

Estimating the Impact of School Vouchers and Private Schooling When the Impact Depends on the Duration of Participation.¹ (Job Market Paper)

Jeffrey Yau²

Department of Economics, University of Pennsylvania

November 2004

Abstract

This paper analyzes the impact on academic achievement of school vouchers and private education when the impact depends on the duration of attending private schools, using data from the New York School Voucher Experiment. This experiment randomly assigned vouchers for attending private schools to families that met the eligibility criteria. However, voucher recipients decided whether or not to send their children to private school, and once in the private sector, decided whether or not to stay in that sector. A substantial fraction of families either did not use the vouchers or stopped using them before the end of the program, and only about half of the voucher recipients stayed in the private schools for the entire program period. The impact of offering vouchers has dominated the empirical literature of school vouchers as the central question of interest. However, this conventional impact parameter by itself does not provide a complete picture of the effectiveness of the voucher program in raising academic achievement due to the dynamic aspect of self-selection into the program, and this piece of information alone is insufficient to answer many important policy questions. For example, if the impact of offering vouchers is found to be low, should the program be terminated? If not, how should it be restructured so as to achieve a higher impact in the future? Should it be targeted to a certain population? To answer these types of questions, it is desirable to know the impact of using vouchers for the entire program period and the impact of staying in private schools for part of the program period. As the decision about how long to use the vouchers is nonrandom, non-experimental estimators are required to estimate these impacts. To account for self-selection, I estimate these impacts using the nonparametric, difference-in-difference propensity-score matching estimators recently developed in the econometrics literature and modifying them to additionally take into account selective non-response, a problem that plagues all of the voucher experiments implemented in the U.S. Semiparametric methods are employed to estimate the propensity scores and the results are compared with those using parametric methods. The results shed light on why the impact on achievement varies across racial/ethnic groups. I find that public school improvements over time and whether or not English is the main language spoken at home are important determinants of program impacts.

JEL Classification: I29, I21, C14, C19.

Keywords: Randomized Field Experiment, School Vouchers, Propensity Scores, Difference-in-Difference Matching, Treatment Impact, Semiparametric.

¹ I am truly indebted to Petra Todd for her guidance and support, and I am also very grateful to Jere Behrman, Gregory Kordas, and Ken Wolpin for their advices. In addition, I sincerely thank Rebecca Maynard, John Deke, Paul Decker, Daniel Meyer, John Homrighausen, Mark Dynarski, Melissa Clark, and several other researchers at Mathematica Policy Research, Inc. for their comments and answering many of my questions regarding the data set. Mathematica Policy Research, Inc. is gratefully acknowledged for providing me with the data and a summer fellowship for this project. I also thank Melissa Tartari and the participants at the 38th Annual Meeting of the Canadian Economics Association, the 79th Annual Conference of the Western Economic Association International, and the 2004 Inter-University Graduate Student Conference at Yale for their comments.

² Email: yau@econ.upenn.edu. Phone: (215) 518-8119. Department of Economics, University of Pennsylvania, Room 551 McNeil Building, 3718 Locust Walk, Philadelphia, PA 19104.

1 Introduction

Since the inception of the Milwaukee Parental Choice voucher program fourteen years ago, thirty-eight states have implemented school voucher programs and their importance in educational reform policies continues to grow. For example, the Senate recently approved the establishment of an experimental school-voucher program in the District of Columbia, DC, that will cost \$40 million dollar annually for the next five years.¹ President George W. Bush has proposed to spend \$50 million for a nation-wide, pilot school voucher program for children to attend private and religious schools at federal taxpayers' expense starting October 1, 2004.² In addition to publicly funded school voucher programs, privately-funded voucher programs are also on the rise. In 2002, 47 companies contributed \$46.6 million to school voucher programs, and in the first half of 2003, 29 companies already contributed \$22 million.³

School vouchers can potentially help children from disadvantaged families to improve their academic achievement. The earlier empirical literature evaluating the effectiveness of school voucher programs focuses on quantifying the *impact of offering school vouchers* on students' academic achievement. The earlier studies use observational data (Witte et. al. (1995), Greene, Peterson, and Du (1997), Rouse (1998), Metcalf (2000)) and are often criticized for lacking a valid control group of students that are "comparable" to the group that received vouchers.⁴ Without a valid comparison group, estimates of the impacts of offering school vouchers are subject to selection bias. Responding to these criticism, some researchers have designed randomized social experiments where vouchers are allocated through a lottery. Several voucher experiments have been conducted in the U.S. since 1997. The New York School Choice Scholarship Program (NYSCSP) represents one of the largest such experiments and is the only school voucher experiment with data available for public

¹The Washington Post, Jan 23, 2004.

²CNN, February 13, 2004

³PalmBeachPost.com, July 19, 2003.

⁴A comparable group is one whose only difference from the treatment group is not receiving the treatment; they otherwise would have been the same.

use.⁵

This study analyzes the impact on students' academic achievement of the NYSCSP which ran from 1997 to 2000. I focus on (1) analyzing the impact of using vouchers to attend private schools for various length of time where the impact may depend on the duration of attending private school, (2) identifying variables that may explain the variation of the program impact, and (3) examining the evolution of the impact over time. Experiments in school vouchers provide the "gold-standard" estimate for the impact of offering vouchers on students' academic performance.⁶ The *impact of offering vouchers* is often termed the "Intent-to-Treat (ITT)" effect in the literature. ITT has dominated the empirical literature on school vouchers as the central parameter of interest, and it is the parameter estimated by the experiment. If a substantial fraction of students do not take advantage of the vouchers or quit using the vouchers prematurely, then the impact of offering vouchers is not the only interesting parameter. In the NYSCSP program, more than 20% of the voucher recipients did not use the vouchers at all, another 25% switched back to public schools before the program ended, and only about half of the students who received vouchers stayed in private schools for the entire program period. This dynamic selection aspect in families' school choice decisions is manifested in all of the voucher experiments implemented in the U.S. In the voucher experiments in Washington, DC and Dayton, Ohio, students who used the vouchers until the end of the programs reached only 29% and 60%, respectively (Howell et. al. (2002)).

When the take-up rate is low or when duration of participation in the voucher program varies across individuals, then ITT does not provide enough information to answer some important policy questions. For example, if the ITT effect is low, should the program be continued? If so, should the entire program structure remain intact, or should some of the features be altered? How should the program be redesigned to generate a higher average

⁵It is made possible by *Mathematica Policy Research, Inc.*

⁶It is because the counter-factual average test scores of the voucher recipients if they had not received the vouchers can be estimated unbiasedly using the mean test scores of the students who do not receive the vouchers.

impact in the future? Should the program be targeted at certain populations? Answering these kinds of questions demands a more in-depth analysis of voucher take-up decisions and of the timing of their effects. Many different treatment impacts can be distinguished such as the average impact of offering vouchers, of ever attending private schools, of using the vouchers for only part of the program period, and of using the vouchers for the entire program period.

This study addresses the following questions regarding different treatment effects of the NYC School Voucher Experiment: 1) What is the impact on achievement of using vouchers to attend private schools for the entire program period for the families that used vouchers to the fullest extent? 2) What is the impact on the subgroup of students who had switched back to public schools before the program ended? 3) How do program impacts evolve over time? 4) How does the performance of the public education system affect the estimated impacts? 5) Are program impacts heterogeneous with respect to baseline test scores and demographic characteristics?

Because selection on how long to use the vouchers is nonrandom, I adopt a non-experimental estimation strategy, the method of propensity-score matching which explicitly selects a group of individuals that are “comparable” to the treatment group.⁷ The outcomes (e.g. test scores) of the matched-comparison⁸ group are used to construct the counter-factual out-

⁷Two exceptions emerge recently: Howell et. al. (2002a, 2002b) estimate *the impact of being induced (by the vouchers) to ever attend private schools*. The researchers apply the local average treatment effect (LATE) estimator (Angrist and Imbens (1994)), which belongs to the instrumental variable framework, to compute this impact because whether or not to attend private schools is a choice variable and cannot be identified using OLS regressions. Using the NYC Voucher Experiment, Mayer et. al. (2002) also employ the LATE estimator to compute this impact as well as *the impact of being induced (by the vouchers) to stay in the private schools for the entire 3-year program period*. However, the LATE estimator is insufficient to identify this particular treatment impact using the data from NYC Experiment even though it can be used to identify the impact of being induced to ever attend private schools using randomization on school vouchers as an instrument.

⁸The procedure of constructing the matched-comparison group is developed in great detail in the “Econometric Strategy” Section

comes for the treated individuals, corresponding to their outcome had they not received the treatment. The NYSCSP data set provides a key advantage over observational data sets from other school voucher programs (e.g. Milwaukee) in that the randomized-out control group can serve as a reservoir of students for constructing the relevant comparison group. To implement the matching estimators, I employ recently developed nonparametric, difference-in-difference propensity-score matching methods. I modify them to additionally account for selective non-response. Also, I estimate the propensity scores using both parametric and semiparametric methods.

My empirical results indicate that among the students who fully use the vouchers, only African American students have statistically and economically significant, positive program impacts. The treatment impact for Hispanic students varies by whether English is spoken in the household. I establish that the estimated impacts of the voucher program on Hispanic students are much lower than African American students, in part because of the large test score gains among Hispanic students (but not among African American students) in public schools. Surprisingly, the significant, positive impact experienced by the African American students already appeared in the first year of the program, meaning the program impact stays flat after the first year. I also show the improvement of public school over time and initial academic performance are important determinants of the program impact. On the other hand, mother's education level, the number of school-age children living in the same household, and whether or not students initially attended a "good school"⁹ do not affect the variation of the treatment impact. Among the students who do not use the vouchers until the end of the program, my results show that the magnitude of the impact on mathematics (but not reading) test score for African American students is comparable to that of the African American students who fully used the vouchers, but Hispanic students do not experience any impact. Finally, my results call into question the validity of the LATE estimators that have been applied to this data, which assume the students who do not use the vouchers for the entire duration of the program receive zero program impact.

⁹Good schools are public schools whose students, on average, score higher than the citywide average.

The paper is structured as follows: Section 2 reviews the empirical literature on school vouchers. Section 3 explains why the *LATE* estimator, using the randomization as the instrument, cannot identify the impact of using vouchers for different lengths of time. Section 4 defines the evaluation problem, a fundamental problem encountered in any program evaluation studies. Section 5 discusses the econometric methods employed to solve the evaluation problem. I formulate a dynamic model of school choice to guide the matching variables used in the propensity score model, which plays a key role in implementing the propensity-score matching estimators. Section 6 describes the data from New York School Choice Scholarship Program. Section 7 presents the empirical findings, and section 8 concludes.

2 Empirical Literature on School Vouchers

The empirical literature on school vouchers has heavily emphasized on quantifying *the impact of offering school vouchers* on test scores. Nonetheless, the empirical results on this impact under various voucher programs are mixed. Most of the studies that evaluate publicly funded programs focus on the voucher programs in Milwaukee and Cleveland. Empirical evidence from Milwaukee Parental Choice Program, which started in 1990 and has received the most attention in the literature, ranges from no impact (Witte et. al., 1995) to positive impact on mathematics (but not reading) test scores (Rouse, 1998) to positive impact on both mathematics and reading test scores (Greene, Peterson, and Du 1997). This program, however, differed dramatically before and after 1997. The program enrolled only 341 students who enrolled in 7 non-sectarian, private schools in the 1990-91 school year. During the first five years of the program, enrollment was capped at 1% of the total enrollment in the Wisconsin public schools. After 1997, the Wisconsin legislature raised the enrollment cap up to 15,000 students and allowed religiously affiliated schools to participate. Unfortunately, data on students and their test scores are not collected after the change in 1997. Therefore, the treatment impact on students' academic achievement of this expanded voucher program in Milwaukee is not known and cannot be generalized from the impact of the early version

of the program.

The official evaluation of the Cleveland Scholarship and Tutoring Program (CSTP), which began in the 1996-97 academic year and still operates, has been conducted by Kim Metcalf's evaluation team at the Indiana Center for Evaluation. This study is criticized for methodological problems (Gill et. al. (2001)). Several studies (McEvan (2000), Peterson et. al. (1999)) also caution the interpretation of the Indiana Team's findings. It is not clear whether or not the CSTP can raise the average academic performance of the students who receive the vouchers.

Instead of using non-experimental data, many recent studies evaluate randomized field experiments on private school vouchers. The privately funded voucher programs in Dayton, Ohio, and Washington, DC started in 1998¹⁰. These two privately funded voucher programs and the one in New York City provided partial tuition subsidy to children from low income families. According to Howell and his coauthors (2002a), student applicants in Dayton needed to have family income less than 2 times the federal poverty line while the income eligibility in DC was capped at 2.7 times the poverty line. In both of these programs, the value of vouchers decreased as family income increased. Families were never given vouchers that covered the full tuition. The vouchers in Dayton could cover up to \$1200 or 60% of tuition, whichever was less, and the corresponding figures were \$1700 and 60% in DC. The response rates in these two cities after two years reached only about 50%. The researchers find that the impact on test scores of *offering voucher* in Dayton is not significantly different from zero but that in DC is statistically significant and positive after 2 years. They also examine the impact of *switching to private schools* and find that only African American students experience statistically significant, large positive gains in test scores in both cities (Howell et. al. (2002a)).

