Identification with Latent Choice Sets:  
The Case of the Head Start Impact Study  

Vishal Kamat  
Departments of Economics  
Northwestern University  
v.kamat@u.northwestern.edu  
November 9, 2017  
(Job Market Paper)  

Abstract  

This paper studies identification of program effects in settings with latent choice sets. Here, by latent choice sets, I mean the unobserved heterogeneity that arises when the choice set from which the agent selects treatment is heterogeneous and unobserved by the researcher. The analysis is developed in the context of the Head Start Impact Study, a social experiment designed to evaluate preschools as part of Head Start, the largest early childhood education program in the United States. In this setting, resource constraints limit preschool slots to only a few eligible children through an assignment mechanism that is not observed in the data, which in turn introduces unobserved heterogeneity in the child’s choice set of care settings. I propose a nonparametric model that explicitly accounts for latent choice sets in the care setting enrollment decision. In this model, I study various parameters that evaluate Head Start in terms of policies that mandate enrollment and also those that allow voluntary enrollment into Head Start. I show that the identified set for these parameters given the information provided by the study and by various institutional details of the setting can be constructed using a linear programming method. Applying the developed analysis, I find that a significant proportion of parents voluntarily enroll their children into Head Start if provided access and that Head Start is effective in terms of improving short-term test scores across multiple policy dimensions.

KEYWORDS: Latent choice sets, program evaluation, head start impact study, partial identification, revealed preference analysis, linear programming, social experiments.

JEL classification codes: C14, C31, I21.

*I am extremely grateful to my committee chair, Ivan Canay, and my other committee members, Chuck Manski and Alex Torgovitsky, for their extensive guidance and feedback. I am thankful to Gideon Bornstein, Nicolas Inostroza, Lola Segura and Max Tabord-Meehan for several helpful comments and discussions. I also thank Eric Auerbach, Joel Horowitz, Sam Norris, Matt Notowidigo, Rob Porter and Pedro Sant’Anna, and seminar participants at Northwestern University for useful comments. I finally thank Research Connections for providing data on the Head Start Impact Study. Financial support from the Robert Eisner Memorial Fellowship is gratefully acknowledged.
1 Introduction

This paper studies identification of program effects in settings with latent choice sets. Here, by latent choice sets, I mean the unobserved heterogeneity that arises when the choice set from which the agent selects treatment is heterogeneous and unobserved by the researcher. Such unobserved heterogeneity in choice sets is common to data on public programs, where resource constraints limit program access to only a few eligible agents through an assignment mechanism that is not observed by the researcher. In these settings, the agent’s decision to select into treatments is based not only on the preferences over all the treatments but also on the choice set of available treatment options. In this paper, I propose a nonparametric model that explicitly accounts for latent choice sets in the treatment selection decision and study the identification of various program effect parameters in the context of this model.

The analysis in this paper is developed in the context of the Head Start Impact Study (HSIS), which was a social experiment designed to evaluate Head Start. Head Start is the largest early childhood education program in the United States, which provides free preschool education to three- and four-year-old children from low income eligible households. As noted previously in Kline and Walters (2016), preschools are resource constrained to the extent that available slots are limited to only a few eligible children. In this setting, the choice set of care settings that the parents face for their child is determined by their applications to various preschools and by how these preschools allocate slots. However, the HSIS does not provide data on these decisions leading to the parents’ possibly constrained choice set of care settings. Instead, it only provides data on the resulting care setting selected from the given choice set. Despite this, as further discussed in Section 3.1, previous studies on the HSIS employ choice models that do not account for the unobserved heterogeneity that may be present in the choice sets, and implicitly assume all preschools are available to every child.

This paper develops a nonparametric framework that aims to explicitly account for the latent choice set of care settings in the care setting selection decision. The proposed model treats the observed selected care setting as simply the product of the parents’ utility maximization decision given the parents’ choice set of care settings. In particular, the model is entirely nonparametric and only assumes that parents behave rationally and have a strict preference relationship over all care settings. Importantly, the parents’ obtained choice set is permitted to include only a subset of preschools and to be correlated in an unrestricted manner with their preferences and their child’s potential outcomes under the various care settings. Moreover, by treating this choice set as a latent variable, the model assumes that the researcher does not observe the parents’ obtained choice set. As further discussed in Remark 3.1, the proposed model builds on the discrete choice framework of Manski (2007) by allowing for latent choice sets of care settings and potential outcomes associated with each care setting.
In the proposed model, I study a range of parameters that evaluates Head Start in comparison to alternate care settings across multiple policy dimensions. The first class of parameters evaluates the effect on a child’s short-term test score outcome of Head Start in comparison to home care, i.e. no preschool. This class of parameters corresponds to evaluating policies that mandate Head Start enrollment versus no preschool enrollment. As noted in early work by Heckman et al. (1997) and Manski (1996, 1997a), many policies however do not mandate attendance but rather allow agents to voluntarily select into their preferred treatment option. In turn, I also study a class of parameters that evaluates the effect of providing Head Start access to parents. That is a Head Start option from which the parents can voluntarily choose whether to enroll their child.

The identification analysis begins by studying what we can learn about the parameters of interest using only the restrictions imposed on the model by the HSIS experiment. These restrictions correspond to those that the observed data and the experimental design of the HSIS place on the underlying model distribution. The parameters of interest are in general not point identified but rather partially identified. Deriving the identified set analytically for these parameters is difficult due to the complicated structure of the model and the imposed restrictions. Instead, I propose a general computational method to obtain the identified set for the various parameters given restrictions imposed on the model. For computational tractability, the proposed method requires that the underlying model variables are discrete, or transformed into discrete variables, and that the imposed restrictions are linear in the model distribution. Under these requirements, I show that the identified set can be computed by solving linear programming problems. I illustrate that the restrictions imposed by the HSIS experiment satisfy these requirements. An important benefit of the computational method is the flexibility by which restrictions imposed by additional assumptions on the model can also be studied. With the aim of obtaining stronger conclusions, I discuss several additional nonparametric assumptions motivated by unique institutional details of the HSIS setting.

The developed identification analysis is applied using the empirical distribution of the HSIS sample data to estimate the identified sets for the parameters of interest. In order to construct confidence intervals for these parameters, I discuss how the profiled subsampling method proposed by Romano and Shaikh (2008) can be applied in a computationally tractable manner. The estimated identified sets are informative. For example, using only the information provided by the experimental design of the HSIS, I find that between 79.9% to 91.7% of parents prefer to enroll their child into Head Start when they do not have any alternate preschool option. Amongst this subgroup of parents who prefer to enroll their child, I find under an additional nonparametric monotonicity assumption and a nonparametric assumption on how the experiment affects the choice set of care settings that between 4.9% to 42.7% of their children strictly benefit in terms of improving their short-term test scores. Furthermore, these estimated identified sets are tighter for some specific subgroups and under additional nonparametric assumptions. In summary, the findings suggest
towards the benefits of Head Start and qualitatively corroborate those of previous studies on the
HSIS under weak nonparametric assumptions.

The remainder of the paper is organized as follows. Section 2 describes the experimental design
of the HSIS, specifically focusing on how the parents may face heterogeneous choice sets of care
settings. Based on this description, Section 3 introduces the formal model used to analyze the
HSIS and also provides a comparison to previously used models in this setting. Section 4 illustrates
various parameters of interest and the formal identification analysis. Section 5 discusses performing
statistical inference on the parameter of interest using the HSIS sample data. Section 6 presents the
empirical results. Here I first present results using only the restrictions imposed by the HSIS ex-
periment and then under restrictions imposed by additional nonparametric assumptions motivated
by the details of the HSIS setting. Section 7 concludes.

2 Experimental Design of the HSIS

Head Start is the largest early childhood education program in the United States. The program
provides free preschool education to three- and four-year-old children from disadvantaged house-
holds. Program eligibility for households is primarily determined by the federal poverty line, yet
certain exceptions qualify additional low income households. As part of a congressional mandate,
the Head Start Impact Study (HSIS) was a social experiment implemented in the beginning of Fall
2002 with the aim of evaluating the impact of the program.

The experiment acquired a sample of Head Start preschool centers and participating children
using a multistage stratified sampling scheme, an extensive description of which is provided in
Puma et al. (2010). The centers and the children were sampled not from the entire population, but
rather from specific sub-populations. More specifically, centers were first sampled from a so-called
saturated population of Head Start centers, which referred to centers where the number of available
slots was strictly smaller than the number of applicants. From each sampled center, children were
then sampled separately from the sub-population of newly entering three- and four-year-old eligible
applicants in Fall 2002, where newly entering referred to children who were not previously enrolled
in any Head Start services. As noted in Puma et al. (2010), the sample of three-year-old children
differed considerably from that of four-year-old children in terms of observed covariates. Following
previous studies, the analysis in this paper is hence separately performed for the two age samples.
Note that the sub-population restrictions introduced on the selected sample may raise possible
concerns on the external validity of the empirical results using this sample.

Next, I provide a brief description of the HSIS experimental design. In order to clearly emphasize
the aspects of this design that this paper focuses on, I organize this description into the following
stages:
Stage 1: At each sampled center, the experiment randomly offered center access to the children sampled from that center. To be precise, let $Z$ be an indicator denoting the randomized center offer to a given child. If $Z = 1$, the child was granted access to that center for two consecutive years if the child was a three-year-old and for a single year if the child was a four-year-old. Whereas, if $Z = 0$, the child’s parents were told access to that center would not be granted for that year, but if the child was a three-year-old the parents could choose to reapply for access in the following year, i.e. Fall 2003. This asymmetry in the experimental offer for the two age groups naturally arose as Head Start services are unavailable to five year old children. However, the experiment did not control the child’s chances of obtaining an offer from other preschools, both Head Start centers from the one the child was sampled from and alternate non Head Start preschools. In particular, the child’s parents could choose to apply to these other preschools and in turn possibly obtain access, which primarily depended on whether the corresponding applied preschools too were saturated or not. As a fallback, the parents also always had the option to care for their child at home by themselves, a relative or some known individual.

Stage 2a: After obtaining their respective choice set of care settings based on the experimental offer and their application decisions to other preschools, the parents then decided in which care setting to enroll their child for in Fall 2002.

Stage 2b: The experiment further allowed the parents’ enrollment decision to vary over the school year across different care settings, i.e. the child could attend multiple care settings for varying duration between Fall 2002 and Spring 2003.

Stage 3: After the child completed an entire school year enrolled in various selected care settings, the experiment collected data on a number of outcomes in Spring 2003. The experiment further continued to collect outcomes for upto four additional years in the child’s corresponding care setting for those years.

