Identification of Causal Effects when Program Participation is Partially Determined by Lotteries: The Case of Oversubscribed Magnet Programs*

John Engberg, Dennis Epple, Jason Imbrogno, Holger Sieg, and Ron Zimmer

November 19, 2008

*John Engberg is a research scientist at the RAND Corporation, Pittsburgh, PA 15213. Dennis Epple is the Thomas Lord Professor of Economics, Holger Sieg is Professor of Economics, Jason Imbrogno is a Ph.D. student at Carnegie Mellon University, Tepper School of Business, Pittsburgh, PA 15213. Ron Zimmer is an Associate Professor at Michigan State University, College of Education, East Lansing, MI 48824. We would like to thank Mark Roosevelt, Superintendent of the Pittsburgh Public Schools, for supporting this research and granting us access the PPS database. We would also like to thank Guido Imbens and Ken Wolpin for comments. Financial support for this research is provided by the Institute of Education Sciences (IES R305A070117).

Abstract

The purpose of this paper is to study identification and estimation of causal effects when program participation is partially determined by lottery outcomes. This research design arises in many applications in education when access to oversubscribed programs is partially determined by randomization. The paper treats program participation as the outcome of a decision process with four latent household types that can be interpreted as complying and non-complying "stayers", "leavers", and "at risk households." We show that the parameters of the underlying model of program participation are identified. Our proofs of identification are constructive and can be used to design a GMM estimator for all parameters of interest. We also show that the probability limits of commonly used linear estimators are functions of the parameters of our framework and thus partially identify the relevant treatment effects. We apply our new methods to study the effectiveness of magnet programs in a large urban school district. Our findings show that magnet programs help the district to attract and retain students from households that are at risk of leaving the district. These households have higher incomes, are more educated, and have children that score higher on standardized tests than households that stay in district regardless of the outcome of the lottery.

Keywords: Causal Effects, Noncompliance, Program Evaluation, Randomized Experiments, Instrumental Variables.

1 Introduction

The purpose of this paper is to study identification and estimation of treatment effects when program participation is partially determined by lottery outcomes. In a standard experimental design, each subject agrees to participate in the experiment and randomization completely determines whether the individual is assigned to the treatment or the control group.[1] In our design, randomization gives potential participants the option to participate in the program. Individuals that win a lottery can chose whether or not to participate in the program. This type of research design arises in many applications in education.[2] Many school districts use lotteries to determine access to over-subscribed educational programs. Lottery winners are accepted into the program, with the ultimate choice of attendance left to the student and his family. Lottery losers do not have the option to participate in the program. Program participation then depends on lottery outcomes as well as on household decisions.

To study the causal effects of such programs we develop a new method to estimate the treatment effect associated with participation in a given program. The basic idea is to treat program participation as the outcome of a decision process with four latent types of households. The first type is a "complier" and chooses the magnet program if it wins. The second type is a "non-complier" and does not chose the magnet program even if he wins the lottery. Both of these types stay in the district if they lose the lottery[3] The third and forth type leave the district if they lose the lottery. The third type is a "non-complier" and will not enroll its child in the district independently of

[1]See, for example, Heckman and Vytlacil (2007) for an overview of the program evaluation literature.

[2]Angrist (1990) introduced the use of lotteries to study the impact of military service on earnings. Of course, in his application program participation is mandatory: the penalties of avoiding the draft are quite significant.

[3]The district offers a standard education program to all households that do not win the lottery.

the outcome of the lottery.[4] Finally, there is a forth household type who complies with the lottery and will participate in the magnet program if it wins the lottery. Given that many urban school districts are experiencing declining enrollment, this type is most interesting and important from a policy perspective. From the perspective of the school district, we can think of these latent types as complying and non-complying stayers, leavers and at-risk households. The household types are latent, i.e. unobserved by both the researcher and the school district administrators.

One key objective of the analysis is then to identify and estimate the proportions of these four latent types and to characterize differences in observed characteristics among these types. We also need to understand under what additional assumptions we can estimate the causal effect of the program on a a variety of potential outcomes. We formalize the ideas that underlie our decomposition into latent types. We show that the parameters of the underlying framework of program participation are non-parametrically identified. Our proofs of identification are constructive and can be used to design a GMM estimator for all parameters of interest (with respect to retention.) We can thus study the effectiveness of various programs that try to attract and retain students. Evaluating the effectiveness of the program on other outcomes is more difficult due to the underlying selection problems. We provide conditions that allow us to identify and estimate (local) average treatment effects for "complying stayers." We also show that it is impossible to identify the effects for "students at risk" without imposing additional assumptions on the selection process.

Our approach is related to recent work by Angrist, Imbens, and Rubin (1996) who

[4]Households have incomplete information and need to gather information to learn about the features of different programs. Households have to sign up for lotteries months in advance. At that point, they dot not have accumulated all relevant information. Once they have accumulated all relevant information, they may decide to opt out of the public school system since their preferred choice dominates the program offer by the district. Note that there are typically no penalties in participating in the lottery and declining to participate in the the program.

also study an experimental design in which compliance is not perfect. Individuals are assigned to the treatment, randomly. However, some individuals assigned to treatment do not take it, and some not assigned to treatment do take it. They refer to those two types as "never-takers" and "always-takers." There is also a third type who does exactly what their assignment requires. These are referred to as "compliers." They show that the standard instrumental variables regression using random assignment as an instrument gives the local average treatment effect for compliers. The notions in that paper are closely related to your concepts of "complying and non-complying stayers", "leavers' and "at risk households." However, there are important differences in the underlying assumptions. In our framework, individuals have access to two outside options (stay in the school district non-magnet program and leave) while the standard program evaluation framework typically only allows for one outside option. Angrist et al. (1996) assume that they observe program participation decisions and outcomes for all individuals in the sample. We also assume that program participation decisions are observed for everybody, but other educational outcomes are only observed for individuals that stay in the school district. As a consequence, our results are different. In our research design the standard IV estimator only yields a consistent estimator of the (local) average treatment effect, if the fraction of "at risk households" is negligible, i.e. if we only have one type of "compliers."

