ESSAYS IN EDUCATION ECONOMICS

BILLIE S. DAVIS

APRIL 25^{TH} 2013

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

(ECONOMICS)

FROM CARNEGIE MELLON UNIVERSITY

Doctoral Committee:

Co-Chair Dennis Epple, Thomas Lord Professor of Economics, Carnegie Mellon University

Co-Chair Holger Sieg, J. M. Cohen Term Chair in Economics, University of Pennsylvania

Ron Zimmer, Associate Professor of Public Policy and Education, Vanderbilt University

Daniele Coen-Pirani, Associate Professor of Economics, University of Pittsburgh

Abstract

My dissertation examines the causes of, and one possible program to help alleviate, urban public school district enrollment decline. Enrollment decline has plighted many cities in the Northeast and Midwest over the past two decades, leading to financial struggles and, sometimes, large school closures or reorganizations.

In my first essay, “Determinants of Urban Enrollment Decline”, I describe how changing birth rates, migration out of the Northeast and Midwest, suburban migration, and substitution to charter and private schools have led to disproportionate enrollment losses in large urban public school districts. Specifically, I use a panel of data on public, charter, and private school enrollment and city characteristics to analyze the impact of metropolitan enrollment losses and substitution to charter and private schools on the urban core district’s education “market share”.

My second essay[1], "Bounding the Impact of a Gifted Program On Student Retention Using a Modified Regression Discontinuity Design", examines whether gifted programs can help urban districts retain students with higher SES backgrounds. Gifted programs often employ IQ thresholds for admission, with those above the threshold being admitted. These types of admission rules are often mandated by state rules and create strong incentives to manipulate the IQ score of students to increase access to the program. We propose two new tests that can be used to detect local manipulation of IQ scores. In the presence of local manipulation, the standard regression discontinuity estimator does not identify the local average treatment effect of the program. We show how to modify the approach to construct a lower bound for the effectiveness of the program. This lower bound can be estimated using a modified RD estimator. Our application uses a new and unique data set that is based on applications and admissions to a gifted program of an anonymous urban school district. Our point estimates suggest that there is a favorable effect on retention for students in higher SES households.

[1] Co-authored with John Engberg, a senior researcher at the RAND Corporation; Dennis Epple, the Thomas Lord Professor at Carnegie Mellon University; Holger Sieg, the J.M. Cohen Term Professor of Economics at the University of Pennsylvania; and Ron Zimmer, Associate Professor of Public Policy and Education at Vanderbilt University.

Acknowledgements

Contents

Forthcoming

Chapter 1

Introduction

Forthcoming

Chapter 2: Factors of Urban Enrollment Decline

2.1 Introduction

In many large metropolitan areas throughout the United States, particularly in the Northeast and Midwest (NEMW), urban public school districts have experienced large enrollment declines over the past several decades.[2,3,4] For example, from 1990 to 2009, urban district enrollment fell by 47% in Kansas City, MO and Detroit, 39% in St. Louis, 34% in Cincinnati, 30% in Pittsburgh and Cleveland, and 18% in Philadelphia. (CCD [2012]). Such enrollment losses have led to schools operating below capacity. In response to this underutilization as well as to address academic concerns, many districts have chosen to close school buildings (Sunderman & Payne [2009]). For example, the urban core district in Pittsburgh closed 22 schools in 2006 and Kansas City closed 29 schools in 2010. In March 2013, media sources[5] reported that the Chicago Public School District (CPS), having already closed about 100 schools from 2001 to 2012, would close 54 schools in 2013, pending board approval. If approved, the closures would affect around 30,000 students and would be "one of the largest closures of schools ever seen in America". CPS argues that they need to close "underutilized, under-resourced schools" to help reduce huge budget deficits. Parents and teachers though are outraged and the local teachers union plans to "sponsor training for its members and for parents in non-violent protest such as occupations, demonstrations and other forms of disruption".

