Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia

By JOSHUA ANGRIST, ERIC BETTINGER, AND MICHAEL KREMER*

Demand-side subsidies for education are increasingly common in developing countries. Chile and Colombia have both offered educational vouchers for private secondary schools, while Brazil, India, Israel, and Mexico have introduced student stipends that reward attendance and performance. Interest in demand-side subsidies in developing countries parallels interest in the United States, where publicly funded vouchers for private schools have been distributed in a number of cities.

Previous research on primary and secondary school vouchers typically focuses on the short-run effects of vouchers on test scores. The results so far suggest that vouchers benefit some groups of recipients, though the extent of test-score gains is disputed.[1] Missing from most studies of voucher effects is an assessment of impacts on longer-term outcomes—such as high-school graduation rates—that are more clearly tied to economic success.

This paper examines the longer-run effects of Colombia's PACES program, one of the largest voucher initiatives ever implemented.[2] Between 1991 and 1997, PACES awarded nearly 125,000 vouchers to low-income high-school students. Since vouchers were renewable annually conditional on satisfactory academic progress as indicated by scheduled grade promotion, the program provided incentives for students to work harder as well as widening their schooling options. PACES vouchers may therefore have effects similar to merit-based college scholarships and test-based achievement awards (see, e.g., Angrist and Victor Lavy, 2002; Kremer et al., 2004; Susan Dynarski, 2002).

In Bogotá as well as a number of other large cities, PACES vouchers were awarded by lottery. The random assignment of vouchers facilitates a natural-experiment research design in which losers provide a comparison group for winners. Our previous research (Angrist et al., 2002) used the voucher lotteries to show that in the three years after random assignment, PACES winners completed more years of school and had lower grade repetition, higher test scores, and a lower probability of working than did losers.

This paper examines the impact of winning the voucher lottery on outcomes seven years later. In particular, we use administrative data from Colombia's centralized college entrance examinations, the ICFES, to obtain information on high-school graduation status and academic achievement.[3] ICFES registration is a good proxy for high-school graduation, since 90 percent of all graduating high-school seniors take the exam (World Bank, 1993). The principle advantage of administrative records in this context, in addition to providing longer-term

* Angrist: Department of Economics, Massachusetts Institute of Technology, E52-353, 50 Memorial Drive, Cambridge, MA 02142 (e-mail: angrist@mit.edu); Bettinger: Department of Economics, Case Western Reserve University, 11119 Bellflower Road, Cleveland, OH 44106 (e-mail: bettinger@cwru.edu); Kremer: Department of Economics, Harvard University, 207 Littauer Center, Cambridge, MA 02138 (e-mail: mkremer@fas.harvard.edu). Special thanks goes to Cristina Estrada, Claudia Gonzalez, Marcela Monsalvo, and Ana Gomez for research assistance. We are also grateful to Jorge Estrada for help interpreting Colombian ID numbers and to the staff at ICFES for providing data. We thank the National Institutes of Health and the World Bank for funding this research, and Victor Chernozhukov, Bernd Fitzenberger, and seminar participants at the NBER Summer Institute, Universitat Pompeu Fabra, and the ZEW Evaluation Workshop for helpful discussions and comments.

[1] See, e.g., Cecilia Elena Rouse (1998), William G. Howell and Paul E. Peterson (2002), Angrist et al. (2002), and Alan B. Krueger and Pei Zhu (2003).

[2] PACES is an acronym for Programa de Amplicación de Cobertura de la Educación Secundaria.

[3] ICFES is an acronym for Colombia's college admissions testing service, the Instituto Colombiano Para El Fomento de la Educación Superior.

outcomes, is that our measure of high-school graduation status suffers no loss to follow-up and that ICFES test-score data are much less expensive to obtain than survey data or scores from a specialized testing program.

Our analysis shows that voucher winners have substantially higher high-school graduation rates than losers. Since more lottery winners than losers took the ICFES exam, direct comparisons of test scores for winners and losers are subject to selection bias. We therefore discuss a number of solutions to this selection problem, including parametric methods and nonparametric quantile-specific bounds. After adjusting for selection bias, voucher winners appear to have learned more than losers. The fact that the program increased test scores even fairly high in the score distribution suggests that the program increased learning not only by increasing incentives for students at risk of repeating grades, but also through other mechanisms, such as increasing school choice.

Section I provides additional background on the PACES program and voucher lotteries. Section II presents estimates of the effect of PACES vouchers on high-school graduation rates, as measured by ICFES registration rates. Section III discusses the problem of selection bias in analyses of test scores and presents estimates of effects on scores using alternative approaches to the selection problem. Section IV concludes the paper.

I. Background

A. *The PACES Program*

The PACES program, established in late 1991, offered vouchers to children in low-income neighborhoods. To qualify for a voucher, applicants must have been entering the Colombian secondary school cycle, which begins with grade six, and have been 15 years of age or less. Prior to applying, students must already have been admitted to a participating secondary school (i.e., one that would accept vouchers) in a participating town. The list of participating towns included all of Colombia's largest cities.[4]

PACES vouchers were worth about US$190 in 1998. Our survey data show matriculation and tuition fees for private schools attended by voucher applicants in 1998 averaged about $340, so most voucher recipients supplemented the voucher with private funds. By way of comparison, the average annual per-pupil public expenditure in Colombia's public secondary school system in 1995 was just over $350 (Colombia DNP, 1999), and public school parents in our sample typically paid tuition or fees of roughly $58. Per capita GNP in Colombia was then around $2,280 (World Bank, 1999).

Participating schools tended to serve lower-income pupils and to have lower tuition than nonparticipating private schools. Schools with a vocational curriculum were also overrepresented among those participating in the program. Many elite private schools opted out, so just under half of private schools in the ten largest cities accepted PACES vouchers in 1993. In 1995, there were approximately 3.1 million secondary school pupils in Colombia. Almost 40 percent attended private schools, and about 8 percent of these used PACES vouchers.

Pupil-teacher ratios and facilities were similar in public and participating private schools, and test scores in participating private schools were close to those in public schools, though significantly below those in nonparticipating private schools. Public schools and private voucher schools had similar access to technology (King et al., 2003). Public school and private voucher school students in Bogotá also performed similarly on the ICFES exam, though public schools tended to have a larger number of students take the exams than private voucher schools. In contrast, the median score in private nonvoucher schools in Bogotá was much higher

[4] PACES was meant to increase secondary school enrollment and was motivated in large part by the fact that many secondary schools were clearly very crowded, especially in large cities. The average Colombian school day was four hours and many of the school buildings in Bogotá hosted multiple sessions per day. In fact, according to data available from ICFES, less than 2 percent of public schools in Bogotá hosted only one session per day, while 17 percent hosted three sessions per day. Private schools that accepted the voucher, by contrast, hosted fewer sessions. Almost 15 percent of private voucher schools had only one session per day and only 4 percent had three sessions per day. This likely allowed for a longer school day in private schools. Multiple sessions in public schools also made it easier for teachers to teach simultaneously in both types of schools.

than the median score at both public and voucher-accepting private schools in Bogotá. Across the country, both private and public schools in Bogotá had higher ICFES scores than other areas of Colombia.

