \mathcal{W}$ , i.e. a generic value in the sample space of the summary random variable.

The objective of the analysis is to learn about the various parameters of interest described in Section 3.3. The analysis exploits the fact that each of these parameters can be rewritten in terms of functions of  $Q$ . In particular, we can show that each parameter can be written as  $\theta(Q)$  such that

$$\theta(Q) = \frac{\sum_{w \in \mathcal{W}} a_{\text{num}}(w) \cdot Q(w)}{\sum_{w \in \mathcal{W}} a_{\text{den}}(w) \cdot Q(w)} , \quad (14)$$

where  $a_{\text{num}} : \mathcal{W} \rightarrow \mathbf{R}$  and  $a_{\text{den}} : \mathcal{W} \rightarrow \mathbf{R}$  are known (or estimated) functions, i.e. we can write each parameter as a fraction of linear functions of  $Q$ . For example, this can be straightforwardly observed for the parameter in (8) that can be re-written as a linear function, a special case of linear-fractional functions where the denominator takes a value of one by taking  $a_{\text{den}}(w) = 1$  for all  $w \in \mathcal{W}$ , as follows

$$\text{ATE}^D(C_+, C_-)(Q) = \sum_{z \in \mathcal{Z}} \sum_{w \in \mathcal{W}_z^h} \text{Prob}\{Z = z\} \cdot Q(w) , \quad (15)$$

where  $\mathcal{W}_z^h = \{w \in \mathcal{W} : d(u, c(z) \cup \{h\}) = h\}$  for each  $z \in \mathcal{Z}$ . With some algebra, we can similarly derive these functions for the remaining parameters as I illustrate in Appendix S.7.3. Indeed, note that the parameters that do not condition on any event such as that in (8) can be written as linear parameters, while those that condition on some event can be written more generally as linear-fractional parameters.

Given that the function  $\theta$  is known for each parameter, what we can learn about that parameter is determined by the possible values that  $Q$  can take. These values are restricted by the distribution of the data and the imposed assumptions. The distribution of the data imposes restrictions on the conditional probability mass functions  $Q_z$  for  $z \in \mathcal{Z}$  through the structure of the model. These restrictions can formally be stated as

$$\sum_{w \in \mathcal{W}_x} Q_z(w) = \text{Prob}\{Y = y, D = d | Z = z\} \quad (16)$$

for all  $x = (y, d, z) \in \mathcal{Y} \times \mathcal{D} \times \mathcal{Z} \equiv \mathcal{X}$ , where  $\mathcal{W}_x$  is the set of all  $w$  in  $\mathcal{W}$  such that  $c = c(1)$  if  $z = 1$  and  $c = c(0)$  if  $z = 0$ ,  $d(u, c) = d$  and  $y = y(d)$ . More specifically,  $\mathcal{W}_x$  is the set of all underlying values in  $\mathcal{W}$  that could have possibly generated the observed value  $x \in \mathcal{X}$  by the outcome equation in (4) and the selection equation in (3). Intuitively, these restrictions capture the information the data reveals through the structure of the model about the child's underlying choice sets as well as on their underlying test scores and preferences.

Assumption HSIS(i) imposes restrictions on how  $Q_z$  for  $z \in \mathcal{Z}$  are related to  $Q$ . Specifically, since Assumption HSIS(i) states that the treatment group assignment indicator  $Z$  is statistically independent of the underlying latent random variables summarized by  $W$ , it imposes that

$$Q_z(w) = Q(w) \quad (17)$$

for all  $w \in \mathcal{W}$  and  $z \in \mathcal{Z}$ , i.e. the distribution of the underlying random variable is invariant to the treatment assignment status. The restrictions in (16) and (17) can then together be used to state the following restrictions imposed on  $Q$  by the observed data

$$\sum_{w \in \mathcal{W}_x} Q(w) = \text{Prob}\{Y = y, D = d | Z = z\} \quad (18)$$

for all  $x = (y, d, z) \in \mathcal{X}$ , where  $\mathcal{W}_x$  is defined as in (16). To see the restrictions imposed by Assumption HSIS(ii), note first that it can be equivalently stated as  $\text{Prob}\{h \notin C(1)\} = 0$ , i.e. zero probability on the occurrence of events where Head Start is not in the potential choice set had the child been assigned to the treatment group. It can straightforwardly be seen that this statement can then be re-written in terms of a restriction on  $Q$  as

$$\sum_{w \in \mathcal{W}_{\text{HSIS}}} Q(w) = 0, \quad (19)$$



where  $\mathcal{W}_{\text{HSIS}} = \{w \in \mathcal{W} : h \notin c(1)\}$  denotes the set of all underlying values such that potential choice set under the treatment group does not contain Head Start.

The various restrictions imposed on  $Q$  by the data and the baseline model may not be generally sufficient to restrict the values that  $\theta(Q)$  can take and, in turn, may not generally allow us to reach informative conclusions. To this end, it may be useful to consider additional assumptions on the baseline model that can further restrict the values that  $Q$  takes. I consider several such assumptions in Section 3.5 below. Denoting by  $\mathcal{S}$  a finite set of restrictions imposed on  $Q$  by these various assumptions, each of the restrictions  $s \in \mathcal{S}$  can be given by

$$\sum_{w \in \mathcal{W}_s} Q(w) = 0 , \quad (20)$$

where  $\mathcal{W}_s$  is some known subset of  $\mathcal{W}$ , i.e. there are only a finite number of restrictions imposed and that each restriction imposes zero probability on the occurrence of certain events. Observe that the restriction imposed by Assumption **HSIS(ii)** stated in (19) corresponds to a special case of (20) taking  $\mathcal{W}_s = \mathcal{W}_{\text{HSIS}}$ . To this end, more generally, let  $\mathcal{S}$  correspond to the set of all restrictions that are imposed on  $Q$  by the various assumptions including Assumption **HSIS(ii)**.

Given the above restrictions imposed by the distribution of the data and by the assumptions on the unknown  $Q$ , what we can learn about  $\theta(Q)$  can formally be captured by the identified set, i.e. the set of parameter values that can be generated by all  $Q$  satisfying the various imposed restrictions. Denoting by  $\mathbf{Q}_{\mathcal{W}}$  the set of all probability mass functions on the sample space  $\mathcal{W}$ , the identified set can be defined as follows

$$\Theta = \{\theta_0 \in \mathbf{R} : \theta(Q) = \theta_0 \text{ for some } Q \in \mathbf{Q}\} , \quad (21)$$

where

$$\mathbf{Q} = \{Q \in \mathbf{Q}_{\mathcal{W}} : Q \text{ satisfies (18), and (20) for each } s \in \mathcal{S}\} \quad (22)$$

is the set of all  $Q$  that satisfy the various restrictions imposed by the data and the assumptions. Note that, by definition, the identified set sharply captures all that we can learn about the parameter given the data and the imposed assumptions. In turn, if the parameter is point identified then the identified set corresponds to the point identified parameter value, whereas if the parameter is partially identified then the identified set corresponds to the sharpest set of parameter values that can be generated by the model given the data and imposed assumption.

In general, providing a sharp characterization of the identified set in analytical manner for each parameter  $\theta(Q)$  can be difficult due to the large number of restrictions that the data and the assumptions impose on  $Q$  and due to the structure of the function  $\theta(Q)$ . In the following proposition, I describe how a linear programming procedure can be used to computationally do so. In this proposition, I introduce the following additional quantity  $\tilde{\theta}(Q) \equiv \sum_{w \in \mathcal{W}} a_{\text{num}}(w) \cdot Q(w)$ .

**Proposition 3.1.** Suppose that  $\mathbf{Q}$  in (22) is non-empty and the parameter characterized by (14) is such that

$$\sum_{w \in \mathcal{W}} a_{\text{den}}(w) \cdot Q(w) > 0 \quad (23)$$

holds for every  $Q \in \mathbf{Q}$ . Then the identified set in (21) can be written as  $\Theta = [\theta_l, \theta_u]$ , where the endpoints of this interval are solutions to the following two linear programming problems

$$\theta_l = \min_{\gamma, \{Q(w)\}_{w \in \mathcal{W}}} \tilde{\theta}(Q) \quad \text{and} \quad \theta_u = \max_{\gamma, \{Q(w)\}_{w \in \mathcal{W}}} \tilde{\theta}(Q) , \quad (24)$$

subject to the following constraints:

- (i)  $\gamma \geq 0$  .
- (ii)  $0 \leq Q(w) \leq \gamma$  for every  $w \in \mathcal{W}$  .
- (iii)  $\sum_{w \in \mathcal{W}} Q(w) = \gamma$  .
- (iv)  $\sum_{w \in \mathcal{W}_x} Q(w) = \gamma \cdot \text{Prob}\{Y = y, D = d | Z = z\}$  for every  $x = (y, d, z) \in \mathcal{X}$  .
- (v)  $\sum_{w \in \mathcal{W}_s} Q(w) = 0$  for every  $s \in \mathcal{S}$  .
- (vi)  $\sum_{w \in \mathcal{W}} a_{\text{den}}(w) \cdot Q(w) = 1$  .

Given the structure of the linear-fractional parameters and the linear restrictions imposed on the probability mass function  $Q$ , Proposition 3.1 illustrates that the identified set for each parameter is an interval and that the bounds of this interval can be computed by solving two linear programming problems. Computing the identified set using Proposition 3.1 requires two conditions. First, it requires  $\mathbf{Q}$  to be non-empty, i.e. the model is correctly specified, to ensure that the identified set is a non-empty interval. If this is not the case, the linear programs automatically terminate. Second, it requires the denominator of the parameter of interest to be positive for every  $Q \in \mathbf{Q}$  to ensure that the parameter is well-defined for all feasible model distributions. This condition can easily be verified in practice. To see how, note that the denominator is also a linear-fractional parameter and more specifically a linear parameter. The above proposition can hence be employed to compute the lower bound for this auxiliary parameter to check if it is strictly positive.

Linear programming procedures have also been previously proposed in related treatment effect and discrete choice models as a convenient and practical approach for obtaining the identified set for various parameters of interest and, specifically, in cases where obtaining analytical bounds are difficult to derive—see, for example, Balke and Pearl (1997), Demuyneck (2015), Freyberger and Horowitz (2015), Kline and Tartari (2016), Laffers (2013), Manski (2007, 2014), Mogstad et al. (2018) and Torgovitsky (2019a,b). However, note that the underlying arguments

that justify these procedures in these papers do not immediately apply to all the parameters defined in the previous section. In particular, while these arguments apply to parameters that can be written as linear functions of  $Q$ , they do not apply to those that can be written as only linear-fractional functions of  $Q$ . In the latter case, as further illustrated in the proof of Proposition 3.1, one needs to first make the observation that the identified set can generally be instead written as solutions to two so-called linear-fractional programs. This observation then allows invoking results from Charnes and Cooper (1962) which imply that the following transformation

$$\tilde{Q}(w) = \gamma \cdot Q(w) \quad \text{where} \quad \gamma = \frac{1}{\sum_{w \in \mathcal{W}} a_{\text{den}}(w) \cdot Q(w)} ,$$

can be used to transform the linear-fractional programs to equivalent linear programs—see also Russell (2021) who independently makes a related observation in an alternative treatment effect model.

### 3.5 Additional Assumptions on the Baseline Model

The baseline model proposed in Section 3.1 imposed only Assumption HSIS and left the dependence between the variables across the three stages completely unrestricted. However, as we will observe in the empirical analysis in Section 4, the bounds for some of the parameters under the baseline model can be wide in terms of reaching informative conclusions regarding the effects of providing a Head Start offer. To this end, as previously highlighted above, an additional benefit of the proposed model is that it allows considering easy-to-interpret assumptions on how the various variables may be related. I conclude this section with describing several such reasonable nonparametric assumptions, which I find allow us to reach stronger conclusions. As I illustrate in Appendix S.7.4, each of these assumptions, similar to Assumption HSIS(ii), can be re-written as restrictions on  $Q$  in the form of (20). As a result, the general linear programming procedure already described in Proposition 3.1 can continue to be used to characterize the identified sets for the various parameters in the presence of these additional assumptions.

I consider the following additional assumptions that impose restrictions on the dependence between the variables across the three stages of the model in various ways.

**Assumption UA.** (Unaltered Alternative)  $a \in C(0)$  if and only if  $a \in C(1)$  .

**Assumption NC.** (Nested Choice)  $d(U, \{n, h\}) = h$  if and only if  $d(U, \{n, a\}) = a$  .

**Assumption MTR.** (Monotone Treatment Response)  $Y(h) \geq Y(n)$  and  $Y(h) \geq Y(a)$  .

**Assumption Roy.** For each  $d, d' \in \mathcal{D}$ , if  $Y(d') > Y(d)$  then  $d(U, \{d, d'\}) = d'$  .

Assumption **UA** states that assignment to either the treatment or control group does not affect whether the child received an offer to an alternative preschool. This is in contrast to the baseline model that allowed whether such an offer was received or not to vary across the two experimental groups. As a result, Assumption **UA** rules out cases, for example, where being assigned to the control group induces parents to apply and receive an offer from alternative preschools. Assumption **NC** states that the parents prefer Head Start to home care if and only if they also prefer alternative preschools to home care. It is motivated by the idea that parents may have preferences where they prefer to enroll their child in any preschool, Head Start or an alternative one, to home care or vice versa. Assumption **MTR**, based on the monotone treatment response assumption from [Manski \(1997\)](#), states that the test scores under Head Start cannot be worse off than that under an alternative preschool or home care. It is motivated by the notion that since Head Start is considered to be a high quality care setting, it may be reasonable to believe that it is at least as good as alternative preschools and home care with respect to test scores. Finally, Assumption **Roy** states that parents prefer to enroll their child in the care setting with the higher test score. In particular, it takes preferences to be based solely on maximizing outcomes and is in the spirit of the Roy selection model ([Heckman and Honore, 1990](#); [Mourifie et al., 2020](#)).