For the purposes of this paper, it is most important to note that the experiment did not provide any information on the parents’ application decisions to other preschools in Stage 1, which implies that we do not observe the parents’ choice set. To the extent that the parents’ preferred care setting may not have been available in their choice set, it is important to account for this fact when studying the parents’ enrollment decisions in Stage 2a. The aim of this paper is to develop a framework that accounts for this feature of the experiment when analyzing the experimental data.

Note that certain aspects of the experiment involved important dynamic features. In particular, Stage 2b introduced a dynamic decision process and a duration treatment, and Stage 3 introduced a panel of outcomes under differing care settings every year. However, the formal model proposed in the following section is static in the sense that it does not account for either of these dynamic
features. Instead, following previous studies on the HSIS, the analysis in this paper uses an administratively coded focal care setting as the unique enrollment decision for the entire year, i.e. it essentially combines Stage 2a and Stage 2b - see Table 1 for some descriptive statistics on this enrollment decision. Moreover, following Kline and Walters (2016), the analysis in this paper considers a single outcome of interest taken to be the average of the Woodcock Johnson III (WJIII) test score and the Peabody Picture and Vocabulary Test (PPVT) test score in Spring 2003, which is then standardized with respect to the corresponding baseline test score average in Fall 2002.

Furthermore, note that the above description had no mention of observed covariates with respect to sampled Head Start centers and sampled children. Bloom and Weiland (2015) and Walters (2015) have previously noted that Head Start center covariates are related to the variation in child outcomes across centers. In order to account for this in the empirical analysis, I focus on two covariates that have been regarded as important measures of preschool quality: (i) $HC$ denotes an indicator for whether the Head Start center had the HighScope curriculum, which was part of the influential Perry Preschool program; and (ii) $CS$ denotes an indicator for whether the Head Start center had a low class size ratio.

## 3 Model Framework

In this section, I propose a nonparametric model that generates the parents’ observed selected care setting for their child and the child's observed test score outcome under the selected care setting. The proposed model is tightly connected to the various stages of the experimental design presented in the previous section. Before proceeding to these stages, following the analysis in Kline and Walters (2016), I begin by assuming that the set of all possible care settings, i.e. the set of treatments, is categorized into the following

$$\mathcal{D} = \{0, 1, 2\},$$

where 0 denotes home care or no preschool, 1 denotes an alternate or non Head Start preschool and 2 denotes a Head Start preschool.
In Stage 1, parents apply to both Head Start and alternative preschools and as a product obtain their choice set of care settings. This choice set implicitly encompasses the costs that parents face to acquire a preschool in their choice set. Let $C(1)$ denote the parents’ (potential) choice set had access been granted to the Head Start center as part of the experiment, and let $C(0)$ denote the parents’ (potential) choice set had access not been granted. As parents always had the fallback option of home care, these choice sets are assumed to possibly take values in the following set

$$C = \{\{0\}, \{0, 1\}, \{0, 2\}, \{0, 1, 2\}\},$$

i.e. the set of all possible choice sets that contain home care. Let $C$ denote the parents’ obtained choice set, which is given by the following relationship

$$C = C(1)Z + C(0)(1 - Z).$$

In Stage 2, parents decide in which care setting to enroll their child given their choice set obtained from Stage 1. Let $U(d)$ denote the indirect utility that the parents would obtain had their child been enrolled in care setting $d \in D$. In particular, this utility corresponds to the parents’ benefits and costs that they face if their child was enrolled in a given care setting. Let $D$ denote the observed enrollment decision, which is assumed to be given by the following utility maximization relationship

$$D = \arg \max_{d \in C} U(d). \quad (1)$$

Note that 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 hence more useful to directly refer to the parents’ underlying preference type that these utilities represent. To this end, assume that each parent has a strict preference relation over the set of care settings, which then implies that the utilities can be strictly ordered, i.e.

$$U(0) > U(1) > U(2)$$

where $U(k)$ denotes the $(k + 1)$th largest value of the utilities $\{U(d) : d \in D\}$. Denoting by $d_k$ the care setting in $D$ with the $(k + 1)$th largest utility for the parents, i.e. $U(k) = U(d_k)$, the parents’ underlying preference type is then simply given by the following label

$$U = (d_0 \succ d_1 \succ d_2),$$

which corresponds to the preference ordering on $D$ that the utilities $\{U(d) : d \in D\}$ represent. Note that since there are three possible care settings, this implies that there are six possible preference orderings over them. To be concrete, this set can explicitly be stated as follows

$$\mathcal{U} = \{(2 \succ 1 \succ 0), \ (2 \succ 0 \succ 1), \ (1 \succ 2 \succ 0), \ (1 \succ 0 \succ 2), \ (0 \succ 2 \succ 1), \ (0 \succ 1 \succ 2)\}.$$
Using the above notation, the utility maximization relationship in (1) can be re-written in terms of the preference types and the obtained choice set through the following relationship

$$D = \sum_{u \in U, c \in C} d_{u,c} I\{U = u, C = c\} \equiv d_{U,C},$$  \hspace{1cm} (2)$$

where

$$d_{U,c} = \arg \max_{d \in c} U(d)$$

denotes the preferred care setting for the preference type $U$ under a non-empty subset $c \subseteq D$, i.e. under a given possible choice set.

In Stage 3, the child’s observed test score outcome is realized under the enrolled care setting from Stage 2. Let $Y(d)$ denote the (potential) test score outcome had the child been enrolled in care setting $d \in D$. Let $Y$ denote the observed test score outcome under the enrolled care setting. The observed test score is related to the potential test scores and the enrolled care setting through the following relationship

$$Y = \sum_{d \in D} Y(d) I\{D = d\} \equiv Y(D).$$  \hspace{1cm} (3)$$

Next, I emphasize two important aspects of the above described model. First, the model placed no restrictions on the dependence between the underlying variables in the three model stages, which implies that the potential outcomes, the preference type and the choice sets are allowed to be statistically dependent. As a result, this allows the obtained choice sets to be dependent on the preference type and the potential test scores, and further the selected care setting to be dependent on the potential outcomes.

Second, as the experiment did not provide data on the parents’ application decisions, the value of the obtained choice set in Stage 1 of the model is not observed. Recall however that the experimental design randomly offered Head Start access to children, which implies that Head Start would be present in the obtained choice set had an offer been made. This partial information formally imposes some restrictions on the model, which are stated in the form of an assumption below.

**Assumption HSIS.** Let the underlying model variables be such that

(i) $2 \in C(1)$.

(ii) $(Y(0), Y(1), Y(2), U, C(0), C(1)) \perp Z \mid (HC, CS)$.

Assumptions HSIS(i) states that being assigned an offer guaranteed that a Head Start preschool was in the parents’ choice set. To be concrete, it ensures that the choice set $C(1)$ can possibly only take values in the following set

$$C_1 = \{\{0, 2\}, \{0, 1, 2\}\}.$$
Assumption HSIS(ii) states that conditional on the Head Start center, access to that center was randomly assigned to the children.

**Remark 3.1.** In the context of observed choice data, using an agent’s underlying preference type $U$ was first proposed by Marschak et al. (1959) for the purposes of testing an agent’s rationality. Manski (2007) extended this idea to propose a general nonparametric discrete choice framework by formally introducing the selection equation in (2). In the setting of observed and statistically independent choice sets, Manski (2007) then developed the analysis to predict choice probabilities under various counterfactual choice sets. See also Kline and Tartari (2016) and Manski (2014) for applications of this framework to study labor supply decisions. The model discussed in this section builds on that of Manski (2007) in two directions essential to the experimental design of the HSIS. First, the model allows for the obtained choice set in Stage 1 to be unobserved by the researcher and also to be correlated with the preferences. Second, the model introduces a Stage 3 with the outcome equation in (3), where the selection equation in (2) is an intermediate step.

### 3.1 Previous Models on the HSIS

Kline and Walters (2016) have previously proposed a parametric structural model to evaluate the HSIS. The most important distinction between their choice model and that proposed in the previous section is the conceptual differences in the description of the parents’ decision process. In particular, their choice model assumes that the utility from enrolling in the various care settings is given by

$$U(2) \equiv U(2, Z), \; U(1), \; \text{and} \; U(0),$$

i.e. the randomized offer affects the utility under Head Start, and that the selected care setting is then given by the following relationship

$$D = \arg \max_{d \in D} U(d),$$

i.e. utility is maximized over all possible care settings. By imposing that parents maximize enrollment utility from the set of all care settings, this choice model implicitly assumes that parents face all care settings when making enrollment decisions. More specifically, it does not explicitly account for Stage 1 of the experimental setting, i.e. the fact that parents may not face all preschools in their choice set when making enrollment decisions. As a result, the underlying choice model may possibly mis-specify the choice set of obtained care settings - see Stopher (1980) and Williams and Ortúzar (1982) for early criticisms on mis-specifying choice sets in parametric discrete choice models.

In contrast, their choice model implicitly combines Stage 1 and Stage 2 into a single stage. To be specific, the choice model does not treat the utility as solely the parents’ utility from enrolling their child in a given care setting in Stage 2, i.e. the parents’ benefits and costs of enrolling their
child in a given care setting. Instead, the utility implicitly also encompasses the costs from Stage 1 that the parents face to acquire a preschool in their choice set, as noted by how the experimental offer affects the utility of enrolling in a Head Start preschool. This treatment of the utility may however not capture the structure present in the decision problem of the parents. Moreover, it does not permit the study of counterfactuals that relate to the parents’ preferences over the care settings in Stage 2.

A further distinction between the model in Kline and Walters (2016) and that proposed in the previous section are the various additional assumptions imposed. In particular, their model imposes assumptions such as separability and parametric specifications on both the utilities and potential outcomes. Unlike, for example, Assumption HSIS, such assumptions are not motivated by nonparametric arguments from the empirical setting. Instead, these arguments are based on conditions required to point identify different parameters in the model. This is in contrast to the partial identification direction pursued in this paper, which is valid under weaker assumptions.

Remark 3.2. Walters (2015) also proposes a parametric structural model to evaluate the variation in parameters across different Head Start centers sampled as part of the HSIS. This model focuses on a setting with only two treatments consisting of a Head Start preschool and no Head Start preschool, i.e. alternative preschools and home care are grouped into a single treatment. The discussion in this section also conceptually applies to this model.

Remark 3.3. Previous work on discrete choice models has proposed parametric approaches that allow for latent choice sets such as, for example, Ben-Akiva and Boccara (1995), which also provides a discussion on the various approaches. These approaches in essence augment the standard parametric discrete choice models with a parametric mixture model over the unobserved choice sets. The approach pursued in this paper is instead nonparametric and is conceptually very different.