Vouchers were renewed automatically through eleventh grade, when Colombian high school ends, provided the recipient's academic performance warranted promotion to the next grade. In practice, vouchers appeared to have had an incentive effect that increased grade promotion rates. For example, approximately 86 percent of voucher winners were promoted in sixth grade (compared to 80 percent of voucher losers). The incentive effects that led to this increase in promotion rates were probably strongest for relatively weak students, who were on the margin of failure had they not won a voucher.

Our earlier results (Angrist et al., 2002) suggest that three years after entering the lottery, voucher winners were 16 percentage points more likely to be attending a private school. Fee payments for voucher winners were $52 higher than for losers, suggesting that some winners may have used their vouchers to trade up to higher-priced schools. Although lottery winners and losers had similar enrollment rates, winners had completed 0.12 additional years of school, partly because they were 6 percentage points less likely to have repeated a grade. But winners also appear to have learned more. Among a subsample of lottery applicants who agreed to take a standardized test, winners scored 0.2 standard deviations more than losers, the equivalent of a full grade level. However, the sample of test takers was small (only 283), and hence these differences are only marginally significant at conventional levels. Moreover, since only 60 percent of those invited to take the test did so, sample selection issues remain a concern.

B. *Data and Descriptive Statistics*

Table 1 reports summary statistics drawn from 4,044 application forms completed by applicants who applied in 1994 to enter private school in sixth grade in 1995 in Bogotá. Of these applicants, 59 percent were awarded vouchers. Applicants were almost 13 years old, on average, and about evenly split between boys and girls. Roughly 88 percent of applicants came from households with a telephone or access to a telephone.

We matched PACES applicants with 1999–2001 ICFES records using national ID numbers, an identification number consisting of 11 digits, the first 6 of which show date of birth.[5] A final "check digit" in the ID number bears a mathematical relationship to the other digits. We used the embedded check digit and birth dates to determine whether ID numbers were valid. About 9.5 percent of applicants had invalid birth dates.[6] Among applicants with valid birth dates, 97 percent reported valid ID numbers.

There is no evidence that voucher winners are more likely to be matched with ICFES records because they have more accurately recorded ID numbers.[7] In fact, voucher winners were 1 percentage point *less* likely to have a valid ID, although this difference is not significantly different from zero, as can be seen in column 3, row 2, of Table 1. The results of restricting the sample to those with valid birth dates embedded in their ID numbers, reported in column 4, show an even smaller voucher effect, which is also statistically insignificant.

Voucher winners and losers had similar demographic characteristics, except possibly for a small age difference. These contrasts can be seen in the remaining rows of Table 1. The age differences by voucher status appear to be driven by a few outlying observations, probably due to incorrectly coded ID numbers among losers. The age gap falls when the sample is limited to those with valid ID numbers, though it remains marginally significant. We therefore control for age when estimating voucher effects.

II. Effects on High-School Graduation

As noted in the introduction, we use ICFES registration status as a proxy for high-school

[5] Angrist et al. (2004) provide a detailed description of the matching procedure and the ICFES.

[6] Birth dates are considered valid when they imply applicants were aged 9 to 25.

[7] We found some evidence of differential record-keeping in the first cohort of Bogotá applicants from 1992, before the lottery process was computerized. Because of this and other data problems, the 1992 applicant cohort was omitted from this study.

TABLE 1—CHARACTERISTICS OF ICFES MATCHING SAMPLE BY VOUCHER STATUS

	Means		Difference by voucher status (winners vs. losers)			
	Full sample (1)	Sample w/valid age (2)	Full sample (3)	Sample w/ valid age (4)	Valid ID and age (5)	Valid ID and age and has phone (6)
Won voucher	0.588	0.585				
Valid ID	0.876	0.967	−0.010	0.001	—	—
			(0.010)	(0.006)		
Age at time of application	12.7	12.7	−0.137	−0.086	−0.085	−0.091
	(1.8)	(1.3)	(0.064)	(0.045)	(0.044)	(0.047)
Male	0.487	0.493	0.004	0.011	0.012	0.008
			(0.016)	(0.017)	(0.017)	(0.018)
Phone	0.882	0.886	0.013	0.008	0.008	—
			(0.010)	(0.011)	(0.011)	
N	4,044	3,661	4,044	3,661	3,542	3,139

Notes: Robust standard errors are reported in parentheses in columns 3–6. Regression estimates of differences by voucher status in column 4 are for the sample with valid age data embedded in the national ID number. A valid age must be between 9 and 25. Column 5 reports results for a sample limited to those with a valid ID check digit and column 6 shows results for a sample further limited to those with a phone. The sample includes applicants from the 1995 lottery cohort.

graduation status because 90 percent of graduating seniors take the ICFES exam. Estimates of voucher effects on high-school graduation rates were constructed using the following regression model:

$$T_i = X_i'\beta_0 + \alpha_0 D_i + \varepsilon_i, \tag{1}$$

where T_i is an indicator of ICFES registration, D_i is an indicator for whether applicant i won a voucher, and X_i is a vector of controls for age and sex. We also report estimates without covariates. Students in the 1995 applicant cohort who were promoted on schedule should have registered to take the ICFES exam at one of two opportunities in the 2000 school year. Because some students may also have skipped or repeated grades, we also checked those registered for the exams offered in 1999 and 2001. If a student was found to have been tested more than once, we retained the first set of test scores.

About 35 percent of PACES applicants were matched to ICFES records using ID numbers, a result that can be seen in the first row of Table 2. This rate, the dependent-variable mean for equation (1), falls to 33 or 34 percent when matches are validated using city of residence or the first seven letters of students' last names, and to 32 percent when matches are validated using both city of residence and the first seven letters of students' last names.

Results from models with no controls, reported in column 1, show that vouchers raised ICFES registration rates about 7 percentage points, a highly significant difference.[8] Because of the slight differences between voucher winners and losers reported in Table 1, the estimated effect falls to about 6 percentage points with demographic controls but remains significantly different from zero. There is no clear pattern of differences in voucher effects by sex, though the base rate is lower for boys. Using city of residence to validate matches leads to slightly smaller treatment effects for girls and overall, but the change is not substantial. Validation by matching on names as well as ID numbers leads to treatment effects almost identical to those without validation, and validation using both city and name generates estimates similar to those using city only.[9]

[8] The standard errors reported in Table 2 and elsewhere are heteroskedasticity-consistent.

[9] Our previous estimates show voucher winners were less likely to have repeated a grade and, hence, as time goes on, we might expect the gap in ICFES registration rates between winners and losers to decline as more losers finish secondary school. But voucher winners who had previously repeated a grade may have been more likely than losers to take the ICFES, so it is unclear whether the gap in ICFES registration rates should have increased or decreased over time. In any case, our analysis shows little change in ICFES match rates from 1999 to 2001.