Sunderman and Payne (2009) note that the small number of studies researching the impact of school closures have found negative or zero effects on displaced students' achievement (de la Torre & Gwynne [2009], Young et al [2009], Kirshner et al [2009]). More recently, Engberg et al (2011) found that students from closed schools experience adverse effects on test scores and attendance, though these negative impacts can be minimized if students are moved to higher-performing schools. A related body of research on the impact of student mobility,

---

[2] To define metropolitan areas, I utilize the 2003 Metropolitan and Micropolitan statistical area definitions from the U.S. Census Bureau, Population Division, Internet Release Date: June 10, 2003. Each metropolitan area is defined by its component counties. While definitions of metro areas or of component counties may change periodically due to demographic (e.g. new population centers) or legal (e.g. county consolidations) changes, for comparability over time, I fix the definitions. Note that given these definitions, a metropolitan area refers to a large population center, possibly some smaller population centers, and the surrounding areas.

[3] See Appendix B for a map showing the boundaries of each US Census Region.

[4] Throughout the paper, unless otherwise specified, "enrollment" refers to elementary and secondary school students. This includes kindergarten to 12th grades as well as "ungraded" students.

[5] Sources: The Economist: "Education: Class dismissed: The city plans to close lots of schools", Mar 30, 2013, Chicago, From the print edition, accessed online.
New York Times: "Chicago Says It Will Close 54 Public Schools", March 21, 2013 by Yaccino, S. & Rich, M.. Accessed on NYTimes.com on March 22, 2013

outside the context of school closings, has consistently found adverse effects on student outcomes (Hanushek et al. [2004], Booker et al. [2007], Xu et al. [2009], Ozek [2009]). The population changes and school closings also raise concerns about increasing racial and economic segregation in schools and communities (Fischer, et al., [2004]), particularly given "tipping point" behavior in which all whites leave after the minority share reaches a certain level (Card, Mas, and Rothstein [2008]).

Several factors have driven urban enrollment decline. Demographic changes such as lower birth rates and migration out of the Northeast and Midwest have led to decreases in the overall population of school age children in these areas. In metropolitan areas, this has often resulted in "empty desks" in affluent suburban school districts and subsequent migration from the urban core to the suburbs. This is an example of parents "voting with their feet", an illustration of Tiebout (1956) sorting. Thus, cities and their school districts have borne the bulk of population and enrollment losses caused by changing demographics. Recently, substitution to charter schools has also become an increasingly important factor in urban district enrollment decline. Substitution to private school is another potential factor, though the popularity of private schools has declined in recent years.

First, I describe the relevant demographic changes. In the U.S., following a rise from the mid-1970s to 1990, the number of births fell from 1990 to 1996 before rising again through 2009[6]. The number of births directly determines the number of children who will attend first grade approximately 6 years later.[7] Based just on the birth pattern, schools should have had fewer first graders enrolling in 1996 (1990 + 6 years) through 2002 (1996 + 6 years) before enrollment began to rebound. This pattern for 1st graders is very evident in the data, as can be seen in the left panel of Figure 2.1. Figure 2.1 shows the yearly enrollment of all students[8] by grade from 1989 to 2009. The impact seen in 1st grade enrollment "ripples" through the other grades, with the peak moving up by one year for each grade. The effect becomes milder through the grades and the peak in 12th grade enrollment, which should be seen in 2007 (1990 + 17 years), is absent. This may be due to students dropping out or being held back more in later grades.

The birth pattern from 1991 to 2009, though, varied across census regions, as shown in Figure 2.2[9]. In the

[6] Data is through 2009. The rise in births through 1990 is referred to as the "Echo Boom". See Appendix C for more detail.
[7] International migration and deaths also contribute to a small degree.
[8] This includes enrollment in public, charter, and private schools with all designations including regular, medical, special education, alternative education, vocational education, etc.
[9] The number of zero year olds is used as a proxy for the number of births.

South, the number of births declined until the mid-1990s before climbing steadily well beyond the 1990 level. In the West, Midwest, and Northeast, the number of births declined through the late 1990s. In the West, births then gradually climbed again, surpassing the 1990 level in the early 2000s. In the Midwest, births did not reach near the 1990 level until the late 2000s. Finally, in the Northeast, births rose and fell periodically but remained well below the 1990 level.