4 Identification Analysis

In this section, I first describe various parameters of interest based on the model proposed in the previous section and then develop the subsequent identification analysis. In particular, as further discussed in Remark 4.3 below, I note that the developed identification analysis is based on a general finite dimensional computational approach, which specifically requires all the observed and underlying variables used in the analysis to be discrete in nature. For the model described for the HSIS in the previous section, this is the case except for the test score outcome of interest. Hence, for the purposes of the analysis, I take a simple binary transformation of the test score outcome that corresponds to an intuitive summary indicator to evaluate whether a given test score is viewed as high or low. To be formal, denote by

\[ Y^\dagger \equiv 1\{Y > 0\} \]
a binary transformation of the observed test score outcome, and, similarly, denote by

$$Y^\dagger(d) \equiv 1\{Y(d) > 0\}$$

(5)

a binary transformation of the potential test score outcome under each care setting $d \in \mathcal{D}$. Recall that the test scores are standardized with respect to the corresponding baseline test scores in Fall 2002. In turn, under these binary transformations, a value of one signifies a “high” test score in the sense of being above the baseline mean test score and a value of zero signifies a “low” test score in the sense of being below the baseline mean test score. However, note that in principle alternate discrete transformations of the outcomes are also permitted by the identification analysis.

In Section 4.1 below, I begin by describing parameters of interest that are formally defined as functions of the model distribution under the transformed binary outcomes. In order to be formal in the description of these parameter, I first introduce some additional notation here. To this end, denote by

$$W = (Y^\dagger(0), Y^\dagger(1), Y^\dagger(2), U, C(0), C(1), Z, HC, CS) \sim Q$$

(6)

the random variable that summarizes the underlying model variables under the transformed potential test scores for each child, where $Q$ denotes the distribution of this random variable defined on a discrete sample space $\mathcal{W} = \{0,1\}^3 \times \mathcal{U} \times \mathcal{C} \times \mathcal{C} \times \{0,1\}^3$. Due to the discreteness, note that $Q$ is a probability mass function with support contained in $\mathcal{W}$, i.e. $Q : \mathcal{W} \to [0,1]$ such that

$$\sum_{w \in \mathcal{W}} Q(w) = 1.$$

Given a parameter of interest based on the unknown probability mass function $Q$, I formally state in Section 4.2 below the identification problem, i.e. what we can learn about the parameter given the known distribution of the observed data and the assumptions imposed on the model. Similar to the above notation, denote by

$$X = (Y^\dagger, D, Z, HC, CS) \sim P$$

(7)

the random variable that summarizes the observed random variables under the transformed observed outcome for each child, where $P$ denotes the distribution of this random variable defined on a discrete sample space $\mathcal{X} = \{0,1\} \times \mathcal{D} \times \{0,1\}^3$. Similar to $Q$, note that $P$ is a probability mass function with support contained in $\mathcal{X}$, i.e. $P : \mathcal{X} \to [0,1]$ such that

$$\sum_{x \in \mathcal{X}} P(x) = 1.$$
motivate in part the choice of parameters and assumptions studied in Section 4.1 and Section 4.2. For expositional reasons, in these sections below, I only make note of these choices when they are formally presented as assumptions and leave the discussion on their importance in terms of computation after I present the identification result in Proposition 4.1 in Section 4.3. In these sections below, I repeatedly use \( w \) to refer to 
\[
(y(0), y(1), y(2), u, c(0), c(1), z, hc, cs) \in W,
\]
i.e. a generic value in the sample space of the underlying random variables.

4.1 Parameters of Interest

In the proposed model with the transformed outcomes, one could possibly consider a number of interesting parameters with the aim of evaluating Head Start. The developed identification analysis allows for the flexibility to study a number of such parameters as long as they satisfy a specific condition. As discussed in Section 4.3, this condition is motivated in part by those required to ensure that the final optimization problem is a linear program. I formally state this condition in the following assumption.

**Assumption 4.1.** The parameter of interest \( \theta(Q) \) can be written as
\[
\theta(Q) = \frac{\sum_{w \in W} a_{\text{num}}(w) \cdot Q(w)}{\sum_{w \in W} a_{\text{den}}(w) \cdot Q(w)},
\]
where \( a_{\text{num}}, a_{\text{den}} : W \to \mathbb{R} \) are known functions.

Assumption 4.1 states that the parameters of interest are required to be fractions of linear functions of the probability mass function of the model random variables. Note that a special case of this class of functions are linear functions, where by construction the denominator takes a value of one. I next describe several interesting parameters that I study in the empirical results, which satisfy this assumption and also correspond to policies evaluating Head Start across multiple dimensions. These policies in particular correspond to those that mandate Head Start enrollment and also those that allow voluntary enrollment into Head Start - see Heckman et al. (1997) and Manski (1996, 1997a) for early studies, and also Heckman et al. (2006, 2008) and Heckman and Vytlacil (2007) for examples of more recent studies evaluating the latter policies.

Note however that Assumption 4.1 is flexible in the sense that it allows a researcher to also choose and study parameters that correspond to other policy dimensions. In order to emphasize this flexibility, I state below the class of parameters that I study in the form of examples of Assumption
4.1. In these examples below, I denote by 
\[ D_c = \sum_{u \in \mathcal{U}} d_{u,c} I \{ U = u \} \]
the care setting in which the parent would enroll their child had their choice set of care settings been exogenously pre-specified to a set \( c \subseteq \mathcal{D} \).

**Example 4.1. (Head Start versus Home care)** This class of parameters aims to evaluate Head Start in comparison to home care, which corresponds to comparing policies mandating Head Start enrollment versus mandating home care. I begin by considering
\[ \text{PB}_{2|0}(Q) = \text{Prob}_{Q}[Y^\dagger(2) = 1, \ Y^\dagger(0) = 0] \]
which denotes the proportion of children who strictly benefit (in terms of the binary test score) from Head Start in comparison to home care. Note that this parameter can be re-written to satisfy Assumption 4.1 as follows
\[ \text{PB}_{2|0}(Q) = \frac{\sum_{w \in \mathcal{W}_{2|0}} Q(w)}{\sum_{w \in \mathcal{W}} Q(w)} , \]
where \( \mathcal{W}_{2|0} = \{ w \in \mathcal{W} : y(2) = 1, y(0) = 0 \} \) is the set of all underlying values that correspond to a strictly higher outcome under Head Start in comparison to home school. As the converse of this parameter, I also consider
\[ \text{PL}_{2|0}(Q) = \text{Prob}_{Q}[Y^\dagger(2) = 0, \ Y^\dagger(0) = 1] \]
which denotes the proportion of children who strictly lose from Head Start in comparison to home care. Moreover, to simultaneously account for both these proportions, I also consider the difference in these quantities
\[ \text{ATE}_{2|0}(Q) = \text{PB}_{2|0}(Q) - \text{PL}_{2|0}(Q) \]
which corresponds to the average treatment effect (in terms of the binary test score) of Head Start versus home care. Note that in a manner similar to the re-writing of (9), the latter two parameters can also be re-written to satisfy Assumption 4.1.

**Example 4.2. (Option value of Head Start without an alternate preschool)** This class of parameters aims to evaluate the option value of Head Start when an alternate preschool option is absent, which corresponds to comparing policies allowing parents to freely select between Head Start and home care, i.e. from the choice set \( \{2, 0\} \), versus mandating parents to select home care by not providing any preschool choices, i.e. through the choice set \( \{0\} \). Note that, under this comparison, children whose parents do not exercise the Head Start option are left unaffected. To this end, denoting by
\[ \mathcal{U}_{20|0} = \{ u \in \mathcal{U} : d_{u,\{2,0\}} = 2 \} \]
the set of preference types that prefer Head Start over home care, I consider
\[ \text{PBOE}_{20|0}(Q) = \text{Prob}_Q[Y^\dagger(2) = 1, Y^\dagger(0) = 0 \mid U \in \mathcal{U}_{20|0}], \] (12)
which denotes the proportion who strictly benefit from the Head Start option conditional on the parents exercising the option in the setting where an alternate preschool option is absent. Note that this parameter can be re-written to satisfy Assumption 4.1 as follows
\[ \text{PBOE}_{20|0}(Q) = \frac{\sum_{w \in \mathcal{W}_{20|0}} Q(w)}{\sum_{w \in \mathcal{W}_2} Q(w)}, \]
where \( \mathcal{W}_{20|0} = \{w \in \mathcal{W} : y(0) = 0, y(2) = 1, u \in \mathcal{U}_{20|0}\} \) and \( \mathcal{W}_2 = \{w \in \mathcal{W} : u \in \mathcal{U}_{20|0}\} \). Further, in order to evaluate the proportion of children whose parents exercise the option, I also consider
\[ \text{PE}_{20|0}(Q) = \text{Prob}_Q[U \in \mathcal{U}_{20|0}] \] (13)
which denotes the proportion of children whose parents prefer Head Start over home care. Note that in a manner similar to the rewriting of (9) and (12), this parameter can also be re-written to satisfy Assumption 4.1.

**Example 4.3. (Option value of Head Start with an alternate preschool)** The previous example focused on parameters that evaluated the option value of a Head Start preschool in the absence of an alternate preschool option. Here I list analogous parameters that evaluate this option value when an alternate preschool option is also available, i.e. the choice set \( \{2, 1, 0\} \) versus the choice set \( \{1, 0\} \). To this end, denoting by
\[ \mathcal{U}_{210|10} = \{u \in \mathcal{U} : d_u,\{2,1,0\} = 2\} \]
the set of preference types that prefer Head Start over an alternate preschool and home care, let
\[ \text{PBOE}_{210|10}(Q) = \text{Prob}_Q[Y^\dagger(2) = 1, Y^\dagger(D_{\{1,0\}}) = 0 \mid U \in \mathcal{U}_{210|10}] \] (14)
denote the proportion who strictly benefit from the Head Start option conditional on the parents exercising the option in the setting where an alternate preschool option is available. Further, let
\[ \text{PE}_{210|10}(Q) = \text{Prob}_Q[U \in \mathcal{U}_{210|10}] \] (15)
denote the proportion of children whose parents prefer Head Start over an alternate preschool and home care. Similar to those in Example 4.2, note that these parameters can also be re-written to satisfy Assumption 4.1.