TABLE 2—VOUCHER STATUS AND THE PROBABILITY OF ICFES MATCH

	Exact ID match		ID and city match		ID and 7-letter name match		ID, city, and 7-letter match	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	A. All applicants (N = 3542)							
Dependent var. mean	0.354		0.339		0.331		0.318	
Voucher winner	0.072	0.059	0.069	0.056	0.072	0.059	0.068	0.056
	(0.016)	(0.015)	(0.016)	(0.014)	(0.016)	(0.014)	(0.016)	(0.014)
Male		−0.052		−0.053		−0.043		−0.045
		(0.014)		(0.014)		(0.014)		(0.014)
Age		−0.160		−0.156		−0.153		−0.149
		(0.005)		(0.005)		(0.005)		(0.005)
	B. Female applicants (N = 1789)							
Dependent var. mean	0.387		0.372		0.361		0.348	
Voucher winner	0.067	0.056	0.069	0.057	0.071	0.060	0.073	0.062
	(0.023)	(0.021)	(0.023)	(0.021)	(0.023)	(0.021)	(0.023)	(0.021)
Age		−0.168		−0.164		−0.160		−0.156
		(0.006)		(0.006)		(0.006)		(0.006)
	C. Male applicants (N = 1752)							
Dependent var. mean	0.320		0.304		0.302		0.288	
Voucher winner	0.079	0.063	0.071	0.055	0.074	0.059	0.065	0.050
	(0.022)	(0.020)	(0.022)	(0.020)	(0.022)	(0.020)	(0.022)	(0.020)
Age		−0.153		−0.148		−0.146		−0.141
		(0.007)		(0.007)		(0.007)		(0.006)

Notes: Robust standard errors are shown in parentheses. The sample used to construct this table includes all Bogotá applicants with valid ID numbers and valid age data (i.e., ages 9 to 25 at application). The sample is the same as in Table 1, column 5.

On balance, the estimated effects of voucher status on ICFES registration are remarkably robust to changes in sample, specification, and the definition of a match. It therefore seems fair to say that PACES vouchers increased the likelihood of ICFES registration, and probably high-school graduation, by 5 to 7 percentage points for both boys and girls. This amounts to an increase of 15 to 20 percent in the probability students took the ICFES exam.

III. Effects on College-Entrance-Exam Scores

A. *The Selection Problem*

Because PACES voucher winners were more likely than losers to take the test, the test-score distributions of winners and losers are not directly comparable. To see the consequences of differential test-taking rates for comparisons of scores among test-takers, it helps to introduce notation for the potential outcomes (scores) that would be revealed under alternative treatment (voucher) assignments. Let y_{1i} be the ICFES score student i would obtain after winning a voucher, and let y_{0i} denote the score student i would obtain otherwise. We assume that both of these potential outcomes are well-defined for all pupils, whether they actually took the test or not, and whether they won the lottery or not. The average causal effect of winning the voucher on the scores of all winners is $E[y_{1i} - y_{0i}|D_i = 1]$. Of course, in practice, we observe scores only for those who were tested. Moreover, among tested pupils, we observe only y_{1i} for winners and y_{0i} for losers.

Using a notation paralleling the notation for potential test scores, let T_{1i} and T_{0i} denote potential test-taking status. That is, T_{1i} is a dummy for whether a student would have taken the ICFES after winning the lottery and T_{0i} is a dummy for whether a student would have taken the ICFES after losing the lottery. By virtue of the random assignment of D_i, the vector of all potential outcomes $\{y_{1i}, y_{0i}, T_{1i}, T_{0i}\}$ is jointly independent of D_i, though the elements of this vector are probably correlated with each other. Observed test-taking status, the dependent variable in the previous section, is linked to potential outcomes by the equation

$$T_i = T_{0i} + (T_{1i} - T_{0i})D_i.$$

Similarly, the latent score variable (i.e., what we would observe if all students were tested) is

$$y_i = y_{0i} + (y_{1i} - y_{0i})D_i.$$

The observed win/loss contrast in test scores among those who were tested can be now written

$$\begin{aligned} &E[y_i|T_i = 1, D_i = 1] - E[y_i|T_i = 1, D_i = 0] \\ &= E[y_{1i}|T_{1i} = 1, D_i = 1] - E[y_{0i}|T_{0i} = 1, D_i = 0] \\ &= E[y_{1i}|T_{1i} = 1] - E[y_{0i}|T_{0i} = 1], \end{aligned}$$

where the second equality is because D_i is randomly assigned. This contrast does not have a causal interpretation, since students with $T_{1i} = 1$ and $T_{0i} = 1$ are not drawn from the same population unless D_i has no effect on the probability of being tested. In fact, we can expand this contrast further to write

$$\begin{aligned} (2)\quad &E[y_i|T_i = 1, D_i = 1] - E[y_i|T_i = 1, D_i = 0] \\ &= E[y_{1i} - y_{0i}|T_{0i} = 1] \\ &\quad + E[y_{1i}|T_{1i} = 1] - E[y_{1i}|T_{0i} = 1]. \end{aligned}$$

Equation (2) shows that the win/loss contrast among test-takers is equal to the average causal effect on those who would have been tested anyway, $E[y_{1i} - y_{0i}|T_{0i} = 1]$, plus a term that captures the selection bias due to the fact that we are conditioning on ICFES registration status, itself an outcome that is affected by treatment.

The bias term in equation (2) is likely to be negative if PACES vouchers increased test scores. To illustrate this, suppose that $y_{1i} = y_{0i} + \alpha$, where $\alpha > 0$, and that students chose to be tested if their potential scores exceeded a constant threshold, μ. Then $T_{ji} = 1[y_{ji} > \mu]$ for $j = 0, 1$; and the selection bias is

$$(3)\quad E[y_{0i}|y_{0i} > \mu - \alpha] - E[y_{0i}|y_{0i} > \mu],$$

which is clearly negative. Of course, this example presumes that vouchers are never harmful. If vouchers were harmful, selection bias could (by the same argument) mask a negative treatment effect.

To provide an empirical basis for the claim that selection bias in the sample of ICFES takers is likely to be negative, we used test scores from our earlier *random* sample of Bogotá students, the sample used by Angrist et al. (2002) to assess the effect of vouchers on learning. Our sample of 259 tested students is somewhat more likely to have taken the ICFES test than the overall average test rate for the 1995 Bogotá cohort (about 44 percent versus 35 percent overall). Importantly, however, and in contrast with the ICFES test, the likelihood of taking our test is the same for voucher winners and losers. Thus, our earlier sample of test-takers is not contaminated by self-selection bias of the sort affecting ICFES-takers (though there are missing score data for other reasons).

A regression using stacked math and reading scores from the earlier testing sample generates a voucher effect of 0.186, with effects measured in standard deviation units (and with a standard error adjusted for student clustering of 0.105). Limiting this sample to the roughly 44 percent of tested students who also took the ICFES generates a voucher effect of 0.044 (s.e. = 0.157). The pattern of substantial (and usually marginally significant) positive effects in the full sample and considerably smaller and insignificant treatment effects when this sample is limited to those who also took the ICFES test appears for all dependent variables and specifications. This finding illustrates the fact that conditioning on ICFES testing status almost certainly drives positive treatment effects toward zero.

B. *Parametric Strategies*

In a first attempt to adjust for selection bias, we used a modified Tobit procedure. In particular, we fit parametric models to artificially completed score data constructed by censoring observed scores at or above a particular value or quantile, with all those below this point *and nontakers* assigned the censoring point. Subject to the normality assumption, this provides consistent estimates of treatment effects on the latent scores of all students, assuming those not tested would have scored below the artificial censoring point. Moreover, a comparison of Tobit results using different censoring points provides a natural specification test for this procedure since, if correctly specified, results using different censoring points should be similar. A key drawback in this case is the need to assume normality of the uncensored latent score distribution. The quality of the normal approximation may be especially poor, given the relatively discrete nature of the score data. We therefore discuss alternative approaches in the next section.