Figure 2.1: Yearly Enrollment By Grade, United States

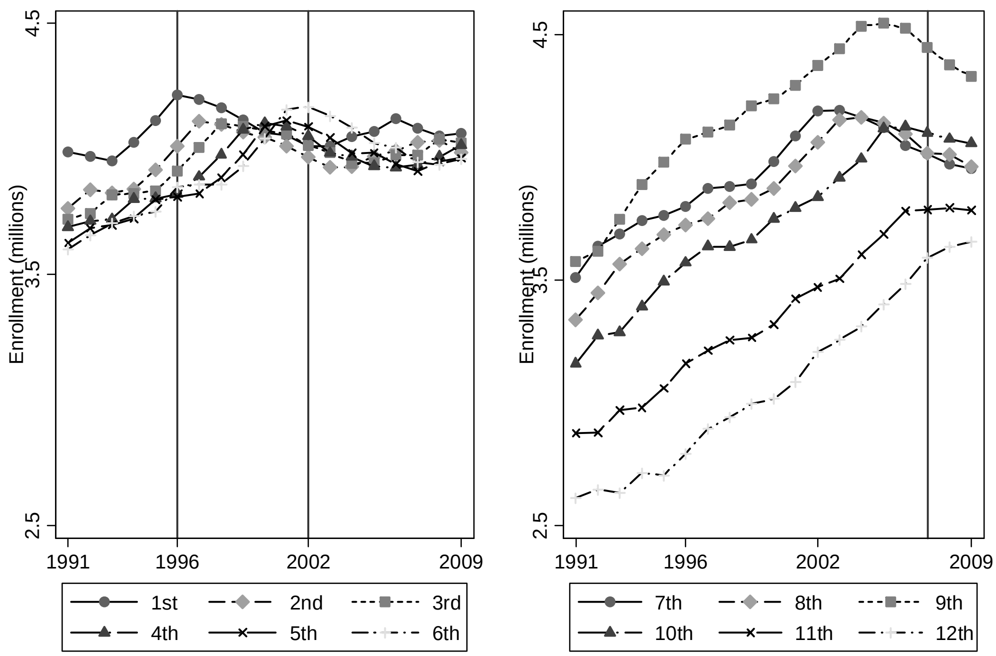
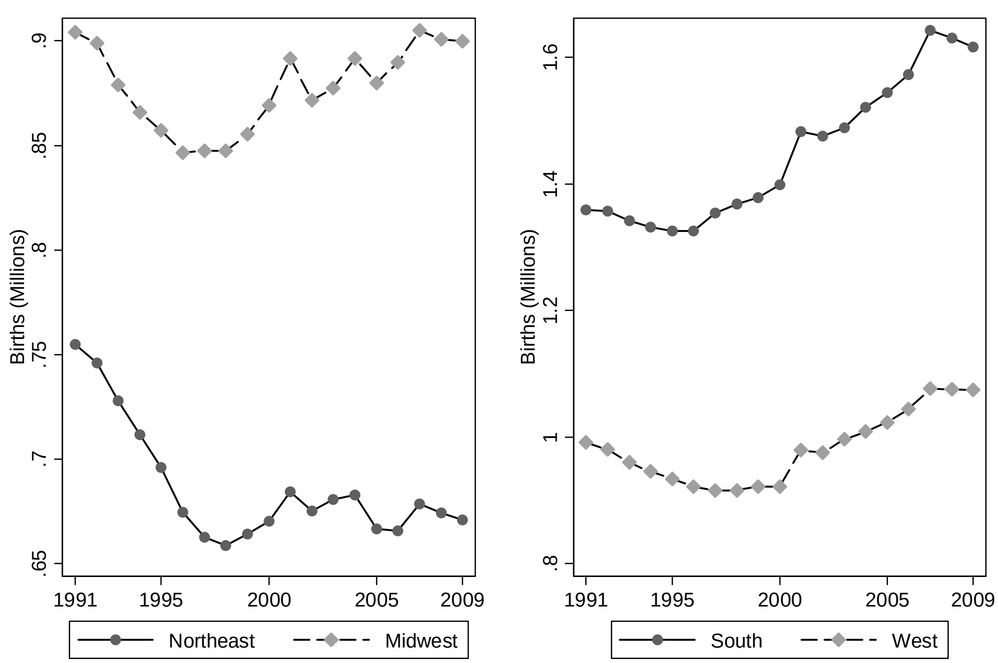