**Remark 4.1.** Previous studies on the HSIS have evaluated local average treatment effects in the framework of Imbens and Angrist (1994) - see Kline and Walters (2016) for arguments under which
these parameters evaluate relevant policy effects. Using the notation in this paper, we could also consider parameters evaluated conditional on the so-called compliers such as, for example,

$$\text{PB}_{2|\text{late}}(Q) = \text{Prob}_{Q}[Y^\dagger(2) = 1, \ Y^\dagger(D(0)) = 0 \mid D(1) = 2, \ D(0) = d \in \{0, 1\}]$$,

which is a local counterpart of (9), where \(D(z) = d_{u,c}(z)\) for \(z \in \{0, 1\}\). Similar to the conditional parameter in (12), such parameters can also be re-written to satisfy Assumption 4.1. As shown in Kline and Walters (2016), some of these local parameters can be nonparametrically point identified under some conditions - see, for example, Angrist and Imbens (1995), Heckman and Pinto (2017), Heckman et al. (2006, 2008), Heckman and Vytlacil (2007), Hull (2015) and Kirkeboen et al. (2016) for further point identification results on such local parameters in multiple treatment settings.

**Remark 4.2.** A class of parameters that evaluate the spread of the distribution cannot be written as fractions of linear functions of the probability mass function \(Q\) as in Assumption 4.1. Examples of such parameters include the interquartile range and the variance. See Blundell et al. (2007) and Stoye (2010) for examples of studies that provide analytical bounds on such parameters in alternate settings.

### 4.2 Identification Problem

Given a pre-specified parameter of interest from the previous section, the identification problem studies what we can learn about it given the restrictions imposed on the unknown distribution of the model by the known distribution of the observed data and by the assumptions imposed on the model. Note that the identification problem abstracts away from sampling uncertainty, i.e. it supposes that the distribution of the observed data \(P\) is known rather than an estimate of it. The latter is addressed in Section 5.

I begin by describing the restrictions imposed by the distribution of the observed data \(P\) on the distribution of the model. These restrictions can formally be stated as

$$\sum_{w \in \mathcal{W}_x} Q(w) = P(x)$$

(16)

for all \(x = (y, d, z, hc, cs) \in \mathcal{X}\), where \(\mathcal{W}_x\) is the set of all \(w\) in \(\mathcal{W}\) such that \(c = c(1)z + c(1)(1 - z), d_{u,c} = d\) and \(y = y(d)\). That is \(\mathcal{W}_x\) 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 (3) and by the selection equation in (2).

Next, I describe the restrictions imposed by the experimental design of the HSIS, i.e. Assumption HSIS, on the distribution of the model. Before proceeding, I note that the developed identification analysis is flexible in the sense that it also allows for restrictions imposed by additional assumptions. More specifically, the identification analysis only requires that all these restrictions
formally satisfy a specific condition. Similar to Assumption 4.1 and as discussed in Section 4.3, this condition is also motivated by those required to ensure that the final optimization problem is a linear program. In what follows, I first formally state this condition in the following assumption, and then show that the restrictions imposed by Assumption HSIS satisfy this condition.

**Assumption 4.2.** Let $S$ be a finite set of restrictions imposed on $Q$ such that each restriction $s \in S$ satisfies

$$M_s(Q) \leq (\text{or } =) b_s ,$$

where $b_s$ is a known or identified value in $\mathbb{R}$, and

$$M_s(Q) = \sum_{w \in W} a_s(w) \cdot Q(w) ,$$

such that $a_s : W \to \mathbb{R}$ is a known or identified function.

Assumption 4.2 states that there are only a finite number of restrictions imposed and that each restriction imposes a linear constraint on the distribution of the model. Moreover, the assumption allows these restrictions to be based on known values and on values identified by features of the observed distribution of the data. As discussed in Section 5, this distinction is important when performing statistical inference.

In the following lemma, I show that Assumption HSIS imposes restrictions that satisfy the requirement in Assumption 4.2. Before proceeding, similar to Assumption 4.1, note that Assumption 4.2 is flexible in the sense that it also allows a researcher to impose additional assumptions on the model. In Section 6.1, I describe several additional nonparametric assumptions motivated by the details on the HSIS setting that satisfy this requirement.

**Lemma 4.1.** Assumption HSIS imposes restrictions on $Q$ that satisfy Assumption 4.2.

**Proof:** In order to see the restriction implied by Assumption HSIS(i), note that this assumption can be written as

$$\text{Prob}_Q[2 \notin C(1)] = 0 .$$

Equivalently, this can be re-written as a linear restriction on $Q$ in the form of Assumption 4.2 as

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

where $\mathcal{W}_{\text{HSIS}} = \{w \in W : c(1) \in \{\{0\}, \{0,1\}\}\}$ is the set of all underlying values such that choice set with an offer does not contain a Head Start preschool.
In order to see the restrictions imposed by Assumption HSIS(ii), I first introduce some additional shorthand notation for the underlying values. To this end, denote by
\[ \bar{y} = (y(0), y(1), y(2)) \in \{0, 1\}^3 \]
values of the potential outcomes and by
\[ \bar{c} = (c(0), c(1)) \in C^2 \]
values of the potential choice sets, and by
\[ \bar{z} = (hc, cs) \in \{0, 1\}^2 \]
values of the Head Start center characteristics. Using this notation, the conditional independence assumption in Assumption HSIS(ii) imposes the following restriction
\[
Q(\bar{y}, u, \bar{c}, 0, \bar{z}) \sum_{\bar{y} \in \{0, 1\}^3, u \in U, \bar{c} \in C^2} Q(\bar{y}, u, \bar{c}, 0, \bar{z}) = Q(\bar{y}, u, \bar{c}, 1, \bar{z}) \sum_{\bar{y} \in \{0, 1\}^3, u \in U, \bar{c} \in C^2} Q(\bar{y}, u, \bar{c}, 1, \bar{z})
\]
for all values of \( \bar{y} \in \{0, 1\}^3, u \in U, \bar{c} \in C^2 \) and \( \bar{z} \in \{0, 1\}^2 \). Since the restrictions imposed by the distribution of the observed data imply that both the denominators can be written in terms of the known data distribution by
\[
\sum_{\bar{y} \in \{0, 1\}^3, u \in U, \bar{c} \in C^2} Q(\bar{y}, u, \bar{c}, z, \bar{z}) = \sum_{y \in \{0, 1\}, d \in D} P(y, d, z, \bar{z})
\]
for each \( z \in \{0, 1\} \), it then follows that above restriction can be re-written as a linear restriction on \( Q \) in the form of Assumption 4.2 as
\[
\sum_{y \in \{0, 1\}, d \in D} P(y, d, 1, \bar{z}) \cdot Q(\bar{y}, u, \bar{c}, 0, \bar{z}) - \sum_{y \in \{0, 1\}, d \in D} P(y, d, 0, \bar{z}) \cdot Q(\bar{y}, u, \bar{c}, 1, \bar{z}) = 0 \quad (20)
\]
for all values of \( \bar{y} \in \{0, 1\}^3, u \in U, \bar{c} \in C^2 \) and \( \bar{z} \in \{0, 1\}^2 \).

Given the restrictions imposed by the distribution of the observed data and by additional assumptions on the model distribution, I next ask what we can learn about the parameter of interest. This is formally defined by the identified set, i.e. the set of feasible parameter values such that the distribution of the model satisfies the various imposed restrictions. In order to formally state the identified set, denote first by \( Q_W \) the set of all probability mass functions on the sample space \( W \). The identified set can then be stated as follows
\[
\Theta = \{ \theta_0 \in \mathbb{R} : \theta(Q) = \theta_0 \text{ and } Q \in Q \}, \tag{21}
\]
where
\[
Q = \{ Q \in Q_W : Q \text{ satisfies (16) and Assumption 4.2 } \}. \tag{22}
\]
is the set of all model distributions that satisfy the restriction imposed by the data and the assumptions. Note that if the identified set is empty then model is said to be mis-specified, i.e. the data is incompatible with the imposed assumptions on the model. Further, if the identified set contains a single point then the parameter is said to be point identified, and if it is a proper subset of all possible values that the parameter can take then the parameter is set to be partially or set identified.

4.3 Identified Set

Due to the complicated structure of the underlying model, analytically characterizing the identified set in (21) is a difficult task. Instead, in the proposition below, I state an alternate more useful characterization of the identified set. This characterization in particular permits the use of tractable linear programming methods to obtain the identified set. Moreover, in the proof of this proposition presented below, I emphasize the specific role that conditions stated in Assumptions 4.1-4.2 play in ensuring that the optimization problem is a linear program. In this proposition below, I introduce the following additional quantity

\[ \tilde{\theta}(Q) = \sum_{w \in W} a_{\text{num}}(w) \cdot Q(w) . \]

Proposition 4.1. Suppose that \( Q \) in (22) is non-empty and the parameter of interest satisfies Assumption 4.1 such that

\[ \sum_{w \in W} a_{\text{den}}(w) \cdot Q(w) > 0 \]  

holds for every \( Q \in Q \). Then the identified set in (21) can be written as

\[ \Theta = [\theta_l, \theta_u] , \]  

where the lower and upper bounds of this interval are solutions to the following two linear programming problems

\[ \theta_l = \min_{\gamma, \{Q(w)\}_{w \in W}} \tilde{\theta}(Q) \text{ and } \theta_u = \max_{\gamma, \{Q(w)\}_{w \in W}} \tilde{\theta}(Q) , \]  

subject to the following constraints:

(i) \( \gamma \geq 0 . \)

(ii) \( 0 \leq Q(w) \leq \gamma \) for every \( w \in W . \)

(iii) \( \sum_{w \in W} Q(w) = \gamma . \)

(iv) \( \sum_{w \in W_x} Q(w) = \gamma \cdot P(x) \) for every \( x \in X . \)
\[
\sum_{w \in W} a_s(w) \cdot Q(w) \leq \gamma \cdot b_s \text{ for every } s \in S .
\]

\[
\sum_{w \in W} a_{\text{den}}(w) \cdot Q(w) = 1 .
\]

**Proof:** To begin, note that \( Q_W \) is closed and convex. Further, note that \( Q \) is a set of distributions in \( Q_W \) that is obtained by placing linear equality and inequality constraints imposed by the data in (16) and by restrictions characterized in Assumption 4.2. This in turn implies that \( Q \) is a closed and convex set as well.

Next, it follows from Assumption 4.1 that \( \theta(Q) \) is a linear-fractional function of \( Q \) where as stated in the theorem the denominator is required to be positive, i.e. (23) holds, for every \( Q \in Q \). Along with \( Q \) being a closed and convex set, this in turn implies that \( \theta(Q) \) is a closed and convex set in \( R \). More specifically, it follows that the identified set in (21) can be written as the closed interval in (24), where the lower bound and upper bound are given by

\[
\theta_l = \min_{Q \in Q} \theta(Q) \text{ and } \theta_u = \max_{Q \in Q} \theta(Q) .
\]

In order to complete the proof, note that both the optimization problems in (26) have linear-fractional objectives as guaranteed by Assumption 4.1 and a finite number of linear constraints as guaranteed by the data restrictions in (16) and by Assumption 4.2. Such optimization problem are commonly referred to as linear fractional programs. This structure required to ensure that the optimization problems in (26) can stated as linear-fractional programs is the fundamental value in considering Assumptions 4.1-4.2. This is useful because, as shown by Charnes and Cooper (1962), linear-fractional programs can be equivalently re-stated as linear programs of the form in (25) from which the result then directly follows.