There exist several studies that evaluate the voucher experiment in New York City: Howell et. al (2002a, 2002b), Mayer, Peterson, Myers, Tuttle, and Howell (2002), Krueger

¹⁰The information about the privately funded school programs in Dayton, Ohio, and Washington, DC draw exclusively from Howell et. al (2002a).

and Zhu (2003), Peterson and Howell (2003)¹¹. Howell et. al. (2002a, 2002b) evaluate all of the voucher experiments in New York City, Washington, DC, and Dayton, Ohio. Their main results for the NYC case are similar to those of Mayer et. al. (2002), so I do not review Howell et. al. (2002a, 2002b) here.

In their report, Mayer and his coauthors examine the effects on test scores and on other qualitative responses of different dimensions of the NYSCSP program. In terms of test scores, they estimate (i) the impact of offering a voucher, (ii) of being induced to ever attend private schools, and (iii) of being induced to stay in the private schools for the entire three-year program period. As mentioned in the introduction, using regression approach on this experimental data can only guarantee an unbiased estimates on the impact of offering vouchers while using the LATE estimator that uses the randomization of voucher assignment as the only instrumental variable can also identify the impact of being induced to ever attend private schools. However, a strong assumption has to be made when using the voucher assignment status as the only instrumental variable to quantify the impact of being induced to stay in the private schools for a l program periods, where $2 \leq l \leq L = \text{duration of the program}$ (e.g. 3 years). The assumption imposes zero test score impact on those who attend private schools for strictly less than l periods, and the assumption grows stronger as l increases.

Unlike Howell et. al. (2002a,b) and Mayer et. al. (2002), Krueger and Zhu (2003) concentrates almost entirely on estimating the average effect of offering vouchers on test scores¹². These authors question whether or not the large, positive impact on the test scores of African American student found in Mayer's study is generated by a particular sample and/or a particular definition of race used by Mayer and his coauthors. More specifically, Krueger and Zhu are concern that the estimated treatment impacts in Mayer's study are sensitive with respect to the inclusion of students who lack baseline test scores

¹¹Barnard et al (2003) develop a Bayesian framework to estimate the program impacts of the NYSCSP program, and only the 1st year follow-up information is used in their study.

¹²In section 5 of their paper, they also used instrumental variables to estimate the impact on test scores of the number of years in private school.

and the definition of students' race. They argue that the students without baseline test scores should also be included in estimations. With an augmented sample that includes the students without baseline test scores, the researchers find that the estimated effect of offering vouchers on African American students' mathematics test score is equal to 4 national percentile rank (NPR) (Table 3a, Krueger and Zhu)¹³, which is much smaller than the 7.03 NPR estimated by Mayer (Table 22, Mayer et. al.). In fact, all of Krueger and Zhu's estimated impacts on African America students' test score are uniformly smaller than those reported in Mayer's report.

In addition, Krueger and Zhu demonstrate that altering the definition of race can result in a drop in the estimated impacts on test scores for all African American students (Krueger and Zhu, Table 5). In Mayer et. al's study, a student's race follows that of his/her mother. Using this definition, Krueger and Zhu estimate an improvement of the combined mathematics and reading test scores to be 4.1 NPR and is statistical significant. Substituting this definition by the one in which a student's race follows that of either his/her mother or father, the estimated impact drops to 1.52 NPR and is no longer statistically significant.

Despite the two studies disagree on the results regarding the test score impacts for American Americans, they reach a similar conclusion regarding the treatment impact on Hispanic students' test scores: they are not statistically different from zero. (See table 3A and 3B in Krueger and Zhu). These findings raise a puzzle: Why did this particular school vouchers benefit African American students but not Hispanic students? Howell and Peterson (2002) propose that the differential treatment impacts between the two racial groups are caused by the differential characteristics in the public schools initially attended by these students. Krueger and Zhu test and reject this hypothesis, but no alternative explanation is provided. I investigate this issue and present some empirical evidences in the empirical result section in this paper.

¹³This larger sample composes of students who the kindergarten cohort as well as cohort 1 to cohort 4. The former group makes up of 71% of these additional individuals whereas the latter group fills up the remaining 29%.

Responding to Krueger and Zhu’s critique, Peterson and Howell (2003) re-estimate the program impacts by (1) using four different classifications of ethnicity¹⁴, (2) adding explanatory variables based on Krueger and Zhu’s various econometric specifications, and (3) including students who did not have baseline test scores. This gives rise to 120 different ways to re-access the evidence of large, positive, and statistically significant impacts on African American students’ test scores found by Mayer et. al. (2002). Peterson and Howell conclude that African American students who received vouchers do experience positive test score impact and dispute Krueger and Zhu’s criticism.

In sum, even focusing on the *impact on achievement of offering vouchers*, the experimental and non-experimental studies combined offer a wide range of estimates. Most puzzling is the fact that only African American students experience statistically significant and positive impact, and students from other racial/ethnic groups do not.

3 Neither OLS nor LATE is Sufficient to Identify The Impact when It Depends on the Duration of Attending Private Schools

Before discussing the methods of propensity-score matching, it would be useful to see (i) why OLS may fail to identify the impact of attending private schools and (ii) why the LATE estimator requires a strong assumption when estimating the impact of using vouchers to attend private schools for l years, where l is at least two. Recall that the conventional impact of interest is the average impact on test scores of offering vouchers. With experimental data, this average treatment effect can be unbiasedly estimated by running an OLS regression of test scores on the treatment indicator of having a voucher or not and the randomization-block variables. This regression can be expressed as:

¹⁴The four ethnicity classification scheme: A child was considered a African American if (1) the mother was a African Americans, (2) either mother or father was a African American, (3) both mother and father were African Americans, and (4) the parental caretaker of the child was a African American

$$Y_i = \phi_0 + \phi_1 D_i^V + Z_i \gamma + u_i \quad (1)$$

where $i = 1, \dots, n$, indexes the students in the sample; Y_i measures student i 's test score; D_i^V represents an indicator variable equal to *one* if student i receives a voucher and equal to *zero* otherwise; and Z_i is a vector of randomization block variables, which are composed of the variables used to perform the randomization. The parameter of interest is the coefficient of D_i^V , ϕ_1 , which measures the mean impact of offering school vouchers. Therefore, data from randomization on the assignment of vouchers can be used to unbiasedly identify the average effect on test scores of offering vouchers.

However, this framework cannot be applied to identify the mean impact on test scores of using the vouchers for the entire program period or for any l periods where $l \geq 2$. While vouchers are assigned through a lottery system, using the vouchers is a decision. Howell (2004) showed that the students who chose to use vouchers to attend private schools (the “takers”) in the NYSCSP program differ along many dimensions from those students who declined the vouchers (the “decliners”). Therefore, while the original voucher group and the no-voucher group are identical (by virtue of randomization), the voucher takers and the no-voucher group are not equivalent in terms of their characteristics. The endogeneity problem encountered here is analogous to that in the return-to-schooling literature in that the choice of going to college or not is endogenous in the wage equation. The impacts of using vouchers cannot be identified in an OLS regression framework.

Next, consider using the Local Average Treatment Effect (LATE) estimator¹⁵ to identify the impact of using vouchers to attend private schools until the end of the program¹⁶. This approach involves two regression equations: the selection equation and the outcome equation.

¹⁵Angrist, Imbens, Rubin (1996) also name the LATE estimator as Complier Average Causal Effect (CACE) estimator.

¹⁶Note again that in principle, this argue applies to attending private schools any particular number of years, so long as the number is at least two.

$$S_i = \alpha_0 + \alpha_1 D_i^V + X_i' \alpha_2 + \varepsilon_{p,i} \quad (2)$$

$$Y_i = \beta_0 + \beta_1 S_i + X_i' \beta + \varepsilon_{y,i} \quad (3)$$

where $S_i = 1$ if student i attends private school for the entire 3-year program period; X_i include the randomized blocks and the characteristics of child i , his/her family and the school that he/she attended; Y_i measures student i 's test score at the end of the program; and $(\varepsilon_{p,i}, \varepsilon_{y,i})$ are the random error terms that capture the unobservables that affect the public/private school choice and student achievement. Note that the indicator variable S_i is endogenous because the school choice decision (choosing private schools versus public schools) is made based on both the observed (child, family, and school) characteristics as well as the unobserved (to the econometricians) characteristics such as the innate ability of the child. D_i^V indicates the voucher assignment status based on randomization and is independent of the two error terms $(\varepsilon_{p,i}, \varepsilon_{y,i})$ but is correlated with S_i ¹⁷. This can serve as a valid instrument. Even though it can produce an asymptotically unbiased and efficient estimate of β_1 , which captures the impact of being induced to attend private schools of the entire duration of the program, the LATE setup applies in this setting requires a strong assumption about the impact of those students who attended private schools less than the 3-year program period in order to identify this parameter. To see this, note that $S = 1$ if student attended private schools for the duration of the program and $S = 0$ if the students did not attend at all or attended for some years but then switched back to public schools before the program ended. The assumption requires that exposure to private schools for less than the 3-year program period had no impact on the 3rd follow-up test scores¹⁸. Using the propensity-score matching method to estimate the impact of using vouchers for a particular number of years does not require this assumption. I also use the propensity-score matching

¹⁷Students with vouchers had much a higher proportion of attending private schools for the entire 3-year program period whereas only a negligible number of students without vouchers attended private schools until the end of the program. The school vouchers do alter the parental school choice decision.

¹⁸Mayer et. al. (2002) are also aware this assumption and state that in their paper.

method to estimate the impact on the test scores of these partial-attendees and show that this assumption does not hold.

4 The Evaluation Problem

The central problem in any economic program evaluation exercises is the evaluation problem. Let Y_{kit} be individual i 's outcome measure in period t if he/she receives treatment k , where $k = \{s, s'\}$. The individual treatment impact of moving from treatment state s to treatment state s' with outcome measured at time t is

$$\Delta_{it}(s, s') = Y_{sit} - Y_{s'it} \quad (4)$$

If the outcome of interest is test scores, then the individual treatment impact measures the test score gain (or loss) at period t for student i when the student receive treatment s relative to treatment s' . Because each individual can only receive one and only one treatment in each time period, one of the two terms in this individual treatment impact is not observed. Missing data is a fundamental problem in any impact evaluation study.

Instead of computing the impact for every single individual, a more popular impact parameter of interest measures the *mean impact of treatment on the treated (TT)*, which represents the average effect on the outcome for individuals who receive treatment s :

$$\begin{aligned} \Delta_{TT,t}(X_{t'} = x_{t'}) &= E(Y_{st} - Y_{s't} \mid X_{t'} = x_{t'}, D_s = 1) \\ &= E(Y_{st} \mid X_{t'}, D_s = 1) - E(Y_{s't} \mid X_{t'}, D_s = 1) \end{aligned} \quad (TT)$$

where $D_s = 1$ if treatment s is received; $D_s = 0$ otherwise, conditional on some observed characteristics $X_{t'}$ with t' denoting the pre-program period and t denoting the post-program period. The second term in the above equation is not observed, for it is the counterfactual average outcome in the treatment state s' for the individuals who receive treatment s . Propensity-score matching methods provide a way to solve this missing data problem by imputing the counterfactual mean using the outcomes of a “matched-comparison” group¹⁹.

¹⁹The details of forming the matched-comparison group is provided later in this section.

An average version of equation (TT), which integrates over the region of support of $X_{t'}$, $S_{X_{t'}}$, can be written as

$$\begin{aligned}\Delta_{TT,t}(X_{t'}) &= E(Y_{st} - Y_{s't} \mid X_{t'}, D_s = 1) \\ &= \int_{S_X} E(Y_{st} - Y_{s't} \mid X_{t'} = x_{t'}, D_s = 1) f_X(X_{t'} = x_{t'} \mid D_s = 1) dx\end{aligned}\tag{5}$$

where $f_X(X_{t'} = x_{t'} \mid D_s = 1)$ is the conditional density of $X_{t'}$, conditional on receiving treatment s .

The current study focuses on quantifying this impact parameter. It is important to note that the definition of treatment depends on the question of interest. In this study, there are two relevant definitions of treatment: (1) using vouchers for entire 3-year program period, and (2) using vouchers for strictly less than the 3-year program period²⁰. This parameter in general measures the average difference in outcomes between the individuals who receive treatment s and the same set of individuals had they received treatment s' . The counterfactual outcome of this latter group is not observed. Next section discusses the methods and assumptions needed to estimate the counterfactual outcomes of this latter group.