Figure 2.2: Births by Region

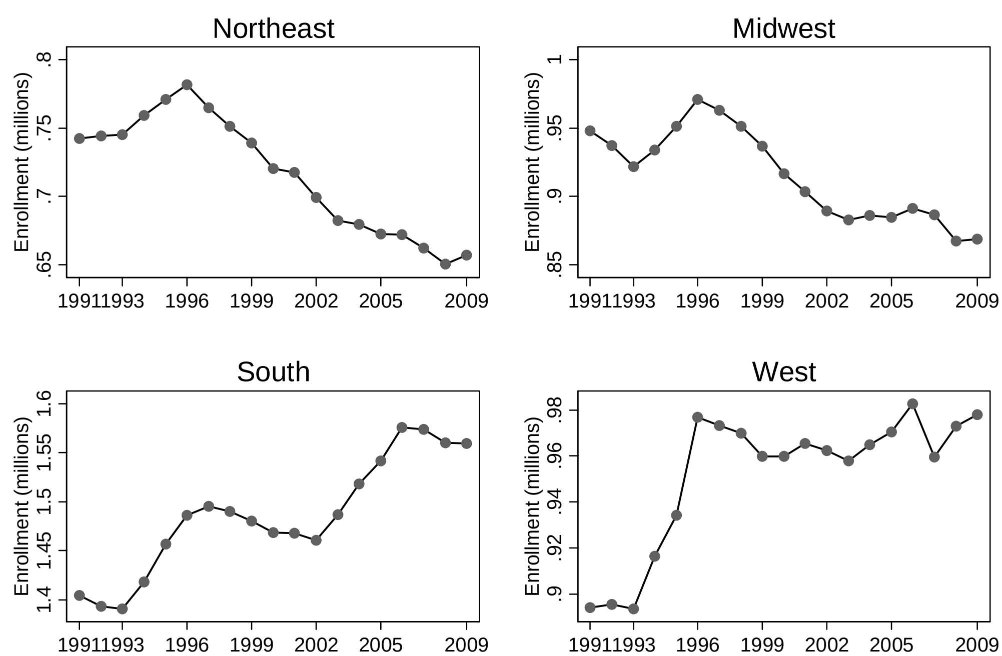


At the regional level, migration is another important factor affecting the number of school age children. Given the birth patterns, absent net in-migration, schools in the Northeast, and somewhat in the Midwest, necessarily experienced enrollment declines because there were simply fewer children being born there. In fact, there was net migration out of these regions. From 1995 to 2000, the Northeast region lost 1.27 million more people than it gained and the Midwest region lost 0.54 million more than it gained. Concurrently, the South gained 1.8 million more people than it lost and the West region ended up about even. See Table 2.1 for more details. This combination of declining births and out-migration translated into steep first grade enrollment losses in the Northeast and Midwest after the 1996 peak. The South and West saw slight losses before rebounding. Figure 2.3 shows these patterns.

Table 2.1 Regional Domestic Migration, 1995 to 2000

	Domestic Migration: 1995 to 2000 (millions)		
	Moved to another region	Moved in from another region	Net
Northeast Region	2.808	1.537	-1.271
Midwest Region	2.951	2.41	-0.541
South Region	3.243	5.042	1.799
West Region	2.654	2.666	0.012

Figure 2.3: Yearly Enrollment in 1st Grade By Region



This evidence shows that demographic changes, specifically lower birth rates and regional migration, led to a wide-spread decline in the number of school-age children in the Northeast and Midwest. In metropolitan areas, however, this decline was concentrated in the urban core due to increased migration into more affluent suburban neighborhoods. The level of substitution from the urban core to the suburbs depends on many factors: relative housing prices, crime rates, and racial, income, and age compositions; physical features such as waterways and highways which may restrict suburban expansion; proximity to green space or cultural attractions (e.g. museums, theaters, stadiums); the relative ease of commuting to the urban core; the expansion of suburban job opportunities; and the location, extent, and quality of school options. The decline in central city population and movement to the suburbs, or "city flight," has been analyzed by demographers, economists, and education researchers. Boustain and Schertzer (2010) give an overview of the relevant literature. Cullen and Levitt (1999), for instance, found evidence that rising crime rates contributed to city depopulation in the 1970s and 1980s. Baum-Snow (2011) found that reduced costs of commuting because of new highways contributed markedly to central city population decline from 1950 to 1990.