Our estimation approach is also closely related to linear IV estimators that have been commonly used in the related literature to study attraction and retention effects. Cullen, Jacob, and Levitt (2006), for example, have advocated in a recent, influential study the use of linear estimators to analyze open enrollment school choice in the Chicago Public Schools.[5] We show in this paper that two of the most popular

[5]Lotteries were also used by Rouse (1998) to study the impact of the Milwaukee voucher program. Hoxby and Rockoff (2004) also use lotteries to study Chicago charter schools. These estimators have been used by Ballou, Goldring, and Liu (2006) to examine a magnet program. Hastings, Kane, and Staiger (2006) estimate a model of school choice based on stated preferences for schools in Charlotte.

linear estimators have well-defined interpretations within our framework of program participation. We derive the probability limits of the standard "intend-to-treat" OLS estimator and the IV estimator, that uses the outcome of the lottery as an instrument for program participation.[6] We show that the probability limits of these estimators are functions of the parameters of our framework. The GMM estimator that we develop is more comprehensive and provides full identification of all parameters of interest. Our approach thus provides a unified interpretation of most commonly used linear estimators. More importantly, it also provides additional insights that are outside the scope of traditional linear estimators.

We apply the techniques developed in this paper to study the effectiveness of magnet programs. While debates surrounding the effectiveness of other school choice options such as charter schools and educational vouchers have grabbed much attention from researchers and policymakers, magnet programs have gotten less attention despite the fact that they are much more prevalent than charter schools or educational voucher programs. A second objective of this paper is to provide new research to understand the causal effects of magnet programs. Our application focuses on magnet programs operated by Pittsburgh Public Schools (PPS). Our findings show that magnet programs help the district to attract and retain students from households that are at risk of leaving the district. These households have higher incomes, are more educated, and have children that score higher on standardized tests than households that stay in district regardless of the outcome of the lottery. These households have many options outside the public school system, but apparently, they view the exist-

---

Since school attendance was partially the outcome of a lottery, they use the lottery outcomes as instruments to estimate the impact of attending the first choice school. Angrist, Bettinger, Bloom, King, and Kremer (2002) study the effects of vouchers when there is randomization in selection of recipients from the pool of applicants.

[6] Angrist and Imbens (1994) discuss identification and estimation of local treatment effects. Heckman and Vytlacil (2005) provide a general framework for econometric policy evaluation.

ing programs as desirable programs for their children. We also find evidence that the market for elementary school competition is more competitive than the market for middle and high school education. The fraction of households at risk declines with age of the students. Magnet programs are most effective in attracting households that have young school-age children.

The rest of the paper is organized as follows. Section 2 develops our new methods for estimation of treatment effects when program participation is partially determined by lotteries. We discuss identification and estimation. We also show that commonly used linear IV estimators can be interpreted as partially identifying different components of our framework. Section 3 provides some institutional background for our application and discusses our main data sources. Section 4 reports the empirical findings of our paper. Finally, we offer some conclusions and discuss the policy implications of our work in Section 5.

2 Identification and Estimation of Causal Effects

2.1 The Research Design

We consider a research design in which program participation is only partially determined by randomization, i.e. a design with imperfect compliance. These designs arise when randomization occurs at the application stage. An applicant that receives a favorable random draw in the lottery has the option to participate in the program. But he is not required to participate and hence can opt out before the program begins. This design thus differs from the standard experimental design in which randomization occurs after individuals have already committed to participate in the program. Since our application focuses on magnet school, we will develop our methods within this context. However, the methods derived in this paper apply quite broadly and are not restricted to the application that we study.

Consider the problem of a household that has to decide whether or not to enroll a student in a magnet program offered by a school district.[7] We only consider households that have decided to participate in a lottery which determines access to an educational program. Let W denote a discrete random variable which is equal to 1 if the student wins the lottery and 0 if it loses. Let w denote the fraction of households that win the lottery. If the student wins the lottery, he has three options: he can participate in the program, he can participate in a different program offered by the same school district, or it can leave the district and pursue educational opportunities outside the district. If it loses, it only has the last two options. Let M be 1 if a student attends the (magnet) program and 0 otherwise. Finally, let A denote a random variable that is one if a student attends a school in the district and 0 otherwise.

The key idea behind our method is that we use four latent types to classify households. We make the following assumption

Assumption 1

1. *Let s denote the fraction of "stayers." These households will remain in the district regardless of whether they are admitted to the magnet program. This group has two subgroups. A fraction s_m will attend the magnet if admitted. A fraction s_n will not attend the magnet if admitted. Note that $s = s_n + s_m$*

2. *Let l denote the fraction of "leavers." These are households that will leave the district regardless of whether they are admitted to the magnet program.*

3. *Let r denote the fraction that is at risk. They will remain in the district and attend the magnet program if admitted to the magnet program, and they will*

[7] We use the terms "household" to describe the decision maker and "student" to describe the person that participates in the program.

leave the district otherwise.

Since the household type is latent, one key empirical problem is identifying and estimating the proportions of each type in the underlying population. Moreover, we are often interested in how these types of households differ along observed characteristics. For example, we would like to test the hypothesis that households that are classified to be "at risk" are more likely to have higher levels of income than "stayers." To formalize these ideas, consider a random variable X that measures an observed household characteristic such as income or socio-economic status. Appealing to our decomposition, let μ_r, μ_{s_m}, μ_{s_n}, μ_l denote the mean of random variable X conditional on belonging to group r, s_n, s_m, and l respectively. The goal of the first part of the analysis is then to identify and estimate the following nine parameters $(w, r, s_n, s_m, l, \mu_r, \mu_{s_m}, \mu_{s_n}, \mu_l)$.[8]

In addition to studying the effectiveness of magnet programs on attraction and retention of students, we also like to studying the effects of the program on other student outcomes. Let T be an outcome measure of interest, for example, the score on a standardized achievement test. Following standard notation in the program evaluation literature, we have

$$T \;=\; M\,T_1 \;+\; (1-M)\,T_0 \tag{1}$$

where T_1 denotes the outcome if the student attends the magnet school and T_0 if he attends a different program in the district.[9]

Conceptually, we can define four different average treatment effects, one for each

[8] It is straight forward to allow X to be a vector.