The above assumptions impose easy-to-interpret restrictions that may reasonably hold for certain children in the population. However, one can argue that it seems potentially strong to assume that they hold for every child in the population. This is specifically the case with Assumptions **MTR** and **Roy**, which are arguably stronger than Assumptions **UA** and **NC**. To this end, I also consider weaker versions of these assumptions, where I assume that they hold for only some pre-specified proportion of children. More formally, I allow the probability mass function of the summary random variable to be given by

$$Q(w) = \lambda \cdot H_1(w) + (1 - \lambda) \cdot H_0(w) \quad (25)$$

for each  $w \in \mathcal{W}$ , where  $\lambda \in [0, 1]$  corresponds to the known proportion of children for which either Assumptions **MTR** or **Roy** hold, and  $H_1$  and  $H_0$  correspond to the probability mass functions of the children for which these assumptions hold and may not necessarily hold, respectively. In particular, I suppose that the restrictions imposed by these assumptions of the form (20) are imposed only on  $H_1$  and not on  $H_0$ , which are otherwise identical in term of the mass assigned to preferences and choice sets. In Appendix [S.3](#), I illustrate how the procedure in Proposition [3.1](#) can be straightforwardly modified to account for these weaker versions of Assumptions **MTR** or **Roy**.

Indeed, if  $\lambda$  is taken to be equal to one, these weaker versions are equivalent to Assumptions **MTR** or **Roy**. However, by taking this proportion to be equal to values strictly smaller than one, one can make these assumptions flexibly weaker to a given level. Alternatively, these flexible relaxations can be viewed as a way to analyze the sensitivity of the empirical conclusions to the various additional assumptions—similar in spirit to the breakdown point analysis considered in [Horowitz and Manski \(1995\)](#), [Kline and Santos \(2013\)](#) and [Masten and Poirier \(2020\)](#).

Table 1: Summary statistics

| Panel (a): Proportion in each care setting by experimental group |                  |                  |
|--|------------------|------------------|
| Care setting   | Treatment        | Control          |
| Home care  | 0.093            | 0.542            |
| Alternative preschool  | 0.082            | 0.318            |
| Head Start   | 0.825            | 0.140            |
| Sample size  | 2290             | 1337             |
| Panel (b): ITT and IV Estimands                                  |                  |                  |
| Estimand   | Discretized      | Raw              |
| ITT <sup>D</sup>   | 0.685<br>(0.018) |                  |
| ITT <sup>Y</sup>   | 0.170<br>(0.035) | 0.151<br>(0.030) |
| IV   | 0.249<br>(0.051) | 0.220<br>(0.044) |

*Notes:* Parentheses report standard errors clustered at Head Start center level.

## 4 Empirical Analysis

In this section, I apply the framework developed in the previous section to the HSIS data to analyze the average effects of a policy providing an offer to participate in Head Start.

### 4.1 Data and Summary Statistics

As noted in Section 2, the HSIS collected data for two age cohorts in Spring 2003. In the empirical analysis, following Kline and Walters (2016), I use a version of the data that pools the two cohorts together. Following Feller et al. (2016) and Kline and Walters (2016), I take the participation decision to be a categorized version of an administratively coded focal care setting variable that evaluated care setting attendance for the entire year. I take the test score outcome to be a discretized version of that used in Kline and Walters (2016). They use a summary index of cognitive test scores as measured by the average of the Woodcock Johnson III and the Peabody Picture and Vocabulary Test test scores, where each score is standardized to have mean zero and variance one in the control group for each cohort. Given this summary index, I discretize it using the quantiles of its empirical distribution to take ten support points.<sup>3</sup> See Appendix S.5 for further details on how the analysis sample was constructed and how the discretization was performed.

I begin by presenting some summary statistics of interest. Table 1 Panel (a) reports the participation shares across the different types of care settings by treatment and control groups. A large

<sup>3</sup>In unreported results, I compute the bounds in Table 2 using 15 and 20 support points, and find that the numerical results are relatively similar and that the qualitative conclusions remain unchanged.

proportion (82.5%) of the treatment group participated in Head Start, which suggests that many prefer to participate in Head Start over their available options. Comparing this to that participating in Head Start in the control group suggests that many control children may not have had a Head Start offer. However, given that this proportion is not zero, this suggests that at least some proportion (14%) in fact received an offer even when assigned to the control group.

Table 1 Panel (b) reports estimates for the intention to treat estimands on Head Start participation ( $ITT^D$ ) and test scores ( $ITT^Y$ ), i.e. the difference in the Head Start participation rates and average test scores between the treatment and control groups, respectively, as well as the instrumental variable (IV) estimand that takes their ratio, i.e.  $ITT^Y/ITT^D$ . As  $ITT^Y$  does not require the test scores to be discrete, I also report results under the raw undiscretized version of the test scores. The results under both versions are relatively similar. Since some children in the control group received a Head Start offer, recall that these estimands, and specifically the IV estimand as shown in Appendix S.1, allow us to evaluate only the local average effect of providing a Head Start offer for the subgroup of compliers who take up the offer. We can observe that an offer results in a positive effect of 0.249 standard deviation points for this specific subgroup. These effects are highlighted in the original report in Puma et al. (2010), and further analyzed in Feller et al. (2016) and Kline and Walters (2016).

As noted in Section 2, these local effects generally capture solely the policy effects of providing a Head Start offer through assignment to the treatment and control groups of the experiment. Below, I apply the developed framework to analyze the average effects of providing an offer unconditional of the complier group, which allows us to learn about the policy effects of providing an offer beyond the experimental assignment.

## 4.2 Effects of a Head Start Offer

Table 2 reports estimated identified sets and 95% confidence intervals for the various parameters evaluating the average effects of a Head Start offer defined in Section 3.3. Each column corresponds to a specification of the model determined by which of the auxiliary assumptions from Section 3.5, or their weaker versions in the form of (25), are imposed on the baseline model. The identified sets are estimated by applying the linear programming procedure from Proposition 3.1, or the more general version from Appendix S.3 in the case of the weaker versions of the assumptions, to the empirical distribution of the data. The confidence intervals are constructed using a subsampling procedure from Kalouptsi et al. (2020) described in Appendix S.4.

$ATE^D(C_+, C_-)$  under the baseline specification in Column (1) reveals that providing a Head Start offer affects a large percentage of children, between 70.8% and 85.5%, who take up the offer and participate in Head Start. This highlights that a large proportion of parents prefer Head Start over their other available options. Under Assumption UA from Column (2) onwards, this

Table 2: Average effects of a Head Start offer

| Assumption        | (1)    | (2)    | (3)    | (4)   | (5)   | (6)    | (7)   | (8)   | (9)    |
|-------------------|--------|--------|--------|-------|-------|--------|-------|-------|--------|
| UA                |        | ✓      | ✓      | ✓     | ✓     | ✓      | ✓     | ✓     | ✓      |
| NC                |        |        | ✓      | ✓     | ✓     | ✓      | ✓     | ✓     | ✓      |
| MTR( $\lambda$ )  |        |        |        | 1     | 0.9   | 0.8    |       |       |        |
| Roy( $\lambda$ )  |        |        |        |       |       |        | 1     | 0.9   | 0.8    |
| Parameter         |        |        |        |       |       |        |       |       |        |
| $ATE^D(C_+, C_-)$ | 0.688  | 0.805  | 0.805  | 0.805 | 0.805 | 0.805  | 0.805 | 0.805 | 0.805  |
|                   | 0.708  |        |        |       |       |        |       |       |        |
|                   | 0.855  | 0.825  | 0.825  | 0.825 | 0.825 | 0.825  | 0.825 | 0.825 | 0.825  |
|                   | 0.875  | 0.845  | 0.845  | 0.845 | 0.845 | 0.845  | 0.845 | 0.845 | 0.845  |
| $ATE^Y(C_+, C_-)$ | -1.233 | -0.213 | -0.213 | 0.070 | 0.018 | 0.026  | 0.070 | 0.018 | 0.026  |
|                   | -1.183 | -0.113 | -0.113 | 0.170 | 0.118 | 0.076  | 0.170 | 0.118 | 0.076  |
|                   | 1.329  | 0.410  | 0.410  | 0.410 | 0.410 | 0.410  | 0.410 | 0.410 | 0.410  |
|                   | 1.379  | 0.460  | 0.460  | 0.460 | 0.460 | 0.460  | 0.460 | 0.460 | 0.460  |
| $ATOP(C_+, C_-)$  | -1.669 | -0.237 | -0.237 | 0.106 | 0.043 | -0.008 | 0.106 | 0.043 | -0.008 |
|                   | -1.619 | -0.137 | -0.137 | 0.206 | 0.143 | 0.092  | 0.206 | 0.143 | 0.092  |
|                   | 1.788  | 0.497  | 0.497  | 0.497 | 0.497 | 0.497  | 0.497 | 0.497 | 0.497  |
|                   | 1.838  | 0.547  | 0.547  | 0.547 | 0.547 | 0.547  | 0.547 | 0.547 | 0.547  |

Notes: For each parameter, the inner (outer) panel reports the lower and upper bounds (lower and upper endpoints of the confidence interval), respectively. Lower and upper bounds are not repeated when they coincide. Assumptions **MTR**( $\lambda$ ) and **Roy**( $\lambda$ ) refer to Assumptions **MTR** and **Roy** imposed on  $\lambda$  proportion of children as in (25).

parameter becomes point identified to 82.5%, the proportion participating in Head Start in the treatment group from Table 1 Panel (a). This arises because, under Assumption **UA**, the child's outside options are the same under the treatment and control groups and, in turn, the proportion participating in Head Start in the treatment group exactly corresponds to the proportion that choose to take up a Head Start offer.

Turning to the effects on test scores,  $ATE^Y(C_+, C_-)$  and  $ATOP(C_+, C_-)$  have estimated bounds that are relatively wide when only Assumptions **UA** or **NC** are imposed. Intuitively, this is because under these specifications the relationship between the test scores and the other model variables remains completely unrestricted. However, we can reach informative conclusions when Assumptions **MTR** or **Roy** are imposed that restrict the relationship between these variables.  $ATOP(C_+, C_-)$  from Columns (4) and (7) reveals that the offer can benefit the children who take it up. In particular, their test scores on average can increase between 0.206 and 0.497 standard deviation points. As noted in Section 3.5, imposing Assumptions **MTR** or **Roy** for all children can be arguably strong. Nonetheless, from Columns (5)-(6) and (8)-(9), we can observe that these conclusions on the positive effects of a Head Start offer can also continue to hold under more flexible specifications that impose these additional assumptions not on all children but on only some given proportion, namely 80% and 90% in this case.



Table 3: Heterogeneity in the availability of an alternative preschool offer

| Parameter              | Specification |        |        |       |        |        |       |        |        |
|------------------------|---------------|--------|--------|-------|--------|--------|-------|--------|--------|
|                        | (1)           | (2)    | (3)    | (4)   | (5)    | (6)    | (7)   | (8)    | (9)    |
| $ATE_{+a}^D(C_+, C_-)$ | 0.000         | 0.535  | 0.535  | 0.535 | 0.535  | 0.535  | 0.535 | 0.535  | 0.535  |
|                        | 0.000         | 0.575  | 0.575  | 0.575 | 0.575  | 0.575  | 0.575 | 0.575  | 0.575  |
|                        | 0.941         | 0.910  | 0.822  | 0.822 | 0.822  | 0.822  | 0.822 | 0.822  | 0.822  |
|                        | 0.961         | 0.930  | 0.842  | 0.842 | 0.842  | 0.842  | 0.842 | 0.842  | 0.842  |
| $ATOP_{+a}(C_+, C_-)$  |               | -1.542 | -1.474 | 0.000 | -0.374 | -0.628 | 0.000 | -0.374 | -0.628 |
|                        | -             | -1.392 | -1.324 | 0.000 | -0.351 | -0.578 | 0.000 | -0.351 | -0.578 |
|                        |               | 1.554  | 1.415  | 1.028 | 1.292  | 1.397  | 1.028 | 1.292  | 1.397  |
|                        |               | 1.704  | 1.565  | 1.178 | 1.442  | 1.547  | 1.178 | 1.442  | 1.547  |
| $ATE_{-a}^D(C_+, C_-)$ |               |        | 0.808  | 0.808 | 0.808  | 0.808  | 0.808 | 0.808  | 0.808  |
|                        | -             | -      | 0.828  | 0.828 | 0.828  | 0.828  | 0.828 | 0.828  | 0.828  |
|                        |               |        | 1.000  | 1.000 | 1.000  | 1.000  | 1.000 | 1.000  | 1.000  |
|                        |               |        | 1.000  | 1.000 | 1.000  | 1.000  | 1.000 | 1.000  | 1.000  |
| $ATOP_{-a}(C_+, C_-)$  |               |        | -0.841 | 0.022 | -0.269 | -0.398 | 0.022 | -0.269 | -0.398 |
|                        | -             | -      | -0.641 | 0.072 | -0.169 | -0.298 | 0.072 | -0.169 | -0.298 |
|                        |               |        | 1.061  | 0.717 | 0.856  | 0.930  | 0.717 | 0.856  | 0.930  |
|                        |               |        | 1.261  | 0.917 | 1.056  | 1.130  | 0.917 | 1.056  | 1.130  |

Notes: For each parameter, the inner (outer) panel reports the lower and upper bounds (lower and upper endpoints of the confidence interval), respectively. ‘-’ indicates (23) is not satisfied. Specifications correspond to columns from Table 2.