To be specific, this restatement follows in two steps. First, it introduces the so-called Charnes-Cooper transformation

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

which is well defined given that (23) holds for every \( Q \in Q \). Then, by transforming the restrictions and parameters written in terms of \( Q \) to those in terms of \( \gamma \) and \( \tilde{Q} \), the minimization and maximization problem in (26) can be re-written as that in (25).

In the proof of Proposition 4.1 above, it is worth re-emphasizing that the main role that Assumptions 4.1-4.2 play is in ensuring that the optimization problems in (26) can be restated as linear programs. In principle, even in the absence of such linear assumptions, the identified set can still be characterized as solutions to optimization problems - see Torgovitsky (2016) for an example of such a result. Such optimization problems may however not share the same reliability
and tractability as linear programs for the dimensions present in this application, noted in Remark 4.5.

In addition to Assumptions 4.1-4.2, note that computing the identified set using Proposition 4.1 requires two further conditions. First, it requires \( Q \) to be non-empty, i.e. it requires the imposed model to be not mis-specified, to ensure that the identified set is an interval. If this is not the case, the linear program automatically terminates, which in turn provides a simple specification check for whether the assumptions imposed on the model are compatible with the observed data. Second, it requires the denominator of the parameter of interest to be positive, i.e. (23) holds, for every \( Q \in Q \). This is to ensure that the parameter of interest 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 as its corresponding denominator by construction is always one. The above proposition can then be employed to compute the lower bound for this auxiliary parameter to check if it is strictly positive given the restrictions imposed by the observed data and additional assumptions imposed on the model.

As noted when describing Assumptions 4.1-4.2, it is important to emphasize that an attractive feature of the linear programming method in Proposition 4.1 is its generality. In particular, this generality flexibly allows the study of various parameters under restrictions imposed by Assumption HSIS and also combinations of other assumptions. This permits in the empirical results in Section 6 to study various parameters for various specifications of restrictions without the need to analytically re-derive the identified set in each case.

The benefits of using computational procedures for obtaining the identified set has also been noted in previous studies on settings where deriving analytical bounds is difficult - see, for example, Manski (2007) in a related setting discussed in Remark 3.1 and Torgovitsky (2016) in a dynamic potential outcome setting. These studies have also proposed to focus attention on parameters and restrictions that are linear in the model distribution in order to use linear programs to compute the identified set. More specifically, with respect to the class of parameter, these previous proposals limit attention to only those that could be written as linear functions of the underlying model distribution. In contrast, by leveraging the result from Charnes and Cooper (1962), I also focus on a larger class of linear-fractional functions. In the setting of this paper, this in turn permits the study of an additional important class of conditional parameters such as the option value parameters described in Examples 4.2-4.3.

**Remark 4.3.** In absence of the binary transformations in (4) and (5), the model distribution is an infinite dimensional object as the test score outcome is continuously distributed. In this case, the identified set for similar parameters could also possibly be written as solutions to infinite dimensional programs under related conditions - see Mogstad et al. (2017) and Torgovitsky (2017) for examples of such results. However, computational methods that maximize, or minimize, over
infinite dimensional spaces are in general intractable. To this end, the binary transformation considered here, and more generally any discrete transformation, translates the underlying space to a finite dimensional one, which allows for the use of tractable computational methods.

Remark 4.4. The identified set stated in (21) may not be sharp as only information from the distribution of the binary transformed test score is used rather than from underlying continuous test score. However, had the outcome of interest been binary, the identified set as stated would be sharp. Furthermore, had the outcome of interest been discrete, the analysis as previously noted can also be extended to obtain the sharpest set.

Remark 4.5. When employing Proposition 4.1 under the restrictions imposed by Assumption HSIS, there are 6144 unknown variables determined by the probability mass function of the model along with 12288 restrictions determined by (ii), 48 restrictions determined by (iv), and 3073 restrictions determined by (v).

5 Statistical Inference

The identification analysis in the previous section obtained the identified set assuming that the population distribution $P$ of the observed data was known. The empirical results apply this analysis to obtain an estimate of the identified set using the empirical distribution of the HSIS sample data. Here, to be precise, the HSIS sample data is given by

$$X^{(n)} = \{X^{(n)}_{g} : 1 \leq g \leq G\}$$

(27)

where $n$ denotes the total sample size, and

$$X^{(n)}_{g} = \{X_{ig} : 1 \leq i \leq n_{g}\}$$

denotes the cluster of observations of the random variable $X$ in (7) for all the $i$th sampled children from the $g$th sampled Head Start center from the experiment.

In this section, I describe how I construct confidence intervals for the parameters of interest in the empirical results. Before proceeding, I note that deriving formal results for constructing confidence intervals for partially identified parameter that are uniformly asymptotically valid over a large class of distributions is outside the scope of this paper. As noted in Imbens and Manski (2004), ensuring uniform validity is in particular important in partially identified settings. In this section, I simply provide a description of how an existing method from the literature can practically be applied for the empirical results in order to sensibly capture the sampling uncertainty in the estimates.

I begin by noting that Kaido et al. (2016) and Mogstad et al. (2017) have recently proposed methods to obtain uniformly asymptotically valid confidence intervals for partially identified parameters obtained as solutions to linear programming problems. These methods specifically exploit
the geometrical structure present in linear programming problems. Implementing these methods
require solving non-linear optimization problems (over many bootstrap draws) that critically re-
quire that the obtained solution to be the global optimum rather than a local one. However, given
the dimensions of the computational program noted in Remark 4.5 and the computational limita-
tions noted in Remark 5.1, repeatedly solving such optimization problems reliably in this empirical
application is not feasible.

Instead, I use a computationally tractable version of the profiled subsampling procedure pro-
posed by Romano and Shaikh (2008), which is shown to be uniformly asymptotically valid under
some high level conditions - see also Politis and Romano (1994) and Romano et al. (2012) for
related results on the validity of subsampling. The sampling framework that the described method
aims to capture is one where: (i) the random variables \( X_{ig} \) are identically distributed by \( P \)
and uncorrelated across centers, i.e. independent across the index \( g \); and (ii) the number of centers is
large, i.e. \( G \to \infty \), and the number of applicants per center is small, i.e. \( n_g \) is fixed for each \( g \). The
description of this method follows the exposition in Torgovitsky (2016), which restates the feasible
set of underlying model distributions in (22) in terms of a moment equalities model.

To this end, first note that the restriction imposed by the data on the model distribution in
(16) can be re-written as

\[
E_P[m_{dat,x}(X,Q)] = 0 \text{ for all } x \in \mathcal{X},
\]

where

\[
m_{dat,x}(X,Q) = \sum_{w \in W} Q(w) - I\{X = x\}.
\]

Next, suppose that the set of imposed restrictions \( S \) in Assumption 4.2 can be partitioned into \( S_1 \)
and \( S_2 \), where each of the restrictions in these partitions satisfy some specific conditions stated in
the following assumptions.

**Assumption 5.1.** For all \( s \in S_1 \), the linear restriction characterized by (17) is such that \( a_s : W \to \mathbb{R} \) and \( b_s \) are known and do not depend on \( P \).

**Assumption 5.2.** For all \( s \in S_2 \), the linear restriction characterized by (17) is an equality, and
further is such that \( a_s(w) \) for each \( w \in W \) and \( b_s \) are linear functions of \( P \).

Assumption 5.1 states that all restrictions imposed by \( s \in S_1 \) are deterministic in the sense that
they do not depend on features of the data and are hence not estimated. The restriction imposed
by Assumption HSIS(i) in (19), for example, satisfies this assumption. The restrictions satisfying
this assumption determine the feasible “parameter” space of the model, i.e.

\[
\mathcal{Q}' = \{Q \in \mathcal{Q}_W : M_s(Q) \leq b_s \text{ for } s \in S_1\}.
\]
In contrast, Assumption 5.2 states all restrictions imposed by \( s \in S_2 \) are stochastic in the sense that they depend linearly on the distribution of the data and are hence estimated using the sample data. The restrictions imposed by Assumption HSIS(ii) in (20), for example, satisfy this assumption. Moreover, this assumption implies that each restriction (17) for \( s \in S_2 \) can be re-written as a moment equality, i.e.

\[
E_P[m_s(X, Q)] = 0 \quad \text{for all } s \in S_2 ,
\]

which can be formally shown using the same arguments as used for showing the moment equalities imposed by the data restrictions. Using this notation, we can see that the feasible set of underlying model distributions is defined by the parameter space in (29) and by the moment equalities in (28) and (30).

Using the restated moment equalities model, I now describe a test at a pre-specified level \( \alpha \in (0, 1) \) based on Romano and Shaikh (2008) for a null hypothesis that the parameter of interest is equal to a given value \( \theta_0 \in \mathbb{R} \), i.e. \( \theta(Q) = \theta_0 \) - see Canay and Shaikh (2017) for a review of other methods that can be applied in moment equalities models. Confidence intervals can then be obtained by inverting the test, i.e. by collecting the set of all values of \( \theta_0 \) for which the test does not reject at level \( \alpha \). The test requires a choice of test statistic that rejects the null hypothesis for large values. I propose using the following profiled test statistic

\[
TS_n(\theta_0) = \sqrt{G} \min_{Q \in Q(\theta_0)} \sum_{x \in X} |\hat{m}_{\text{dat},x}(Q)| + \sum_{s \in S_1} |\hat{m}_s(Q)| ,
\]

where \( \hat{m}_{\text{dat},x}(Q) \) is an empirical analogue of the the moment in (28), \( \hat{m}_s(Q) \) is an empirical analogue of the moment in (30), and \( Q(\theta_0) = \{ Q \in Q' : \theta(Q) = \theta_0 \} \). As shown in Appendix A, an important practical benefit of this choice of test statistic is that it can be solved as a solution to a linear programming problem. In order to obtain a critical value, the test draws \( B \) subsamples of size \( b \) drawn randomly without replacement from the clusters of Head Start center observations in (27) and computes \( TS_{n,b,j}(\theta_0) \), which is the test statistic in (31) using the \( j \)th subsample of size \( b \). Denoting by \( \hat{c}_n(1 - \alpha, \theta_0) \) the \( (1 - \alpha) \)-quantile of the subsampling distribution

\[
L_n(t, \theta_0) = \frac{1}{B} \sum_{j=1}^{B} 1\{TS_{n,b,j}(\theta_0) \leq t \} ,
\]

the test can be then be denoted by

\[
\phi_{n}^{SS}(\theta_0) = 1\{TS_n(\theta_0) \geq \hat{c}_n(1 - \alpha, \theta_0) \} .
\]

The \( (1 - \alpha) \) confidence interval for the parameter of interest is in turn given by

\[
C_n = \{ \theta_0 \in \mathbb{R} : \phi_{n}^{SS}(\theta_0) = 0 \} .
\]
Note that the above described test requires computing the test statistic over multiple subsamples and further requires inverting the test to construct confidence intervals. Despite this, by ensuring that all the performed optimization problems are linear programs, I find that constructing confidence intervals using this test can still be relatively tractable.