[9] This model is typically referred to as the switching regression model due to Quandt (1972) and Maddala (1983). It also known in the statistical literature as the Rubin Model. See Rubin (1974, 1978). It also shares many similarities with the Roy Model as discussed in Heckman and Honore (1990).

latent group:

$$\begin{aligned} ATE_{S_n} &= E[T_1 - T_0 | S_n = 1] \\ ATE_{S_m} &= E[T_1 - T_0 | S_m = 1] \\ ATE_R &= E[T_1 - T_0 | R = 1] \\ ATE_L &= E[T_1 - T_0 | L = 1] \end{aligned} \tag{2}$$

We also study whether we can identify and estimate these types of treatment effects.

2.2 Identification

To establish identification, it is necessary to establish the information set of the researcher. We assume:

Assumption 2 *The researcher observes probabilities and conditional means for the following eight outcomes:*

Table 1: Observed Outcomes

	W	M	A	$Pr\{W, M, A\}$	$E[X\|W, M, A] = E[X\|M, A]$
I	1	1	1	$w\ (r + s_m)$	$\frac{r\mu_r + s_m\mu_{s_m}}{r + s_m}$
II	1	1	0	0	
III	1	0	1	$w\ s_n$	μ_{s_n}
IV	1	0	0	$w\ l$	μ_l
V	0	1	1	0	
VI	0	1	0	0	
VII	0	0	1	$(1 - w)\ (s_n + s_m)$	$\frac{s_n\mu_{s_n} + s_m\mu_{s_m}}{s_n + s_m}$
VIII	0	0	0	$(1 - w)\ (r + l)$	$\frac{r\mu_r + l\mu_l}{r + l}$

Note that only five outcomes occur with non-zero probability.

Identification of the parameters can be established sequentially. First, we discuss identification of the five probabilities. We have the following result.

Lemma 1 *The parameters (w, r, s_n, s_m, l) are identified by the five nonzero probabilities in Table 1.*

Proof: Parameter w is the fraction that wins the lottery:

$$\begin{aligned} w &= Pr(W=1, M=1, A=1) + Pr(W=1, M=1, A=0) \\ &+ Pr(W=1, M=0, A=1) + Pr(W=1, M=0, A=0) \end{aligned} \tag{3}$$

Given w, s_n is identified from (1,0,1):

$$s_n = Pr(W=1, M=0, A=1)/w \tag{4}$$

l is identified from (1,0,0):

$$l = Pr(W=1, M=0, A=0)/w \tag{5}$$

Given w and s_n, s_m is identified form (0,0,1):

$$s_m = Pr(W=0, M=0, A=1)/(1-w) - s_n \tag{6}$$

Given l, s_n, s_m, r is identified of the identity:

$$r = 1 - l - s_m - s_n \tag{7}$$

Q.E.D.

Note that there is no over-identification at this stage since the five probabilities in Table 1 add up to one, and the last two probabilities add up to $1 - w$. Next we discuss identification of the four conditional means. We have the following result.

Lemma 2 *Given* (w, r, s_n, s_m, l)*, the parameters* $(\mu_r, \mu_{s_m}, \mu_{s_n}, \mu_l)$ *are identified by the five non-degenerate conditional expectations observed in Table 1.*

Proof: μ_l is identified from (1,0,0):

$$\mu_l \;=\; E(X|W=1, M=0, A=0) \;=\; E(X|M=0, A=0) \tag{8}$$

Similarly μ_{s_n} is identified from (1,0,1):

$$\mu_{s_n} \;=\; E(X|W=1, M=0, A=1) \;=\; E(X|M=0, A=1) \tag{9}$$

Given μ_{s_n}, μ_{s_m} is identified from (0,0,1):

$$\mu_{s_m} \;=\; [(s_n + s_m)E(X|W=0, M=0, A=1) \;-\; s_n\mu_{s_n}]/s_m \tag{10}$$

Given μ_{s_m}, μ_r is identified from (1,1,1):

$$\mu_r \;=\; [(r + s_m)E(X|W=0, M=0, A=1) \;-\; s_m\mu_{s_m}]/r \tag{11}$$

Q.E.D.

Note that there is one over-identification condition at this stage. This arises due to the condition that W is orthogonal to X.[10]

Lemma 1 and Lemma 2 then imply the following result:

Proposition 1 *The parameters* $(w, r, s_n, s_m, l, \mu_r, \mu_{s_m}, \mu_{s_n}, \mu_l)$ *are identified.*

We now turn to an analysis of identification issues that arise when studying the effects of magnet programs on student outcomes. We make the following assumption regarding observables:

[10] The lotteries are assumed to be fair and blind in the sense that the district does not pre-select winners and losers based on beliefs about attendance or any socio-economic or student characteristic found in X.

Assumption 3 *We assume that the researcher only observes outcomes T for students that remain in the school district, i.e. we do not observe outcomes for "leavers" and "at risk households" that lose the lottery.*

It is useful to assume initially that we observe the latent household. Table 2 provides a summary of the relevant conditional expectations.[11] Conditioning on lottery outcomes, there are eight conditional expectations. Three of these pertain to outcomes that are not observed if we observed the latent type since students in these latent groups leave the school district. The remaining five conditional expectations relate to household types that remain in the district.

Table 2: Mean Outcomes

	Stayers		At Risk	Leavers
	$A = 1$ and $M = 1$	$A = 1$ and $M = 0$		
Win Lottery $W = 1$	$E[T_1 \vert S_m = 1]$	$E[T_0 \vert S_n = 1]$	$E[T_1 \vert R = 1]$	leave
Lose Lottery $W = 0$	$E[T_0]S_m = 1]$	$E[T_0 \vert S_n = 1]$	leave	leave

From Table 2, it is evident that even if we observed the latent type, there is little hope in identifying ATE_{S_n}, ATE_R or ATE_L without imposing some additional assumptions on the underlying selection process. For stayers that never attend the magnet program, we cannot identify $E[T_1|S_m = 1]$. For students at risk, we cannot identify $E[T_0|R = 1]$. For leavers, we can neither identify $E[T_1|L = 1]$ nor $E[T_0|L = 1]$. We thus have the following result:

[11]Note that we are implicitly assuming that the mean performance of stayers who would decline lottery admission is the same whether they win or lose the lottery, i.e. $E[T_0|S_n = 1, W = 1] = E[T_0|S_n = 1, W = 0] = E[T_0|S_n = 1]$.