### 4.3 Heterogeneity Based on an Alternative Preschool Offer

The above results reveal that a Head Start offer affects a large number of children who take up the offer, and can have a positive effect on their subsequent test scores. To better understand these effects and their driving factors, I next analyze heterogeneity based on whether a child had an offer or not to an alternative (non-Head Start) preschool. In particular, I do so using the parameters in (10)-(13), which analyze the effects of an offer conditional on the subgroup of children who have and do not have an offer to an alternative preschool.

Table 3 reports estimates of the identified set and 95% confidence intervals for these parameters, computed as in Table 2. Note here that the parameters evaluating the effects for the group who do not have an alternative preschool offer may not be well-defined under all the specifications. This is because the probability of this subgroup, corresponding to the conditioning event of these parameters, can potentially be zero given the data and specification. In turn, since this implies that (23) is not satisfied, the linear programming procedure cannot be applied to obtain the bounds for these parameters. However, when Assumption NC is imposed, this is not the case anymore as the probability of the conditioning event is bounded away from zero.

The results reveal that providing a Head Start offer affects more children who don’t have an alternative preschool offer than those who do. In particular, from Column (3),  $ATE_{+a}^D(C_+, C_-)$  and  $ATE_{-a}^D(C_+, C_-)$  reveal that between 57.5% and 82.2% of children with an alternative preschool

offer take up the Head Start offer in comparison to between 82.8% and 100% of those without an alternative preschool offer. Moreover, under Assumptions [MTR](#) or [Roy](#), the results reveal that providing an offer can potentially have positive test score gains for only those who don't have an alternative preschool offer and not for those who do. These results together highlight that the effects of providing an offer are primarily driven by the subgroup of children who do not have an alternative preschool offer and hence any outside preschool option.

#### 4.4 Cost-Benefit Analysis of a Head Start Offer

The motivation to study the above effects of a Head Start offer was to capture the potential costs and benefits of a policy providing an offer. I conclude the empirical analysis by using the various effects to perform a cost-benefit analysis of such a policy, similar in spirit to that performed by [Kline and Walters \(2016\)](#) using the local effects. The analysis can be done by attaching an income value to the test score gains to evaluate the benefits as well as one to the take up of Head Start offer to evaluate the costs. To this end, let  $\rho$  denote the income value for the test score gains and  $\phi_h$  denote the cost of a student participating in Head Start. Moreover, as emphasized in [Kline and Walters \(2016\)](#), it is also important when measuring the costs to account for the potential cost savings associated with the movement of children out of alternative preschools, as these preschools are often publicly funded as well. To this end, let  $\phi_a$  analogously denote the cost of participating in an alternative preschool.

Using these values, the difference in average benefits between a policy that provides a Head Start offer to one that does not can be given by the average difference in test scores under the choice set with and without Head Start, i.e.  $C_+$  and  $C_-$ , times the income value, i.e.

$$AB(C_+, C_-) = \rho \cdot ATE^Y(C_+, C_-) .$$

Similarly, the difference in average costs can be given by the proportion who take up the Head Start offer times the cost of participation along with subtracting the costs associated with those who were induced into Head Start from an alternative preschool, i.e.

$$AC(C_+, C_-) = \phi_h \cdot ATE^D(C_+, C_-) - \phi_a \cdot \text{Prob}\{D_{C_+} = h, D_{C_-} = a\} .$$

Taking the difference of the benefit and cost parameters, we can then define the average surplus

$$AS(C_+, C_-) = AB(C_+, C_-) - AC(C_+, C_-) ,$$

which can be used to perform a cost-benefit of a policy providing an offer. Given values of  $\rho$ ,  $\phi_h$  and  $\phi_a$ , these parameters, similar to [\(8\)-\(13\)](#), can be rewritten as [\(14\)](#). In turn, the linear programming procedure can be continued to be applied to characterize the identified set for these parameters. Following [Kline and Walters \(2016\)](#), I take  $\rho = \kappa \cdot E$ , where  $E$  corresponding to the

Table 4: Cost-benefit analysis of providing a Head Start offer

| Parameter           | Specification |        |          |        |          |        |          |        |          |        |          |        |
|---------------------|---------------|--------|----------|--------|----------|--------|----------|--------|----------|--------|----------|--------|
|                     | (4)           |        | (5)      |        | (6)      |        | (7)      |        | (8)      |        | (9)      |        |
|                     | $\kappa$      |        | $\kappa$ |        | $\kappa$ |        | $\kappa$ |        | $\kappa$ |        | $\kappa$ |        |
|                     | 0.1           | 0.2    | 0.1      | 0.2    | 0.1      | 0.2    | 0.1      | 0.2    | 0.1      | 0.2    | 0.1      | 0.2    |
| AC( $C_+$ , $C_-$ ) | 4,146         | 4,146  | 4,146    | 4,146  | 4,146    | 4,146  | 4,146    | 4,146  | 4,146    | 4,146  | 4,146    | 4,146  |
|                     | 4,346         | 4,346  | 4,346    | 4,346  | 4,346    | 4,346  | 4,346    | 4,346  | 4,346    | 4,346  | 4,346    | 4,346  |
|                     | 5,185         | 5,185  | 5,185    | 5,185  | 5,185    | 5,185  | 5,185    | 5,185  | 5,185    | 5,185  | 5,185    | 5,185  |
|                     | 5,385         | 5,385  | 5,385    | 5,385  | 5,385    | 5,385  | 5,385    | 5,385  | 5,385    | 5,385  | 5,385    | 5,385  |
| AB( $C_+$ , $C_-$ ) | 3,850         | 7,900  | 2,254    | 4,509  | 808      | 1,817  | 3,850    | 7,900  | 2,254    | 4,509  | 808      | 1,817  |
|                     | 5,850         | 11,700 | 4,054    | 8,109  | 2,608    | 5,217  | 5,850    | 11,700 | 4,054    | 8,109  | 2,608    | 5,217  |
|                     | 14,073        | 28,146 | 14,073   | 28,146 | 14,073   | 28,146 | 14,073   | 28,146 | 14,073   | 28,146 | 14,073   | 28,146 |
|                     | 15,673        | 31,146 | 15,673   | 31,146 | 15,673   | 31,146 | 15,673   | 31,146 | 15,673   | 31,146 | 15,673   | 31,146 |
| AS( $C_+$ , $C_-$ ) | -1,335        | 2,715  | -2,931   | -677   | -4,377   | -3,569 | -1,335   | 2,715  | -2,931   | -677   | -4,377   | -3,569 |
|                     | 665           | 6,515  | -1,131   | 2,923  | -2,577   | 31     | 665      | 6,515  | -1,131   | 2,923  | -2,577   | 31     |
|                     | 9,727         | 23,800 | 9,727    | 23,800 | 9,727    | 23,800 | 9,727    | 23,800 | 9,727    | 23,800 | 9,727    | 23,800 |
|                     | 11,327        | 26,800 | 11,327   | 26,800 | 11,327   | 26,800 | 11,327   | 26,800 | 11,327   | 26,800 | 11,327   | 26,800 |

Notes: For each parameter, the inner (outer) panel reports the lower and upper bounds (lower and upper endpoints of the confidence interval), respectively. Specifications correspond to columns from Table 2.

average present discounted value of lifetime earnings for Head Start applicants is taken to be \$34,339 and  $\kappa$  corresponding to the relationship between earnings and a one standard deviation increase in test scores is taken to be either 0.1 or 0.2, which is towards the lower and higher end of the values for this relation noted in the literature (Kline and Walters, 2016, Table A.IV). Moreover, following Kline and Walters (2016), I take  $\phi_h = \$8,000$  and  $\phi_a = 0.75 \cdot \phi_h = \$6,000$ .

Table 4 reports estimated identified sets and 95% confidence intervals for these parameters for a subset of the specifications from Table 2 that impose restrictions on test scores, where the estimates and confidence intervals are computed as in Table 2. The results reveal that the average net costs of providing an offer is relatively low in comparison to the absolute value of participating in Head Start. Similar to that highlighted by Kline and Walters (2016), this arises due to relatively high cost savings from children who enroll out of alternative preschools when provided with a Head Start offer. Taking the benefits into account, the average surplus reveals that whether average benefits outweigh costs depends on the strength of the assumptions and the relation between test scores and earnings. In particular, when Assumptions MTR or Roy are imposed on all children in Column (4) and (7), the benefit outweighs the cost. However, when these assumptions are imposed only on some proportion of the children in the remaining columns, the benefit generally outweighs the cost only when  $\kappa = 0.2$ , i.e. the relation between test scores gains and earnings is taken to be towards the higher end of estimates noted in the literature.

## 5 Conclusion

I develop a framework to evaluate the average effects of providing an offer to participate in a program given data from an experiment that randomizes offers. I allow for the fact that individuals may not comply with their assigned treatment status and that the data may only provide partial information on the receipt of an offer across individuals. The main idea behind the framework is to use a novel selection model to relate whether an offer was received to the observed data, and to then exploit the structure of the model to define and learn about various parameters evaluating the average effects of an offer. I illustrate the framework by analyzing the policy effect of providing a Head Start offer using the Head Start Impact Study (HSIS).

I conclude by highlighting the flexibility of the developed framework. Specifically, it can be broadly applied to different experiments where it can accommodate several institutional details unique to that experiment. In the paper, I demonstrate this feature by developing the framework in the context of the HSIS where I show how the framework accommodated several details such as the presence of alternative preschools and the specific information that the experiment provided on whether an offer was received. I demonstrate this flexibility further in Appendix [S.6](#), where I present the generalized version of the framework and show how it can be applied to accommodate alternative details in two other experiments.

## References

- ABALUCK, J. and ADAMS, A. (2018). What do consumers consider before they choose? identification from asymmetric demand responses. Tech. rep.
- ANGELUCCI, M., KARLAN, D. and ZINMAN, J. (2015). Microcredit impacts: Evidence from a randomized microcredit program placement experiment by compartamos banco. *American Economic Journal: Applied Economics*, **7** 151–82.
- BALKE, A. and PEARL, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, **92** 1171–1176.
- BARSEGHYAN, L., COUGHLIN, M., MOLINARI, F. and TEITELBAUM, J. (2018). Heterogeneous consideration sets and preferences. Tech. rep., Cornell University.
- CATTANEO, M. D., MA, X., MASATLIOGLU, Y. and SULEYMANOV, E. (2020). A random attention model. *Journal of Political Economy*, **128** 2796–2836.
- CHARNES, A. and COOPER, W. W. (1962). Programming with linear fractional functionals. *Naval Research Logistics (NRL)*, **9** 181–186.
- CRAWFORD, G. S., GRIFFITH, R. and IARIA, A. (2020). A survey of preference estimation with unobserved choice set heterogeneity. *Journal of Econometrics*.
- DEMUYNCK, T. (2015). Bounding average treatment effects: A linear programming approach. *Economics Letters*, **137** 75–77.
- FELLER, A., GRINDAL, T., MIRATRIX, L. and PAGE, L. C. (2016). Compared to what? variation in the impacts of early childhood education by alternative care type. *The Annals of Applied Statistics*, **10** 1245–1285.
- FREYBERGER, J. and HOROWITZ, J. L. (2015). Identification and shape restrictions in nonparametric instrumental variables estimation. *Journal of Econometrics*, **189** 41–53.
- HECKMAN, J. J. and HONORE, B. E. (1990). The empirical content of the roy model. *Econometrica: Journal of the Econometric Society* 1121–1149.
- HECKMAN, J. J. and PINTO, R. (2018). Unordered monotonicity. *Econometrica*, **86** 1–35.
- HECKMAN, J. J., URZUA, S. and VYTLACIL, E. (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics*, **88** 389–432.
- HECKMAN, J. J., URZUA, S. and VYTLACIL, E. (2008). Instrumental variables in models with multiple outcomes: The general unordered case. *Annales d’Economie et de Statistique* 151–174.