The idea behind the parametric approach is spelled out in more detail below. We assume that the causal effect of interest could be estimated by regressing latent scores, y_i, on D_i and covariates, $\boldsymbol{X}_i$. That is, the regression of interest is

$$(4) \qquad y_i = \boldsymbol{X}_i'\beta + \alpha D_i + \eta_i,$$

where η_i is a normally distributed error. An artificially censored dependent variable is constructed using

$$(5) \qquad Y_i(\tau) \equiv 1[T_i y_i \geq \tau] y_i + 1[T_i y_i < \tau]\tau$$

for some positive threshold, τ. Assuming any untested student would have scored at or below this threshold if they had been tested, the parameters in (4) can be consistently estimated by applying Tobit to $Y_i(\tau)$. This is not realistic for $\tau = 0$ but it may be for, say, the tenth percentile of the score distribution among test-takers. Finally, note that if Tobit using $Y_i(\tau_0)$ identifies α, then Tobit using $Y_i(\tau_1)$ will also work for any threshold value τ_1, such that $\tau_1 > \tau_0$.[10]

As a benchmark for this procedure, we again report estimates using the sample of test-takers, without adjusting for censoring. Among test-takers, winners scored about 0.7 points higher on the language exam, with similar though less precise effects in samples of boys and girls. These results are reported in column 1 of Table 3.[11] The estimated effects for math scores are smaller though still positive. Including all students and censoring both nontakers and low scorers at the first percentile of the score distribution generates a voucher effect of 1.1 (s.e. = 0.24) for language and 0.79 (s.e. = 0.18) for math. This can be seen in column 2 of Table 3.

Tobit estimates using data censored at the first and tenth percentiles among test-takers suggest much larger effects than those that arise without correcting for selection bias. The Tobit estimates are on the order of two to four points for language and two to three points for math, in all cases significantly different from zero (reported in columns 3 and 4). Effects are at the lower end of this range, around two points, when the data are censored at the tenth percentile. The estimates using artificially censored data tend to be somewhat larger for boys than for girls.

Assuming the Tobit model applies to data censored at the first percentile, the Tobit coefficient estimates should be the same when estimated using data censored at the tenth. The decline in estimates moving column 3 to column 4 of Table 3 therefore suggests the first percentile is too low a threshold for the Tobit model to apply. On the other hand, Tobit estimates of α are remarkably stable when the distribution is artificially censored with a cutoff that removes the lower 10 to 80 percent of scores. This can be seen in Figures 1A and 1B, which plot the estimated Tobit coefficients and confidence bands for alternative censoring points. The estimated treatment effects are fairly stable, at around two points, turning down slightly when the lower 90 percent of scores among takers are censored. It should be noted, however, that the confidence intervals widen at this point. Moreover, normality may be a worse approximation for the upper tail of the score distribution.

[10] As a partial check on this, we compared the scores of ICFES takers and nontakers on our earlier achievement tests. Assuming percentile scores on the two tests are similar, this comparison is informative about the assumption invoked here. Indeed, the two test scores are highly correlated. Moreover, a comparison of earlier test results by ICFES-taker status shows markedly lower average scores and a clear distribution shift to the left for ICFES nontakers relative to ICFES takers.

[11] The same covariates and sample were used to construct the results in Tables 2 to 5.

TABLE 3—OLS AND TOBIT ESTIMATES OF THE EFFECTS OF THE VOUCHERS ON ICFES SCORES

	OLS with score > 0 (1)	OLS censored at 1% (2)	Tobit censored at 1% (3)	Tobit censored at 10% (4)
	A. Language scores			
Full sample				
Dep var mean	47.4	37.3	37.3	42.7
	(5.6)	(8.0)	(8.0)	(4.7)
Voucher effect	0.70	1.14	3.29	2.06
	(0.33)	(0.24)	(0.70)	(0.46)
Girls				
Dep var mean	47.0	37.6	37.6	42.8
	(5.7)	(8.1)	(8.1)	(4.7)
Voucher effect	0.74	1.04	2.88	1.86
	(0.45)	(0.34)	(0.91)	(0.59)
Boys				
Dep var mean	47.8	37.0	37.0	42.5
	(5.5)	(7.9)	(7.9)	(4.6)
Voucher effect	0.66	1.25	3.77	2.29
	(0.48)	(0.34)	(1.10)	(0.71)
	B. Math scores			
Full sample				
Dep var mean	42.5	35.7	35.7	37.6
	(4.9)	(5.8)	(5.8)	(4.6)
Voucher effect	0.40	0.79	2.29	1.98
	(0.29)	(0.18)	(0.51)	(0.45)
Girls				
Dep var mean	42.3	35.9	35.9	37.8
	(4.8)	(5.8)	(5.8)	(4.6)
Voucher effect	0.18	0.62	1.84	1.60
	(0.38)	(0.25)	(0.66)	(0.58)
Boys				
Dep var mean	42.8	35.4	35.4	37.5
	(5.0)	(5.7)	(5.7)	(4.5)
Voucher effect	0.70	0.95	2.82	2.43
	(0.44)	(0.25)	(0.79)	(0.69)

Note: Robust standard errors are reported in parentheses below the voucher effects. The censoring point used to construct the estimates in columns 2 to 4 is the percentile of the test-score distribution among test-takers indicated in the column heading. Standard deviations are reported below the dependent variable means. Sample sizes in column 1, panel A (language scores), are 1,223 for the whole sample, 672 for girls, and 551 for boys. An additional boy and girl took the math test. The samples in the other columns are 3,541 overall, 1,788 girls and 1,753 boys. Covariates include age and gender.

On balance, the model described by equations (4) and (5) seems to provide a reasonably coherent account of the voucher impact on latent scores. Effects in this range amount to a score gain of about 0.2σ, where σ is the standard deviation of the latent residual in equation (4). This is consistent with our earlier estimates of effects on achievement for a 1988 random sample of Bogotá eighth graders.[12]

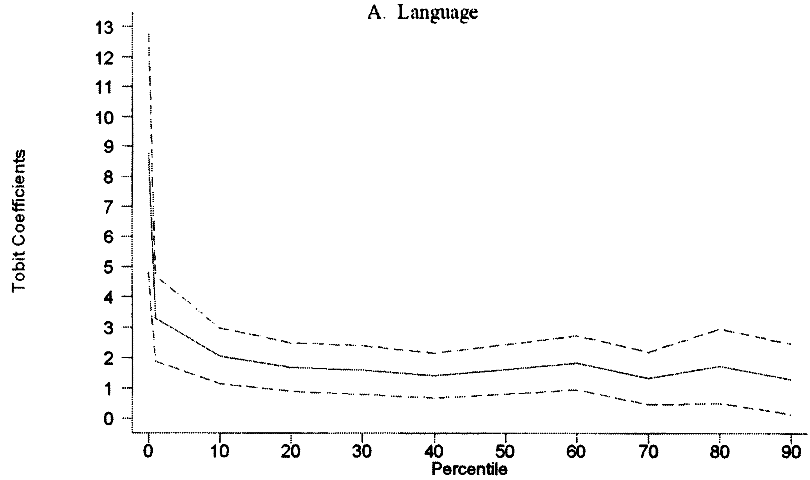
C. *Nonparametric Bounds*

The strategies discussed in the previous section rely on strong functional-form and distributional assumptions. This section builds on the discussion above to derive a set of nonparametric bounds for quantile-specific program impacts on the distribution of test scores. Because selection bias is most likely negative when treatment effects are positive, selection-contaminated comparisons provide a lower bound on the impact of vouchers on achievement. Here, we also develop an upper bound by adapting a theoretical result from our earlier paper (Angrist et al., 2002). Related studies discussing nonparametric bounds on selection bias include Charles F. Manski (1989) and David S. Lee (2002).