5 Econometric Strategy: Propensity-Score Matching Estimators with Exact Match

The data on the NYSCSP voucher students who attend private schools for three years identifies the average treated-state test scores measured after the program, $E(Y_{1t} \mid X_{t'}, D = 1)$, whereas the data on the NYSCSP students who were not offered vouchers gives a direct

²⁰In principle, I could define treatment as using vouchers to attend private schools for l years, where $l = 1, 2, \dots, L$. L stands for some positive integer. In the New York School Voucher Program, L is 3. However, because the sample size would become too small once I split into attending one year, attending two years, and so on, I separate the voucher takers into full attendents (those used the vouchers for the entire 3-year program period) and the partial attendents (those used the vouchers for less than 3 years).

estimate of $E(Y_{0t} | X_{t'}, D = 0)$. Two assumptions are required for matching estimators to identify the parameter of interest, the mean impact of treatment on the treated (TT).

Assumption 1: *Conditional mean independence*

$$E(Y_{0t} | X_{t'}, D = 1) = E(Y_{0t} | X_{t'}, D = 0) = E(Y_{0t} | X) \quad (\text{A1a})$$

Assumption 2: *Existence of comparable non-participants:*

$$\Pr(D = 1 | X_{t'}) < 1 \quad (\text{A1b})$$

The conditional mean independence condition assumes that the mean outcomes without participation between the treatment participants and the non-participants are equal, conditioning on a vector of variables $X_{t'}$. In other words, conditioning on $X_{t'}$, the mean outcome of the non-participants can be used to impute the mean non-participating outcome for the participants. Assumption (2) simply rules out the possibility of an empty set of nonparticipants at all values of $X_{t'}$. These are weak assumptions for identifying TT . Nonetheless, if unobservables are important determinants of program participation, then the mean independence assumption would not hold. With these two assumptions, the average impact of treatment on the treated can be identified:

$$\begin{aligned} \Delta_{TT,t}(X_{t'}) &= E(Y_{1t} - Y_{0t} | X_{t'}, D = 1) \\ &= E(Y_{1t} | X_{t'}, D = 1) - \underbrace{E(Y_{0t} | X_{t'}, D = 1)}_{\text{not observed}} \quad (6) \\ &= E(Y_{1t} | X_{t'}, D = 1) - \underbrace{E(Y_{0t} | X_{t'}, D = 0)}_{\text{directly estimated from data}} \quad (\text{by A1a}) \end{aligned}$$

where the third equality follows from assumption (A1a).

If the set of matching variables $X_{t'}$ is large, matching on them becomes problematic because either many cells are left empty or the convergence rate may be very slow when $E(Y_{0t} | X_{t'}, D = 1)$ is estimated nonparametrically. Instead, I apply Rosenbaum and Rubin

(1983)'s seminal result and employ the propensity score (p-score) matching. In this case, assumption (A1a) needs to be modified as

$$E(Y_{0t} | P(X_{t'}), D = 1) = E(Y_{0t} | P(X_{t'}), D = 0) = E(Y_{0t} | P(X_{t'})) \quad (\text{CS1a})$$

where $P(X_{t'}) \equiv P(D = 1 | X_{t'})$ is the conditional probability of receiving treatment and is called as propensity score (p-score). This condition means that the average no-treatment state outcome of the recipients is the same as that of the treatment non-recipients. As such, conditioning on the p-score, participation decision is independent of the average outcome in the absence of treatment. The simplest-form, cross-section matching estimator for $\Delta_{TT,t}(X_{t'})$ can be derived as follow:

$$\widehat{\alpha}_{TT,t}^{CS} = N_1^{-1} \sum_{i \in I_1 \cap S_p} \left\{ Y_{it}(D = 1) - \widehat{Y}_{it}(D = 0) \right\} \quad (\text{CSTT})$$

$$= N_1^{-1} \sum_{i \in I_1 \cap S_p} Y_{it}(D = 1) - N_1^{-1} \sum_{i \in I_1 \cap S_p} \widehat{Y}_{it}(D = 0) \quad (7)$$

$$= N_1^{-1} \sum_{i \in I_1 \cap S_p} Y_{1it} - N_1^{-1} \sum_{i \in I_1 \cap S_p} \left\{ \sum_{j \in I_0 \cap S_p} \widehat{W}(i, j) * Y_{0jt} \right\} \quad (8)$$

$$= N_1^{-1} \sum_{i \in I_1 \cap S_p} \left\{ Y_{1it} - \sum_{j \in I_0 \cap S_p} \widehat{W}(i, j) * Y_{0jt} \right\} \quad (9)$$

where $Y_{it}(D = 1) = Y_{1it}$ and $Y_{it}(D = 0) = Y_{0it}$; N_1 refers to the number of participants; (I_1, I_0) stand for the set of the participants and the non-participants, respectively; S_p represents the region of common support of the p-score²¹; and $\widehat{W}(i, j)$ is the estimated weight assigned to non-participant j when matching with participant i . These weights sum to one for each i in I_1 . Different functional forms taken by the weighting functions $\widehat{W}(i, j)$ define different methods of p-score matching. I use four different types of p-score matching in this paper, and the next sub-section explains them in detail²².

²¹See the precise specification of the region of overlapping support in the section *Empirical Implementation of the Propensity Score Model*.

²²The results using all of these other matching estimators are available upon requested.

For difference-in-difference matching estimators, which I use to quantify the main results presented in this study, assumption (A1a) once again has to be modified. An equivalent version of (CS1a) for the difference-in-difference matching estimator can be expressed as

$$\begin{aligned}
& E(Y_{0t} - Y_{0t'} \mid P(X_{t'}), D = 1) \\
= & E(Y_{0t} - Y_{0t'} \mid P(X_{t'}), D = 0) \tag{DID1a} \\
= & E(Y_{0t} - Y_{0t'} \mid P(X_{t'}))
\end{aligned}$$

I adopt the conditional nonparametric, difference-in-difference, local linear regression matching estimator (Heckman et. al (1997, 1998) and modify it to additionally account for selective non-response. This estimator has not been applied to evaluate school voucher programs in the U.S. The general form of this estimator can be written as

$$\begin{aligned}
\widehat{\alpha}_{TT}^{DID} = & N_{1t}^{-1} \sum_{i \in I_{1t} \cap S_{pt}} \left\{ Y_{1it} - \sum_{j \in I_{0t} \cap S_{pt}} \widehat{W}(i, j) * Y_{0jt} \right\} \\
& - N_{1t'}^{-1} \sum_{i \in I_{1t'} \cap S_{pt'}} \left\{ Y_{1it'} - \sum_{j \in I_{0t'} \cap S_{pt'}} \widehat{W}(i, j) * Y_{0jt'} \right\}
\end{aligned} \tag{DDTT}$$

5.1 Different Methods of (Propensity-Score) Matching

Different methods of p-score matching are used to check the sensitivity of the estimates. The main results reported in this study is estimated using the difference-in-difference, local linear matching estimator. The four p-score matching estimators considered in this paper are

1. Local Linear Regression Matching
2. Kernel Matching
3. M -Nearest Neighborhood Matching, where M represents the number of neighbors

4. Caliper Matching

For notational simplicity, I omit the time subscript in this sub-section. Each of these matching estimators can be implemented in either cross-section version or difference-in-difference version. The former requires data only in the post-program period whereas the latter demands data in both the pre- and post-program periods. As mentioned in the previous section, these methods differ in their weighting function $\widehat{W}(i, j)$ used in equations (CSTT) and (DDTT). Suppose that the propensity score is already estimated for each person in the entire sample²³. Let $P_i \equiv P(X_i|\widehat{D} = 1)$ and $P_j \equiv P(X_j|\widehat{D} = 1)$ where $(i, j) \in \{1, 2, \dots, N\}$, $N = N_1 + N_0$.

1. Local Linear Regression Matching:

$$\widehat{W}(i, j)_{LLR} = \frac{K_{ij} \sum_{k=1}^{N_0} K_{ik} \left(\frac{P_k - P_i}{h_{N_0}} \right)^2 - \left[K_{ij} \left(\frac{P_j - P_i}{h_{N_0}} \right) \right] \left[\sum_{k=1}^{N_0} K_{ik} \left(\frac{P_k - P_i}{h_{N_0}} \right) \right]}{\sum_{j=1}^{N_0} K_{ij} \sum_{k=1}^{N_0} K_{ik} \left(\frac{P_k - P_i}{h_{N_0}} \right)^2 - \left[\sum_{j=1}^{N_0} K_{ij} \left(\frac{P_j - P_i}{h_{N_0}} \right) \right]^2} \quad (10)$$

where $K_{ij} \equiv K \left(\frac{P_i - P_j}{h_{N_0}} \right)$ and $K(\cdot)$ represents a kernel function.

2. Kernel Matching:

$$\widehat{W}(i, j)_{KM} = \frac{K \left(\frac{P_i - P_j}{h_{N_0}} \right)}{\sum_{j=1}^{N_0} K \left(\frac{P_i - P_j}{h_{N_0}} \right)} \quad (11)$$

where $K(\cdot)$ is a kernel function, and h_{N_0} is a bandwidth.

3. M -Nearest Neighborhood Matching:

$$\widehat{W}(i, j)_{NN} = \frac{1}{\#Neighbors} * \begin{cases} 1 & \text{if } |P_i - P_j| \leq \min_M |P_i - P_j| \\ 0 & \text{if } |P_i - P_j| > \min_M |P_i - P_j| \end{cases} \quad (12)$$

where $\min_M |\cdot|$ is the M^{th} shortest distance between P_i and P_j (for some j) and $|P_i - P_j|$ denotes the absolute value between P_i and P_j .

²³The details of how to estimate the propensity scores are discussed in the last subsection of this section.

4. Caliper Matching:

$$\widehat{W}(i, j)_{CAL} = \frac{1}{\#(\text{matched persons})} \begin{cases} 1 \text{ if } |P_i - P_j| \leq \rho \\ 0 \text{ if } |P_i - P_j| > \rho \end{cases} \quad (13)$$

where ρ is the maximum tolerable distance between P_i and P_j .

5.2 Propensity-Score Matching Estimators that Account for non-response

In the presence of non-response, the mean effect of treatment on the treated becomes conditional only on those individuals who have data available²⁴:

$$\Delta_{TT,t}(A = 0) = \int_{S_p} E(Y_{1t} - Y_{0t} \mid A = 0, P(X_{t'}) = p, D = 1) f_p(p \mid A = 0, D = 1) dp \quad (14)$$

where $A = 1$ if a person did not take the test in period t (the “non-respondents”) and equal to zero otherwise. Notice that a student did not participate in the testing session in period t does not mean that the student did not receive the treatment (e.g., attended private schools) between the period t and t' ²⁵. Without further assumptions, Δ_{TT} and $\Delta_{TT}(A = 0)$, in general, are different from each other. So, an extra assumption is required in order to identify Δ_{TT} .

To simplify notations, let $P(D = 1 \mid X_{t'}) \equiv P(X_{t'}) \equiv P$. To control for non-response explicitly in the estimator and to identify $\Delta_{TT,t}$, one more assumption is imposed.

Conditional mean independence assumption:

²⁴An alternative way that requires a much weaker assumption about the outcomes of the non-respondents is to produce an interval estimate rather than a point estimate (Manski (1990)). This idea is discussed in the appendix.

²⁵Information on private school attendance is included in each of the survey wave during the program period (1997/8 - 1999/2000).

$$\begin{aligned}
& E(Y_{1t} - Y_{0t} \mid A = 1, P(X_{t'}), D = 1) && \text{(A3)} \\
& = E(Y_{1t} - Y_{0t} \mid A = 0, P(X_{t'}), D = 1) \\
& = E(Y_{1t} - Y_{0t} \mid P(X), D = 1)
\end{aligned}$$

This assumption states that conditional on the propensity score, the average no-treatment outcomes of the responded, treatment-recipients are equal to that of the non-responded, treatment-recipients.