- HECKMAN, J. J. and VYTLACIL, E. (2001). Policy-relevant treatment effects. *American Economic Review*, **91** 107–111.
- HECKMAN, J. J. and VYTLACIL, E. (2005). Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica*, **73** 669–738.
- HECKMAN, J. J. and VYTLACIL, E. J. (2007). Econometric evaluation of social programs, part i: Causal models, structural models and econometric policy evaluation. *Handbook of econometrics*, **6** 4779–4874.
- HO, K. and ROSEN, A. M. (2017). *Partial Identification in Applied Research: Benefits and Challenges*, vol. 2 of *Econometric Society Monographs*. Cambridge University Press, 307–359.
- HOROWITZ, J. L. and MANSKI, C. F. (1995). Identification and robustness with contaminated and corrupted data. *Econometrica: Journal of the Econometric Society* 281–302.
- KALOUPTSIDI, M., KITAMURA, Y., LIMA, L. and SOUZA-RODRIGUES, E. A. (2020). Partial identification and inference for dynamic models and counterfactuals. Tech. rep., National Bureau of Economic Research.
- KIRKEBOEN, L. J., LEUVEN, E. and MOGSTAD, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, **131** 1057–1111.
- KLINE, P. and SANTOS, A. (2013). Sensitivity to missing data assumptions: Theory and an evaluation of the us wage structure. *Quantitative Economics*, **4** 231–267.
- KLINE, P. and TARTARI, M. (2016). Bounding the labor supply responses to a randomized welfare experiment: A revealed preference approach. *The American Economic Review*, **106** 971–1013.
- KLINE, P. and WALTERS, C. R. (2016). Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics*, **131** 1795–1848.
- LAFFÉRS, L. (2013). A note on bounding average treatment effects. *Economics Letters*, **120** 424–428.
- LEE, S. and SALANIÉ, B. (2018). Identifying effects of multivalued treatments. *Econometrica*, **86** 1939–1963.
- MANSKI, C. F. (1997). Monotone treatment response. *Econometrica: Journal of the Econometric Society* 1311–1334.
- MANSKI, C. F. (2007). Partial identification of counterfactual choice probabilities. *International Economic Review*, **48** 1393–1410.
- MANSKI, C. F. (2014). Identification of income–leisure preferences and evaluation of income tax policy. *Quantitative Economics*, **5** 145–174.

- MARSCHAK, J. (1960). Binary choice constraints on random utility indicators. In *Stanford Symposium on Mathematical Methods in the Social Sciences*. Stanford University Press.
- MASTEN, M. A. and POIRIER, A. (2020). Inference on breakdown frontiers. *Quantitative Economics*, **11** 41–111.
- MOGSTAD, M., SANTOS, A. and TORGOVITSKY, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, **86** 1589–1619.
- MOGSTAD, M. and TORGOVITSKY, A. (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics*.
- MOURIFIE, I., HENRY, M. and MÉANGO, R. (2020). Sharp bounds and testability of a roy model of stem major choices. *Journal of Political Economy*, **128** 3220–3283.
- PINTO, R. (2019). Noncompliance as a rational choice: A framework that exploits compromises in social experiments to identify causal effects. Tech. rep.
- PUMA, M., BELL, S., COOK, R., HEID, C., SHAPIRO, G., BROENE, P., JENKINS, F., FLETCHER, P., QUINN, L., FRIEDMAN, J. ET AL. (2010). Head start impact study. final report. *Administration for Children & Families*.
- RUSSELL, T. M. (2021). Sharp bounds on functionals of the joint distribution in the analysis of treatment effects. *Journal of Business & Economic Statistics*, **39** 532–546.
- TORGOVITSKY, A. (2019a). Nonparametric inference on state dependence in unemployment. *Econometrica*, **87** 1475–1505.
- TORGOVITSKY, A. (2019b). Partial identification by extending subdistributions. *Quantitative Economics*, **10** 105–144.



# Supplementary Appendix to “Identifying the Effects of a Program Offer with an Application to Head Start”

Vishal Kamat  
Toulouse School of Economics  
University of Toulouse Capitole  
[vishal.kamat@tse-fr.eu](mailto:vishal.kamat@tse-fr.eu)

December 14, 2024

## Abstract

This document provides proofs and additional details pertinent to the author’s paper “Identifying the Effects of a Program Offer with an Application to Head Start.” Section [S.1](#) provides a motivating discussion for the developed framework. Section [S.2](#) describes additional parameters capturing alternative policies of potential interest. Section [S.3](#) describes the modified procedure used to compute bounds under weaker forms of the assumptions. Section [S.4](#) describes the procedure used to compute confidence intervals in the empirical analysis. Section [S.5](#) provides additional details pertinent to the empirical analysis. Section [S.6](#) describes the generalized framework and how it applies to two other experiments. Section [S.7](#) provides all proofs and other mathematical derivations.

## S.1 Motivating Discussion for Framework

In this section, I provide motivation for the framework developed in Section 3 by illustrating what standard treatment effect tools can and cannot evaluate with respect to the effects of a Head Start offer in the HSIS. For the purposes of this section, I introduce, with some slight abuse, alternative notation to that introduced in Section 3. Specifically, since the objective is to analyze the average “treatment” effect of a Head Start offer, I introduce potential outcome notation where the “treatment” of interest corresponds to the receipt of an offer. As we will observe below in Section S.1.1, this notation will help show, using more familiar arguments without the concept of choice sets, how standard intention-to-treat and instrumental variable estimands can evaluate only so-called local average effects of an offer—see, for example, Imbens and Rubin (2015, Chapter 23) for a textbook exposition on such notation. Moreover, in Section S.1.2, it will allow us to emphasize why not observing from where an offer is received corresponds to not observing the treatment of interest and, hence, complicates the application of available tools to move beyond local effects. For simplicity, the setup and notation introduced in this section does not account for the possible heterogeneity in the availability of an alternative (non-Head Start) preschool offer.

For a given child, let  $C$  denote the treatment of interest corresponding to an indicator for whether they received a Head Start offer or not, let  $D$  denote an indicator for whether they participated in Head Start or not, and let  $Y$  denote their test score. Let  $D_1$  and  $D_0$  respectively denote the child’s potential indicator for whether they would participate in Head Start with and without an offer, which are related to the realized indicator of Head Start participation by

$$D = D_1C + D_0(1 - C) . \tag{S.1}$$

Analogously, let  $Y_1$  and  $Y_0$  respectively denote the child’s potential test scores with and without an Head Start offer, which are related to the realized test score by

$$Y = Y_1C + Y_0(1 - C) . \tag{S.2}$$

Furthermore, let  $Z$  denote an indicator for whether the child was assigned to the treatment group or not, and let  $C_1$  and  $C_0$  respectively denote the potential indicators for whether an offer was received in the treatment and control group. These potential indicators are related to the realized indicator by

$$C = C_1Z + C_0(1 - Z) . \tag{S.3}$$

The structure of setup and the experimental design introduce certain natural restrictions between the above described variables. I formally state these restrictions in the following assumptions:

**Assumption M1.**  $D_0 = 0$  .

**Assumption M2.**  $D_0 = D_1 \implies Y_0 = Y_1$  .

**Assumption M3.**  $(Y_0, Y_1, D_0, D_1, C_0, C_1) \perp Z$  .

**Assumption M4.**  $C_1 = 1$  .

Assumption **M1** states that if the child did not receive a Head Start offer then they cannot participate in Head Start. Assumption **M2** states that receiving an offer can affect test scores only if it affects the enrollment decisions, i.e. an offer does not directly affect outcomes but only indirectly through affecting the participation decision. Assumption **M3** states that assignment to either the treatment or control group was random. Assumption **M4** states that if the child was assigned to the treatment group then they received an offer. Note that Assumption **M1** and Assumption **M2** capture logical restrictions that naturally follow from the structure of the setup, whereas Assumption **M3** and Assumption **M4** capture restrictions that follow from the design of the experiment.

### S.1.1 Interpreting the ITT and IV Estimands

In the above described setup, the effect of a Head Start offer on participation decisions and test scores for a given child corresponds to  $D_1 - D_0$  and  $Y_1 - Y_0$ , i.e. the difference in potential responses with and without an offer. I begin by illustrating what the standard intention-to-treat (ITT) and instrumental variable (IV) estimands can evaluate with respect to these effects. The ITT estimands on participation decisions and test scores are defined by

$$\text{ITT}_D = E[D|Z = 1] - E[D|Z = 0] , \quad (\text{S.4})$$

$$\text{ITT}_Y = E[Y|Z = 1] - E[Y|Z = 0] , \quad (\text{S.5})$$

i.e. the difference in mean responses between the treatment and control group, whereas the IV estimand is defined by

$$\text{IV} = \frac{\text{ITT}_Y}{\text{ITT}_D} , \quad (\text{S.6})$$

i.e. the ratio of the ITT estimand on test scores to that on participation. Using arguments from [Imbens and Angrist \(1994\)](#) modified to the above setting, the following proposition rewrites the ITT and IV estimands in terms of the underlying potential variables of the setup.

**Proposition S.1.1.** Suppose that Assumptions **M1-M4** hold. Then

$$\begin{aligned} \text{ITT}_D &= E[D_1 - D_0 | C_1 = 1, C_0 = 0] \cdot \text{Prob}\{C_1 = 1, C_0 = 0\} , \\ \text{ITT}_Y &= E[Y_1 - Y_0 | C_1 = 1, C_0 = 0] \cdot \text{Prob}\{C_1 = 1, C_0 = 0\} , \\ \text{IV} &= E[Y_1 - Y_0 | D_1 = 1, C_1 = 1, C_0 = 0] . \end{aligned}$$

The above proposition states that the ITT estimands on participation and test scores respectively evaluate the average effect of a Head Start offer on participation and test scores conditional

on the compliers times the proportion of compliers, i.e. the subgroup of children who comply with their assigned status in the control group and do not receive an offer. Similarly, the IV estimand evaluates the average effect of an offer on test scores for those who would in fact take up the offer and participate but again conditional on the compliers.

In summary, the above proposition shows that the ITT and IV estimands can help evaluate the conditional average effects of a Head Start offer which is conditional on the subgroup of compliers. In some cases, this subgroup is indeed the one of interest. For example, if we wanted to learn the policy effects of providing an offer through assignment to the treatment group relative to the control group in the experiment then the complier subgroup is the only one of interest as they are the only ones whose receipt of an offer is affected by such a policy. However, in other cases, we may instead be interested in the unconditional effect for all the children. For example, if we wanted to learn the effects of a policy providing an offer to all children more generally beyond the experimental assignment. In such cases, the objective is then to learn about  $E[D_1 - D_0]$  and  $E[Y_1 - Y_0]$ , i.e. the (unconditional) average effect of an offer on participation and test scores, and  $E[Y_1 - Y_0 | D_1 = 1]$ , i.e. the (unconditional) average effect of an offer on test scores for those who in fact take up the offer. But, unless every control parent complies with their assigned status and in turn the complier subgroup coincided with the entire population, i.e.  $\text{Prob}\{C_1 = 1, C_0 = 0\} = 1$ , the ITT and IV estimands cannot evaluate the average effects of an offer unconditional of the complier subgroup. In the setting of the HSIS, it is indeed the case that some control parents do not comply with their assigned status, which is indirectly revealed by the fact that some parents enroll their child into Head Start even when assigned to the control group as noted when presenting the summary statistics in Section 4.1.

### S.1.2 Moving Beyond Local Effects

More generally, following the language of the treatment effect literature, we can observe that the ITT and IV estimands allow us to only evaluate local average treatment effects (LATE) but not average treatment effects (ATE) of providing a Head Start offer. In this literature, a number of tools have been developed to move beyond the LATE and evaluate the ATE—see, for example, [Mogstad and Torgovitsky \(2018, Section 6\)](#) for a recent overview. Nonetheless, due to certain observational problems that arise in the setting of the HSIS, these tools generally cannot be applied to evaluate the average effects of a Head Start offer in the above described setup. To be specific, when applied to the above described setup, these tools typically require the distribution of  $(Y, D, C, Z)$  to be observed. However, in the setting of the HSIS, this is not the case. Specifically, while the experiment collected data on the test score, the participation decision and the treatment assignment status, it did not collect data on the receipt of an offer. As a result, while we observe the values of  $Y$ ,  $D$  and  $Z$  for every child, we do not observe the value of  $C$  for any child. In other words, we do not observe the value of the “treatment” of interest whose average effect on participation decisions

and test scores we want to evaluate.

The objective of the framework developed in Section 3 is to study what we can learn about the average effects of an offer in such cases. In particular, it aims to show how we can exploit only the distribution of  $(Y, D, Z)$  to do so. The framework is based on the idea that while the offer receipt is not directly observed, the structure of the model implies partial information on whether an offer is received across children. For example, in the above described model, the relation in (S.1) and Assumption M1 imply that  $D = 1 \implies C = 1$ , i.e. the child’s participation in Head Start reveals that they received an offer. Similarly, the relation in (S.3) and Assumption M4 imply that  $Z = 1 \implies C = 1$ , i.e assignment to the treatment group reveals that the child received an offer. The framework then aims to show how to exploit this partial information through these various implications and learn about the average effects of an offer.

As mentioned, the above model did not account for the presence of an alternative preschool offer for simplicity. The model presented in the framework in Section 3 shows how to expand on the above described setup and notation, and introduce their presence to study the average effects of a Head Start offer based on their availability. Similar to a Head Start offer, we do not directly observe whether children received an offer to alternative preschools and instead only observe whether they participated in an alternative preschool. The developed framework shows how the structure of the extended model can be similarly used to exploit this partial information and learn about the average effects of a Head Start offer based on the availability of an alternative preschool offer.