Proposition 2 *Without imposing additional assumptions on the selection of students into latent groups,* ATE_{S_n}*,* ATE_R *and* ATE_L *are not identified.*

Attention, therefore, focuses on identification of ATE_{S_m}.

Since we do not observe the latent type, Assumption 3 implies that we only observe mean outcomes for the students conditional on W and A. For students who win the lottery and attend the magnet school, we observe

$$E[T_1|W=1, M=1, A=1] \;=\; \frac{s_m E[T_1|S_m=1] + rE[T_1|R=1]}{s_m + r} \tag{12}$$

We observe mean performance of stayers who lose the lottery:

$$E[T_0|W=0, M=0, A=1] \;=\; \frac{s_m E[T_0|S_m=1] + s_n E[T_0|S_n=1]}{s_m + s_n} \tag{13}$$

Finally we also observe the mean performance of stayers who win the lottery and decline to enroll in the magnet program:

$$E[T_0|W=1, M=0, A=1] \;=\; E[T_0|S_n=1] \tag{14}$$

Equations (13) and (14) imply that we can identify $E[T_0|S_m=1]$ and $E[T_0|S_n=1]$, since s_n and s_m have been identified before. However equation (12) then implies that it is difficult to separately identify $E[T_1|S_m=1]$ and $E[T_1|R=1]$. As a consequence we have the following result:

Proposition 3 $E[T_0|S_m = 1]$ *and* $E[T_0|S_n = 1]$ *are identified. If* $r \neq 0$ *and* $E[T_1|S_m = 1] \neq E[T_1|R = 1]$*, then* ATE_{s_m} *is not identified without imposing further assumptions.*

Before we proceed, we offer three observations. First, the lack of identification arises from the at risk group. If for example stayers and at risk students differ by unobservable, we have a standard selection problem. Second, if controlling on observables is sufficient to deal with the selection problem, a matching approach can be justified.[12]

[12] For a discussion of matching estimators, see, among others, Rosenbaum and Rubin (1983), Heckman, Ichimura, and Todd (1997), and Abadie and Imbens (2006).

Third, attrition *per se* is not the problem. If membership in the at risk group is negligible (i.e., $r \approx 0$), identification is achieved even if the attrition rate, l, is large.

2.3 A GMM Estimator

We observe a random sample of N applicants to a magnet program, indexed by i. We view these as N independent draws from the underlying population of potential applicants to this magnet program. Let W_i, M_i, A_i, X_i now denote the random variables that correspond to observation i.

Note that the proofs of identification are constructive. Replacing population means by sample means thus yields consistent estimators for the parameter of interests. Nevertheless it is useful to place the estimation problem with a well defined GMM framework. This allows us to estimate simultaneously all parameters and compute asymptotic standard errors. Based on Table 1 we can construct the following five orthogonality conditions based on the relevant choice probabilities

$$\frac{1}{N}\sum_{i=1}^{N}\Big[W_i M_i A_i - w(r+s_m)\Big] \tag{15}$$

$$\frac{1}{N}\sum_{i=1}^{N}\Big[W_i(1-M_i)A_i - w\; s_n\Big] \tag{16}$$

$$\frac{1}{N}\sum_{i=1}^{N}\Big[W_i(1-M_i)(1-A_i) - w\; l\Big] \tag{17}$$

$$\frac{1}{N}\sum_{i=1}^{N}\Big[(1-W_i)(1-M_i)A_i - (1-w)(s_n+s_m)\Big] \tag{18}$$

$$\frac{1}{N}\sum_{i=1}^{N}\Big[(1-W_i)(1-M_i)(1-A_i) - (1-w)(r+l)\Big] \tag{19}$$

Moreover we can also construct orthogonality conditions, based on the condition means.

$$\frac{1}{N}\sum_{i=1}^{N}\Big[W_i M_i A_i X_i - w(r+s_m)\ \frac{r\mu_r + s_m\mu_{s_m}}{r+s_m}\Big] \tag{20}$$

$$\frac{1}{N}\sum_{i=1}^{N}\Big[W_i(1-M_i)A_i X_i - w\ s_n\ \mu_{s_n}\Big] \tag{21}$$

$$\frac{1}{N}\sum_{i=1}^{N}\Big[W_i(1-M_i)(1-A_i)X_i - w\ l\ \mu_l\Big] \tag{22}$$

$$\frac{1}{N}\sum_{i=1}^{N}\Big[(1-W_i)(1-M_i)A_i X_i - (1-w)(s_n+s_m)\ \frac{s_n\mu_{s_n} + s_m\mu_{s_m}}{s_n+s_m}\Big] \tag{23}$$

$$\frac{1}{N}\sum_{i=1}^{N}\Big[(1-W_i)(1-M_i)(1-A_i)X_i - (1-w)(r+l)\ \frac{r\mu_r + l\mu_l}{r+l}\Big] \tag{24}$$

Hence the parameters of the model can be estimated using a GMM estimator (Hansen, 1982).

There are, of course, other feasible estimators. For example, one can replace equations (10) with the orthogonality conditions:

$$\frac{1}{N}\sum_{i=1}^{N}\Big[W_i - w\Big] \tag{25}$$

One could also eliminate equation (14) and substitute the identity $1 = r + l + s_n + s_m$ into the other orthogonality conditions. Moreover, the estimator above easily generalizes to the case in which X is vector of random variables. We simply stack all orthogonality conditions to obtain a simultaneous estimator that exploits all relevant orthogonality conditions.

2.4 Interpreting Commonly Used OLS and IV Estimators

There exists a close relationship between the GMM estimator discussed above and some more commonly used OLS and IV estimators.

First, we consider an estimator that is also some called the "intend-to-treat" estimator since it does not account for actual program participation. We have the following result:

Proposition 4

Consider the linear regression model:

$$A \;=\; \beta_0 \;+ \beta_1 W \;+ \varepsilon \tag{26}$$

Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the probability limits of the least squares estimators. Under standard regularity assumptions $\hat{\beta}_0 = s$ and $\hat{\beta}_1 = r$.