**Remark 5.1.** Constructing confidence intervals can be an expensive computational exercise due to the fact that the test statistic needs to be computed over many bootstrap draws or subsamples and further due to the inversion of the test. For the purposes of the HSIS data, due to a restricted access data agreement, an important computational limitation was that this exercise needed to be performed on a personal computer without any assistance from a computing cluster. ■

**Remark 5.2.** The subsampling test requires a choice of block size $b$. Formally, the only requirements are that $b \to \infty$ and $\frac{b}{G} \to 0$ as $G \to \infty$. In the empirical results, I take $b = G^{2/3}$ following Bugni (2016). Moreover, I take the number of drawn subsamples $B$ to be 100 due to the computational limitations noted in Remark 5.1. ■

**Remark 5.3.** By drawing subsamples from the clusters of Head Start centers, the described test accounts for the dependence present between observations sampled from a given Head Start center. Randomized experiments may however introduce additional dependence due to various randomization schemes used to assign treatment that the described test does not capture. See Bugni et al. (2017a,b) for formal results on performing valid inference for point identified parameters under such dependence introduced by a general class of randomization schemes. ■

## 6 Empirical Results

In this section, I present the empirical results using the formal analysis developed in the preceding sections on the data from the HSIS. Table 2 begins by presenting the estimated identified sets and 95% confidence intervals for the parameters discussed in Section 4.1 using only the restrictions imposed by the observed data and Assumption HSIS.

Observe in Table 2 that the lower bound of the parameter $\text{PE}_{210|10}$, the proportion of children whose parents exercise the Head Start option when an alternate preschool is available, is equal to zero. Since this parameter corresponds to the denominator of the parameter $\text{PBOE}_{210|10}$ in Example 4.3, the identified set for the latter parameter hence cannot be computed. In particular, the linear programming method in Proposition 4.1 cannot be employed since, as noted in the previous section, it requires that the denominator for the parameter of interest is strictly positive under the feasible model distributions. In order to still measure the benefits of a Head Start option when an alternative preschool is present, using the notation from Section 4.1, I additionally report results for

$$
\text{PB}_{210|10}(Q) = \text{Prob}_Q[\text{Y}^\dagger(\text{D}_{(2,1,0)}) = 1, \text{Y}^\dagger(\text{D}_{(1,0)}) = 0],
$$

(32)
Table 2: Estimated worst case bounds and confidence intervals for each age group.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Age 3</th>
<th>Age 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Head Start versus Home care</td>
<td></td>
</tr>
<tr>
<td><strong>PB&lt;sub&gt;20&lt;/sub&gt;</strong></td>
<td>[0.000, 0.537]</td>
<td>[0.000, 0.627]</td>
</tr>
<tr>
<td></td>
<td>[0.000, 0.689]</td>
<td>[0.000, 0.693]</td>
</tr>
<tr>
<td><strong>PI&lt;sub&gt;20&lt;/sub&gt;</strong></td>
<td>[0.000, 0.307]</td>
<td>[0.000, 0.432]</td>
</tr>
<tr>
<td></td>
<td>[0.000, 0.460]</td>
<td>[0.000, 0.572]</td>
</tr>
<tr>
<td><strong>ATE&lt;sub&gt;20&lt;/sub&gt;</strong></td>
<td>[-0.149, 0.386]</td>
<td>[-0.280, 0.420]</td>
</tr>
<tr>
<td></td>
<td>[-0.239, 0.506]</td>
<td>[-0.400, 0.578]</td>
</tr>
<tr>
<td><strong>Option value of Head Start without an alternate preschool</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>**PBOE&lt;sub&gt;20</td>
<td>0&lt;/sub&gt;**</td>
<td>[0.000, 0.594]</td>
</tr>
<tr>
<td></td>
<td>[0.000, 0.694]</td>
<td>[0.000, 0.796]</td>
</tr>
<tr>
<td><strong>PE&lt;sub&gt;20&lt;/sub&gt;</strong></td>
<td>[0.866, 0.917]</td>
<td>[0.799, 0.911]</td>
</tr>
<tr>
<td></td>
<td>[0.786, 0.959]</td>
<td>[0.729, 0.960]</td>
</tr>
<tr>
<td><strong>Option value of Head Start with an alternate preschool</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>**PB&lt;sub&gt;2</td>
<td>10&lt;/sub&gt;**</td>
<td>[0.000, 0.670]</td>
</tr>
<tr>
<td></td>
<td>[0.000, 0.730]</td>
<td>[0.000, 0.622]</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>**PBOE&lt;sub&gt;2</td>
<td>10&lt;/sub&gt;**</td>
<td>-</td>
</tr>
<tr>
<td>**PE&lt;sub&gt;2</td>
<td>10&lt;/sub&gt;**</td>
<td>[0.000, 0.866]</td>
</tr>
<tr>
<td></td>
<td>[0.000, 0.930]</td>
<td>[0.000, 0.893]</td>
</tr>
</tbody>
</table>

Notes: For each parameter, the upper panel denotes the estimated identified set and the lower panel denotes the 95% confidence interval. Appendix C provides details on how the data used for these results was constructed.

which denotes the proportion of children who strictly benefit from a Head Start option when an alternate preschool option is also present. However, as further noted in Section 6.2, the lower bound of the denominator can be strictly positive under restrictions imposed by additional assumptions, which then permits the study of PBOE<sub>2|10</sub> under these assumptions.

The so-called worst case bounds in Table 2 are substantially informative on inferring how parents choose between the a Head Start preschool and home care when a Head Start offer is made, i.e. the parameter denoted by PE<sub>20|0</sub>. Using only the data and the experimental design of the HSIS, the estimated identified sets implies that between 79.9% to 91.7% of the parents across both age cohorts exercise their Head Start option when only home care is the outside option. However, the data and the experimental design are generally uninformative with respect to reaching stronger conclusions based on the remaining parameters. For example, one cannot infer whether Head Start positively or negatively affects test scores as measured across various parameters as zero is included in the identified set for these parameters.

In the remainder of this section, I illustrate additional identifying assumptions with the aim of tightening the worst case bounds and reaching stronger conclusions regarding the efficacy of
Head Start. In order to easily employ the linear programming method in Proposition 4.1, note that these assumptions impose restrictions that satisfy Assumption 4.2. In Section 6.1 below, I first describe these assumptions and provide a discussion on their plausibility in the context of the HSIS. In Section 6.2 below, I then report the estimated bounds under these assumptions and provide a discussion on their identifying power.

6.1 Additional Identifying Assumptions

An important feature of the below described nonparametric assumptions is their transparency in terms of the restrictions that they impose on the empirical setting. This allows us to easily assess whether these restrictions are justified in the context of the HSIS - see Manski (2003) for a general discussion on the importance of such an assessment. In the description of these assumptions below, I focus attention to this assessment and leave the explicit mathematical derivations of their imposed restrictions satisfying Assumption 4.2 for Appendix B. Furthermore, from the explicit forms of these restrictions, it is apparent that they satisfy either Assumption 5.1 or Assumption 5.2, which in turn allows constructing confidence intervals using the subsampling procedure discussed in Section 5.

6.1.1 Unaltered Alternatives

Recall that Assumption HSIS required that the experimental offer guaranteed that a Head Start preschool was in the choice set. The assumption however imposed no restrictions on the choice set had an experimental offer not been made. In the following assumption, I state a reasonable assumption on the relation between the choice set with and without an offer.

Assumption UA. \( C(1) = C(0) \cup \{2\} \).

Assumption UA states that apart from introducing a Head Start preschool in the child’s choice set of care setting, it leaves the choice set completely unaltered. This assumption makes an implicit restriction on the application decisions of the parents to obtain preschool slots, i.e. the unobserved process that generates the choice sets. For example, it rules out cases such as not obtaining an experimental offer induced the parents to apply to additional alternate preschools and successfully obtain an offer, i.e. a case where there is an alternate preschool present in \( C(0) \) but not in \( C(1) \).

6.1.2 Site Level Instruments

Previous studies on the HighScope curriculum and on class size ratios suggest that the considered Head Start preschool characteristics would affect the child’s test scores under Head Start. Moreover, due to competition between preschools, it is likely that they would also affect the child’s test scores
under an alternative preschool. However, it is reasonable to believe that it would play no role on the child’s test score under home care, which motivates the following assumption.

**Assumption SLI.** $Y(0) \perp (HC, CS)$ .

Assumption SLI states that the considered baseline covariates of a Head Start preschool are independent of the child’s test scores had the child attended home care. Such an assumption rules out settings where the characteristics of a Head Start preschool are endogenously determined by the quality of families applying there, and hence the outcome under home care.

**Remark 6.1.** Kline and Walters (2016) use Head Start preschool characteristics with additional assumptions to point identify local average treatment effects on sub-populations other than the compliers, described in Remark 4.1. These additional assumptions essentially require that the local average treatment effects conditional on values of the Head Start characteristics are homogeneous across these values.

**Remark 6.2.** It may also be reasonable to assume that the test scores under Head Start weakly improve with the described Head Start characteristics in the sense of the monotone instrumental variable assumption of Manski and Pepper (2000, 2009). I find that this assumption does not have much identifying power in this setting. Note however that such an assumption may have more identifying power in alternate applications.

### 6.1.3 Monotone Treatment Response

A large body of research on early childhood interventions suggest that preschool enrollment may be beneficial in comparison to home care. In turn, it may be reasonable to believe that both Head Start and alternate preschools are at least as good with respect to a child’s test score in comparison to home care, which motivates the following assumption.

**Assumption MTR.** For each $d \in \{1, 2\}$, $Y(d) \geq Y(0)$ .

Assumption MTR states that the potential test scores under both preschools is weakly greater than the potential test score under home care. This assumption is a version of the Monotone Treatment Response (MTR) assumption proposed by Manski (1997b). In particular, note that such an assumption imposes that no child can possibly lose from attending any of the two preschools in comparison to home care. To the extent that some children may strictly be better off under home care, such an assumption may be subject to suspect.