While the above described treatment effect model was useful to motivate the problem behind learning the average effects of an offer, note that the framework in Section 3 proposes an alternative more convenient representation of the model. In particular, it introduces a structural, more economically-grounded choice model to capture the various above implications on whether an offer was received and learn about the effects of an offer as well as consider extensions such as accounting for the availability of alternative preschool offers or imposing additional assumptions. More specifically, the framework begins with a distinct set of care settings and the insight that receiving an offer from a given preschool can be framed as receiving that preschool in the choice set of care settings from which participation decisions are made. By exploiting this insight, it then shows how a selection model can be used to relate from where an offer was received to the observed treatment assignment status, participation decision and test score. As highlighted in Section 3.2, the proposed model has several benefits towards straightforwardly analyzing the effects of an offer. First, in this model, the relationship between the various variables such as those in (S.1)-(S.3) and the logical restrictions from the setup such as Assumption M1 and Assumption M2 can be automatically captured through the structure of the model, and the restrictions imposed by the experimental design such as Assumption M3 and Assumption M4 can be intuitively captured through Assumption HSIS. Second, the model allows defining a range of parameters evaluating the average effects of a Head Start offer by comparing choice sets with and without Head Start. Finally, it also allows one to

consider a range of additional easy-to-interpret identifying assumptions that allows reaching more informative conclusions on the effects of an offer.

## S.2 Parameters Capturing Alternative Policies

As previously highlighted, the parameters in Section 3.3 compared responses under choice sets  $C_+$  and  $C_-$  with the aim of capturing the effects of a policy that provides a Head Start offer relative to one that does not. In this section, I define some other parameters based on comparing alternative choice sets that capture alternative policies of potential interest. Similar to those in Section 3.3, note that each of these parameters can be rewritten in terms of (14). In turn, the linear programming procedure described in Section 3.4 can be continued to be applied to characterize the identified set for these parameters.

Instead of comparing the responses under choice sets  $C_+$  and  $C_-$  to each other, we can compare them to those under the child's obtained choice set  $C$ . Since  $C$  corresponds to the child's choice set under the status quo regime present in the data, these comparisons can be viewed as capturing the policy effects of providing a Head Start offer or not relative to the status quo regime. In particular, analogous to (7)-(9), we can define

$$\text{ATE}^D(C_+, C) = \text{Prob}\{D_{C_+} = h, h \notin C\} , \quad (\text{S.7})$$

$$\text{ATE}^Y(C_+, C) = E[Y_{C_+} - Y_C] , \quad (\text{S.8})$$

$$\text{ATOP}(C_+, C) = E[Y_{C_+} - Y_C | D_{C_+} = h, h \notin C] \quad (\text{S.9})$$

to capture the effects of providing an offer an offer relative to the status quo, and

$$\text{ATE}^D(C, C_-) = \text{Prob}\{D_C = h, h \in C\} , \quad (\text{S.10})$$

$$\text{ATE}^Y(C, C_-) = E[Y_C - Y_{C_-}] , \quad (\text{S.11})$$

$$\text{ATOP}(C, C_-) = E[Y_C - Y_{C_-} | D_C = h, h \in C] \quad (\text{S.12})$$

to capture the effects of the status relative to not providing an offer. Note that relative to (7)-(9) we additionally have the expression  $h \notin C$  in (S.7)-(S.9) as  $1\{D_{C_+} = h\} - 1\{D_C = h\} = 1\{D_{C_+} = h, h \notin C\}$ , which captures that fact that if the child already had an offer under the status quo regime they would not be affected. Similarly, in (S.10)-(S.12), we additionally have the expression  $1\{h \in C\}$  as  $1\{D_C = h\} - 1\{D_{C_-} = h\} = 1\{D_C = h, h \in C\}$ , which captures the fact that if the child did not have an offer under the status quo they would be not be affected when an offer is not provided.

The above parameters did not change whether the child received an offer or not to an alternative (non-Head Start) preschool. This was based on the idea that while the policymaker has control on providing an offer or not to Head Start, they may not have so for alternative preschools. However,

since these alternative schools are often partially government funded, it is possible that providing an offer to them could also be a policy variable. To this end, one could also be interested in evaluating parameters that evaluate the effects of a Head Start offer with and without an alternative school to capture the policy effect of a Head Start when no offer to an alternative school is made available and when it is. To this end, when an offer to an alternative is available, the choice sets with and without a Head Start offer are respectively given by  $C_+ \cup \{a\} = \{h, a, n\}$  and  $C_- \cup \{a\} = \{a, n\}$ . In turn, we can define

$$\text{ATE}^D(\{h, a, n\}, \{a, n\}) = \text{Prob}\{D_{\{h, a, n\}} = h\} , \quad (\text{S.13})$$

$$\text{ATE}^Y(\{h, a, n\}, \{a, n\}) = E[Y_{\{h, a, n\}} - Y_{\{a, n\}}] , \quad (\text{S.14})$$

$$\text{ATOP}(\{h, a, n\}, \{a, n\}) = E[Y_{\{h, a, n\}} - Y_{\{a, n\}} | D_{\{h, a, n\}} = h] \quad (\text{S.15})$$

to capture the effects of providing a Head Start offer when an alternative preschool offer is available. Similar, when an offer to an alternative preschool is not available, the choice sets with and without a Head Start offer are respectively given by  $C_+ \setminus \{a\} = \{h, n\}$  and  $C_- \setminus \{a\} = \{n\}$ , and, in turn, we can define

$$\text{ATE}^D(\{h, n\}, \{n\}) = \text{Prob}\{D_{\{h, n\}} = h\} , \quad (\text{S.16})$$

$$\text{ATE}^Y(\{h, n\}, \{n\}) = E[Y_{\{h, n\}} - Y_{\{n\}}] , \quad (\text{S.17})$$

$$\text{ATOP}(\{h, n\}, \{n\}) = E[Y_{\{h, n\}} - Y_{\{n\}} | D_{\{h, n\}} = h] \quad (\text{S.18})$$

to capture the effects of providing a Head Start offer when an alternative preschool offer is absent.

### S.3 Modified Linear Program

In this section, I describe how the linear programming procedure from Proposition 3.1 can be modified to allow for the weaker versions of Assumptions MTR and Roy in the form of (25). The arguments behind this modified procedure are similar to those of Proposition 3.1 presented in Section 3.4 with a few differences. Instead of the unknown quantity being only the probability mass function  $Q$ , there are two additional auxiliary unknown probability mass functions  $H_0$  and  $H_1$  with support similarly contained in  $\mathcal{W}$ , i.e.  $H_0 : \mathcal{W} \rightarrow [0, 1]$  and  $H_1 : \mathcal{W} \rightarrow [0, 1]$  such that  $\sum_{w \in \mathcal{W}} H_0(w) = 1$  and  $\sum_{w \in \mathcal{W}} H_1(w) = 1$ .

Given these additional unknown quantities, we need to introduce the additional restrictions over them which then indirectly introduce restrictions on  $Q$  through their relationship to it. To this end, from Section 3.5, recall first that  $H_1$  and  $H_0$  are identical in terms of mass associated to preferences and choice sets, i.e.

$$\sum_{(y(n), y(a), y(h)) \in \mathcal{Y}^3} H_0(w) = \sum_{(y(n), y(a), y(h)) \in \mathcal{Y}^3} H_1(w) \quad (\text{S.19})$$



for each  $(u, c(0), c(1)) \in \mathcal{U} \times \mathcal{C} \times \mathcal{C}$ , where  $w = (y(n), y(a), y(h), u, c(0), c(1))$ . Second, note that when Assumptions **MTR** or **Roy** are additionally imposed on the baseline model, the restrictions it introduces are imposed only on  $H_1$ . More specifically, let  $\mathcal{S}_1$  denote the set of restrictions imposed on the  $H_1$  such that each restriction  $s \in \mathcal{S}_1$  satisfies

$$\sum_{w \in \mathcal{W}_s} H_1(w) = 0, \quad (\text{S.20})$$

where  $\mathcal{W}_s$  is a known subset of  $\mathcal{W}$ . Given these restrictions on the auxiliary probability mass functions, the relationship between them and  $Q$  then captures how they indirectly impose restrictions on  $Q$ . To this end, recall that the relationship between them and  $Q$  is captured through the following restrictions

$$Q(w) = \lambda \cdot H_1(w) + (1 - \lambda) \cdot H_0(w) \quad (\text{S.21})$$

for each  $w \in \mathcal{W}$ , where  $\lambda \in [0, 1]$  is a pre-specified known proportion. To summarize, what we know about  $Q$  from these restrictions imposed through the auxiliary distributions can then be captured by the following set

$$\mathbf{Q}_{\lambda, \text{aux}} = \{Q \in \mathbf{Q}_{\mathcal{W}} : Q \text{ satisfies } (\text{S.21}); H_0, H_1 \in \mathbf{Q}_{\mathcal{W}} \text{ satisfy } (\text{S.19})\text{--}(\text{S.20}) \text{ for } s \in \mathcal{S}_1\}, \quad (\text{S.22})$$

where recall that  $\mathbf{Q}_{\mathcal{W}}$  is the set of all probability mass functions on the sample space  $\mathcal{W}$ .

Using the notation from Section 3.4, the identified set for a pre-specified parameter  $\theta(Q)$  and for a pre-specified  $\lambda$  can then be written as follows

$$\Theta_\lambda = \{\theta_0 \in \mathbf{R} : \theta(Q) = \theta_0 \text{ for some } Q \in \mathbf{Q}_\lambda\}, \quad (\text{S.23})$$

where

$$\mathbf{Q}_\lambda = \mathbf{Q} \cap \mathbf{Q}_{\lambda, \text{aux}} \quad (\text{S.24})$$

is the intersection of the set of feasible distributions determined by restrictions directly imposed on  $Q$  by the data and assumptions as described in Section 3.4 and indirectly imposed on  $Q$  through the restrictions imposed on the auxiliary distributions as described above.

Given the linear-fractional structure of the parameter and the linear structure of all the restrictions, the following proposition states the linear program that can be used to compute the identified sets in this case. Indeed, given this structure, the same arguments used in the proof of Proposition 3.1 can be used to formally show this result. Moreover, note that the only difference between the linear program in this proposition and that in Proposition 3.1 is the introduction of additional variables and additional linear restrictions on these variables.

**Proposition S.3.1.** Suppose that  $\mathbf{Q}_\lambda$  in (S.24) is non-empty and the parameter characterized by (14) is such that

$$\sum_{w \in \mathcal{W}} a_{\text{den}}(w) \cdot Q(w) > 0 \quad (\text{S.25})$$

holds for every  $Q \in \mathbf{Q}_\lambda$ . Then the identified set in (S.23) can be written as  $\Theta_\lambda = [\theta_l, \theta_u]$ , where the lower and upper bounds of this interval are solutions to the following two linear programming problems

$$\theta_l = \min_{\gamma, \{Q(w), H_0(w), H_1(w)\}_{w \in \mathcal{W}}} \tilde{\theta}(Q) \quad \text{and} \quad \theta_u = \max_{\gamma, \{Q(w), H_0(w), H_1(w)\}_{w \in \mathcal{W}}} \tilde{\theta}(Q), \quad (\text{S.26})$$

subject to the following constraints:

- (i)  $\gamma \geq 0$ .
- (ii)  $0 \leq Q(w) \leq \gamma$  for every  $w \in \mathcal{W}$ .
- (iii)  $\sum_{w \in \mathcal{W}} Q(w) = \gamma$ .
- (iv)  $\sum_{w \in \mathcal{W}_x} Q(w) = \gamma \cdot \text{Prob}\{Y = y, D = d | Z = z\}$  for every  $x = (y, d, z) \in \mathcal{X}$ .
- (v)  $\sum_{w \in \mathcal{W}_s} Q(w) = 0$  for every  $s \in \mathcal{S}$ .
- (vi)  $\sum_{w \in \mathcal{W}} a_{\text{den}}(w) \cdot Q(w) = 1$ .
- (vii)  $0 \leq H_0(w) \leq \gamma$  for every  $w \in \mathcal{W}$ .
- (viii)  $\sum_{w \in \mathcal{W}} H_0(w) = \gamma$ .
- (ix)  $0 \leq H_1(w) \leq \gamma$  for every  $w \in \mathcal{W}$ .
- (x)  $\sum_{w \in \mathcal{W}} H_1(w) = \gamma$ .
- (xi)  $\sum_{w \in \mathcal{W}_s} H_1(w) = 0$  for every  $s \in \mathcal{S}_1$ .
- (xii)  $Q(w) = \lambda \cdot H_1(w) + (1 - \lambda) \cdot H_0(w)$  for every  $w \in \mathcal{W}$ .
- (xiii)  $\sum_{(y(n), y(a), y(h)) \in \mathcal{Y}^3} H_0(y(n), y(a), y(h), u, c(0), c(1)) = \sum_{(y(n), y(a), y(h)) \in \mathcal{Y}^3} H_1(y(n), y(a), y(h), u, c(0), c(1))$   
for every  $(u, c(0), c(1)) \in \mathcal{U} \times \mathcal{C} \times \mathcal{C}$ .

## S.4 Statistical Inference

In this section, I describe the procedure used to construct the confidence intervals, for the various parameters of interest, presented in the empirical results in Section 4. To this end, let

$$X^{(N)} = \{X_g^{(N)} : 1 \leq g \leq G\} \quad (\text{S.27})$$

denote the HSIS sample data, where  $N$  denotes the total sample size, and

$$X_g^{(N)} = \{(Y_{ig}, D_{ig}, Z_{ig}) : 1 \leq i \leq N_g\}$$

denotes the cluster of observations for all the  $i$ th sampled children from the  $g$ th sampled Head Start preschool from the experiment.