The average (over the propensity scores) version of $\Delta_{TT,t} = E(Y_{1t} - Y_{0t} \mid D = 1)$ can be written as

$$\begin{aligned}
& \Delta_{TT,t} \\
& = E(Y_{1t} - Y_{0t} \mid D = 1) \\
& = \int_{S_p} E(Y_{1t} - Y_{0t} \mid P(X_{t'}) = p, D = 1) f_p(p \mid D = 1) dp \\
& = \int_{S_p} E(Y_{1t} - Y_{0t} \mid A = 0, P(X_{t'}) = p, D = 1) f_p(p \mid D = 1) dp && \text{(by A3)} \\
& = \int_{S_p} E(Y_{1t} - Y_{0t} \mid A = 0, P(X_{t'}) = p, D = 1) f_p(p \mid A = 0, D = 1) \left(\frac{f(p \mid D = 1)}{f(p \mid A = 0, D = 1)} \right) dp && \text{(TT2)}
\end{aligned}$$

where $f_p(p|\cdot)$ denotes the conditional density of the propensity score, conditional on the responded, treatment-recipients. Notice that each of the three pieces of this last equation can be estimated using the available data because the propensity score is estimated for every student at the baseline, including those student who lack test scores at the baseline and in the 3rd year. According to equation (TT2), the mean impact of treatment on the treated can be consistently estimated by

$$\begin{aligned}
& \widehat{\alpha}_{TT, \text{non-response}}^{DID} \tag{DDTT} \\
= & \sum_{i \in I_1 \cap S_p \cap \{A=0\}} \left\{ \omega_i * \left[Y_{1it} - \sum_{j \in I_0 \cap S_p \cap \{A=0\}} W(\widehat{i}, j) * Y_{0jt} \right] * \left(\frac{f(P_i | \widehat{D}_i = 1)}{f(P_i | \widehat{A}_i = 0, D_i = 1)} \right) \right\} \\
& - \sum_{i \in I_1 \cap S_p \cap \{A=0\}} \left\{ \omega_i * \left[Y_{1it'} - \sum_{j \in I_0 \cap S_p \cap \{A=0\}} W(\widehat{i}, j) * Y_{0jt'} \right] * \left(\frac{f(P_i | \widehat{D}_i = 1)}{f(P_i | \widehat{A}_i = 0, D_i = 1)} \right) \right\}
\end{aligned}$$

where t is the post-program period (the 3rd year); t' is the pre-program period (baseline); S_p is the region of overlapping support of the propensity scores; $\{A = 0\}$ is the set of responded students, who have 3rd year test scores; I_1 is the set of the treatment group members; I_0 is the set of the control group members; w_i is the ‘‘original baseline’’ sampling weights, obtained directly from the data²⁶; $w_i = \frac{w_i}{\sum_{i \in I_1 \cap S_p \cap \{A=0\}} w_i}$; $(Y_{1it}, Y_{1it'})$ are the test

²⁶The reason I emphasize the term ‘‘original baseline’’ is that the adjusted sample weights are also provided in the NYSCSP data set and are used by Mayer et al (2002), Krueger and Zhu (2003), and Peterson and Howell (2003). These weights are adjusted to control for the non-responses in the each of the follow up surveys and test sessions. However, these adjusted weights are not used in the current study because the estimation equation specified in equation (DDTT) uses only the original baseline weights.

The (baseline) sampling weights are used to capture the experimental design. In particular, it reflects (1) the different probabilities of being offered an educational voucher to each eligible family and (2) the composition of the population of eligible applicants. The baseline weights are computed as follow

$$w_i = \frac{1}{f_i * p_i} \tag{15}$$

where f_i includes the adjustment factors (1. five discrete points at which families applied for scholarships; 2. whether a child originally attended a public school with below-average achievement; 3. the number of eligible children in a family $\{1, 2, 3 \text{ or more}\}$. So, there were $30 = 5 * 2 * 3$ randomized blocks), and p_i represents the probability of being selected for a voucher.

The dataset provided to me does not include the first two adjustment factors nor the individual probability of being selected for a voucher p_i ; it only includes the original and the revised versions of the 30 randomized blocks. For this reason, I am not able to figure out whether or not a child originally attended a below-citywide-average public elementary school. For the details of how to construct the baseline weights and

scores of person i who belongs to the treatment group; $(Y_{0jt}, Y_{0jt'})$ are the test score of person j who belongs to the control group; D_i is the treatment indicator that is equal to *one* if person i receives the treatment and equal to *zero* otherwise; $\widehat{W}(i, j)$ is an estimated weight applied to Y_{0j} . Both $f(P|D = 1)$ and $f(P|A = 0, D = 1)$ are estimated nonparametrically. I_1 is the set of individuals who receive treatment. S_p is the region of common support of the p-scores, and it is not indexed by t or t' for the same reasons as those of I_1 . Computing the ratio of the propensity score of each individual $\frac{f(P_i|D=1)}{f(P_i|A=0, D=1)}$ involves two steps. The first step estimates a model of propensity score p_i using all of the students with baseline test scores. With p_i estimated for each student, $f(P_i|D = 1)$ and $f(P_i|A = 0, D = 1)$ are then estimated by kernel density methods.

5.3 The Propensity Scores

A Propensity score represents the conditional probability of receiving treatment, conditioning on some characteristics. The definition of treatment depends on the question of interest. For example, if the question of interest is “what is the average impact of using vouchers for the entire duration of the experiment?”, then treatment is defined as “using vouchers for the entire duration of the experiment.” Similarly, if the question of interest is “what is the impact of using vouchers to attend private schools for l years where l is some integer smaller than the length of the program, then the treatment should be defined as “using vouchers to attend private schools for x years .”

The evidences in the data section show that school vouchers greatly altered families’ incentive to choose private schooling. For this reason, the propensity score for the voucher students is defined as *the probability of using vouchers to attend private schools for the entire program period, conditional on being offered the vouchers and some baseline characteristic X_{it}* , and the propensity score for the no-voucher students is defined as *the probability of using vouchers to attend private schools for the entire three-year program period had they been offered been the school vouchers, conditional on some baseline characteristics X_{it}* .

their adjusted versions, see Mayer et al (2002).

5.4 Empirical Implementation of the Propensity Score Model

The propensity score is estimated using a generalized additive model (*GAM*) of the form

$$\log \left(\frac{P(X_{t'})}{1 - P(X_{t'})} \right) = \alpha_0 + \sum_{k=1}^{K_1} f(X_{k,t'}) + \sum_{k=1}^{K_2} X_{k,t'} \alpha_k \quad (16)$$

where the $f(\cdot)$'s are smooth functions that allow a flexible modeling of the continuous matching (explanatory) variables. *GAM* is a semiparametric method, and in the literature of program evaluation a parametric model (such as logit) is usually applied to estimate the propensity-scores.

I develop a simple dynamic model of school choice to capture the parents' decision on choosing the type of schools (public and private) for their children in each of the children's schooling year before attending college and to provide a guidance for the variables used in the matching procedure (See appendix B). To implement the matching procedure, variables that depend on the whether or not treatment is received cannot be used as the matching variables. It is because the propensity score matching procedure requires to use the average outcome of the matched-comparison group students who do not receive the treatment as an estimate of the mean "no-treatment outcome" of the treatment group²⁷. This procedure demands that the density of the matching variables not to vary with the treatment status of the treatment group. All of the variables used in the matching procedure are observed before the private school attendance decision is made. The matching (or conditioning) variables include two set of variables: Child characteristics: initial test scores, initial grade, age, gender, race, place of birth, whether or not English is the main language at home; and Family/parental characteristics: the number of school age children living in the same family, house income, mother's religion, mother's education level, mother's employment status at the baseline and a set of dummy variables indicating whether or not the child's family receives food stamp and Medicaid.

I estimate the propensity score in two steps. First, using only the students who received

²⁷The average no-treatment outcome of the treatment group is the mean outcome of this group if they had not receive the treatment.

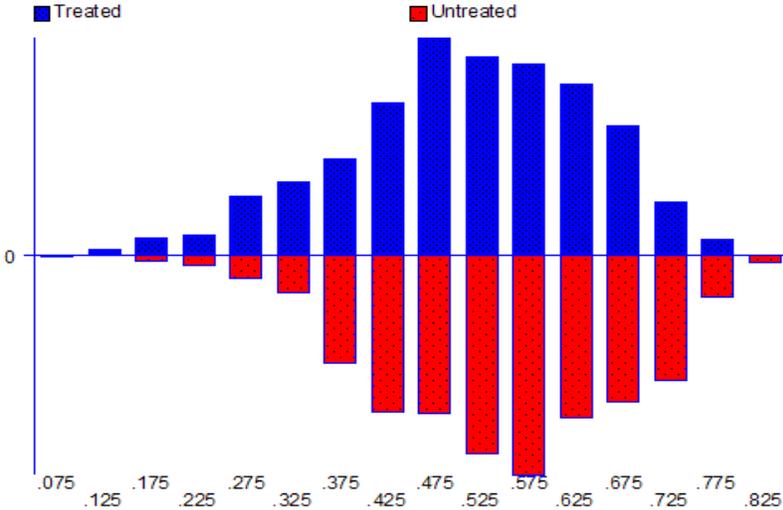
vouchers, I estimate a semi-parametric, binary regression model (using GAM) with the indicator that takes value one if a student attended a private school for the entire 3-year period and zero otherwise as the independent variable and the variables mentioned in the previous paragraph as the matching (or explanatory) variables. Second, the estimated coefficients from the first-stage estimation are used to simulate the probability of receiving the treatment for the individual in the no-voucher group.

The methods of propensity-score matching require a “matched-comparison” group. Students in the no-voucher group can be extracted to form the “matched-comparison” group. As emphasized in the introduction, the no-voucher group as a whole cannot be used as a control group when estimating the impact of using vouchers for different lengths of the time; the no-voucher group no longer has the same distribution of (both observed and unobserved) characteristics as the group that used the vouchers until the program ended or the group that used the vouchers for only part of the program period. In other words, randomization in the assignment of school vouchers can only guarantee the equivalence of characteristics between the group with vouchers and the group without. As how long to use the vouchers is a decision, randomization in the assignment of school vouchers does not guarantee the equivalence of characteristics among the group who use the vouchers, the group who do not use the vouchers, and the group without vouchers.

In practice, I combine propensity-score matching with exact matching on two discrete variables: Race (African Americans, Hispanics, Other) and Cohort (1,2,3,4, or 5). This strategy avoids matching two individuals who have different races and grade level even though they may have similar propensity scores.

When implementing the propensity-score matching procedure, an overlapping support requirement is imposed. Denote the estimated minimum and the maximum propensity scores for the treatment group as (P_1^{\min}, P_1^{\max}) and the corresponding probabilities for the matched-comparison groups as (P_0^{\min}, P_0^{\max}) . The overlapping region of propensity scores between the treatment and the matched-comparison groups is defined as $(\min \{P_0^{\min}, P_1^{\min}\}, \min \{P_0^{\max}, P_1^{\max}\})$. Observations with propensity scores fall outside this range are not used in the estimation.

The elimination of observations sitting outside of the region of common support potentially could create large evaluation bias (Heckman et. al. (1997)). The non-overlapping support problem, however, does not raise any concern in the current study. The following graph shows the estimated propensity score for all of the voucher recipients who attended private schools for the duration of the experiment (denoted as the “treated” in the top half of the graph) and that of the entire no-voucher group students (denoted as the “untreated” in the bottom half of the graph). The no-voucher group provides a set of individuals who are extracted to form the matched-comparison group. Clearly, the range of the estimated propensity score of the “treated” overlaps the entire range of that of the “untreated”. This is one of the key advantages of having the no-voucher group as a potential comparison group from which the matches are chosen according the propensity scores.



Evaluation biases generated from the mismatch of geographic locations between the treatment and the comparison individuals and the mismatch of the survey instrument used to collect data on the two groups could be serious²⁸. These sources of bias do not occur in the current study because the group students who used the vouchers and the group of students who did not receive the vouchers had parents who showed interests in private education;

²⁸Heckman et. al. (1997) empirically show the importance of these two sources of biases, using their data from a job-training program.

lived in the same metropolitan area – New York City; had “low” household income at the time of the application; were administered the same questionnaire; and took the same tests (ITBS) at the baseline and in each of the follow-up years.

Cross-section matching can eliminate biases due to (I) the difference in the regions of support of the propensity-scores²⁹ and (II) difference in the distributions of the conditioning variables X_{ν} (Heckman et. al. (1997)). Only the bias that is caused by the differences in unobservables between the two groups cannot be wiped out through the cross-sectional matching procedures. Nonetheless, if the unobserved factors are time invariant, then difference-in-difference matching estimators can also eliminate them.

6 Data from New York School Choice Scholarship Program

The NYSCSP program represents one of the four randomized social experiments on private school vouchers implemented in the U.S.³⁰ Privately funded by the School Choice Scholarships Foundation, the program randomly selected 1374 scholarships to children from low-income families that met the eligibility criteria. Each scholarship/voucher provided the eligible families with a maximum value of \$1400 annually for the cost of attending private schools for up to three years. There was no restriction on the religious orientation of the private schools. To satisfy the eligibility rules, the students (i) had to be residents in New York City, (ii) enrolled in and attended public schools in New York City, (iii) entered either the 1st, 2nd, 3rd, 4th, or 5th grade in fall 1997, and (iv) qualified for the federal school lunch program.³¹ The applicants were randomly selected by a lottery to receive the vouchers.

²⁹However, as I showed in the histogram above, potential evaluation bias caused by non-overlapping support of propensity scores should not raise any concern.

³⁰The other three were conducted in Washington, D.C., Dayton, Ohio, and nation-wide by the Children’s Scholarship Foundation. The data collected in these three randomized trials, however, are not publicly available.

³¹These rules are listed in the eligibility checklist in the application/questionnaire of the NYSCSP program.