### 6.1.4 Roy Model

The parents’ preferences for care settings may be related to the potential test score their child may receive under each of these settings. In the case where it is assumed that these preferences are
based solely on optimizing these potential test scores, we obtain a version of the Roy model stated as follows.

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

Assumption Roy states that parents are omniscient in the sense that they exactly know their child’s potential test scores under each setting and then prefer the care setting where their child would attain the highest test score. To be more clear, note that the above assumption can equivalently be restated in terms of utilities as

$$Y(d') > Y(d) \implies U(d') > U(d)$$

for every $d, d' \in \mathcal{D}$, i.e. if the potential test score under a given treatment is higher than that under another then the utility under that treatment is also higher than the utility under the other. The credibility of this assumption can be suspect for multiple reasons of which I state two important ones relevant to this setting. First, parents may not exactly know their child’s potential test score under each care setting but may only have an expectation of what it may be. In turn, parents’ preferences may be based on selecting care settings that maximize expected rather than actual potential test scores. To the extent that these two values differ, the Roy model may not reasonable capture how preferences may be related to the potential test scores. Second, parents may make enrollment decisions based not only on test scores but also on additional factors such as possible costs that may be incurred from attending a given preschool.

Assumption Roy can be weakened to partially address the second concern noted above. Since the preschools may possible be closely related, it is reasonable to believe that costs incurred at either are similar. Hence, it may be the case such that at least the preferences between the two preschools are based solely on potential test scores as stated in the following assumption.

**Assumption SemiRoy.** For each $d, d' \in \{1, 2\}$, if $Y(d') > Y(d)$ then $d_{U,(d,d')} = d'$.

**Remark 6.3.** Assumption Roy has an important distinction from versions of the Roy model previously studied in the literature such as, for example, in Heckman and Honore (1990) and Mourifie et al. (2015). These versions do not account for the possibility that the treatment is not selected from the set of all treatments, i.e. it does not account for unobserved heterogeneity in choice sets.

### 6.2 Bounds under Additional Identifying Assumptions

Table 3 and Table 4 report for the age 3 cohort and the age 4 cohort, respectively, the estimated identified set for the parameters presented in Table 2 under the additional restrictions imposed by the various assumptions described in the previous section. I organize the discussion of the results by the various assumptions:
Table 3: Estimated identified sets for the age 3 cohort

<table>
<thead>
<tr>
<th>Assumption</th>
<th>UA ✓ ✓ ✓ ✓ ✓ ✓</th>
<th>SLI ✓ ✓ ✓</th>
<th>MTR ✓ ✓ ✓ ✓</th>
<th>Roy ✓ ✓ ✓</th>
<th>SemiRoy ✓ ✓</th>
</tr>
</thead>
</table>

Parameter

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Head Start versus Home care</th>
<th>Option value of Head Start without an alternate preschool</th>
<th>Option value of Head Start with an alternate preschool</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB₂₀éro</td>
<td>0.099 0.029 0.029 0.000 0.070 0.096 0.096 0.070 0.096 0.537 0.386 0.352 0.537 0.386 0.320 0.309 0.352 0.286</td>
<td>0.010 0.032 0.032 0.000 0.076 0.105 0.105 0.076 0.105 0.561 0.414 0.407 0.594 0.396 0.335 0.326 0.388 0.326</td>
<td>0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 - - - 0.000 0.000 0.000 0.000 0.000 0.000</td>
</tr>
<tr>
<td>PL₂₀éro</td>
<td>0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000</td>
<td>0.000 - - - 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000</td>
<td>0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000</td>
</tr>
<tr>
<td>ATE₂₀éro</td>
<td>-0.149 0.029 -0.031 -0.149 0.070 0.096 0.096 -0.031 0.027 0.386 0.386 0.352 0.375 0.386 0.320 0.309 0.352 0.286</td>
<td>0.386 0.386 0.352 0.375 0.386 0.320 0.309 0.352 0.286</td>
<td></td>
</tr>
</tbody>
</table>

Notes: For each parameter and combination of assumptions denoted by the various check marks, the upper panel denotes the lower bound of the identified set and the lower panel denotes the upper bound of the identified set. Appendix C provides details on how the data used for these results was constructed.

**UA**: Assumption UA has identifying power in terms of the proportion who benefit, and proportion who are induced to enroll into Head Start when provided the option in the presence of an alternate preschool option. It provides evidence that there is a strictly positive proportion of three-year-old children, at least approximately 1%, who benefit from Head Start in comparison to home care. It is also provides evidence that a strictly positive proportion of parents from 20.7% to 86.6% across the two age cohorts are induced to enroll their child into Head Start when provided the option when an alternate preschool option was available to them.

**SLI**: In unreported results, I find that Assumption SLI does not have much identifying power in this setting in terms of reaching stronger conclusions over what can already be learned.
Table 4: Estimated identified sets for the age 4 cohort

<table>
<thead>
<tr>
<th>Assumption</th>
<th>UA</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLI</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MTR</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Roy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SemiRoy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Start versus Home care</td>
</tr>
<tr>
<td><strong>PB</strong>&lt;sub&gt;20&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

| Option value of Head Start without an alternate preschool                |
| **PBOE**<sub>20</sub>  | 0.000  | 0.000  | 0.000  | 0.000  | 0.049  | 0.049  | 0.000  | ✓  | 0.000  |
|                       | 0.654  | 0.473  | 0.456  | 0.695  | 0.427  | 0.399  | 0.402  | ✓  | 0.402  |
|                       | 0.799  | 0.799  | 0.799  | 0.799  | 0.799  | 0.799  | 0.799  | ✓  | 0.799  |
|                       | 0.911  | 0.911  | 0.911  | 0.911  | 0.911  | 0.911  | 0.911  | ✓  | 0.911  |

| Option value of Head Start with an alternate preschool                  |
| **PB**<sub>210</sub>  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | 0.000  | ✓  | 0.000  |
|                       | 0.523  | 0.333  | 0.333  | 0.532  | 0.262  | 0.262  | 0.310  | ✓  | 0.310  |
|                       | 0.738  | -  | -  | -  | 0.000  | 0.000  | -  | ✓  | -  |
|                       | 0.799  | 0.799  | 0.797  | 0.799  | 0.799  | 0.799  | 0.799  | ✓  | 0.799  |

Notes: For each parameter and combination of assumptions denoted by the various check marks, the upper panel denotes the lower bound of the identified set and the lower panel denotes the upper bound of the identified set. Appendix C provides details on how the data used for these results was constructed.

from the worst-case bounds. However, when combined with other assumptions, they can have identifying power as discussed below.

**MTR**: Assumption MTR provides evidence that there is a strictly positive proportion of three-year-old children, from 2.9% to 38.6% and from 3.2% to 41.4%, who benefit from Head Start as measured across two alternate policy dimensions by PB<sub>210</sub> and PBOE<sub>20</sub>, respectively. Moreover, since by construction this assumption imposes that there is a zero proportion of children who lose from Head Start, it implies that PB<sub>210</sub> is equal to ATE<sub>210</sub>, the average treatment effect of Head Start in comparison to home care, which is then also strictly positive.

**Roy + SemiRoy**: Similar to Assumption MTR, Assumption Roy also provides evidence
that there is a strictly positive proportion of three-year-old children, from 2.9% to 35.2%
and from 3.2% to 40.7%, who benefit from Head Start as measured by $\text{PB}_{2|0}$ and $\text{PBOE}_{20|0}$, respectively. However, in contrast, it allows for at most 10% of children who are strictly better off under home care as observed in $\text{PL}_{2|0}$. Comparing the results between Assumption Roy and Assumption SemiRoy implies that using the total preference ordering between the care settings is important in this setting in terms of identifying power.

**Combinations:** Combining Assumptions UA-MTR strengthens the estimated bounds for the three-year-old children from 7.0% to 38.6% and from 7.6% to 39.6% for the parameters $\text{PB}_{2|0}$ and $\text{PBOE}_{20|0}$, respectively. This combination additionally provides evidence that there is strictly positive proportion of four-year-old children, from 4.5% to 42.0% and from 4.9% to 42.7%, who benefit from Head Start as measured by $\text{PB}_{2|0}$ and $\text{PBOE}_{20|0}$, respectively. This additional evidence however is absent when the assumptions are specified separately. These estimated identified sets are further tightened when Assumptions SLI-SemiRoy are also sequentially imposed. When combining these assumptions with Assumption Roy instead of Assumption MTR, similar results are found for the age three cohort. However, for the age four cohort, a combination of Assumption UA, and Assumption Roy or Assumption SemiRoy is considered infeasible by the linear program, i.e. the data is not compatible with these combinations of assumptions.

As noted above, the estimated identified sets across various specifications of assumptions are fairly informative for the majority of the parameters. An exception is the estimated identified set for the parameters that measure the proportion of children who benefit from the option value of Head Start when an alternative preschool is available, i.e. $\text{PB}_{210|10}$ and $\text{PBOE}_{210|10}$. This is primarily due to the fact that no assumptions are imposed on whether an alternative preschool is present in a child’s choice set by the experimental setting or by the additional identifying assumptions.

Overall, across various specifications of assumptions, the evidence suggested by the estimated identified sets can be qualitatively summarized as follows: (i) Head Start is more effective in comparison to home school in terms of improving short-term test scores; (ii) providing a Head Start option is effective in terms of inducing people to attend Head Start and further benefiting those children who attend in terms of short-term test scores; (iii) these summarized results are more pronounced for the age three cohort than the age four cohort; and (iv) results on the whether there is a strictly positive effect of providing a Head Start option when an alternate is present is inconclusive, though an upper value that bounds the possible effect is fairly tight.

Recall from Section 5 that due to the practical limitations computing confidence intervals for all the combinations of parameters and specifications of assumptions reported above is difficult. For the purposes of illustration, I only report and discuss 95% confidence intervals for some choices of parameters and assumptions. For the parameters $\text{PB}_{2|0}$ and $\text{PBOE}_{20|0}$, I find that: (i) under
the combination of Assumption UA and Assumption MTR, the confidence interval is $[0.000, 0.506]$ and $[0.000, 0.526]$, respectively; whereas (ii) under the combination of Assumption UA, Assumption MTR and Assumption SLI, the confidence interval is $[0.000, 0.570]$ and $[0.000, 0.505]$, respectively.