Suppose we are prepared to assume that winning the lottery was never harmful, i.e., that $y_{1i} \geq y_{0i}$ for all i. The identifying power of this *monotone treatment response* assumption is discussed by Manski (1997). In this context, monotone treatment response seems reasonable, since lottery winners were free to turn down their vouchers and attend public school if they felt continued voucher use was harmful.[13] It also seems reasonable to assume $T_{1i} \geq T_{0i}$, since test-taking status is probably determined by,

[12] Effect sizes calculated using the distribution of the latent Tobit residual seem like an appropriate standard of comparison, since the testing strategy used in our earlier study can be thought of as providing estimates of effects on latent scores (i.e., scores when all applicants in the relevant sample are tested whether or not they registered for ICFES). While the positive treatment effect found here is consistent with our previous results, it is unclear whether the magnitude of the voucher effect should have increased or decreased over time. On one hand, the private school attendance gap was growing in seventh and eighth grades. On the other, within three years of the voucher lottery, 50 percent of voucher winners were no longer using the voucher, so effects may have waned.

[13] This condition need not hold if some students who chose to use vouchers anticipated gains that did not materialize, with a subset ending up having been harmed by voucher use. In practice, however, the "never-harmful" assumption is made more plausible by the fact that vouchers typically did not cover the entire cost of private school. This means that students using vouchers presumably expected gains large enough to outweigh the financial costs of attending private school, with a low risk of adverse academic effects.



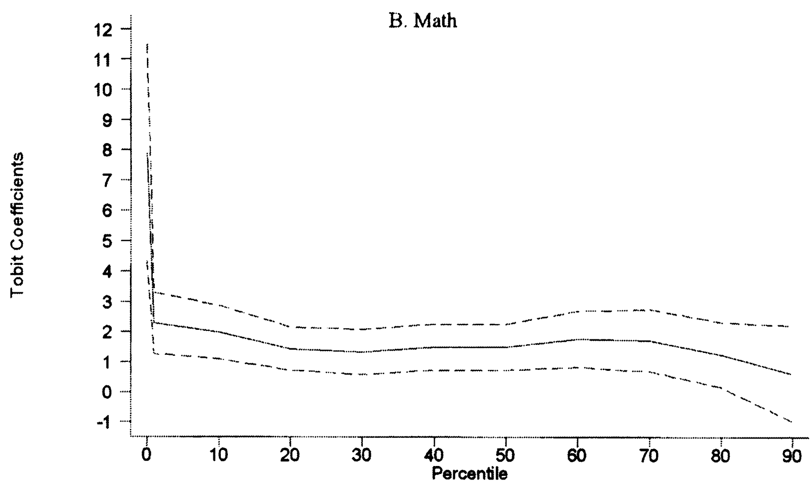


FIGURE 1. TOBIT COEFFICIENTS BY CENSORING PERCENTILE IN SCORE DISTRIBUTION

Notes: The figure plots Tobit estimates of the effect of vouchers on test scores, using data censored at the point indicated on the X-axis (i.e., values below the indicated percentile are assigned a value of zero). For the purposes of this exercise, nontakers are also coded as having a score of zero.

among other things, expected scores. Finally, it is useful to define a score variable that equals zero for those not tested:

$$Y_{ji} = T_{ji} y_{ji} \qquad \text{for} \qquad j = 0, 1.$$

Note that given our "never-harmful" assumptions, we also have $Y_{1i} \geq Y_{0i}$ for all i, and that

$$E[Y_{1i} - Y_{0i} | T_{0i} = 1] = E[y_{1i} - y_{0i} | T_{0i} = 1].$$

The observed outcome, Y_i, is linked to potential outcomes by

$$\begin{aligned} Y_i &= Y_{0i} + (Y_{1i} - Y_{0i}) D_i \\ &= T_{0i} y_{0i} + (T_{1i} y_{1i} - T_{0i} y_{0i}) D_i . \end{aligned}$$

To simplify notation, drop subscripts for individuals and let $q_0(\theta)$ be the θ-quantile of the distribution of Y_0 and let $q_1(\theta)$ be the θ-quantile of the distribution of Y_1.

For the development that follows, it's useful to define a rank-preservation restriction on the joint distribution of (Y_0, Y_1):

DEFINITION: *The random variable Y_1 is said to be a θ-quantile-preserving transformation (θ-QPT) of the random variable Y_0 if $P(Y_1 \geq q_1(\theta)|Y_0 \geq q_0(\theta)) = 1$.*

Note that Y_1 is a θ-QPT of Y_0 if the two potential outcomes are linking by a weakly increasing function, or if, for any two draws from the joint distributions of Y_1 and Y_0, the orderings of Y_1 and Y_0 are the same. The θ-QPT concept extends the idea of rank-preservation to quantile-specific comparisons.[14]

The following proposition establishes a set of quantile-specific bounds on average treatment effects in the presence of sample selection bias (the proof appears in the Appendix):

PROPOSITION 1: *Suppose that $y_1 \geq y_0$ and $T_1 \geq T_0$. Choose $\theta \geq \theta_0$ where $q_0(\theta_0) = 0$. Then*

$$E[Y|D = 1, Y > q_1(\theta)] - E[Y|D = 0, Y > q_0(\theta)]$$

$$\geq E[y_1 - y_0|y_0 > q_0(\theta), T_0 = 1]$$

$$\geq E[Y|D = 1, Y > q_0(\theta)]$$

$$- E[Y|D = 0, Y > q_0(\theta)].$$

Furthermore, if Y_1 is a θ-quantile-preserving transformation of Y_0, then the first inequality is an equality.

Note that we can choose a quantile, θ_0, such that $q_0(\theta_0) = 0$, and then drop the lower θ_0 percent of the Y_1 distribution to obtain an upper bound on $E[y_1 - y_0|T_0 = 1]$. At the same time, the unadjusted conditional-on-positive contrast in test scores provides a lower bound. Moreover, if Y_1 and Y_0 are linked by a θ-QPT, the upper bound provides an estimate of $E[y_1 - y_0|T_0 = 1]$.[15] We can use this fact to estimate or bound average treatment effects at a number of points in the score distribution.