In each year from 1997 to 2000, both voucher group and the randomized-out control group students were tested using the Iowa Test of Basic Skills (ITBS), and a surveys were implemented to both the parents and students who attended fourth grade or above.³² To quantify different treatment impacts of the voucher program, I use both mathematics and reading test scores and the survey data at the baseline (1997) and in each of the three follow-up years (1998,1999, 2000).

6.1 Evidences of Non-Response

The original sample consisted of 2666 students, 1374 (51.5%) were awarded a voucher and 1292 (48.5%) were not. Of the 2666 students, only 1851 (or 69.4%) had the baseline test scores. The high non-response rate at the baseline is mainly due to the fact that all of the 575 students who entered grade one (**cohort-1**) in fall 1997 were not required to take the test at-the baseline. For students who entered grade 2 through 5 (**cohort-2 - cohort-5**) in Fall 1997, the percentage of students without baseline test scores ranged from 10% to 14%. Reasons for missing the baseline tests include illness, students' refusal to take the tests, and some tests were lost in the administrative process (Howell et. al., 2002). The breakdown of these numbers are presented in the following table:

Entrance Grade in Fall 1997	Has Baseline Scores	No Baseline Scores
1	0	575(100%)
2	473(89.4%)	56(10.6%)
3	513(88.4%)	67(11.6%)
4	490(89.7%)	56(10.3%)
5	375(86.2%)	60(13.8%)
Total	1851	814³³

³²These data were collected by researchers from Mathematica Policy Research, Inc. (*MPR*) and Harvard University's Program on Education Policy and Governance (*PEPG*).

³³One student had missing cohort information. So, the sum of the students with and without baseline

A subject is considered as non-response in a particular follow-up survey if the subject does not have test scores in that follow-up year. The next table shows the non-response rates over the program period from 1997 to 2000.

Number of students with test scores in year t		
	include cohort-1	exclude cohort-1
Baseline (1997)	1851 (69.7%) ³⁴	1851 (88.6%) ³⁵
1 st Follow-up (1998)	2080 (78%)	1632 (78%)
2 nd Follow-up (1999)	1754 (65.8%)	1352 (64.7%)
3 rd Follow-up (2000)	1801 (67.5%)	1369 (65.5%)
Total # Students	2665	2090

The column “include cohort-1” includes all of the students entering grade 1 in fall 1997 while the column “exclude cohort-1” excludes this cohort. As mentioned in the previous subsection, all of the 575 kindergarten students are not tested at the baseline. Therefore, the number of students with test scores in 1997 equals to 1851, whether or not kindergarten students are counted. In the second row of the middle column, there are more students with test scores in the 1st follow-up year (2080) than at the baseline (1851) because the kindergarten cohorts who are require to take the tests every year once they become grade 1 students. Surprisingly, the non-response rate, which is one minus the percentage of the students with test scores in year t , with and without the kindergarten students increase at a similar speed over time. One-third of the original sample is loss by the 3rd year of the program.

The table above shows the non-response pattern, including the students with missing baseline test scores. The next table shows the non-response pattern over time, restricting to students with baseline test scores.

test scores added up to 2665 rather than 2666.

³⁴69.7% of the 2666 original sample have test scores at the baseline year (1997).

³⁵88.6% of the 2090 students from cohort 1 to cohort 4.

Number of students with test scores at the baseline and in year t

	Overall	African Americans	Hispanics
Baseline (1997)	1851 (100%)	806 (100%)	867 (100%)
1 st Follow-up (1998)	1455 (78.6%)	623 (77.3%) ³⁶	709 (80.9%) ³⁷
2 nd Follow-up (1999)	1199 (64.8%)	497 (61.7%)	612 (69.8%)
3 rd Follow-up (2000)	1250 (67.5%)	519 (64.4%)	637 (72.7%)
Total # students with BLTS ³⁸	1851	806	867

When restricting to students with baseline test scores, the entire kindergarten cohort is excluded. The non-response pattern of the students overall (second column of this table) is virtually identical to those in the previous table. On the other hand, African American students have higher non-response rates than do Hispanic students in each of the follow-up years.

6.2 Evidences on Families' Self-Selection on Public/Private Schools

The next table summarizes the patterns of private school attendance of the voucher recipients and the no-voucher group members who have both *baseline* and 3rd year test scores.

Only for students with both *baseline* and 3rd year test scores

Years in Private Schools	Voucher Recipients	No-Voucher Students
0	99 (14%)	499 (86.3%)
1	41 (6%)	22 (3.8%)
2	61 (9%)	43 (7.4%)
3	471 (70%)	14 (2.4%)
Total (Students)	672	578

³⁶77.3% of the African American students with baseline test scores.

³⁷80.9% of the 867 Hispanic students with baseline test scores.

³⁸BLTS stands for baseline test scores.

There are 1250 students with baseline and 3rd year test scores. 672 (53.8%) of these 1250 students receive vouchers, and 578 (46.2%) do not. Among the voucher recipients who have test scores available both at the baseline and in year 3, 70% of them attended private schools for 3 years, and 14% never used the vouchers for private schooling. In contrast, only about 14% of the no-voucher group ever experienced private education. These patterns indicate that vouchers (with the amount of \$1400 maximum per year for up to 3 years) do alter families' school choice behavior; a majority of the voucher recipient families would have sent their children to public schools had they not received the vouchers. The patterns of private school attendance behavior for students with 3rd year test scores (i.e. regardless of whether or not they have baseline test scores), as shown in the next table, are similar to those shown in the previous table.

Students with 3rd year test scores		
Years in Private Schools	Voucher Recipients	No-Voucher Students
0	153 (16%)	730 (86.2%)
1	61 (6.4%)	57 (6.7%)
2	89 (9.3%)	34 (4%)
3	651 (68.2%)	26 (3.1%)
Total (Students)	954	847

7 Empirical Results

This section presents the results of different treatment impacts of the NYSCSP program. These impacts include 1) the average impact on achievement for the students who used vouchers for the entire duration of the program; 2) the average impact on achievement for the students who had switched back to public schools before the program ended; 3) the evolution of the program impacts over the course of the program; and 4) each of these impacts for different racial/ethnic groups. The definition of treatment varies with the impact of interest. For example, the treatment for impact (1) is defined as “using vouchers

to attend private schools for the entire 3-year duration of the program.” As we shall see later in this section, the estimated treatment impact varies across racial/ethnic groups. I examine several factors that may influence the magnitude of the impact: I) The performance of public education system over time; II) whether or not English was the main language spoken at home (primarily for Hispanic students); III) students’ academic performance before they entered the program; (IV) whether or not students originally attended a “high-performing” public school³⁹; (V) whether or not students whose mothers had college education; and (VI) the number of school-age children living in the same household.

It is important to point out that when estimating the program impact for each subgroup such as African American students, Hispanic students, or students whose mothers had college education, the entire matching procedure (except estimating the propensity scores) is implemented. For example, to estimate the mean impact on test scores for all of the African Americans who fully used the vouchers, propensity-score matching is performed using only this subgroup as the treated individuals and the African American students who did not receive the vouchers as a potential comparison group, which is used to construct the matched-comparison group. With a matched-comparison group, I then construct the counterfactual outcome for the treated individuals. The average difference between the outcome of the treated individuals and the counterfactual outcome forms the mean impact of using vouchers for the full 3-year program period (“treatment”) for the “treated” African American students. These steps are carried out for other subgroups.

The results presented in this section are estimated using the non-parametric difference-in-difference local linear regression propensity-score matching strategy with exact matching on race and grade. I modify this estimator to additionally account for non-responses. Propensity-score matching requires a first-step estimation to estimate the conditional probability of receiving the treatment, and I estimate this probability using a generalized additive model, which is a semiparametric method. As a comparison, I also estimate the propen-

³⁹ “High-performing” public schools are those whose had test scores higher than the city-wide average in 1996 academic year.

sity score using a logit model.⁴⁰ I also compare the results with those estimated using alternative propensity-score matching strategies: nearest neighborhood matching (1, and 10 neighbors), caliper matching (0.1 and 0.01 tolerance rates), and kernel matching (using different bandwidths)⁴¹.

7.1 The Impact on Academic Achievement of Using the NYSCSP Vouchers for the Entire Duration of Program

Table 1 shows the estimated average impact on academic achievement of using vouchers to attend private schools for the entire 3-year program period. The definition of treatment changes with the impact of interest. For this impact, the treatment is defined as “using vouchers to attend private schools for the entire duration of the program.”

Table 1
The Impact on Test Scores of Using Vouchers for the Entire 3-Year Program Period

	N_1	N_0	Yr 3 Score	Mathematics	Yr 3 Score	Reading
Third follow-up Tests						
All students	489	667	27.37	-0.74	28.33	0.71
African Americans	206	274	25.7	3.86*	27.6	5.88*
Hispanics	225	331	27.78	-3.80*	28.09	-1.16

Students were tested at the baseline and in each of the follow-up years using the *Iowa Test of Basic Skills (ITBS)*, and achievement here is measured by the reading and mathematics test scores in the 3rd follow-up year. In this and all of the following tables, test scores and impacts are measured using the national percentile ranking (NPR). NPR, which is ranged from 1 to 99, indicates the relative position of a student with respect to other students in the same grade and who are tested at the same time of the year.

I keep the notations unchanged in the and all of the following tables: N_1 = the number of students in the treatment group; N_0 = the number of students in the potential comparison

⁴⁰In the program evaluation literature, parametric methods are often used to estimate propensity scores.

⁴¹These results are presented in appendix D and are available upon request from the author.

group from which the matched-comparison group is extracted; and * indicates statistical significance at the 5% level using a two-tailed test. Standard errors are computed using bootstrap methods, accounting for the dependences across students within the same family. The column “Yr 3 Score” lists the tests scores of the students in the treatment group; the “Mathematics (reading)” column reports the impact on the mathematics (reading) test score measured in the 3rd follow-up survey. The impact is measured using the national percentile ranking (NPR) scores, not absolute scores.

The results indicates that among the students who used the vouchers until the end of the program, only African American students show positive impact on both mathematics and reading scores. These children scored 3.86 NPR higher in mathematics and 5.9 NPR higher in reading had they not been offered the vouchers. Not only is this impact statistically significant, but also it is economically significant; the impact is equivalent to 20% and 27%, respectively, of the standard deviations of the baseline mathematics and reading test score distributions. This creates a puzzle. Why did African American students benefit from using the vouchers for the entire program period but Hispanic students did not? To understand better about the variation in treatment impact across racial/ethnic groups, I investigate the key differences between African American and Hispanic students that may contribute to the variation.⁴² The first variable being examined is “whether or not English is the main language used at home.”

⁴²Krueger and Zhu (2003) and Peterson and Howell (2003) examine the sensitivity of the impact of “offering vouchers” on test scores with respect to the definition of race. Earlier in this project, I used different definition of race and estimated the impact of using vouchers for various length of time. I did not find any substantial difference in the impact with respect to the definition of race adopted.

Table A1: Main Language at Home⁴³

	English	Non-English
African American	504 (97.11%)	15 (2.89%)
Hispanic	268 (42.07%)	369 (57.93%)

Table A2: Used Vouchers for the Duration of the Program

	English	Non-English
African American	191 (98.45%)	3 (1.55%)
Hispanic	100 (43.29%)	131 (56.71%)

Table A1 shows the number and proportion of African American and Hispanic students who mainly speak English (or non-English) at home, conditional on having test scores in both the baseline and the 3rd follow-up years. Table A2 further narrow the group to the students who used the vouchers until the end of the program. These figures indicate that more than half of the Hispanic students do not use English as their main language at home while this proportion is negligible for African American students.

7.2 The Impact on Hispanic Students' Achievement, Conditional on Main Language Spoken at Home

Using only Hispanic students who used vouchers for the entire duration of the program, I (re-)estimate the impact on test scores separately for those who mainly speak English at home and those who do not.⁴⁴

⁴³Only include students who have both test scores in the baseline and 3rd follow-up years.

⁴⁴First notice that there are slight differences in the number of observations (N_1) in this table and those in the lower-right corner of the above table (97 vs 100 and 123 vs 131). This is caused by the overlapping support requirement, which was discussed above in the "propensity scores" section.

Table 2
The Impact on Test Scores of Using Vouchers for the Entire 3-Year Program Period

<i>HISPANIC STUDENTS</i>						
	N_1	N_0	Yr 3 Score	Mathematics	Yr 3 Score	Reading
Third follow-up Tests						
Home Language is English	97	124	31.15	0.62	30.54	0.86
Home Language is not English	123	165	25.21	-6.74	26.20	-1.48

The most notable aspect of this table points to the high test scores of Hispanic students who were English-speaker at home. They scored an average of more than 30 NPR in both mathematics and reading, even several NPRs higher than those of African American students (see Table 1). They ranked almost *six* percentiles higher in mathematics and four percentiles higher in reading than the Hispanic students whose home languages were not English.

Another notable aspect is that the impacts on both of the test scores are positive for the home, English speaking Hispanic students whereas those for the home, non-English speaking Hispanic students are negative. None of these estimated impacts are statistically significant, however. Nevertheless, home language certainly serve as an important factor that explains the variation in impact among Hispanic students.