In summary, at least for these few choice of parameters and assumptions, I find that the positive benefits of the Head Start using estimated identified sets are not statistically significant using the subsampling procedure at a 95% confidence level. However, as previously observed for example by Torgovitsky (2016) in a related setting, it is possible that the subsampling confidence intervals can be highly conservative in this setting. I leave the use of tractable, less conservative and valid statistical inference methods in this empirical application for future work.

7 Conclusion

This paper studies the identification of program effects in settings with latent choice sets. This analysis was developed in the context of the Head Start Impact Study (HSIS) social experiment. In particular, the analysis required a deep understanding of the experimental design and the setting in order to leverage useful information to reach informative conclusions on various parameters of interests.

More specifically, to explicitly account for the latent choice sets, I developed a nonparametric model that was tightly connected to the experimental design of the study. For a large class of parameters in this model, I showed that the identified set using the information provided by the experimental design and various unique institutional details of the setting can be constructed using a flexible linear programming method. Applying the developed analysis, I found that Head Start is effective across various policy dimensions, which corroborates the positive findings of previous studies under weak nonparametric assumptions.

To conclude, I would like to emphasize that though developed in the context of the HSIS the underlying framework applies more generally. In particular, it directly extends to alternate experimental settings with latent choice sets and can also be extended to observational settings with latent choice sets. Indeed, the described parameters of interest and identifying assumptions can be viewed as as illustrative examples with similar counterparts in the alternate empirical settings.
A Restatement of Test Statistic

In this appendix, I describe how the test statistic in (31) can be restated as a solution to a linear program. This restatement in particular follows similar ones stated in Mogstad et al. (2017) and Torgovitsky (2016). To this end, begin by noting that test statistic can be explicitly be written as

\[ TS_n(\theta_0) = \sqrt{G} \min_{\{Q(w)\}_{w \in W}} \sum_{x \in X} |\hat{m}_{\text{dat},x}(Q)| + \sum_{s \in S_2} |\hat{m}_s(Q)| , \]

subject to the following constraints:

(i) \[ \sum_{w \in W} a_{\text{num}}(w) \cdot Q(w) = \theta_0 \cdot \sum_{w \in W} a_{\text{den}}(w) \cdot Q(w) . \]

(ii) \[ 0 \leq Q(w) \leq 1 \text{ for every } w \in W . \]

(iii) \[ \sum_{w \in W} Q(w) = 1 . \]

(iv) \[ \sum_{w \in W} a_s(w) \cdot Q(w) \leq b_s \text{ for every } s \in S_1 . \]

By introducing two additional variables for each data restriction and stochastic restriction, it can be shown that the above program can be re-written as a linear program. More specifically, the above optimization problem is equivalent to

\[ TS_n(\theta_0) = \sqrt{G} \min_{\{Q(w)\}_{w \in W}, \{\mu_1^+(x)\}_{x \in X}, \{\mu_1^-(x)\}_{x \in X}, \{\mu_2^+(s)\}_{s \in S_1}, \{\mu_2^-(s)\}_{s \in S_2}} \sum_{x \in X} (\mu_1^+(x) + \mu_1^-(x)) + \sum_{s \in S_1} (\mu_2^+(s) + \mu_2^-(s)) , \]

subject to the following constraints:

(i) \[ \sum_{w \in W} a_{\text{num}}(w) \cdot Q(w) = \theta_0 \cdot \sum_{w \in W} a_{\text{den}}(w) \cdot Q(w) . \]

(ii) \[ 0 \leq Q(w) \leq 1 \text{ for every } w \in W . \]

(iii) \[ \sum_{w \in W} Q(w) = 1 . \]

(iv) \[ \sum_{w \in W} a_s(w) \cdot Q(w) \leq b_s \text{ for every } s \in S_1 . \]

(v) \[ \mu_1^+(x) \geq 0 \text{ and } \mu_1^-(x) \geq 0 \text{ for every } x \in X . \]

(vi) \[ \mu_2^+(s) \geq 0 \text{ and } \mu_2^-(s) \geq 0 \text{ for every } s \in S_2 . \]

(vii) \[ \hat{m}_{\text{dat},x}(Q) = \mu_1^+(x) - \mu_1^-(x) \text{ for every } x \in X . \]

(viii) \[ \hat{m}_s(Q) = \mu_2^+(s) - \mu_2^-(s) \text{ for every } s \in S_2 . \]

Since, as noted in Section 5, both the moment conditions \( \hat{m}_{\text{dat},x}(Q) \) and \( \hat{m}_s(Q) \) are linear in \( Q \), the above stated problem is a linear program. For a discussion of such a restatement, see (Boyd and Vandenberghe, 2004, Page 294).
B Restrictions Imposed by Additional Identifying Assumptions

Lemma B.1. Assumption Roy imposes restrictions on \( Q \) that satisfy Assumption 4.2.

Proof: In order to see the restriction imposed by Assumption Roy, note first that this assumption can be re-written as

\[
d_{U,\{d,d'\}} = d \implies Y(d') \leq Y(d)
\]

for every \( d, d' \in \mathcal{D} \). Then, denoting by

\[
\mathcal{U}_{\{d,d'\}} = \{ u \in \mathcal{U} : d_{u,\{d,d'\}} = d \}
\]

note the above statement imposes that

\[
\text{Prob}_Q[Y^\dagger(0) = 0, Y^\dagger(d') = 1, U \in \mathcal{U}_{\{d,d'\}}] = 0
\]

for every \( d, d' \in \mathcal{D} \). Equivalently, using the notation introduced in Lemma 4.1, this can re-written as a linear restriction on \( Q \) in the form of Assumption 4.2 as

\[
\sum_{\bar{y} \in \mathcal{Y}_{\{d,d'\}}, u \in \mathcal{U}_{\{d,d'\}}, \bar{c} \in \mathcal{C}^2, z, \bar{z} \in \{0,1\}^2} Q(\bar{y}, u, \bar{c}, z, \bar{z}) = 0
\]

for every \( d, d' \in \mathcal{D} \), where

\[
\mathcal{Y}_{\{d,d'\}} = \{ \bar{y} \in \{0,1\}^3 : y(d) = 0, y(d') = 1 \}.
\]

Lemma B.2. Assumption MTR imposes restrictions on \( Q \) that satisfy Assumption 4.2.

Proof: In order to see the restriction imposed by Assumption MTR, note first that this assumption imposes that

\[
\text{Prob}_Q[Y^\dagger(0) = 1, Y^\dagger(d) = 0] = 0,
\]

for each \( d \in \{1,2\} \). Equivalently, using the notation introduced in Lemma 4.1, this can then be equivalently re-written as a linear restriction on \( Q \) in the form of Assumption 4.2 as

\[
\sum_{\bar{y} \in \mathcal{Y}_d, u \in \mathcal{U}, c \in \mathcal{C}^2, z, \bar{z} \in \{0,1\}^2} Q(\bar{y}, u, \bar{c}, z, \bar{z}) = 0
\]

for each \( d \in \{1,2\} \), where

\[
\mathcal{Y}_d = \{ \bar{y} \in \{0,1\}^3 : y(0) = 1, y(d) = 0 \}.
\]
Lemma B.3. Assumption SLI imposes restrictions on $Q$ that satisfy Assumption 4.2.

Proof: In order to see the restriction imposed by Assumption SLI, note first that this independence assumption imposes that

$$\text{Prob}_Q[Y^+(0) = y, HC = hc, CS = cs] = \text{Prob}_Q[Y^+(0) = y] \cdot \text{Prob}_Q[HC = hc, CS = cs]$$

for $y, hc, cs \in \{0, 1\}$. Equivalently, using the notation introduced in Lemma 4.1, this can then be equivalently re-written as a linear restriction on $Q$ in the form of Assumption 4.2 as

$$\sum_{\bar{y} \in \mathcal{Y}_y, u \in \mathcal{U}, \bar{c} \in \mathcal{C}^2, z \in \{0, 1\}} Q(\bar{y}, u, \bar{c}, z, \bar{z}) - \sum_{y \in \{0, 1\}, d \in \mathcal{D}, z \in \{0, 1\}} P(y, d, z, \bar{z}) \cdot \sum_{\bar{y} \in \mathcal{Y}_y, u \in \mathcal{U}, \bar{c} \in \mathcal{C}^2, \bar{z} \in \{0, 1\}} Q(\bar{y}, u, \bar{c}, z, \bar{z}) = 0$$

for each $y \in \{0, 1\}$ and $\bar{z} \in \{0, 1\}^2$, where

$$\mathcal{Y}_y = \{\bar{y} \in \{0, 1\}^3 : y(0) = y\}.$$

Lemma B.4. Assumption UA imposes a restriction on $Q$ that satisfies Assumption 4.2.

Proof: In order to see the restriction imposed by Assumption UA, note first that this assumption can be written as

$$\text{Prob}_Q[(C(0), C(1)) \in \mathcal{C}_UA] = 1,$$

where

$$\mathcal{C}_UA = \{(\{0\}, \{0, 2\}), (\{0, 2\}, \{0, 2\}), (\{0, 1\}, \{0, 1, 2\}), (\{0, 1, 2\}, \{0, 1, 2\})\}$$

is the set of all combinations of choice with and without offer such that the choice set with an offer is the same as that without an offer except for the inclusion of or the lack of a Head Start preschool. Equivalently, this can be re-written as a linear restriction on $Q$ in the form of Assumption 4.2 as

$$\sum_{w \in \mathcal{W}_{UA}} Q(w) = 1,$$

where $\mathcal{W}_{UA} = \{w \in \mathcal{W} : (c(0), c(1)) \in \mathcal{C}_UA\}$. 

C Data and Variable Construction

The raw data used from the HSIS in this paper is restricted, but access can be acquired by submitting applications to Research Connections at

http://www.researchconnections.org/childcare/resources/19525.
In this appendix, I briefly describe how the raw data was transformed to the final sample used in 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 this description in the following steps which were taken separately for each age cohort:

**Step 1:** I merged all the various data files provided by Research Connections for the HSIS and dropped observations with missing Head Start center IDs, where this center corresponded to that from which the child was sampled. I then made edits to this raw sample as described below.

**Step 2:** I classified a variable for the Head Start HighScope curriculum and class size ratio covariate for each child from a given sampled Head Start center. This classification was required as in practice children from a given Head Start center could have attended different Head Start preschools. The covariate value from the modal attended Head Start preschool was used as the Head Start covariate value for all the children from that center.

**Step 3:** I classified the selected treatment intro the three categories used in the paper using the focal care arrangement variable provided by the HSIS data set.

**Step 4:** All observations where any of the variables used in the analysis were missing were dropped.

**Step 5:** Test score outcomes were then standardized using non-missing baseline test scores of the final sample. Moreover, the class size variable was transformed into a binary variable of high and low class size ratio, where high was taken to be above the median value across centers.
References