Estimates of nonparametric bounds on treatment effects at different quantiles are reported in Table 4 for language and math scores, for θ_0 such that $q_0(\theta_0) = 0$, and for $\theta = 0.75$, 0.85, and 0.95. The largest effects are at $q_0(\theta_0) = 0$, i.e., effects on all pupils who would have been tested if they had not won the lottery. The lower bound for effects on language scores in this population is 0.68 (s.e. = 0.33), while the upper bound is 2.8 (s.e. = 0.31). For $\theta = 0.95$, the bounds fall to an insignificant 0.35 on the low end and still-significant 1.4 (s.e. = 0.34) on the high end. The pattern of bounds by quantile is consistent with either a larger shift in scores for pupils with Y_0 close to the low end of the test-takers' score distribution, or with a tightening of the upper bound at higher quantiles, or both.

A comparison of the entire distribution of test scores for winners and losers supports the notion that the voucher led to an increase in achievement by winners. Panel A of Figure 2, which plots kernel density estimates in the sample of all test-takers, shows slightly flattened and right-shifted distributions for winners. As with the comparisons of means, however, this contrast is contaminated by selection bias, in particular, by the likely introduction of low-scorers into the sample of tested winners. Adjusting for sample selection bias using Proposition 1 leads to a clearer impression of

[14] Susan Athey and Guido W. Imbens (2002), Richard Blundell et al. (2004), and Victor Chernozhukov and Christian Hansen (2005) discuss the identifying power of similar assumptions. Note that θ-QPT does not amount to perfect rank correlation unless it holds for all quantiles. In practice, θ-QPT seems more plausible at upper quantiles of the latent score distribution since jumps or leapfrogging by nontakers who win vouchers is unlikely above high quantiles.

[15] The latter result can also be understood as follows. Angrist (1997) shows that monotonicity in selection status ($T_{1i} \geq T_{0i}$), combined with a constant-effects link between y_1 and y_0, implies that controlling the probability of sample selection eliminates selection bias. Symmetric truncation is equivalent to fixing the probability of sample selection. The proposition generalizes this result to models with a nonconstant but still rank-preserving link between potential outcomes. Krueger and Diane M. Whitmore (2001) used a similar idea to estimate $E[y_1 - y_0|T_0 = 1]$ in a study of class size.

TABLE 4—BOUNDS ON VOUCHER EFFECTS

Loser's value at percentile	Percentile of loser's distribution	Loser's average score above quantile (1)	w/o covariates Lower bound (2)	w/o covariates Upper bound (3)	w/covariates Lower bound (4)	w/covariates Upper bound (5)
		A. Language scores				
0	72nd percentile	46.9	0.68	2.81	0.70	2.80
		(5.5)	(0.33)	(0.31)	(0.33)	(0.31)
41	75th percentile	48.7	0.46	2.47	0.49	2.46
		(3.9)	(0.26)	(0.26)	(0.26)	(0.26)
47	85th percentile	51.2	0.49	2.39	0.50	2.37
		(3.0)	(0.27)	(0.28)	(0.27)	(0.28)
52	95th percentile	55.6	0.35	1.38	0.36	1.39
		(1.7)	(0.31)	(0.34)	(0.31)	(0.34)
		B. Math scores				
0	70th percentile	42.3	0.40	2.40	0.40	2.41
		(4.8)	(0.29)	(0.27)	(0.29)	(0.27)
37	75th percentile	43.7	0.35	1.76	0.34	1.76
		(3.8)	(0.25)	(0.25)	(0.25)	(0.25)
42	85th percentile	46.2	0.24	1.44	0.27	1.48
		(3.2)	(0.28)	(0.28)	(0.28)	(0.28)
47	95th percentile	50.3	−0.09	1.85	−0.11	1.80
		(2.4)	(0.39)	(0.42)	(0.39)	(0.43)

Notes: The table reports bounds computed using the formulas in Proposition 1 in the text. Means and standard deviations are shown in column 1. Estimated bounds and standard errors are shown in columns 2 to 5. Columns 4 to 5 are from models that include controls for age and gender.

a shift. This can be seen in Panel B, which plots score distributions after limiting the distribution of winners to the top 28 percent of the score distribution (including zeros). In other words, panel B plots scores conditional on $Y_0 > q_0(0.72)$, where $q_0(0.72) = 0$, and $Y_1 > q_1(0.72)$. The adjusted figure shows a clearer rightward shift in the distribution for winners, especially in the middle of the density.

The density plots and differences in average treatment effects reported for different quantiles in Table 4 suggest the PACES program had an impact on the distribution of test scores beyond a simple "location shift." As a final exploration of distribution effects (and to quantify the impression left by the figures), we estimated the impact of winning a voucher at different points in the cumulative distribution of test scores. In particular, we estimated voucher effects in equations analogous to equation (1), with the dependent variable given by $1[Y_i > c]$, where c is a quantile in the score distribution among test-takers. This procedure uses a sample where test scores for nontakers are coded as zero. Therefore, assuming that those not tested would have scored below c, the resulting estimates are unaffected by selection bias.[16]

Estimates of effects on the distribution of test scores, reported in Table 5, show the largest impact on the probability test scores exceeded the lowest decile in the score distribution (among test-takers). For example, the probit marginal effects of the impact of a voucher on the probability of crossing the first decile are 0.063 (s.e. = 0.015) for the language score distribution and 0.068 (s.e. = 0.016) for the math score distribution. These estimates appear in columns 3 and 6 of Table 5. The

[16] Quantile regression (QR) is an alternative and perhaps more conventional procedure that captures effects on distribution while avoiding selection bias under the same assumptions. In this case, however, QR is made less attractive by the almost-discrete nature of the test scores. About 80 percent of the mass of the distribution of scores among takers falls into a range of 15 points or less. This near-discreteness causes QR estimates to behave poorly and invalidates standard asymptotic theory for QR, since a regularity assumption for QR is continuity of the dependent variable.