The confidence intervals are constructed by test inversion. In particular, I test at level  $\alpha \in (0, 1)$  the following null hypothesis

$$H_0 : \theta(Q) \equiv \frac{\sum_{w \in \mathcal{W}} a_{\text{num}}(w) \cdot Q(w)}{\sum_{w \in \mathcal{W}} a_{\text{num}}(w) \cdot Q(w)} = \theta_0 \text{ for some } Q \in \mathbf{Q}, \quad (\text{S.28})$$

i.e. the parameter of interest in (14) can be equal to some value  $\theta_0 \in \mathbf{R}$  given the admissible set of distribution  $\mathbf{Q}$ . Confidence intervals of level  $(1 - \alpha)$  for the parameter are then obtained by collecting the values of  $\theta_0$  that are not rejected.

In order to test this null hypothesis, I use a recentered subsampling procedure from Kalouptsi et al. (2020), who show it can have several good theoretical properties. In the context of the setup from Section 3.4, the test can be described as follows:

(i) Compute the test statistic given by

$$TS_N(\theta_0) = N \cdot \min_{\{Q(w)\}_{w \in \mathcal{W}}} \sum_{x \in \mathcal{X}} \left( \hat{P}(x) - \sum_{w \in \mathcal{W}} Q(w) \right)^2 \quad (\text{S.29})$$

subject to the constraints that  $0 \leq Q(w) \leq 1$  for each  $w \in \mathcal{W}$ ,  $\sum_{w \in \mathcal{W}} Q(w) = 1$ , those in (20), and

$$\sum_{w \in \mathcal{W}} \hat{a}_{\text{num}}(w) \cdot Q(w) = \theta_0 \cdot \sum_{w \in \mathcal{W}} \hat{a}_{\text{den}}(w) \cdot Q(w),$$

where  $\hat{P}$  corresponds to the estimated version of  $P$  obtained using the empirical distribution of the data, and, similarly,  $\hat{a}_{\text{num}}$  and  $\hat{a}_{\text{den}}$  correspond to the estimated versions of  $a_{\text{num}}$  and  $a_{\text{den}}$  obtained using the empirical distribution of the data. Indeed, note that the test statistic corresponds to the minimum squared distance from zero of the relation between  $Q$  and the distribution of the data in (18), subject to the various restrictions imposed on  $Q$  and that  $Q$  can generate the parameter value  $\theta_0$ .

(ii) Compute the recentering values

$$\eta(x) = \hat{P}(x) - \sum_{w \in \mathcal{W}_x} Q^*(w)$$

for  $x \in \mathcal{X}$ , where  $Q^*$  corresponds to the minimizer of the minimization problem in (S.29). Then, for  $l = 1, \dots, B$ , compute the recentered subsampling test statistic

$$TS_{l,N}(\theta_0) = b \cdot \min_{\{Q(w)\}_{w \in \mathcal{W}}} \sum_{x \in \mathcal{X}} \left( \hat{P}_l(x) - \eta(x) - \sum_{w \in \mathcal{W}_x^\dagger} Q(w) \right)^2$$

subject to the same restrictions as the problem in (S.29) but with  $\hat{a}_{\text{num}}$  and  $\hat{a}_{\text{den}}$  replaced by their analogues computed using the  $l$ th subsample of size  $b$  drawn without replacement from the data, and, similarly, where  $\hat{P}_l$  corresponds to the analogue of  $\hat{P}$  computed using the  $l$ th subsample. For the empirical results, I draw the subsamples from the clusters of Head Start preschools, i.e. from (S.27), to maintain the dependence between children in the schools. Based on the choice from Kalouptsi et al. (2020), I take  $b \approx 8 \cdot G^{1/2}$ . Moreover, I take  $B = 200$ .

(iii) Compute the critical value of the test  $\hat{c}(1 - \alpha, \theta_0)$  by taking the  $(1 - \alpha)$ -quantile of the distribution of the computed subsample test statistics

$$H(t, \theta_0) = \frac{1}{B} \sum_{l=1}^B 1\{TS_{l,N}(\theta_0) \leq t\}.$$

(iv) Finally, the test is given by  $\phi_N(\theta_0) = 1\{TS_N(\theta_0) > \hat{c}(1 - \alpha, \theta_0)\}$ , i.e. reject if the computed test statistic is larger than the critical value.

Given the above test procedure for a given value of  $\theta_0$ , we can construct confidence intervals by collecting all those values we don't reject, i.e.  $\{\theta_0 \in \mathbf{R} : \phi_N(\theta_0) = 0\}$ .

## S.5 Additional Details for Empirical Analysis

### S.5.1 Details on Discretizing Test Scores

In this section, I outline how I obtain the discrete test score outcome of interest used in the empirical results presented in Section 4. I use the empirical distribution of the observed undiscretized test scores to discretize the test scores to take values in  $\mathcal{Y} = \{y_1, \dots, y_M\}$ . Specifically, for a pre-specified choice of  $M$ , I take each point in  $\mathcal{Y}$  to be given by

$$y_m = \frac{y_m^* + y_{m-1}^*}{2},$$

where  $y_m^*$  denotes the  $m/M$ -quantile of the empirical distribution of the observed undiscretized test score  $Y^*$ , i.e. each point is determined by the midpoint of two specific adjacent quantiles of the empirical distribution of the observed undiscretized test scores. Then, each observed undiscretized test score  $Y^*$  is transformed as follows to its corresponding discretized version

$$Y = \begin{cases} y_m & \text{if } Y^* \in [y_{m-1}^*, y_m^*) \text{ for } m \in \{1, \dots, M-1\} , \\ y_M & \text{if } Y^* \in [y_{M-1}^*, y_M^*] , \end{cases}$$

which is the version of the test score used in the analysis. For the empirical results, I used ten support points for the test scores, i.e.  $M = 10$ , which ensured that all the computational problems were generally tractable. In unreported results, I also computed estimated bounds for the parameters in Table 2 with  $M = 15$  and  $M = 20$ . I find that the numerical results are relatively similar and, moreover, that the qualitative conclusions based on them remain unchanged.

### S.5.2 Data and Variable Construction

The raw data I use from the Head Start Impact Study (HSIS) is restricted, but access can be acquired by submitting applications to Research Connections at <http://www.researchconnections.org/childcare>. In this section, I briefly describe how the raw data was transformed to the final analysis sample used for the empirical results in the paper, which closely followed the publicly available code used to construct the final sample in Kline and Walters (2016). I organize the description in the following steps which were taken:

- (i) I merged all the various data files provided by Research Connections for the HSIS and dropped observations with missing Head Start preschool IDs, where this preschool corresponded to that from which the child was sampled. I then made edits to this raw sample as described below.
- (ii) I classified the participation decision into the three categories used in the paper using the focal care arrangement variable provided by the HSIS data set.
- (iii) All observations where any of the variables used in the analysis were missing were dropped.
- (iv) Test score outcomes were then standardized using the test scores in the control group of the final sample, by test score and age cohort.

## S.6 Generalized Framework and Additional Experiments

In this section, I present the generalized version of the framework which was developed in the context of the HSIS in Section 3 and show how it applies to some examples of alternative experiments. In particular, I consider the following two experiments: a microfinance experiment studied in Angelucci et al. (2015) and the Oregon Health Insurance Experiment.

### S.6.1 Generalized Framework

The framework described below generalizes the one developed in Section 3 in two directions. First, it does not restrict attention to solely three choice alternatives but allows for any finite set of alternatives. Second, it does not restrict attention to the assumption that assignment to the treatment group guaranteed an offer to a given alternative but allows for other assumptions that may arise based on the experiment. Both these generalizations are specifically relevant to ensure that the experiments described in the subsequent two sections fit into this framework.

To this end, suppose that the various alternatives that an individual can choose from take values in the following finite set

$$\mathcal{D} = \{d_1, \dots, d_{|\mathcal{D}|}\} , \quad (\text{S.30})$$

where  $|\mathcal{D}| \geq 2$  to ensure that there are a non-trivial number of alternatives. For a given individual, let the observed variables be denoted by

$$(Y, D, Z) , \quad (\text{S.31})$$

where  $Z$  denotes an indicator for whether the individual is assigned to the treatment group or not,  $D$  denotes their chosen alternative, and  $Y$  denotes their outcome of interest. Suppose that the outcome takes values in the following known discrete set  $\mathcal{Y} = \{y_1, \dots, y_M\}$ . The observed variables are assumed to be generated by several underlying latent variables. Similar to the description of the HSIS setup in Section 3.1, it is convenient to describe these variables and how they are related to the observed variables through various stages.

In Stage 1, the individual obtains their choice set of alternatives by receiving offers to various alternatives. Without loss of generality, let  $d_1$  denote the base alternative in  $\mathcal{D}$  which is always available. Let  $C(1)$  and  $C(0)$  respectively denote the potential choice sets under the treatment and control groups which take values in the following set  $\mathcal{C} = \{c \subseteq \mathcal{D} : d_1 \in c\}$ , i.e. the set of subsets of  $\mathcal{D}$  containing  $d_1$ . Let  $C$  denote the obtained choice set which is related to the potential choice sets by  $C = C(1)$  if  $Z = 1$  and  $C(0)$  if  $Z = 0$ . In Stage 2, the individual chooses their preferred alternative from their obtained choice set. Suppose that each individual has a strict preference relationship over the set of alternatives  $\mathcal{D}$ . Let  $U$  denote the individual's preference type which takes values in  $\mathcal{U}$ , i.e. the set of  $|\mathcal{D}|!$  strict preference types. The observed choice  $D$  is then related to the preference type and obtained choice set through the following relationship

$$D = \sum_{u \in \mathcal{U}, c \in \mathcal{C}} d(u, c) 1\{U = u, C = c\} , \quad (\text{S.32})$$

where  $d(u, c)$  denotes the known choice function that corresponds to what preference type  $u \in \mathcal{U}$  would choose under a non-empty set  $c \subseteq \mathcal{D}$ . Finally, in Stage 3, the individual's outcome is

realized. Let  $Y(d)$  denote the individual's potential outcome had they chosen alternative  $d \in \mathcal{D}$ . The observed outcome  $Y$  is related to the potential outcomes through the following relationship

$$Y = \sum_{d \in \mathcal{D}} Y(d) I\{D = d\} . \quad (\text{S.33})$$

Let the above described underlying latent variables be summarized by the following variable

$$W = (Y(d_1), \dots, Y(d_{|\mathcal{D}|}), U, C(0), C(1)) , \quad (\text{S.34})$$

which takes values on the following discrete sample space  $\mathcal{W} = \mathcal{Y}^{|\mathcal{D}|} \times \mathcal{U} \times \mathcal{C}^2$ . Let  $Q$  denote the probability mass function of this summary variable and let  $Q_z$  denote the probability mass function conditional on  $Z = z \in \mathcal{Z} \equiv \{0, 1\}$ .

The observed data along with the information that the treatment assignment status provides through assumptions on the model restrict the possible values that  $Q$  and  $Q_z$  can take. In particular, the observed data imposes the following restrictions on  $Q_z$ :

$$\sum_{w \in \mathcal{W}_x} Q_z(w) = \text{Prob}\{Y = y, D = d | Z = z\} \quad (\text{S.35})$$

for all  $x = (y, d, z) \in \mathcal{Y} \times \mathcal{D} \times \mathcal{Z} \equiv \mathcal{X}$ , where  $\mathcal{W}_x$  is the set of all  $w$  in  $\mathcal{W}$  such that  $c = c(1)$  if  $z = 1$  and  $c = c(0)$  if  $z = 0$ ,  $d(u, c) = d$  and  $y = y(d)$ . Moreover, the treatment assignment status is assumed to be statistically independent of the underlying variables, which can formally be stated as the following restriction on the probability mass functions:

$$Q(w) = Q_z(w) \quad (\text{S.36})$$

for all  $w \in \mathcal{W}$  and  $z \in \mathcal{Z}$ . In turn, the restriction imposed by the data on  $Q_z$  can then directly be stated as a restriction on  $Q$  as follows:

$$\sum_{w \in \mathcal{W}_x} Q(w) = \text{Prob}\{Y = y, D = d | Z = z\} \quad (\text{S.37})$$

for all  $x = (y, d, z) \in \mathcal{X}$ , where  $\mathcal{W}_x$  is defined as before. The treatment assignment status also provides some information on the potential choice sets, where this information depends on the institutional details of the experiment. Furthermore, similar to the assumptions in Section 3.5, additional information based on the setting can also be imposed on the model. In general, suppose that all this information can be captured through a finite set of restrictions  $\mathcal{S}$  on  $Q$  such that each restriction  $s \in \mathcal{S}$  satisfies

$$\sum_{w \in \mathcal{W}_s} Q(w) = 0 , \quad (\text{S.38})$$

where  $\mathcal{W}_s$  is a known subset of  $\mathcal{W}$ . Given these restrictions, the set of admissible probability mass functions for the underlying latent variables is given by

$$\mathbf{Q} = \{Q \in \mathbf{Q}_{\mathcal{W}} : Q \text{ satisfies (S.37) and (S.38) for each } s \in \mathcal{S}\} , \quad (\text{S.39})$$

where  $\mathbf{Q}_{\mathcal{W}}$  denotes the set of all probability mass functions on the sample space  $\mathcal{W}$ .