7.3 Test Score Improvement of the No-voucher Students

Another reason that Hispanic students did not benefit from using vouchers to attend private schools while African American students did is that the Hispanic students who did not receive vouchers had a much larger improvement in their test performances than did the African American students who did not have vouchers. The huge majority of these no-voucher group students stayed in the public schools. The Hispanic no-voucher group students scored 16.34 NPR in mathematics and 22.15 NPR in reading in the baseline year, and their scores rose to 29.35 NPR and 28.9 NPR, respectively, in the 3rd follow-up year. For the African American no-voucher group students, the mathematics test score went from 14.55 NPR at the baseline to 19.76 NPR in year 3 while the reading test score during this 3-year period dropped from 24.48 NPR to 22.5 NPR.

The academic performances of the no-voucher students provide an important indication

of what the voucher users would have performed had they not been awarded the vouchers, as the counterfactual outcomes are constructed using the no-voucher groups. Because of the dramatic increase in test scores of the Hispanic students who did not receive vouchers, the Hispanic students who did receive and used the vouchers had to show an even more dramatic improvement in test scores in order to experience positive impact on academic achievement. Unless the private schools attended by the Hispanic students provided a much more superior education than did the public schools where they would have stayed had they not received the vouchers, witnessing large and positive impact for the Hispanic students is not very likely.

This finding suggests a further analysis of the difference, if any, in quality of the public schools attended by the African American and Hispanic no-voucher group students. I analyze the following variables but do not find any major difference along each of these dimensions between the two groups: school size; class size; minority in child class; same race in child class; parent's self-reported "feeling" about child school, about whether or not the child's school sets high standard, and about whether or not child's school has high expectation on academic success; whether or not homework is assigned on a daily basis; difficulty of child homework; whether or not child has physical handicap; how well the school satisfies child needs; whether or not child need special help in learning English; and whether or not the child was suspended for discipline reasons. Only African American students showing positive impact from a voucher program is not unique to the NYSCSP program, but a satisfactory explanation to this "puzzle" is yet to be found. Evidence provided in this sub-section suggests that collecting information about quality of the schools (be that private or public) ever attended by *all* of the students in a voucher program is the first step towards solving the "puzzle".

7.4 Do the Performance of Public Schools Over Time Affect the Impact?

The empirical evidences presented so far demonstrate that the estimated impact varies across race and home language spoken at home and highlight that the no-voucher students' per-

formance may serve as an important explanation to the racial difference in program impact. So, I examine whether the temporal differential performance of public schools in different areas of the New York City school district is associated with the variation in impact on test scores. Students participated in the NYSCSP program lived in seven broadly defined areas in New York City at the time of application in Spring 1997. These areas are New York, Lower Bronx, Middle and Upper Bronx, Queens, Western Brooklyn, Upper and Eastern Brooklyn, and Staten Island/Yonkers/Western Middle-Bronx. In the data set, students are reported as living in one of the twelve areas. Due to very small number of observations in some areas, I consolidate them into seven areas based upon the similarity of the average baseline mathematics and reading test scores. The combined mathematics and reading score is called the composite score.

As shown in several places above, most of the no-voucher students stayed in the public schools, and more importantly, a huge majority of the students did not move residential locations during the program period, so residential mobility should not raise any serious concern. After redefining the areas, I rank them by the improvement in the composite test scores of the no-voucher students from 1997 to 2000. The 3-most-improved areas were the three areas in which the students had the largest improvement in composite test scores, and the 3-least-improved areas were defined similarly. Then, I re-estimate⁴⁵ the impact on the composite scores of the students in each of these two groups, using only students with both the baseline and 3rd follow-up test scores. Two estimations are run separately for African American and Hispanic students. The results are shown in table 3.

⁴⁵Similar to the procedure in estimating the impact separately by racial/ethnic groups or by the language group, the re-estimation of the impact on the composite test score of the students living in the 3-most-improved (or the 3-least-improved) areas requires re-matching the students in one of these two groups, re-computing the counterfactual test scores, and re-calculating the average impact. Because the entire matching procedure is re-implemented, the test scores of the matched-comparison group students can be served as the estimated counterfactual test scores of the treatment group students living the 3-most-improved (or the 3-least-improved) areas had they not received the vouchers.

Table 3
The Impact on Test Scores of Using Vouchers for the Entire Program Period

<i>3rd Follow-up Test Scores</i>			
	N_1	N_0	<i>Composite</i>
African Americans			
3 Most Improved Areas	90	98	0.32
3 Least Improved Areas	79	93	6.74*
Hispanic Students			
3 Most Improved Areas	101	140	-3.33
3 Least Improved Areas	64	89	-1.71

A clear pattern emerges: the impact on the composite test score for the full-voucher users⁴⁶ who lived in the 3-most improved areas is much smaller than that of the full-voucher users who lived in the 3-least improved areas. This holds for both African American and Hispanic students. This highlights an important point: When families did not relocate to another school district, they would have sent their children to the public schools where they originally attended had they not used the vouchers to opt out from the public schools. Therefore, if the public schools drastically improve over time, then the impact of attending private schools would not be very large. In contrast, if the public schools only have a slight improvement over time, then students may greatly benefit from switching to private schools.

In fact, when evaluating the impact on achievement of private school voucher programs (especially the impact of using vouchers to attend private schools), we are considering one treatment (staying in public schools) relative to another treatment (attending private schools), rather than receiving a treatment relative to no treatment. Because qualities of the public and the private schools may change over time, it is not necessarily that all students switching to private schools are uniformly benefitted (at least in terms of the impact on test scores) from doing so. It is also important to keep in mind that the quality of the private schools attended by the students matter a great deal. In the NYSCSP program, were the private schools that the participating parents could afford. Most of these were Catholic

⁴⁶For expositional simplicity, I call the students who used vouchers to attend private schools for the entire duration of the program the “full-voucher users.”

schools. The average tuition were \$2100 in year 2000. These schools did not belong to the elite league which people have in mind when they think of private schools. As the public schools also improve over time⁴⁷, the impact on academic achievement of using vouchers for private schooling may not be as large as we expect.

7.5 Do Initial Cognitive Skills Affect the Impact?

Besides the main language spoken at home and the public school improvement over time, initial cognitive skills measured by test scores before the program began may also be an important determinat of program impact. So, I also estimate the impact separately for those who scored above the mean and those who scored below. Table 4 presents these estimates.

Table 4
The Impact on Test Scores of Using Vouchers for the Entire Program Period

<i>3rd Follow-up Test Scores</i>				
	N_1	N_0	<i>Mathematics</i>	<i>Reading</i>
All Students				
Baseline Scores Above the Mean	196	215	-3.99*	-0.24
Baseline Scores Below the Mean	273	355	3.07*	0.68

7.6 The Variation of Impact Along Other Dimensions

I also examine if the impact varies by (i) mother’s education, (ii) the number of school-age children living in the same family, (iii) whether or not the students initially had attended a “good school” before the program began⁴⁸. No variation in impact along these dimensions is detected.

⁴⁷This indeed holds true in New York City from 1999 to 2004. I have also come across some evidences that indicate similar patterns of improvement from 1997 to 1999 but do not have them at the time being. I am in the progress of requesting these evidence from the New York City Department of Education.

⁴⁸“Good schools” are refered to the public schools in which their students’ test scores were higher than the city-wide average.

7.7 The Impact on Achievement for the Voucher Recipients who Attended Private Schools for Less Than the 3-Year Program Period

After studying the impact of using vouchers for the entire program period, I next analyze the impact of using the vouchers to attend private schools for strictly less than 3 years. No statistically significant impact is found except that on the mathematics test score of African American.

Table 5
The Impact on Test Scores of Using Vouchers for those Attended Private Schools for Less Than the 3-Year Program Period

	N ₁	N ₀	Yr 3 Score	Mathematics	Yr 3 Score	Reading
Third follow-up Tests						
All students	200	570	26.8	2.59	24.87	-1.05
African Americans	93	223	25.4	5.47*	24.5	2.73
Hispanics	96	279	27.56	2.19	24.5	-1.40

7.8 Dynamics of the Impact Over the Duration of the Program

Table 6 presents the impacts on the 1st, 2nd, and 3rd follow-up test scores of the students who used vouchers for entire duration of the program. The key pattern found for the 3rd follow-up year already emerges in the 1st follow-up year: no impact is detected on students overall or Hispanic students; only African Americans students benefited from using the vouchers to attend private schools.

Table 6
The Impact on Test Scores of Using Vouchers for the Entire 3-Year Program Period

	N_1	N_0	Mathematics	Reading
First follow-up Tests				
All students	471	574	0.84	1.90
African Americans	192	228	3.72*	5.61*
Hispanics	221	291	0.03	-1.21
Second follow-up Tests				
All students	472	528	-1.45	2.94*
African Americans	199	204	2.40	5.81*
Hispanics	218	273	-4.17*	2.21
Third follow-up Tests				
All students	489	667	-0.74	0.71
African Americans	206	274	3.86*	5.88*
Hispanics	225	331	-3.80*	-1.16

Although African American students already exhibited large, positive impact, which is similar in magnitude to that in the 3rd follow-up year, in the first year of the program, one cannot conclude that the program should then be shortened to one year. The reason is that the impact on year 3 test scores of African American students (as well as students in other racial/ethnic groups) had the program lasted only one year is a counterfactual that is not being estimated. It could be that the year 3 test score dropped had the program run only for one year. Therefore, what I can conclude is that the impact, be that positive or negative, on test scores stays flat from the 1st follow-up year to the 3rd follow-up year.

8 Conclusion and Discussion

In this study, I evaluate the New York School Choice Scholarship Program, one of the best known private school voucher experiments ever implemented in the U.S. I analyze the impact on academic achievement of using vouchers to attend private schools for various length of time, examine the evolution of the program impact, and investigate the variation of the treatment impact along several dimensions including whether or not English is the main language used at home by Hispanic students, the differential performance of public schools in different school districts over time, child's initial cognitive skills measured by the test scores before the program began, mother's education, number of school-age children living

in the same family, and whether or not a student initially attended a low-performing school.

Although the average impact of offering school vouchers (ITT effect) provides an important one-number summary of the program a voucher experiment, knowledge from the ITT effect alone is not enough to answer many important policy questions unless the experiment is “ideal” in the sense that a huge majority of the treatment recipients actually take up the treatment. Information about the impact of using vouchers for various length of time and the factors that are associated with the variation of the impact is needed in order to design a voucher program that may achieve a higher impact in the future.

Estimating these impacts require an alternative (to the conventional instrumental-variable approach) estimation strategy because the IV-approach may not be able to identify these impacts using the randomization of vouchers as the only instrument variable, which is the only valid instrument in the NYSCSP data set. For this reason, I employ the nonparametric, difference-in-difference propensity-score matching estimator and modify it to additionally account for selective non-response, a problem that plagues this as well as the school voucher experiments in Dayton, Ohio and Washington, D.C.

The results found in this study indicate that only African American students who used vouchers for the entire duration of the program as well as those who quited prematurely experience statistically significant and positive impact; other racial groups, especially Hispanic students, do not. This is due to the fact that the Hispanic students who did not receive vouchers and stayed in public schools showed a very large improvement in their national percentile ranking scores over the 3-year program period. Their African American counterparts, on the other hand, did not. I also find that whether or not English is the home language may explain the variation in impact among Hispanic students. Furthermore, I examine other factors that are associated with the variation of the program impact and find that public school improvement over time and initial cognitive skills measured by baseline test scores are also important determinants of the program impact.

Evidence from this study clearly shows that this particular school voucher program in New York City was able to help some (but certainly not all) students to improve their

academic performance. If a school voucher program of similar scale (in terms of the number of vouchers awarded, the amount of each voucher, and the number of years that a voucher can be used) will be implemented in the future, a targeted school voucher program that targets at those students who attend schools that have slow improvement and those students who have very low academic performance may be able to help even more students from disadvantaged families to raise their achievement.

8.1 Generalization of Findings:

8.1.1 The Effects of A Voucher Program on Public Schools

It is important to realize that a voucher program could potentially affects both the choice students as well as the no-choice students who remain in the public schools unless the size of the program is negligible comparing to the total student population in a particular set of school districts. A publicly-funded voucher program can affect the academic performance of no-choice students in two ways: (1) With some of the school budget allocated to the voucher program, the resources of the existing public schools will be reduced (although spending per pupil may not), and (2) if students' academic performance are affected by their peers in the same class or in the same school, and school vouchers may induce a change in the student composition of the existing public schools.

The NYSCSP awarded a voucher to only slightly more than 1,300 students; thus, its scale was very small relative to the total number of the students in public schools in New York City. Moreover, the program was privately funded by the School Choice Scholarship Foundation, so the implementation of the program did not alter New York City's school budget. For these two reasons, little effect should be expected on the no-choice students as a whole in the existing public schools.