Proof: Let a denote the proportion of magnet applicants that attend school in the district. From the definitions given at the outset: $a = s + wr$. Let q denote the fraction of students that win the lottery and attend a school in the district. Then, from the definitions presented at the outset: $q = w(r + s)$.

Consider the normal equations of the least squares estimators. After dividing by total number of students and taking limits, we obtain:

$$\begin{bmatrix} a \\ q \end{bmatrix} = \begin{bmatrix} 1 & w \\ w & w \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \tag{27}$$

or

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 1 & w \\ w & w \end{bmatrix}^{-1} \begin{bmatrix} a \\ q \end{bmatrix} = \frac{1}{1-w} \begin{bmatrix} a - q \\ \frac{1}{w} q - a \end{bmatrix} \tag{28}$$

Consider $\hat{\beta}_1$ in more detail:

$$\hat{\beta}_1 \;=\; \frac{1}{1-w}\Big(\frac{1}{w}q - a\Big) \;=\; \frac{1}{w(1-w)}\,(w(r+s) - w(s+wr)) = r \tag{29}$$

The intercept is:

$$\hat{\beta}_0 \quad = \quad \frac{a-q}{1-w} \quad = \quad \frac{s+wr-w(r+s)}{1-w} \quad = \quad s$$

Q.E.D.

Next we consider a linear IV estimator that uses the outcome of the lottery as an instrument from program participation. This estimator is the preferred estimator in most empirical studies and has, for example, been advocated by Cullen et al. (2006).

Proposition 5 *Consider the linear regression model*

$$A \quad = \quad \gamma_0 + \gamma_1 M + \varepsilon \tag{30}$$

using W as instrument for M. Let $\hat{\gamma}_0$ and $\hat{\gamma}_1$ denote the probability limits of the standard exactly identified linear IV estimator. Under standard regularity assumptions we have $\hat{\gamma}_0 = s$ and $\hat{\gamma}_1 = \frac{r}{(r+s_m)}$.

Proof: Consider the normal equations for the standard IV estimator.[13] After dividing by total number of students and taking limits, we obtain:

$$\begin{bmatrix} a \\ q \end{bmatrix} = \begin{bmatrix} 1 & m \\ w & m \end{bmatrix} \begin{bmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma}_1 \end{bmatrix} \tag{31}$$

Now:

$$\begin{aligned} a &= s + wr \\ m &= wr + ws_m \\ q &= w(r+s) \end{aligned} \tag{32}$$

Substituting these into the above normal equations, we obtain:

$$\begin{bmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma}_1 \end{bmatrix} = \begin{bmatrix} 1 & m \\ w & m \end{bmatrix}^{-1} \begin{bmatrix} a \\ q \end{bmatrix} = \frac{1}{1-w} \begin{bmatrix} a-q \\ \frac{1}{m}q - \frac{w}{m}a \end{bmatrix} \tag{33}$$

[13]In standard notation these are $Z'y = Z'X\widehat{\gamma}$.

Hence,

$$\widehat{\gamma}_1 \;=\; \frac{q - wa}{m(1-w)} \;=\; \frac{wr + ws - ws - wwr}{m(1-w)} \;=\; \frac{r}{(r + s_m)} \tag{34}$$

and

$$\widehat{\gamma}_0 = \frac{q-a}{1-w} = \frac{s(1-w)}{(1-w)} = s \tag{35}$$

Q.E.D.

Note that γ_1 may be interpreted as the probability of an additional student remaining in the district if an additional seat is provided in the magnet program. The coefficient γ_1 will be greater than the coefficient β_1 in regression equation (26). Not all students who win the lottery attend the magnet program. The expected number of lottery wins required to obtain an additional magnet student is $1/(r + s_m)$. In order to obtain an additional student in the magnet program, the coefficient γ_1 is $r/(r + s_m)$.

Finally, we consider the linear model:

$$T_i \;=\; \alpha_0 + \alpha_1 M_i \;+\; \epsilon_i \tag{36}$$

We consider properties of the standard IV estimator that uses W_i as an instrument for M_i. Following Durbin (1954) the IV estimator is defined as the ratio of sample covariances:

$$\begin{aligned} \hat{\alpha_1}^{IV} \;&=\; \frac{\widehat{cov}(T,W)}{\widehat{cov}(M,W)} \qquad (37) \\ &=\; \frac{\sum_{i=1}^N T_i W_i \Big/ \sum_{i=1}^N W_i \;-\; \sum_{i=1}^N T_i(1-W_i) \Big/ \sum_{i=1}^N (1-W_i)}{\sum_{i=1}^N M_i W_i \Big/ \sum_{i=1}^N W_i \;-\; \sum_{i=1}^N M_i(1-W_i) \Big/ \sum_{i=1}^N (1-W_i)} \end{aligned}$$

where $\widehat{cov}$ denotes the sample covariance and the last equality follows from the binary nature of the instrument. We assume that the estimator is implemented using a random sample of students for which we observe (T_i, M_i, W_i). In our application these are the students that are participating in a program offered by the district. The sample thus does not include students that left the district. We have the following result that parallels our finding in Proposition 3:

Proposition 6 *If $r = 0$, then*

$$\hat{\beta}^{IV} \;\; \rightarrow \;\; E[T_1|S_m] \; - \; E[T_0|S_m] \tag{38}$$

If $r \neq 0$, the limit of the IV estimator does not converge to any of the commonly used treatment effects.

Proof:

Let N denote the sample size, N_W the number of students that win the lottery, $N_{W,M}$ the number of students that win the lottery and attend the program, and $N_{W,NM}$ the number of students that win the lottery and do not attend the program. We assume standard regularity conditions apply, so that sample means converge (almost surely) to population means.