In the context of the above described model, suppose that we are interested in learning about a parameter  $\theta(Q)$  that can be written as follows

$$\theta(Q) = \frac{\sum_{w \in \mathcal{W}} a_{\text{num}}(w) \cdot Q(w)}{\sum_{w \in \mathcal{W}} a_{\text{den}}(w) \cdot Q(w)}, \quad (\text{S.40})$$

where  $a_{\text{num}} : \mathcal{W} \rightarrow \mathbf{R}$  and  $a_{\text{den}} : \mathcal{W} \rightarrow \mathbf{R}$  are known (or estimated) functions, i.e. the parameter is a fraction of linear functions of  $Q$ . As described in Section 3.4 in the setting of the HSIS, several parameters that evaluate the average effect of providing an offer correspond to such functions. Given that  $\mathbf{Q}$  stated in (S.39) and  $\theta(Q)$  stated in (S.40) are respectively equivalent to (22) and (14) stated in terms of the HSIS, what we can learn about the parameter can then directly be characterized using the linear programming procedure stated in Proposition 3.1.

As noted, the framework developed in Section 3 corresponds to a special case of the above described framework applied to the setting of the HSIS. In the following two sections, I illustrate two alternative experiments that have different settings to the HSIS, but face conceptually similar noncompliance and observational problems with respect to the receipt of an offer. For each of these experiments, I describe how their setting fits into the above described framework, which can then be applied to evaluate the average effects of a program offer.

### S.6.2 Microfinance Experiment

Angelucci et al. (2015) implemented an experiment in 2009 to analyze the Crédito Mujer microloan product provided to women by Compartamos Banco. Compartamos Banco is a large microfinance program or institution in Mexico, and Crédito Mujer is its group loan product, i.e. where a group of individuals is jointly responsible for the loans of their group. As described in detail in Angelucci et al. (2015), the experiment was implemented by providing an offer to loans from Crédito Mujer to randomly selected geographic areas of north-central part of Sonora, a state in Mexico. Similar to the HSIS, I organize the description into three stages:

**Stage 1:** The experiment randomly assigned a given geographic area to either a treatment group where an offer for Crédito Mujer was provided to individuals living in that area or a control group where it was not. While the experiment verified addresses to ensure that individuals living in control group areas did not receive an offer, they could receive an offer to Crédito Mujer from a treated area if, for example, they had a viable address they could use in a treated area—see Angelucci et al. (2015, Footnote 18). Moreover, individuals living in both treated and control group areas could also potentially receive an offer to alternative microloan products based on their availability in their respective areas.



**Stage 2:** The experiment collected data for a number of individuals in each area on where they borrowed from.

**Stage 3:** The experiment also collected data on a number of outcomes such as, for example, those related to income, labor supply and social well being.

In Stage 1 of this experiment, similar to that of the HSIS, individuals in the treatment group were provided an offer to Crédito Mujer through the experiment and some individuals in the control group were able to receive an offer from it from outside the rules of the experiment. In addition, we generally do not directly observe from where individuals receive offers as data is not collected on the available loan options for any individual, but can only indirectly learn something about it through their treatment assignment statuses and their borrowing decisions. However, unlike the HSIS where individuals could participate in only a single care setting at a given time, individuals in this experiment could be simultaneously borrowing from multiple loan programs at the same time. Below, I describe how this feature can be accommodated into the generalized framework and in turn be used to analyze the average effects of a Crédito Mujer offer.

In this experiment, the set of alternatives in (S.30) corresponds to the set of loan alternatives that an individual can be borrowing from, which can be aggregated such that  $\mathcal{D} = \{n, a, m, ma\}$ , where  $m$  denotes the Crédito Mujer,  $a$  denotes alternative microloan products,  $ma$  denotes both Crédito Mujer and alternative products, and  $n$  denotes no microloan product. Furthermore, the observed variables in (S.31) correspond to the treatment assignment status in Stage 1, the borrowing decision in Stage 2, and the outcome of interest in Stage 3. Furthermore, these observed variables can be assumed to be generated by the choice and outcome equations respectively defined in (S.32) and (S.33) using the corresponding underlying variables in (S.34).

Given the assumed relationship between the observed and underlying variables, the information provided by the data can be captured by restrictions in (S.37) based on the corresponding  $Q$  in this setting. Moreover, the information that the treatment assignment status provides can also be captured through restrictions in terms of those in (S.36) and (S.38). In particular, similar to Assumption HSIS, the information that the treatment assignment status provides can be stated in terms of the following assumption on the underlying variables:

**Assumption MFE.**

- (i)  $(Y(n), Y(a), Y(m), Y(ma), U, C(0), C(1)) \perp Z$ .
- (ii)  $m \in C(1)$ .
- (iii) For  $z \in \{0, 1\}$ ,  $ma \in C(z) \iff m, a \in C(z)$ .

Assumption MFE(i) states that the experiment randomly assigned individuals to the treatment and control groups and in turn implies the restriction in (S.36). Assumption MFE(ii) states that if

an individual lives in an area assigned to the treatment group then the individual receives an offer to Crédito Mujer. Assumption **MFE**(iii) states that, for both in the treatment and control group, an individual receives offers from both Crédito Mujer and alternative products if and only if the individual also has an offer from each product individually. Note that while Assumption **MFE**(iii) is not a feature of the treatment assignment status, it is a logical feature of the way the set of choice alternatives is defined in this setting. Similar to the restatement of Assumption **HSIS**(ii) in (19), it is straightforward to derive that Assumption **MFE**(ii) and Assumption **MFE**(iii) can respectively be restated as  $\sum_{w \in \mathcal{W}_{\text{MFEii}}} Q(w) = 0$  and  $\sum_{w \in \mathcal{W}_{\text{MFEiii}}} Q(w) = 0$ , where  $\mathcal{W}_{\text{MFEii}} = \{w \in \mathcal{W} : m \notin c(1)\}$  and  $\mathcal{W}_{\text{MFEiii}} = \{w \in \mathcal{W} : ma \in c(0), m, a \notin c(0)\} \cup \{w \in \mathcal{W} : m, a \in c(0), ma \notin c(0)\} \cup \{w \in \mathcal{W} : ma \in c(1), m, a \notin c(1)\} \cup \{w \in \mathcal{W} : m, a \in c(1), ma \notin c(1)\}$ , i.e. both of the assumptions can be restated as restrictions in terms of (S.38).

Given that the above described setting of the experiment fits in the generalized model, we can then use the linear programming procedure to learn about various parameters that evaluate the average effects of a Crédito Mujer offer as long as they can be re-written in terms of (S.40) given the corresponding  $Q$  for this setup. In particular, similar to the parameters described in Section 3.3 for the HSIS, we can compare mean borrowing decisions and outcome responses under various choice sets to define parameters analyzing the average effects of a Crédito Mujer offer.

### S.6.3 Oregon Health Insurance Experiment

As described in detail in Finkelstein et al. (2012), the Oregon Health Insurance Experiment (OHIE) was a randomized evaluation of Oregon’s Medicaid program. In particular, the program consisted of two parts: the Oregon Health Plan (OHP) Standard and the OHP Plus. The experiment was the product of a lottery in 2008 that randomly selected individuals from a waiting list to potentially gain a Medicaid offer through OHP Standard. Similar to the description of the HSIS, I organize the description of the OHIE into three stages as follows:

**Stage 1:** The experiment randomly assigned individuals through the lottery to either a treatment group where they received an offer to Medicaid through OHP Standard if they satisfied the eligibility requirements at the time the lottery was conducted, or a control group where they did not receive such an offer. In turn, the lottery did not guarantee Medicaid to every treated individual as some of them may not satisfy the eligibility requirements for OHP Standard. Moreover, individuals in either group could potentially receive an offer to Medicaid from outside the lottery by being eligible for OHP Plus or to alternative non-Medicaid insurance.

**Stage 2:** The experiment collected data on the type of insurance plan in which the individual was enrolled.

**Stage 3:** The experiment collected data on a number of outcomes related to health and health care utilization.

In Stage 1 of this experiment, similar to that of the HSIS, some control individuals were able to receive a Medicaid offer from outside the experiment by being eligible for OHP Plus. In addition, we generally do not observe from where individuals receive offers as data is not collected on the insurance plans for which an individual is eligible.<sup>1</sup> However, unlike the HSIS, every treated individual was not guaranteed Medicaid as they may potentially not satisfy the eligibility requirements. Below, I describe how these features can be accommodated into the generalized framework and in turn be used to analyze the average effects of a Medicaid offer.

In this experiment, the set of alternatives in (S.30) corresponds to the set of insurance plan alternatives, which can be aggregated such that  $\mathcal{D} = \{n, a, m\}$ , where  $m$  denotes Medicaid program insurance,  $a$  denotes alternative non-Medicaid insurance and  $n$  denotes no insurance. Furthermore, the observed variables in (S.31) correspond to the treatment assignment status in Stage 1, the insurance plan enrolled in Stage 2, and the outcome of interest in Stage 3. Furthermore, these observed variables can be assumed to be generated by the choice and outcome equations respectively defined in (S.32) and (S.33) using the corresponding underlying variables in (S.34). Given the assumed relationship between the observed and underlying variables, the information provided by the data can be captured by restrictions in (S.37) based on the corresponding  $Q$  in this setting. Moreover, the information that the treatment assignment status provides can also be captured through restrictions in terms of those in (S.36) and (S.38). In particular, similar to Assumption HSIS, the information that the treatment assignment status provides can be stated in terms of the following assumption on the underlying variables:

**Assumption OHIE.**

- (i)  $(Y(n), Y(a), Y(m), U, C(0), C(1)) \perp Z$  .
- (ii)  $m \in C(0) \implies m \in C(1)$  .

Assumption OHIE(i) states that the experiment randomly assigned individuals to the treatment and control groups and in turn implies the restriction in (S.36). Assumption OHIE(ii) states that if an individual receives a Medicaid offer in the control group then that individual receives it in the treatment group. In particular, it captures the following institutional detail: if an individual receives an offer in the control group then they receive it through OHP Plus and, in turn, also in the treatment group through at least OHP Plus as the lottery does not affect the eligibility

---

<sup>1</sup>The experiment, in fact, collected data on whether treated individuals applied to enroll into OHP Standard and the outcome of their decision on whether they were considered eligible. This observed variable in turn provides additional information on whether individuals had an offer to Medicaid, but falls outside the scope of the developed framework. I leave the extension of the framework to accommodate for such additional information for future work.

for OHP Plus. Similar to the restatement of Assumption **HSIS**(ii) in (19), it is straightforward to derive that Assumption **OHIE**(ii) can be restated as  $\sum_{w \in \mathcal{W}_{\text{OHIE}}} Q(w) = 0$ , where  $\mathcal{W}_{\text{OHIE}} = \{w \in \mathcal{W} : m \in c(0) \text{ and } m \notin c(1)\}$ , i.e. it can be restated as a restriction in terms of (S.38).

Given that the above described setting of the OHIE fits in the generalized model, we can then use the linear programming procedure to learn about various parameters that evaluate the average effects of a Medicaid offer as long as they can be re-written in terms of (S.40) given the corresponding  $Q$  for this setup. In particular, similar to the parameters described in Section 3.3 for the HSIS, we can compare mean enrollment and outcome responses across various choice sets to define various parameters analyzing the average effects of a Medicaid offer.

## S.7 Proofs of Propositions and Additional Derivations

### S.7.1 Proof of Proposition 3.1

Note that  $\mathbf{Q}_{\mathcal{W}}$  is compact and convex. Further, note that  $\mathbf{Q}$  is a set of distributions in  $\mathbf{Q}_{\mathcal{W}}$  that is obtained by placing linear constraints imposed by the data in (18) and by the assumptions in (20). This in turn implies that  $\mathbf{Q}$  is a compact and convex set as well.

Next, it follows from (14) that  $\theta(Q)$  is a linear-fractional function of  $Q$  where the denominator is required to be positive, i.e. (23) holds, for every  $Q \in \mathbf{Q}$ . Along with  $\mathbf{Q}$  being a compact and convex set, this in turn implies that  $\theta(\mathbf{Q})$  is a compact and convex set in  $\mathbf{R}$ . More specifically, it follows that the identified set in (21) can be written as a closed interval, where the lower bound and upper bound are given by

$$\theta_l = \min_{Q \in \mathbf{Q}} \theta(Q) \quad \text{and} \quad \theta_u = \max_{Q \in \mathbf{Q}} \theta(Q) . \quad (\text{S.41})$$

In order to complete the proof, note that the optimization problems in (S.41) have linear-fractional objectives due to the structure of the parameter in (14) and a finite number of linear constraints as guaranteed by the data restrictions in (18) and the structure of the imposed restrictions in (20). Such optimization problems are commonly referred to as linear-fractional programs. For such programs, [Charnes and Cooper \(1962\)](#) show, among other results, that if the feasible set of the program is non-empty and bounded, and if the denominator of the linear-fractional objective is strictly positive for all values in the feasible set, then the linear-fractional program can be transformed to an equivalent linear program—see [Boyd and Vandenberghe \(2004, Section 4.3.2\)](#) for a textbook exposition of this result.