8.1.2 A Large-Scale and Permanent Voucher Program

It should be noted that the results in the current study cannot be generalized to any large-scale, permanent private school voucher programs. Such a program requires consideration of various general equilibrium effects⁴⁹. For such a program, the group of interest does not confine to the group of students who receive vouchers or who actually use the vouchers. The impact of other groups of students are also important: the students without vouchers because they would most likely stay in the public schools where the composition of the students is greatly altered by the flight of the voucher-group students; and the students in the private education sector where the quality of students is also changed due to the entry of the students from public schools.

8.2 An Unintended Consequence of Private School Vouchers

School voucher program is designed to provide parents with the opportunities to choose schools for their children and hopefully to help improving the children's academic performance. Evidence in the NYSCSP data suggests an unintended consequence of the private school voucher program. The following three tables show the descriptive statistics of the grade level the students attended in Fall 1997 when the voucher program began (horizontal) and the grade at which the students should have attended three years later (vertical). The diagonal boxes highlighted with color represent the number of students who progressed normally. The boxes in the upper triangle show the numbers of students who were pulled back by at least one grade. The percentage of students who were pulled back (i.e. repeating a grade) is shown at the bottom of each of these tables. Clearly, the group with vouchers have a much higher percentage of repetition than those without the vouchers.

An increase in the probability of grade retention. Students who were offered vouchers were much more likely to be held back by at least one grade relative to those students

⁴⁹There is a growing theoretical literature of sorting and school choice. For example, see the series papers by Thomas Nechyba.

without vouchers. In fact, a logistic regression⁵⁰ on grade retention shows that being offered a voucher increases the probability of repeating a grade by 7.3 percentage points. Therefore, when evaluating the achievement impact of a school voucher program, any consequences of the program that is closely related to achievement impact should not be ignored.

Table A2A: Cohort (All Students with parent participated in the 3rd year)

3rd Year Grade	1	2	3	4	5	
3	24	1				25
4	212	43	4			259
5	1	194	44			239
6		7	246	47	3	303
7		2	8	276	46	332
8				3	234	237
9					5	5
DK	1		1		1	3
	223	248	290	309	266	1336
% Repeated	0.11	0.18	0.17	0.15	0.18	

Table A2B: Cohort (Voucher Group)

3rd Year Grade	1	2	3	4	5	
3	15	1				16
4	99	27	4			130
5	1	94	26			121
6		3	118	26	2	149
7		2	3	135	33	173
8				2	1	118
9					3	3
DK	1		1			2
	116	127	151	163	155	712
% Repeated	0.13	0.22	0.20	0.16	0.23	

Table A2C: Cohort (Non-Voucher Group)

3rd Year Grade	1	2	3	4	5	
3	9					9
4	113	16				129
5		100	18			118
6		4	128	21	1	154
7			5	141	13	159
8				1	118	119
9					2	2
DK			1			1
	122	120	152	163	134	691
% Repeated	0.07	0.13	0.12	0.13	0.10	

⁵⁰The point here is to look at some preliminary evidence of the association between receiving a voucher and grade retention. It is not intended to uncover the preference parameters of a model in which parents choose whether or not to send their children to private schools and whether or not to pull back their children by one grade. I will explore this unintended consequence of private school vouchers in my future work.

9 References

1. Angrist, J., Imbens, H., and Rubin, D. (1997). “Identification of Causal Effects Using Instrumental Variables.” *Journal of America Statistical Association* 91: 444-455.
2. Angrist, J., and Imbens, H. (1994). “Local Average Treatment Effect.” *Econometrica* 62(2): 467-475.
3. Barnard, J., Frangakis, C., Hill, J., and Rubin, D. (2003). “Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City.” *Journal of America Statistical Association* 98: 299-314.
4. Cohen-Zada, D., and Justman, M. (2002). “The Religious Factor in Private Education.” *working paper*.
5. Cohen-Zada, D., and Justman, M. (2003). “The Political Economy of School Choice: Linking Theory and Evidence.” *J of Urban Economics*. 54(2), 277-308.
6. Dehejia, R., and Wahba, S. (2002). “Propensity Score Matching Methods for Nonexperimental Causal Studies.” *Review of Economics and Statistics* 84(1), 151-161.
7. Ferreyra, M. (2003). “Estimating the Effects of Private School Vouchers in Multi-District Economies.” *working paper*. Carnegie Mellon University.
8. Gill, B., Timpane, P., Ross, K., and Brewer, D. (2001). *Rhetoric Versus Reality: What We Know and What We Need to Know About Vouchers and Charter Schools*. Santa Monica, CA: RAND Corp.
9. Greene, J., Peterson, P., and Du, J. (1997). “The Effectiveness of School Choice: The Milwaukee Experiment.” *Occasional Paper 97-1*. Harvard University Education Policy and Governance.
10. Heckman, J., Ichimura, H., and Todd, P. (1997). “Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme.” *Review*

- of Economic Studies.* 64(4), 605-654.
11. Heckman, J., Ichimura, H., and Todd, P. (1998). "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies.* 65(2), 261-294.
 12. Howell, W., Wolf, P., Campbell, D., and Peterson, P. (2002a). "School Vouchers and Academic Performance: Results from Three Randomized Field Trials." *J of Policy Analysis and Management.* 21(2), 191-217.
 13. Howell, W., Peterson, P., Wolf, P., and Campbell, D. (2002b). *The Education Gap: Vouchers and Urban Schools.* Washington, DC: Brookings Institute Press.
 14. Howell, W. (2004). "Dynamic Selection Effects in Mean-Tested, Urban School Voucher Programs." *J of Policy Analysis And Management.* 23(2): 225-250.
 15. Krueger, A. and Zhu, P. (2003). "Another Look at the New York City School Voucher Experiment." Princeton University. mimeo.
 16. Mayer, D., Peterson, P., Myers, D., Tuttle, C., and Howell, W. (2002). "School Choice in New York City After Three Years: An Evaluation of the School Choice Scholarships Program." Washington, DC: Mathematica Policy Research, Inc.
 17. Metcalf, (1999). "Evaluation of the Cleveland Scholarship and Tutoring Grant Program: 1996-1999." Bloomington, IN: The Indiana Center for Evaluation.
 18. McEwan, P. (2000). "The Potential Impact of Large-Scale Voucher Programs." *Review of Educational Research.* 70, 103-149.
 19. Peterson, P., and Howell, W. (2003). "Efficiency, Bias, and Classification Schemes: Estimating Private-School Impacts on Test Scores in the New York City Voucher Experiment." *The American Behavioral Scientist.* (forthcoming).
 20. Peterson, P., and Howell, W., Greene, J. (1999). "An Evaluation of the Cleveland Voucher Program After Two Years." Cambridge, MA: Program on Education Policy and Governance, Harvard University.

21. Rosenbaum, P., and Rubin, D. (1983). "The Central role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika*. 70, 41-55.
22. Rouse, Cecilia. (1998). "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *Quarterly Journal of Economics* 113: 553-602.
23. Witte, J., Sterr, T., and Thorn, C. (1995). "Fifth Year Report: Milwaukee Parental Choice Program." University of Wisconsin-Madison.

10 Appendices:

10.1 Appendix A: A Dynamic Model of School Choice

I develop a simple dynamic model of school choice to capture the parents' decision on choosing the type of schools (public and private) for their children in each of the children's schooling year before attending college and to provide a guidance for the variables used in the matching procedure.

Consider an economy that lasts for T periods. The economy begins ($t = 1$) with a set of households, each of which consists of a mother, a father and one child, either male or female. Parents' obtain utility from a joint consumption good and the quality of their child. The quality of the child is measured by the child's schooling achievement and his/her religious affiliation. In each period t , the parents decide how much to consume and whether or not to send their child to a private school or a public school.

Each pair of parents belong to one of the two religion types, $k \in \{1, 2\}$, where 1 being "having a religion" and 2 means "not having a religion"⁵¹. The child is born with some inherited cognitive ability, μ_A , which cannot be altered through after-birth training but is likely to affect the child's cognitive achievement later in his/her life. The child also develops some cognitive skills, μ_{ps} , during the preschool years (i.e. from birth to kindergarten). Each child begins school at grade one and finish at grade twelve. Grade retention or promotion are not allowed, so there are 12 decision periods, one corresponding to each grade: $t \in \left\{ \underbrace{1}_{age=6}, 2, \dots, \underbrace{12}_{age=17} \right\}$ where $t = 1$ is the period when the child attends the first year of school (i.e. the First grade) and $t = 12$ is the last school year for the child (i.e. the Twelfth grade).

In each period, the parents receive an earning endowment, Y_t . A constant fraction, θ , of the family earning is spent on the child's learning every period. The remaining fraction, $(1 - \theta)$, is spent on the non-storable consumption good C_t . Because the earning

⁵¹Religious affiliation has been considered to be an important determinant of school choice. See Ferreyra (2003), Cohen-Zada and Justman (2003), and Howell (2004).

endowment differ across families, families invest different amount in child's cognitive capital even though they spend the same fraction of their earnings for this purpose. Before making their first school choice decision for their child, the parents are also endowed with a private education voucher that can be redeemed up to some predetermined, temporal-constant maximum value, v , at a private school of their choice for $s \leq T$ periods⁵². Let $I_v(t) = 1 \{\text{a voucher can be used in period } t\}$, where $1 \{.\}$ is an indicator equal to one if its argument is true, and v be the time-constant maximum value of the school voucher.

While public schools are free, private schools charge an annual tuition and some other costs for items (such as books, school uniform, laboratory fee, and so on). The total cost is denoted by p .

Besides deciding how much to consume, parents make one of the two mutually exclusively school choices for the child in each period⁵³: send the child to (1) a private school, or (2) a public school⁵⁴. The former decision is denoted as $d_1(t) = 1$ while the latter $d_2(t) = 1$. At the beginning of each school year, the child does not belong to any religion, regardless of her religious affiliation in the previous period⁵⁵. She, however, could become religious in that period. The probability that the child becomes religious at a particular age depends only on the religion affiliation of her parents and that of the school in which she attends. In particular, these probabilities are summarized in the following table:

⁵²In the case of NYSCSP, a voucher can be used for three academic years from Fall 1997 to Spring 2000.

⁵³Assume that home schooling is not an option.

⁵⁴I do not distinguish private schools that have religious affiliation and those that have not. I also simply consider charter schools as private schools and Magnet Schools as public schools. Most of the students who received NYSCSP vouchers and chose private education attended Catholic schools.

⁵⁵This assumption greatly simplifies the analysis a great deal. Without it, the religion orientation of the parents, the child, the school, and the entire history of the probability of adopting a particular religion will affect the child's religion in the current period.

Child's probability of becoming religious in a particular age t		
Parents' type	School Type	
	Private	Public
Religious ($k = 1$)	$r_1 = 1$	$0 < r_2$
Not Religious ($k = 2$)	$0 < r_3$	$r_4 = 0$

where $1 = r_1 > r_2 > r_3 > r_4 = 0$. The probability that the child who has religious parents becomes religious is $r_1 = 1$ if she is sent to a private school and is $r_2 < 1$ if she is sent to a public school. If parents are not religious, then the probability that the child becomes religious is r_3 if she is sent to a private school and is $r_4 = 0$ if she attends a public school. The restriction that r_2 takes a larger value than r_3 is made to capture the idea that parents have more influence on their children's religion preferences than schools do.

Parents derive (per-period) utility from consumption, C_t , the quality of the child measured by her test scores at age t , TS_t , and whether or not the child become religious, I_R . Race (rc) – African Americans, Hispanics, or other racial ethnic groups – and the gender of the child are also included in the utility function because different racial ethnic groups may value the religious orientation of their child differently and parents may value the test scores of boys and girls differently. Parents' control variables in period t are $\{d_m(t)\}_{m=1}^{m=2}$ and C_t . So, type- k parents' expected per-period utility (or reward) function in period t is given by

$$EU_k(C_t, TS_t, I_R(t), rc, g; \alpha_m) \tag{17}$$

where the utility function is parameterized by the vector of parameters α_m , and the expectation is taken over the probability the child becomes religious. For each religious type, the utility can be specified as follows:

Type 1 (religious) parents:

If a private school is chosen:

$$\begin{aligned}
& EU_1(C_t, TS_t, I_R(t), rc, g; \alpha_m, d_1(t) = 1) & (18) \\
= & r_1 U_1(C_t, TS_t, I_R(t) = 1, rc, g; \alpha_m, d_1(t) = 1) + (1 - r_1) U_1(C_t, TS_t, I_R(t) = 0, rc, g; \alpha_m, d_1(t) = 1) \\
= & U_1(C_t, TS_t, I_R(t) = 1, rc, g; \alpha_m, d_1(t) = 1)
\end{aligned}$$

because $r_1 = 1$.