Now consider the probability limits of the two key terms in the numerator of the IV estimator:

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} T_i W_i \;\; &= \;\; \frac{N_W}{N}\frac{1}{N_W}\sum_{i=1}^{N_W}[M_i T_{1i} + (1-M_i)T_{i0}] \\
&= \;\; \frac{N_W}{N}\left[\frac{N_{W,M}}{N_W}\frac{1}{N_{W,M}}\sum_{i=1}^{N_{W,M}} T_{1i} \; + \; \frac{N_{W,NM}}{N_W}\frac{1}{N_{W,NM}}\sum_{i=1}^{N_{W,NM}} T_{0i}\right] \\
&\rightarrow \;\; w\left[\frac{r}{r+s_n+s_m}E[T_1|R] \; + \; \frac{s_m}{r+s_n+s_m}E[T_1|S_m] \;\; + \; \frac{s_n}{r+s_n+s_m}E[T_0|S_n]\right] \\
&= \;\; w \; A
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} T_i(1-W_i) \;\; &= \;\; \frac{N-N_w}{N}\frac{1}{N-N_W}\sum_{i=1}^{N-N_w} T_{0i} \\
&\rightarrow \;\; (1-w)\left[\frac{s_m}{s_m+s_n}E[T_0|S_m] + \frac{s_n}{s_m+s_n}E[T_0|S_n]\right] \\
&= \;\; (1-w) \; B
\end{aligned}$$

Next consider the key terms in the denominator and note that:

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} M_i W_i &= \frac{N_W}{N}\frac{1}{N_W}\sum_{i=1}^{N_W} M_i \\ &= w\,\frac{s_m + r}{s_m + s_n + r}\end{aligned}$$

and

$$\frac{1}{N}\sum_{i=1}^{N} M_i(1 - W_i) \;\rightarrow\; 0$$

As a consequence we have:

$$\hat{\beta}^{IV} \;\rightarrow\; \frac{s_m + s_n + r}{s_m + r}(A - B)$$

Suppose $r = 0$, then

$$\hat{\beta}^{IV} \;\rightarrow\; E[T_1|S_m] \;-\; E[T_0|S_m]$$

This effect can also be interpreted as a Local Average Treatment Effect since the students for which $S_n = 1$ are the ones that will be locally affected by a change in the lottery outcome. Q.E.D.

Note that Cullen et al. (2006) consider an application with $r \approx 0$. Hence the IV estimator is consistent and can be used to estimate the relevant treatment effect. Angrist et al. (1996) consider a model with only three latent types. In research design studied in that paper, the IV estimator identifies the Local Average Treatment Effect. In their framework, there is only one type of "compliers." In our models there are two latent "compliers" and two latent "non-compliers." The limit of the IV estimator depends of the mean outcomes of both types of "compliers" and the relative proportions of these latent groups.[14]

[14] Angrist et al. (1996) have a third type called "always takers." These households participate in the program even if they lose the lottery. We assume that these types do not exist in our framework which is consistent with our application.

3 Data

Our application focuses on magnet programs that are operated by Pittsburgh Public Schools. There are magnet programs for all grade levels, and each program has a slightly different focus. For example, there is an elementary school magnet program that focuses on international studies and the French language. We only consider magnet programs that are academically oriented. Every academic year, interested students submit applications for one magnet program of their choice. If the number of applications submitted during registration for any magnet program exceeds the number of available spaces, the district holds a lottery to determine the order in which applicants will be accepted.

In the case of over-subscription, a computerized random selection determines each student's lottery number. The lottery is binding in the sense that students with lower numbers are accepted, higher numbered students are rejected. There is a clear cut-off number that separates the groups. To preserve racial balance in the magnet programs, separate lotteries are held for black students and other students. Some programs also have preferences for students with siblings already attending the magnet programs. Separate lotteries are held for those students with an acceptable preference category for each magnet program. All in all, each lottery is held for a given program, in a given academic year, separately by race, and, finally, separately by preference code.

Lottery winners (lotteried-in) have the option to participate in the magnet program, with the ultimate choice of participation left to the student and his family. Lottery losers (lotteried-out) do not have this option, and thus must make their schooling choice without the availability of the magnet option. When winners decline admission, the students on the wait list become eligible. Again the rank on the wait list is determined by the original lottery. With a fair and balanced lottery, the winners and losers will be determined by chance, thus creating two groups that are similar to

Table 3: Lottery Participant Characteristics

Variable	*Entire Sample*	*Elementary School Applicants Only*	*High School Applicants Only*
PSSA Math	1304.9 (207.9) [1049]	-	1276.5 (195.7) [627]
PSSA Reading	1310.2 (244.4) [1048]	-	1311.5 (258.7) [627]
Sex	0.5123 (0.5000) [2151]	0.5121 (0.5000) [871]	0.5195 (0.5000) [820]
Race	0.7209 (0.4487) [2268]	0.5691 (0.4955) [912]	0.8409 (0.3660) [861]
Free/Reduced Lunch	0.3322 (0.4711) [2161]	0.3375 (0.4731) [871]	0.3195 (0.4666) [820]
Poverty	0.2282 (0.1407) [2162]	0.2177 (0.1358) [872]	0.2344 (0.1456) [820]
Education	0.2949 (0.1933) [2162]	0.3417 (0.2253) [872]	0.2513 (0.1457) [820]
Lottery Win Percentage	64.8	56.3	78.3

Listed as mean, (std dev), [observations]

each other both on observable and unobservable characteristics.

PPS granted us access to its longitudinal student level data from the district's "Real-Time Information" (RTI) database. This database currently has information from the 1998-1999 school year through 2006-2007. In addition to demographic data, the RTI contains detailed information about educational outcomes. This information is linked to each student by a unique ID number. The demographic characteristics for the students include race, sex, free/reduced lunch eligibility, and addresses. Using the addresses, we can assign census tract variables such as poverty and adult education levels to each student. We use two community characteristics that measure the socio-economic composition of the neighborhoods in which students reside. Poverty is the percentage of adults in the student's census tract with an income level below the poverty line. Education is the percentage of adults in the student's census tract with at least a college degree.

As pertaining to student educational outcomes, RTI includes the school of attendance in each year, standardized test scores for the Pennsylvania System of School Assessment (PSSA). The database also contains the outcomes of the magnet lotteries. One of the key features of the RTI is that it contains unusually good information about students residing in Pittsburgh that attend private, charter, and home schools. The reason for this is that PPS operates the school bus transportation network which is also heavily used by students that do not attend PPS. Unfortunately, we do not observe test scores for students outside of PPS.

Table 3 shows descriptive statistics for the entire sample used in this study as well as two important sub-samples that we also consider in estimation. We only consider binding lotteries in this research because treatment and comparison groups are required for the experimental design. In total, over the time frame of the data, there are 203 binding lotteries with 1,396 students lotteried-in and 981 students lotteried-out.