For the linear-fractional programs stated in (S.41), both these conditions are satisfied. Since  $\mathbf{Q}$  is non-empty and bounded, we have that the feasible sets of the programs given by  $\mathbf{Q}$  are indeed non-empty and bounded. Further, since (23) holds for all  $Q \in \mathbf{Q}$ , we also have that the

denominator of the objectives are strictly positive for all values in the feasible set. In turn, the result from [Charnes and Cooper \(1962\)](#) can be invoked to transform the linear-fractional programs in (S.41) to equivalent linear programs which are given by those in (24). Specifically, the equivalent linear programs are obtained by introducing the following so-called Charnes-Cooper transformation

$$\tilde{Q}(w) = \gamma \cdot Q(w) \quad \text{where} \quad \gamma = \frac{1}{\sum_{w \in \mathcal{W}} a_{\text{den}}(w) \cdot Q(w)} , \quad (\text{S.42})$$

which is well-defined given that (23) holds for every  $Q \in \mathbf{Q}$ , and by introducing the additional constraint that  $\gamma \geq 0$ . Then, when the constraints and the objectives stated in terms of  $Q$  for the linear-fractional programs in (S.41) and the relationship between  $Q$  and  $\gamma$  in (S.42) are rewritten in terms of  $\gamma$  and  $\tilde{Q}$ , we obtain the constraints and the objective of the linear programs stated in (24), which concludes the proof.

### S.7.2 Proof of Proposition S.1.1

Since  $C_1 = 1$  by Assumption M4, note that there are two groups of individuals as defined by their potential choice sets in the treatment and control groups: (i)  $C_0 = 0$  and  $C_1 = 1$ ; and (ii)  $C_0 = 1$  and  $C_1 = 1$ . The first group complies with their assigned status in the control group and is hence called the complier group. In contrast, the second group does not comply with their assigned status in the control group and is hence, by analogy, called the noncomplier group. For convenience, let  $T$  denote an indicator for whether the individual is a complier or not.

Next, note that we can rewrite the mean observed test scores of those in the treatment group by

$$E[Y|Z = 1] = E[Y_1 \cdot C_1 + Y_0 \cdot (1 - C_1)|Z = 1] = E[Y_1|Z = 1] = E[Y_1] ,$$

where the first equality follows from expanding  $Y$  using (S.2) and then expanding  $C$  using (S.3), the second equality follows from Assumption M4, and the third equality follows from Assumption M3. Using the law of total probability, we can then rewrite this quantity in terms of the complier and noncomplier groups, i.e.

$$E[Y|Z = 1] = E[Y_1|T = 1] \cdot \text{Prob}\{T = 1\} + E[Y_1|T = 0] \cdot \text{Prob}\{T = 0\} . \quad (\text{S.43})$$

Similarly, note that we can rewrite the mean observed test scores of those in the control group by

$$E[Y|Z = 0] = E[Y_1 \cdot C_0 + Y_0 \cdot (1 - C_0)|Z = 0] = E[Y_1 \cdot C_0 + Y_0 \cdot (1 - C_0)] ,$$

where the first equality follows from expanding  $Y$  using (S.2) and then expanding  $C$  using (S.3), and the second equality follows from Assumption M3. Using the law of total probability, we can

then again rewrite this quantity in terms of the complier and noncomplier groups, i.e.

$$\begin{aligned} E[Y|Z = 0] &= E[Y_1 \cdot C_0 + Y_0 \cdot (1 - C_0)|T = 1] \cdot \text{Prob}\{T = 1\} + \\ &\quad E[Y_1 \cdot C_0 + Y_0 \cdot (1 - C_0)|T = 0] \cdot \text{Prob}\{T = 0\} \\ &= E[Y_0|T = 1] \cdot \text{Prob}\{T = 1\} + E[Y_1|T = 0] \cdot \text{Prob}\{T = 0\} , \end{aligned} \quad (\text{S.44})$$

where the second equality follows from the fact that  $T = 1$  is by construction equivalent to  $C_0 = 0$  and that  $T = 0$  is equivalent to  $C_0 = 1$ . By taking the difference of the quantities in (S.43) and (S.44), we can then show that the ITT estimand on test scores corresponds to the following

$$\text{ITT}_Y = E[Y|Z = 1] - E[Y|Z = 0] = E[Y_1 - Y_0|T = 1] \cdot \text{Prob}\{T = 1\} . \quad (\text{S.45})$$

In an analogous manner, by replacing the test score  $Y$  with the enrollment decision  $D$ , we can show also that the ITT estimand on enrollment into Head Start corresponds to the following

$$\text{ITT}_D = E[D|Z = 1] - E[D|Z = 0] = E[D_1 - D_0|T = 1] \cdot \text{Prob}\{T = 1\} . \quad (\text{S.46})$$

By recalling that  $1\{T = 1\} \equiv 1\{C_0 = 0, C_1 = 1\}$ , we then obtain the expressions for  $\text{ITT}_D$  and  $\text{ITT}_Y$  stated in the proposition.

To obtain what the IV estimand corresponds to, note that we can rewrite  $\text{ITT}_Y$  in (S.45) using the law of total probability by

$$\begin{aligned} \text{ITT}_Y &= E[Y_1 - Y_0|T = 1, D_1 = 1] \cdot \text{Prob}\{D_1 = 1|T = 1\} \cdot \text{Prob}\{T = 1\} + \\ &\quad E[Y_1 - Y_0|T = 1, D_1 = 0] \cdot \text{Prob}\{D_1 = 0|T = 1\} \cdot \text{Prob}\{T = 1\} \\ &= E[Y_1 - Y_0|T = 1, D_1 = 1] \cdot \text{Prob}\{D_1 = 1|T = 1\} \cdot \text{Prob}\{T = 1\} , \end{aligned} \quad (\text{S.47})$$

where the second equality follows from Assumption M1 and Assumption M2 that together imply that  $D_1 = 0 \implies Y_0 = Y_1$ . Furthermore, since  $D_0 = 0$  by Assumption M1, it follows that  $\text{ITT}_D$  in (S.46) can be rewritten as

$$\text{ITT}_D = \text{Prob}\{D_1 = 1|T = 1\} \cdot \text{Prob}\{T = 1\} . \quad (\text{S.48})$$

In turn, by taking the ratio of  $\text{ITT}_Y$  and  $\text{ITT}_D$  as stated in (S.47) and (S.48), we can then show that the IV estimand corresponds to the following

$$\text{IV} = E[Y_1 - Y_0|T = 1, D_1 = 1] .$$

Again, by recalling that  $1\{T = 1\} \equiv 1\{C_0 = 0, C_1 = 1\}$ , we then obtain the expression for IV stated in the proposition, which concludes the proof.

### S.7.3 Rewriting Parameters in Section 3.3 in terms of (14)

Similar to how the parameter in (8) can be rewritten as (14) as illustrated in (15), the remainder of the parameters in Section 3.3 can also be rewritten as (14). For convenience, define  $c_+ = c \cup \{h\}$  and  $c_- = c \setminus \{h\}$  for each  $c \in \mathcal{C}$ . Using this notation, the parameter in (7) can be written as

$$\text{ATE}^Y(C_+, C_-)(Q) = \sum_{z \in \mathcal{Z}} \sum_{w \in \mathcal{W}} \text{Prob}\{Z = z\} \cdot [y(d(u, c_+(z))) - y(d(u, c_-(z)))] \cdot Q(w) ,$$

while the parameter in (9) can be written as the ratio of that above and that in (15). Similarly, the parameters in (10) and (11) can be respectively written as

$$\begin{aligned} \text{ATE}_{+a}^D(C_+, C_-)(Q) &= \frac{\sum_{z \in \mathcal{Z}} \sum_{w \in \mathcal{W}_{z,+a}^h} \text{Prob}\{Z = z\} \cdot Q(w)}{\sum_{z \in \mathcal{Z}} \sum_{w \in \mathcal{W}_{z,+a}} \text{Prob}\{Z = z\} \cdot Q(w)} , \\ \text{ATOP}_{+a}(C_+, C_-)(Q) &= \frac{\sum_{z \in \mathcal{Z}} \sum_{w \in \mathcal{W}_{z,+a}} \text{Prob}\{Z = z\} \cdot [y(d(u, c_+(z))) - y(d(u, c_-(z)))] \cdot Q(w)}{\sum_{z \in \mathcal{Z}} \sum_{w \in \mathcal{W}_{z,+a}^h} \text{Prob}\{Z = z\} \cdot Q(w)} , \end{aligned}$$

where  $\mathcal{W}_{z,+a}^h = \{w \in \mathcal{W} : d(u, c_+(z)) = h, a \in c(z)\}$  and  $\mathcal{W}_{z,+a} = \{w \in \mathcal{W} : a \in c(z)\}$  for  $z \in \mathcal{Z}$ , and the parameters in (12) and (13) can be respectively written as

$$\begin{aligned} \text{ATE}_{-a}^D(C_+, C_-)(Q) &= \frac{\sum_{z \in \mathcal{Z}} \sum_{w \in \mathcal{W}_{z,-a}^h} \text{Prob}\{Z = z\} \cdot Q(w)}{\sum_{z \in \mathcal{Z}} \sum_{w \in \mathcal{W}_{z,-a}} \text{Prob}\{Z = z\} \cdot Q(w)} , \\ \text{ATOP}_{-a}(C_+, C_-)(Q) &= \frac{\sum_{z \in \mathcal{Z}} \sum_{w \in \mathcal{W}_{z,-a}} \text{Prob}\{Z = z\} \cdot [y(d(u, c_+(z))) - y(d(u, c_-(z)))] \cdot Q(w)}{\sum_{z \in \mathcal{Z}} \sum_{w \in \mathcal{W}_{z,-a}^h} \text{Prob}\{Z = z\} \cdot Q(w)} , \end{aligned}$$

where  $\mathcal{W}_{z,-a}^h = \{w \in \mathcal{W} : d(u, c_+(z)) = h, a \notin c(z)\}$  and  $\mathcal{W}_{z,-a} = \{w \in \mathcal{W} : a \notin c(z)\}$  for  $z \in \mathcal{Z}$ .

### S.7.4 Rewriting Assumptions in Section 3.5 in terms of (20)

Similar to Assumption **HSIS**, the assumptions in Section 3.5 can be re-written as restrictions on  $Q$  in the form of (20). In particular, similar to Assumption **HSIS**(ii), it is straightforward to see that Assumption **UA** can be re-written in terms of  $Q$  as

$$\sum_{w \in \mathcal{W}_{\text{UA}}} Q(w) = 0 , \tag{S.49}$$

where  $\mathcal{W}_{\text{UA}} = \{w \in \mathcal{W} : a \in c(0), a \notin c(1) \text{ or } a \notin c(0), a \in c(1)\}$ , and also that Assumption **MTR** can be re-written in terms of  $Q$  as

$$\sum_{w \in \mathcal{W}_{\text{MTR}}} Q(w) = 0 , \tag{S.50}$$

where  $\mathcal{W}_{\text{MTR}} = \{w \in \mathcal{W} : y(n) > y(h) \text{ or } y(a) > y(h)\}$ . In order to see the restrictions imposed by Assumption **Roy**, note first this assumption can be equivalently stated as  $d(U, \{d, d'\}) = d \implies Y(d') \leq Y(d)$  for each  $d, d' \in \mathcal{D}$ . Then, it follows that Assumption **Roy** can also be re-written in terms of  $Q$  as

$$\sum_{w \in \mathcal{W}_{\text{Roy}, \{d, d'\}}} Q(w) = 0 , \quad (\text{S.51})$$

for all  $d, d' \in \mathcal{D}$ , where  $\mathcal{W}_{\text{Roy}, \{d, d'\}} = \{w \in \mathcal{W} : y(d') > y(d), u \in \mathcal{U}_{\{d, d'\}}\}$  and  $\mathcal{U}_{\{d, d'\}} = \{u \in \mathcal{U} : d(u, \{d, d'\}) = d\}$ .



## References

- ANGELUCCI, M., KARLAN, D. and ZINMAN, J. (2015). Microcredit impacts: Evidence from a randomized microcredit program placement experiment by compartamos banco. *American Economic Journal: Applied Economics*, **7** 151–82.
- BOYD, S. and VANDENBERGHE, L. (2004). *Convex optimization*. Cambridge university press.
- CHARNES, A. and COOPER, W. W. (1962). Programming with linear fractional functionals. *Naval Research Logistics (NRL)*, **9** 181–186.
- FINKELSTEIN, A., TAUBMAN, S., WRIGHT, B., BERNSTEIN, M., GRUBER, J., NEWHOUSE, J. P., ALLEN, H., BAICKER, K. and GROUP, O. H. S. (2012). The oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics*, **127** 1057–1106.
- IMBENS, G. W. and ANGRIST, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, **62** 467–475.
- IMBENS, G. W. and RUBIN, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.
- KALOUPTSIDI, M., KITAMURA, Y., LIMA, L. and SOUZA-RODRIGUES, E. A. (2020). Partial identification and inference for dynamic models and counterfactuals. Tech. rep., National Bureau of Economic Research.
- KLINE, P. and WALTERS, C. R. (2016). Evaluating public programs with close substitutes: The case of head start. *The Quarterly Journal of Economics*, **131** 1795–1848.
- MOGSTAD, M. and TORGOVITSKY, A. (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics*.