If a public school is chosen:

$$\begin{aligned}
& EU_1(C_t, TS_t, I_R(t), rc, g; \alpha_m, d_2(t) = 1) & (19) \\
= & r_2 U_1(C_t, TS_t, I_R(t) = 1, rc, g; \alpha_m, d_2(t) = 1) + (1 - r_2) U_1(C_t, T_t, I_R(t) = 0, rc, g; \alpha_m, d_2(t) = 1)
\end{aligned}$$

Type 2 (non-religious) parents:

If a private school is chosen:

$$\begin{aligned}
& EU_2(C_t, TS_t, I_R(t), rc, g; \alpha_m, d_1(t) = 1) & (20) \\
= & r_3 U_2(C_t, TS_t, I_R(t) = 1, rc, g; \alpha_m, d_1(t) = 1) + (1 - r_3) U_2(C_t, TS_t, I_R(t) = 0, rc, g; \alpha_m, d_1(t) = 1)
\end{aligned}$$

If a public school is chosen:

$$\begin{aligned}
& EU_2(C_t, TS_t, I_R(t), rc, g; \alpha_m, d_2(t) = 1) & (21) \\
= & r_4 U_2(C_t, TS_t, I_R(t) = 1, rc, g; \alpha_m, d_2(t) = 1) + (1 - r_4) U_2(C_t, TS_t, I_R(t) = 0, rc, g; \alpha_m, d_2(t) = 1) \\
= & U_2(C_t, TS_t, I_R(t) = 0, rc, g; \alpha_m, d_2(t) = 1)
\end{aligned}$$

When choosing between public and private schools, parents have to predict the current period test score of their child. It is assumed that the parents do not have the knowledge of the “true” technological relationship that connects cognitive achievement to different educational inputs and their child’s endowed cognitive ability. Instead, they predict their child’s current period test scores, TS_t , using current family resources devoted to the child’s

learning activities, θY_t , whether their child attends a private school, $d_m(t)$, and the child's test score in the previous year, TS_{t-1} :

$$TS_t = f(\theta Y_t, d_m(t), TS_{t-1}; \beta) + \varepsilon_{mt} \quad (22)$$

A shock to the test score, ε_{mt} , is realized after making the school choice and schooling input decisions.

The per-period budget constraint (assuming no saving or borrowing) is:

$$C_t + \theta Y_t + d_1(t) * p = Y_t + d_1(t) * (I_v(t) * v) \quad (23)$$

or equivalently,

$$C_t = (1 - \theta) Y_t + d_1(t) * (I_v(t) * v - p) \quad (24)$$

Parents' optimization problem:

In each period t , the parents decide whether to send their child to a private school or a public school. The optimization problem of type- k parent in period t can be expressed as

$$\begin{aligned} & V_k(d_m(t), t \mid S_t) \quad (25) \\ & = \underset{\{d_1(t), d_2(t)\}}{Max} \{U_k(C_t, TS_t, I_R(t), rc, g; \alpha_m) + \beta E_{t+1} [V(d_m(t+1), t+1 \mid S_{t+1})]\} \end{aligned}$$

subject to

$$T_t = f(\theta Y_t, d_m(t), TS_{t-1}; \beta) + \varepsilon_{mt} \quad (27)$$

⁵⁷This is a very parsimonious production function of cognitive skill. Todd and Wolpin (2003,2004) propose that a very general framework to model the current cognitive outcome as a cumulative technological process that depends on the entire history of school inputs, the entire history of family inputs, and the inherited ability of the child.

To put it symbolically, the inputs of the education production function include the child's endowed cognitive ability, μ_A ; the cognitive skills developed during the preschool period, μ_{ps} ; the entire history of family inputs up to the last period (i.e. $t-1$), X_F^{t-1} ; current family input, $X_{F,t}$; the entire history of school inputs up to the

$$C_t = (1 - \theta) Y_t + d_1(t) * (I_v(t) * v - p) \quad (28)$$

where

$V_k(d_m(t), t | S_t)$ denotes the value function at period t ; S_t is the state space in period t ; $U_k(\cdot)$ is type- k parents' period utility function; k = parents' religious orientation. It takes value one if parents have some religious affiliation and two otherwise; C_t = the composite consumption good purchased when the child is at age t ; T_t = a measure of cognitive achievement of the child at age t . Math and Reading test scores are used; Y_t = family income in period t ; $I_R(t) = 1$ if the child becomes religious in period t ; $rc = \text{race} = 1$ if African Americans, $= 2$ if Hispanics, and $= 3$ if other racial/ethnic groups; g = child's gender; β is a discount factor; $f(\cdot)$ is a production function of cognitive achievement; θ = a fraction of the family income that is spent on activities that help the child accumulate cognitive skills; ε_{mt} = school type-specific shock to the production function of cognitive

last period, $X_S^{t-1}(\{d_m(t)\}_{a_1}^{t-1})$; the current school inputs, $X_{S,t}(d_m(t))$; and the measurement error of test scores, ε_{mt} , when school choice m is chosen. The reason of writing X_S^{t-1} as $X_S^{t-1}(\{d_m(t)\}_{a_1}^{t-1})$ and $X_{S,t}$ as $X_{S,t}(d_m(t))$ is to emphasize the fact that families can choose school inputs only through school choice. In other words, not all feasible combinations of school inputs can be purchased in the market of education. This formulation is to capture the idea that school inputs can only be purchased as bundles. For example, a family may not be able to choose the following set of school inputs for its first grade child: (i) a specialist for each subject; (ii) all teachers have at least 15 years of teaching experiences; (iii) the teacher-pupil ratio is 1-to-8; (iv) a laptop computer is provided to each student; and so on. Although parents can always augment the education inputs by hiring private tutors or sending their children to before- and after-school enrichment programs that are academically based, these types of inputs are considered family inputs. With the inputs specified, the education production function can be written as

$$T_t = f\left(X_{F,t}, X_F^{t-1}, X_{S,t}(d_m(t)), X_S^{t-1}(\{d_m(t)\}_{a_1}^{t-1}), \mu_{ps}, \mu_A; \beta\right) + \varepsilon_{mt} \quad (27)$$

where β represents a time-invariant vector of parameters that characterizes the education technology $f(\cdot)$, and $(\varepsilon_{1t}, \varepsilon_{2t})$ are alternative (private vs public schools) specific shocks to the production function. They are assumed to be jointly serially uncorrelated.

achievement if school type $m \in \{1, 2\}$ is chosen in period t and is realized after the school choice decision is made; $I_v(t) = 1$ if a voucher is received in period t ; v = the annual maximum value of the school voucher; a NYSCSP voucher was worth up to \$1400 per year and could be used for a maximum of three years; p = total cost associated with private schooling, which also depends on the residential location of the family. If there are some good private schools in the neighborhood, then the transportation cost is lower. Otherwise, the family may have to incur a higher transportation cost to send its children to a good private school that is further away from where they live. This cost is assumed to be constant over time.

The state space in each period t consists of the following components: income, last year test score, the time period, the fraction of family income spent on child's learning activities; the total cost of private schooling; the availability of a school voucher; the value of the school voucher; religious affiliation of the parents; own race; and the gender of the child.

$$\begin{aligned}
 S_t &= \left\{ \underbrace{(y_t, TS_{t-1}, t)}_{\Omega_t}, \underbrace{\left(\theta, p, I_v(t), v, \underbrace{(r_1, r_2, r_3, r_4)}_r, k, rc, g \right)}_{\Xi} \right\} \\
 &= \{ \Omega_t, \Xi \}
 \end{aligned} \tag{29}$$

Given the state space in period t and the fact that each period's income is either spent on the joint consumption good or child's learning activities, parents' objective is to maximize the expected present discount value of the remaining life-time⁵⁸ utility by choosing one of the two (private or public) school types.

To understand what variables should enter the model of the conditional probability of being offered a voucher and using it to attend a private school for the entire three-year program period, conceptualize how the model is solved. As the problem is one of finite horizon, the model is solved backward recursively starting at the last period.

⁵⁸Recall that the decision horizon has only 12 periods. By remaining lifetime utility, I mean the utility in the remaining decision periods.

The value function of type- k parents in the last period, $T = 12$, is simply

$$V_k(d_m(T), T | S_T) = \max \{U_{k1}(C_T, TS_T, I_R(t), rc, g; \alpha_1), U_{k2}(C_T, TS_T, I_R(t), rc, g; \alpha_2)\} \quad (30)$$

where U_{km} stands for the current utility when the option m is chosen, $m = 1, 2$, and

$$T_T = f(\theta Y_T, d_m(t), TS_{T-1}; \beta) + \varepsilon_{mT} \quad (31)$$

$$C_T = (1 - \theta) Y_T + d_1(T) * (I_v(T) * v + s_T - p) \quad (32)$$

That is, $V_{kT}^* \equiv V_k(d_m(T), T | S_T) =$

$$\begin{aligned} & \max \left\{ \begin{array}{l} U_{k1}((1 - \theta) Y_T + (I_v(T) * v + s_T - p), f(\theta Y_T, d_1(t) = 1, TS_{T-1}; \beta) + \varepsilon_{1T}, I_R(t), rc, g; \alpha_1) \\ U_{k1}((1 - \theta) Y_T, f(\theta Y_T, d_2(t) = 1, TS_{T-1}; \beta) + \varepsilon_{2T}, I_R(t), rc, g; \alpha_2) \end{array} \right\} \\ = & \max \{V_{k1}(S_T), V_{k2}(S_T)\} \end{aligned} \quad (33)$$

where S_T denotes the state space in period T .

In the second to the last period, $T - 1$, the value function becomes

$$\begin{aligned} & V_{k,T-1}^* \\ = & \max \{U_{k1}(S_{T-1}; \alpha_1) + \beta V_{kT}^*, U_{k2}(S_{T-1}; \alpha_1) + \beta V_{kT}^*\} \\ = & \max \{V_{k1}(S_{T-1}), V_{k2}(S_{T-1})\} \end{aligned} \quad (34)$$

Then, in any period $t < T$, the value function is

$$\begin{aligned} & V_{k,t}^* \\ = & \max \{U_{k1}(S_t; \alpha_1) + \beta V_{kt+1}^*, U_{k2}(S_t; \alpha_1) + \beta V_{kt+1}^*\} \\ = & \max \{V_{k1}(S_t), V_{k2}(S_t)\} \end{aligned} \quad (35)$$

In any period t , a family would choose school type m for its children if and only if $V_{k,t}^* (d_m(t) = 1) \geq V_{k,t}^* (d_{m'}(t) = 1)$ where $m \neq m'$. Then, a type k family sending its child to a private school in period t must have their value functions satisfied the following condition:

$$\begin{aligned} & \{V_{k,t}^* (d_1(t) = 1) \geq V_{k,t}^* (d_2(t) = 1)\} \\ \equiv & \{V_{k,t,m=1}^* \geq V_{k,t,m=2}^*\} \end{aligned} \quad (36)$$

and the probability of this event is

$$\Pr \{V_{k,t,1}^* \geq V_{k,t,2}^*\} \quad (37)$$

This probability depends on the following set of variables in the period t state space in that period:

$$\begin{aligned} \Theta_t &= \{\Omega_t, \Xi, (I_R(t), d_m(t))\} \\ &= \left\{ \underbrace{(y_t, T_{t-1}, t)}_{\Omega_t}, \underbrace{(\theta, p, I_v(t), v, \underline{r}, k, rc, g)}_{\Xi}, d_m(t) \right\} \end{aligned} \quad (38)$$

Likewise, the probability that a family sending its child to private school for three consecutive years takes the form:

$$\Pr \{V_{k,t,1}^* \geq V_{k,t,2}^*, V_{k,t+1,1}^* \geq V_{k,t+1,2}^*, V_{k,t+2,1}^* \geq V_{k,t+2,2}^*\} \quad (39)$$

and this probability depends on the union of the following three sets of variables.

$$\Theta_t \equiv \left\{ (y_t, T_{t-1}, t), \underbrace{(\theta, p, I_v(t), v, \underline{r}, k, rc, g)}_{\Xi} \right\} \quad (40)$$

$$\Theta_{t+1} \equiv \{(y_{t+1}, T_t, t+1), \Xi\} \quad (41)$$

$$\Theta_{t+2} \equiv \{(y_{t+2}, T_{t+1}, t+2), \Xi\} \quad (42)$$

Therefore,

$$\begin{aligned} & \Pr \{V_{k,t,1}^* \geq V_{k,t,2}^*, V_{k,t+1,1}^* \geq V_{k,t+1,2}^*, V_{k,t+2,1}^* \geq V_{k,t+2,2}^*\} \\ &= \Pr \{\Theta_t \cup \Theta_{t+1} \cup \Theta_{t+2}\} 1 \end{aligned} \tag{43}$$

where

$$\begin{aligned} & \Theta_t \cup \Theta_{t+1} \cup \Theta_{t+2} \\ &= \{\{y_t, y_{t+1}, y_{t+2}\}, \{T_{t-1}, T_t, T_{t+1}\}, \{t, t+1, t+2\}, \Xi\} \end{aligned}$$