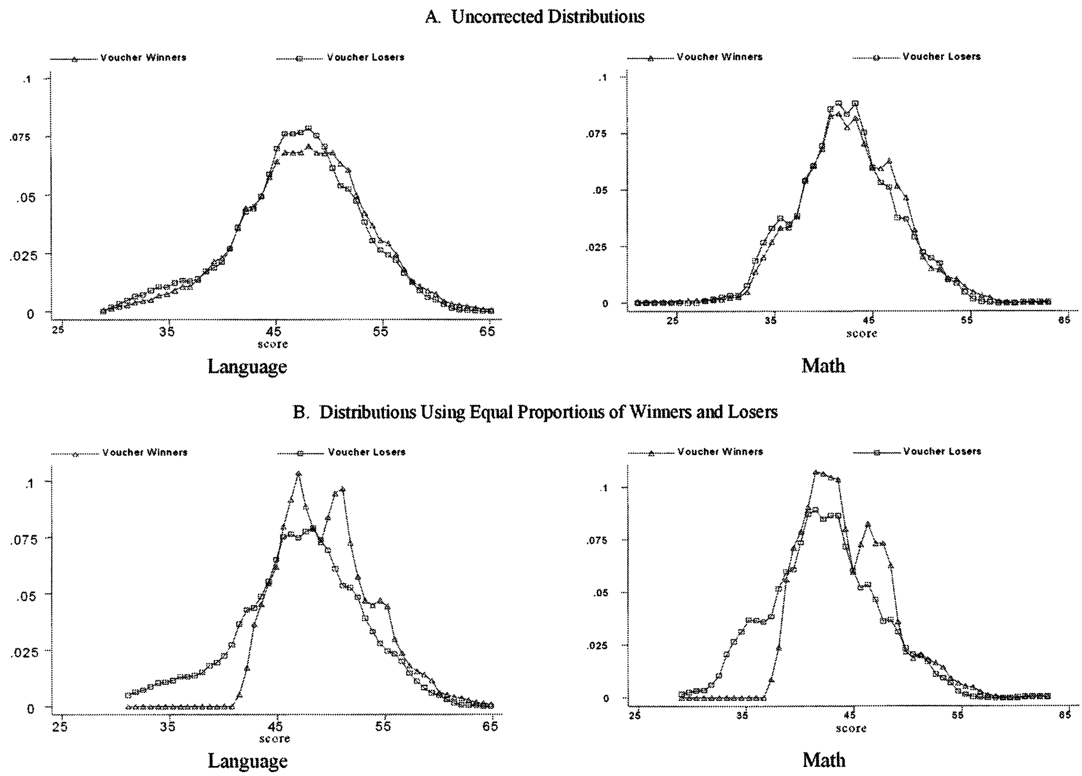


FIGURE 2. TEST-SCORE DISTRIBUTIONS

Notes: Panel A shows uncorrected score distributions for both winners and losers. Panel B shows the same uncorrected distributions of scores for voucher losers, along with a truncated winners' distribution calculated as follows: Suppose π% of losers were tested, while $\pi + \kappa$% of winners were tested; Panel B shows the distribution of scores for winners for the upper π% of the winners' score distribution only.

corresponding estimates fall to a bit over 0.04 at the median, and then to around 0.025 at the seventy-fifth percentile. It seems unlikely that many students at the seventy-fifth percentile of the distribution of test scores among test-takers were in danger of having to repeat a grade. The substantial impact of the program on the likelihood of scoring in the upper quartile of the score distribution among test-takers therefore suggests the program operated through channels other than simply reducing the risk of grade repetition. Moreover, while the estimated distribution shifts at the upper decile in this specification are only 0.01 and 0.003 for language and math, respectively, the former effect is still significantly different from zero.

IV. Summary and Conclusions

This paper presents evidence on the impact of PACES vouchers on relatively long-run educational outcomes for applicants to the Bogotá voucher lottery. PACES vouchers subsidized private school attendance, and were renewable annually, conditional on grade advancement. The random assignment of vouchers facilitates causal comparisons between those who did and not receive vouchers. Administrative data on college entrance exams allow us to estimate the impact of vouchers on high-school graduation rates and scholastic achievement.

The empirical results point to an increase in (proxy) high-school graduation rates of 5 to 7

TABLE 5—EFFECTS ON THE PROBABILITY OF EXCEEDING TEST-SCORE PERCENTILES

Threshold value			Language scores			Math scores		
Language	Math	Score threshold	OLS no covs (1)	OLS w/covs (2)	Probit w/covs (3)	OLS no covs (4)	OLS w/covs (5)	Probit w/covs (6)
40	35	10th percentile	0.069	0.057	0.063	0.073	0.061	0.068
			(0.016)	(0.014)	(0.015)	(0.016)	(0.014)	(0.016)
44	39	25th percentile	0.055	0.045	0.047	0.062	0.052	0.054
			(0.014)	(0.013)	(0.014)	(0.015)	(0.014)	(0.014)
47	42	50th percentile	0.050	0.043	0.041	0.051	0.044	0.044
			(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.011)
51	45	75th percentile	0.033	0.030	0.025	0.034	0.031	0.027
			(0.009)	(0.008)	(0.007)	(0.009)	(0.009)	(0.008)
54	48	90th percentile	0.015	0.013	0.010	0.005	0.003	0.003
			(0.006)	(0.006)	(0.004)	(0.006)	(0.006)	(0.005)

Notes: The dependent variable indicates whether students exceeded various percentiles in the relevant score distribution for test-takers. Marginal effects are reported for probit estimates. Standard errors are reported in parentheses. The sample size used to construct these estimates is 3,541.

percentage points, relative to a base rate of 25 to 30 percent. This is consistent with our earlier results showing a 10-percentage-point increase in eighth-grade completion rates among voucher winners, as well as with the gains on a standardized test we had administered to a small sample of applicants. The magnitude of the test-score gains in our follow-up study turn partly on how selection bias is controlled. Tobit estimates with artificially censored data put the treatment effects at around two points, or roughly 0.2σ relative to the standard deviation of latent scores. Nonparametric bounds bracket this number, with a lower bound that is significantly different from zero. Since the upper bound is tight under the assumption of a rank-preserving treatment effect and the Tobit estimates satisfy a simple overidentification test, something close to the Tobit estimate of two points seems like a good summary estimate.

For the most part, the bounds we estimate on average treatment effects at higher quantiles of the score distribution are smaller than those at the lower end. This may be because program effects were actually greatest at the bottom of the distribution, perhaps due to the incentive effects generated by PACES vouchers. It also seems likely, however, that the upper bounds are tighter higher in the distribution for technical reasons having to do with the relationship between potential outcomes in the treated and nontreated states. In any case, the fact that lottery winners were substantially more likely to score in the top quartile on the national university entrance exam suggests that the PACES program probably improved learning not only by increasing financial incentives to avoid failing a grade, but also by expanding school choice.

On balance, our results suggest a substantial gain in both high-school graduation rates and achievement as a result of the voucher program. Although the benefits of achievement gains per se are hard to quantify, there is a substantial economic return to high-school graduation in Colombia. At a minimum, this suggests demand-side financing efforts similar to the PACES program warrant further study. An unresolved question, however, is how to reconcile the consistently positive voucher effects for Colombia reported here with more mixed results for the United States (see, e.g., Rouse, 1998; Howell and Peterson, 2002). One possibility is that PACES is a better experiment. Among U.S. voucher studies, even those using random assignment were compromised by complex research designs and substantial attrition. As it turns out, the U.S. results are sensitive to how these problems are handled (John Barnard et al., 2003; Krueger and Zhu, 2003). Another possible explanation for divergent effects is a larger gap in the quality of public and private schools in Colombia than in the United States.

Finally, PACES included features not necessarily shared by other voucher programs, such as incentives for academic advancement and the opportunity for those who would have gone to private school anyway to use vouchers to attend more expensive schools.