Before we implement the estimators, we check whether the lotteries are balanced on student observables to ensure a clean experimental design. While assignment within lotteries may be random, participation in a lottery is not. To make use of within-lottery randomness and not the between-lottery non-randomness, we perform a check for balance by running a lottery-fixed effect regression for each observable characteristic as a dependent variable with acceptance as the only independent variable other than the fixed effects. Separate lotteries are held by race, so race is left out of the balance analysis. We test every other observable student characteristic in the data set. Following Cullen et al. (2006) we use equation (39) to determine whether the lottery is balanced:

$$X_i = \beta_1 W_i + \sum_{j=1}^{J} I_{ij}\beta_{2j} + v_i \tag{39}$$

where X_i is the observable characteristic of interest, W_i is a dummy equal to one if student i wins lottery j, I_{ij} is an indicator variable if i participated in lottery j, and v_i is the error term.[15] We estimate a separate regression for each observable. The coefficient β_1 determines the fairness of the lottery system. If we cannot reject the null hypothesis that it is equal to zero, then acceptance into a magnet is not determined by the value of that particular student observable.

Table 4: Balance of Prior Student Characteristics Between Lottery Winners and Losers

Variable	*Entire Sample*	*Elementary School Applicants Only*	*High School Applicants Only*
PSSA Math	21.19 (15.04) [1049]	-	31.76* (19.24)* [627]*
PSSA Reading	28.37 (17.99) [1048]	-	22.26 (26.13) [627]
Sex	0.0034 (0.0267) [2161]	0.0245 (0.0389) [871]	-0.0087 (0.0472) [820]
Race	0.0017 (0.0047) [2268]	0.0058 (0.0073) [912]	0.0009 (0.0083) [861]
Free/Reduced Lunch	0.0036 (0.0234) [2161]	0.0161 (0.0328) [871]	0.0275 (0.0433) [820]
Poverty	-0.0051 (0.0069) [2162]	-0.0024 (0.0093) [872]	-0.0195 (0.0134) [820]
Education	0.0038 (0.0078) [2162]	0.0095 (0.0126) [872]	-0.0007 (0.0123) [820]

*Significant at 10%

The first column of Table 4 shows the results when all students in all binding lotteries are included in the regressions. β_1 is not significant for any tested variable at 10 %. The second and third columns consider the two subsample of interest. The second column includes all students in elementary school while the third column focuses on high school students. The last group of students is similar to the one

[15] Alternatively we could use multivariate Behrens-Fisher type test statistics which require less restrictive assumptions. See, for example, Kim (1992).

chosen by Cullen et al. (2006). This particular subset of students showed low overall attrition and, perhaps more importantly, similar attrition rates for lottery winners and losers. We find that the estimate of β_1 is rarely significant, even at the 10% level. This supports our earlier assumption about the orthogonality of W and X.

The results shown in Table 4 support the notion that the lotteries are fair, creating separate winner and loser groups that are similar in all known characteristics. Any differences between winners and losers are small and statistically insignificant. Therefore, this randomized design theoretically also balances the two groups on unobservable characteristics such as motivation and parental involvement. This holds for the overall population in binding lotteries and for smaller sub-samples that were tested.

4 Empirical Results

Since the lotteries appear to be balanced, our research is valid and we implement our new estimators. The results are summarized in Table 5. We report estimates and estimated standard errors of the fraction of households in each latent class as well as some relevant household characteristics. The characteristics include race, gender, free or reduced lunch, poverty, and college education. Note that the last two measures are based on neighborhood characteristics as reported by the U.S. Census. We report estimates for the full sample of all students as well for the subsample of students that applied to a magnet program that was associated with an elementary school.[16]

Table 5 reveals some interesting new insights into the application process and the decision making process of households. We find that the probability of winning

[16] The sample sizes are too small to precisely estimate causal effects by program for almost all programs, except perhaps the five largest programs. The largest program contributes 491 observations to our sample.

Table 5: Empirical Results

		All Schools		Elementary Schools	
		Coefficient	Std. Error	Coefficient	Std. Error
Fraction	Win	0.65	0.01	0.56	0.02
	At Risk	0.19	0.02	0.22	0.03
	Stay Attend Magnet	0.64	0.02	0.59	0.03
	Stay-Non Magnet	0.09	0.01	0.07	0.01
	Leave	0.09	0.01	0.12	0.01
Black	At Risk	0.58	0.05	0.46	0.06
	Stay-Magnet	0.81	0.01	0.69	0.03
	Stay-Non Magnet	0.67	0.04	0.31	0.08
	Leave	0.43	0.04	0.28	0.06
Female	At Risk	0.55	0.05	0.60	0.07
	Stay-Magnet	0.52	0.02	0.50	0.03
	Stay-Non Magnet	0.53	0.05	0.52	0.09
	Leave	0.37	0.04	0.39	0.06
FRL	At Risk	0.16	0.04	0.18	0.05
	Stay-Magnet	0.43	0.02	0.46	0.03
	Stay-Non Magnet	0.26	0.04	0.29	0.08
	Leave	0.08	0.02	0.05	0.03
Poverty	At Risk	0.21	0.01	0.20	0.02
	Stay-Magnet	0.24	0.01	0.24	0.01
	Stay-Non Magnet	0.23	0.01	0.18	0.02
	Leave	0.15	0.01	0.12	0.01
College	At Risk	0.34	0.02	0.42	0.04
	Stay-Magnet	0.27	0.01	0.30	0.01
	Stay-Non Magnet	0.32	0.02	0.41	0.04
	Leave	0.38	0.02	0.40	0.04
# Observations		2268		905	

the lottery is 0.65 if we average over all programs. The fraction of households that we estimate to be at risk is substantial and consist of 19 percent of the underlying population. Among the 73 percent of households that are not at risk, the vast majority will attend the magnet program if they win they lottery. There are only 9 percent of households that will leave the district regardless of the outcome of the lottery. Overall, these results suggest that most households consider the magnet programs desirable. Our estimates imply that approximately 250 households – 12 percent of the underlying sample – decided to stay in the district because they won the magnet lottery.

Equally interesting are the characteristics of the households that are at risk. For each characteristic, the differences across household types (at risk, leavers, stayer) are statistically significant. We find that these households are on average less likely to be African American and on free or reduced lunch programs than households that are stayers. Moreover, they come from more affluent and better educated neighborhoods.[17] We thus conclude that magnet programs are effective devices for the school district to retain more affluent households. Not surprisingly, the group that are leavers are the most affluent group. These households may just apply to the magnet programs as a back-up option in case their students should unexpectedly not be admitted to an independent, charter, or parochial school.[18]

Table 5 also provides estimates when we restrict attention to households that apply to programs that provide education for children in elementary school. These programs are slightly more competitive as can be seen from the lower probability of winning (0.56 versus 0.65). Moreover, these programs attract a more educated clientele. The fraction of African American families is also lower in this subsample.

[17] Note that the differences in household characteristics are statistically significant from zero at all conventional levels.

[18] It could also be that these households left the district because of job transfers or other issues unrelated to schools.

Not surprisingly, we find that the fraction of at risk families and the fraction of leavers is also higher in this sample. This finding highlights the fact the market for elementary school education is more competitive than the market for high school education.

Table 6: Ability Sorting

		High Schools	
		Coefficient	Std. Error
Fraction	Win	0.86	0.01
	At Risk	0.06	0.04
	Stay Attend Magnet	0.83	0.04
	Stay-Non Magnet	0.06	0.01
	Leave	0.05	0.01
Math	At Risk	1399	169
	Stay-Magnet	1261	13
	Stay-Non Magnet	1262	45
	Leave	1347	45
Reading	At Risk	1380	165
	Stay-Magnet	1299	16
	Stay-Non Magnet	1238	42
	Leave	1413	56
# Observations		530	

Finally, we consider the results for high school applicants. This sub-sample is interesting since we observe test scores for all students in these subsample. We can thus analyze sorting based on ability measured by prior test scores. We implement our GMM estimator using the full set of observed characteristics. In Table 6. we only report the results that pertain to attraction and retention as well as to sorting by ability.[19]

Results are reported in Table 6. We find that the fraction of households that are at risk and leavers are much smaller than in the full sample. Overall Table 6 provides some evidence in favor of ability sorting. Households that are considered to be at risk have on average children with higher math and reading test scores. Households that stay in the district regardless of the outcome of the lottery have the lower test scores. We thus conclude that magnet program retain higher ability students who would leave the district in the absence of these programs.

5 Conclusions

We have considered a research design in which program participation is only partially determined by randomization. These designs arise when randomization occurs at the application stage and potential participants that are randomized into the program can always opt out before the program begins. These designs are commonly used to determine access to oversubscribed program offered by public school systems. We have developed a new empirical method which classifies potential participants in to stayers, leavers, and those that are at risk. The last group of individuals are most interesting form a policy perspective since the decision to participate crucially depends on the outcome of the lottery. We have shown that the parameters that characterize the causal effects are identified and can be estimated using a GMM estimator. Commonly applied IV and OLS estimators partially identify the key components of our framework.

We have applied our new methods to study the effectiveness of magnet program in attracting and retaining students and households in a large urban school district. Our

[19]The full set of results are available upon request from the authors.

findings suggest that magnet programs are useful tools that help the district to attract and retain students from middle class backgrounds. These households have many options outside the public school system. It is considerably more difficult to study the impact of these programs on achievement because of attrition and selection. Some households that do not win the lottery decide to leave the district and pursue other school options for their children. These households have very different observable characteristics then the households that stay behind. It is therefore plausible to assume that households also differ in unobserved characteristics. As a consequence, randomization at the admission stage typically only allows us to identify and estimate the effects associated with retention and attraction of students.

References

Abadie, A. and Imbens, G. (2006). Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, *74 (1)*, 235–67.

Angrist, J. (1990). Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. *American Economic Review*, *80*, 313–36.

Angrist, J., Bettinger, E., Bloom, E., King, E., and Kremer, M. (2002). Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. *American Economic Review*, *92*, 1535–58.

Angrist, J. and Imbens, G. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, *62 (2)*, 467–79.

Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, *91*, 444–56.

Ballou, D., Goldring, E., and Liu, K. (2006). Magnet Schools and Student Achievement. Working Paper.

Cullen, J., Jacob, B., and Levitt, S. (2006). The Effect of School Choice on Student Outcomes: Evidence From Randomized Lotteries. *Econometrica*, *74(5)*, 1191–230.

Durbin, J. (1954). Errors in Variables. *Review of the International Statistical Institute*, *22*, 23–32.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, *50 (4)*, 1029–1053.

Hastings, J., Kane, T., and Staiger, D. (2006). Paternal Preferences and School Competition: Evidence from a Public School Choice Program. Working Paper.

Heckman, J. and Honore, B. (1990). The Empirical Content of the Roy Model. *Econometrica*, *58 (5)*, 1121–49.

Heckman, J., Ichimura, H., and Todd, P. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. *Review of Economic Studies*, *64 (4)*.

Heckman, J. and Vytlacil, E. (2005). Structural Equations, Treatment Effects and Econometric Policy Evaluation. *Econometrica*, *73*.

Heckman, J. and Vytlacil, E. (2007). Econometric Evaluation of Social Programs. In *Handbook of Econometrics 6b*. Elsevier North Holland.

Hoxby, C. and Rockoff, J. (2004). The Impact of Charter Schools on Student Achievement. Working Paper.

Kim, S. (1992). A Practical Solution to the Multivariate Behrens-Fisher Problem. *Biometrica*, *79*, 171–6.

Maddala, G. (1983). *Limited Dependent and Qualitative Response Variables in Econometrics*. Cambridge University Press.

Quandt, R. (1972). A New Approach to Estimating Switching Regression Models. *Journal of the American Statistical Association*, *67*, 306–310.

Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrica*, *70*, 41–55.

Rouse, C. (1998). Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics*, *113*, 553–602.

Rubin, D. (1974). Estimating Causal Effects of Treatment in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, *66*, 688–701.

Rubin, D. (1978). Bayesian Inference For Causal Effects. *Annals of Statistics*, *6*, 34–58.