APPENDIX

PROOF OF PROPOSITION 1:

$E[Y|D = j, Y > q_j(\theta)] = E[Y_j|Y_j > q_j(\theta)]$ by random assignment. Also,

$$E[Y_1|Y_1 > q_1(\theta)] - E[Y_0|Y_0 > q_0(\theta)] = E[Y_1 - Y_0|Y_0 > q_0(\theta)] + \{E[Y_1|Y_1 > q_1(\theta)] - E[Y_1|Y_0 > q_0(\theta)]\}$$

$$= E[y_1 - y_0|y_0 > q_0(\theta), T_0 = 1] + b_0.$$

If Y_1 is a θ-QPT of Y_0, then $b_0 = 0$, so the second part is proved. Otherwise, we need to show that $b_0 \geq 0$. Note that

$$b_0 = E[Y_1 1(Y_1 > q_1) - Y_1 1(Y_0 > q_0)]/P(Y_0 > q_0)$$

since $P(Y_1 > q_1) = P(Y_0 > q_0)$, so

$$b_0 \geq E[Y_1(1(Y_1 > q_1) - Y_1 1(Y_0 > q_0))] = E[Y_1(1(Y_1 > q_1) - 1(Y_0 > q_0))]$$

$$= E[Y_1(1(Y_1 > q_1, Y_0 < q_0) - 1(Y_1 < q_1, Y_0 > q_0))],$$

where the second equality above is a consequence of the facts that $1(Y_1 > q_1) = 1(Y_1 > q_1, Y_0 < q_0) + 1(Y_1 > q_1, Y_0 > q_0)$ and $1(Y_0 > q_0) = 1(Y_1 < q_1, Y_0 > q_0) + 1(Y_1 > q_1, Y_0 > q_0)$. Further simplifying, we have

$$E[Y_1(1(Y_1 > q_1, Y_0 < q_0) - 1(Y_1 < q_1, Y_0 > q_0))] = E[Y_1|Y_1 > q_1, Y_0 < q_0]p_1$$

$$- E[Y_1|Y_1 < q_1, Y_0 > q_0]p_0,$$

where $p_1 = \Pr[Y_1 > q_1, Y_0 < q_0]$ and $p_0 = \Pr[Y_1 < q_1, Y_0 > q_0]$. Clearly,

$$E[Y_1|Y_1 > q_1, Y_0 < q_0] \geq E[Y_1|Y_1 < q_1, Y_0 > q_0].$$

Also, $p_1 = p_0$ because

$$P(Y_1 > q_1) = p_1 + \Pr[Y_1 > q_1, Y_0 > q_0] = \theta = \Pr[Y_1 > q_1, Y_0 > q_0] + p_0 = P(Y_0 > q_0).$$

This establishes the upper bound. The lower bound is a consequence of the fact that

$$E[Y|D = 1, Y > q_0(\theta)] - E[Y|D = 0, Y > q_0(\theta)] = E[Y_1 - Y_0|Y_0 > q_0(\theta)]$$

$$+ \{E[Y_1|Y_1 > q_0(\theta)] - E[Y_1|Y_0 > q_0(\theta)]\}$$

and

$$E[Y_1|Y_0 > q_0(\theta)] = E[Y_1|Y_1 \geq Y_0 > q_0(\theta)] \geq E[Y_1|Y_1 > q_0(\theta)].$$

We can use this proof to get a sense of when the upper bound is likely to be tight. The bias of the upper bound is

$$b_0 \equiv \mathrm{E}[Y_1|Y_1 > q_1(\theta)] - \mathrm{E}[Y_1|Y_0 > q_0(\theta)],$$

which equals zero when Y_1 preserves the θ-quantile of Y_0. Since some of the applicants induced to take the test by winning the lottery presumably scored above the minimum score achieved by applicants who took the test after losing the lottery, the bound is unlikely to be perfectly tight. However, because few of these "leap-frogging" applicants are likely to have scored in the upper quantiles of the distribution, the likelihood that the θ-QPT assumption holds probably increases with θ. We should therefore expect upper bounds estimated using the proposition to be tighter at the top of the distribution than at the bottom.

REFERENCES

Angrist, Joshua D. "Conditional Independence in Sample Selection Models." *Economics Letters*, 1997, *54*(2), pp. 103–12.

Angrist, Joshua D.; Bettinger, Eric; Bloom, Erik; King, Elizabeth and Kremer, Michael. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review*, 2002, *92*(5), pp. 1535–58.

Angrist, Joshua D.; Bettinger, Eric and Kremer, Michael. "Long-Term Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." National Bureau of Economic Research, Inc., NBER Working Papers: No. 10713, 2004.

Angrist, Joshua and Lavy, Victor. "The Effect of High School Matriculation Awards: Evidence from Randomized Trials." National Bureau of Economic Research, Inc., NBER Working Papers: No. 9839, 2002.

Athey, Susan and Imbens, Guido W. "Identification and Inference in Nonlinear Difference-in-Differences Models." National Bureau of Economic Research, Inc., NBER Technical Working Paper: 280, 2002.

Barnard, John; Frangakis, Constatine E.; Hill, Jennifer L. and Rubin, Donald B. "Principal Stratification Approach to Broken Randomized Experiments: A Case Study of School Choice Vouchers in New York City." *Journal of the American Statistical Association*, 2003, *98*(462), pp. 299–311.

Blundell, Richard; Gosling, Amanda; Ichimura, Hidehiko and Meghir, Costas. "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds." Center for Economic Policy Research, CEPR Discussion Papers: No. 4705, 2004.

Chernozhukov, Victor and Hansen, Christian. "An IV Model of Quantile Treatment Effects." *Econometrica*, 2005, *73*(1), pp. 245–61.

Colombia, Departmento Nacional de Planeacion. *Sistema de Indicadores Sociodemográficos para Colombia (SISD) 1980–1997.* Boletin No. 21. Series. Bogotá: Departmento Nacional de Planeacion, June 1999, p. 58.

Dynarski, Susan. "The Consequences of Merit Aid." National Bureau of Economic Research, Inc., NBER Working Papers: No. 9400, 2002.

Howell, William G. and Peterson, Paul E. *The education gap: Vouchers and urban schools.* Washington, DC: The Brookings Institution Press, 2002.

King, Elizabeth M.; Orazem, Peter F. and Wohlgemuth, Darin. "Central Mandates and Local Incentives: The Colombia Education Voucher Program." *World Bank Economic Review*, 1999, *13*(3), pp. 467–91.

King, Elizabeth; Rawlings, Laura; Gutierrez, Marybell; Pardo, Carlos and Torres, Carlos. "Colombia's Targeted Education Voucher Program: Features, Coverage and Participation." World Bank, Working Paper Series on Impact Evaluation of Education Reforms Paper: No. 3, 2003.

Kremer, Michael; Miguel, Edward and Thornton, Rebecca. "Incentives to Learn." National Bureau of Economic Research, Inc., NBER Working Papers: No. 10971, 2004.

Krueger, Alan B. and Whitmore, Diane M. "The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR." *Economic Journal*, 2001, *111*(468), pp. 1–28.

Krueger, Alan B. and Zhu, Pei. "Another Look at the New York City School Voucher Experiment." National Bureau of Economic Research, Inc., NBER Working Papers: No. 9418, 2003.

Lee, David S. "Trimming for Bounds on Treatment

Effects with Missing Outcomes." National Bureau of Economic Research, Inc., NBER Technical Working Paper: No. 277, 2002.

Manski, Charles F. "Anatomy of the Selection Problem." *Journal of Human Resources*, 1989, *24*(3), pp. 343–60.

Manski, Charles F. "Monotone Treatment Response." *Econometrica*, 1997, *65*(6), pp. 1311–34.

Rouse, Cecilia Elena. "Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program." *Quarterly Journal of Economics*, 1998, *113*(2), pp. 553–602.

The World Bank. *Staff appraisal report: Colombia, Secondary Education Project*. Latin America and the Caribbean Region, Report No. 11834-CO. Washington, DC: World Bank, 1993.

The World Bank. *World development report 1998/99*. New York: Oxford University Press, 1999.