In the 18 large NEMW metropolitan areas[10], from 1990 to 2010, the urban core[11] population decreased in eleven of them. In those with urban core population growth, six had urban core growth smaller than suburban growth. In New York City the growth was similar in the city and the suburbs and in Providence the city grew more than the suburbs. The suburban population grew in all of the metropolitan areas except for Pittsburgh. Table 2.2 shows details of these population changes.

Aside from these demographic influences, the introduction and expansion of charter schools has also impacted enrollment in urban traditional public schools. Substitution to charter schools depends on the availability and popularity of such schools in a given metropolitan area. The first US charter school was opened in Minnesota in 1991. As of 2013, nearly 6200 charter schools are operating in 40 states plus the District of Columbia[12], with regulations and popularity varying from state to state (Center for Education Reform [2013]). Charter schools are often located in large metropolitan areas. For example, in Detroit, Minneapolis, and Philadelphia, in 2009, charter school enrollment accounted for more than 5% of total enrollment in the respective metropolitan area.

[10]This includes the metropolitan areas in the Northeast and Midwest census regions with Census 2000 metropolitan area population over 1 million.

[11] The urban core refers to the census-defined "principal city" in each metropolitan area with the largest Census 2000 population as well as any additional principal cities in the metropolitan area with population at least 70% of the largest city. In the large NEMW metropolitan areas, only the Minneapolis-St. Paul metropolitan area includes two principal cities.

[12] As of 2009, Alabama, Kentucky, Maine, Mississippi, Montana, Nebraska, North Dakota, South Dakota, Washington State, West Virginia, and Vermont had no charter schools. Mississippi and Maine have since passed laws allowing charter schools.

Charter schools draw the majority of their enrollment from traditional public schools and the remainder from private schools (Buddin [2012]). Therefore, traditional public schools in metropolitan areas face the most competition for students from the introduction and expansion of charter schools.

Table 2.2: Population Change 1990 to 2010

	Urban Core Population (thousands)			Suburban Population (thousands)		
	1990	Change to 2010		1990	Change to 2010	
		Number	Percent		Number	Percent
Boston	574	43	8%	3,560	375	11%
Buffalo	328	-67	-20%	861	13	1%
Chicago	2,784	-88	-3%	5,398	1367	25%
Cincinnati	365	-68	-19%	1,480	353	24%
Cleveland	505	-109	-21%	1,597	84	5%
Columbus	636	151	24%	769	281	36%
Detroit	1,028	-314	-31%	3,221	362	11%
Hartford	137	-13	-9%	986	101	10%
Indianapolis	732	89	12%	562	373	66%
Kansas City, MO	435	25	6%	1,201	374	31%
Milwaukee	628	-33	-5%	804	157	20%
Minneapolis-St. Paul	641	27	4%	1,898	714	38%
New York City	7,323	853	12%	9,541	1159	12%
Philadelphia	1,586	-60	-4%	3,850	589	15%
Pittsburgh	370	-64	-17%	2,098	-48	-2%
Providence	160	18	11%	1,350	73	5%
Rochester	231	-20	-9%	772	72	9%
St. Louis	397	-77	-20%	2,203	315	14%

Of the 21 states in the Northeast and Midwest, by 2013, all but four (Nebraska, North Dakota, South Dakota, Vermont) had enacted legislation to allow charter schools to operate in the state. The earliest, as noted above, was Minnesota in 1991 while the most recent was Maine in 2011. All of the major metropolitan areas in the NEMW had charter schools by 2009.

Substitution to private schools may also have played a part in urban enrollment decline. Unlike charter schools, private schools are well-established and operate in all states. However, whereas charter schools have steadily gained traction, private schools have generally faced enrollment declines: enrollment in private schools located in metropolitan areas fell by around 6% from 1991 to 2009.[13]

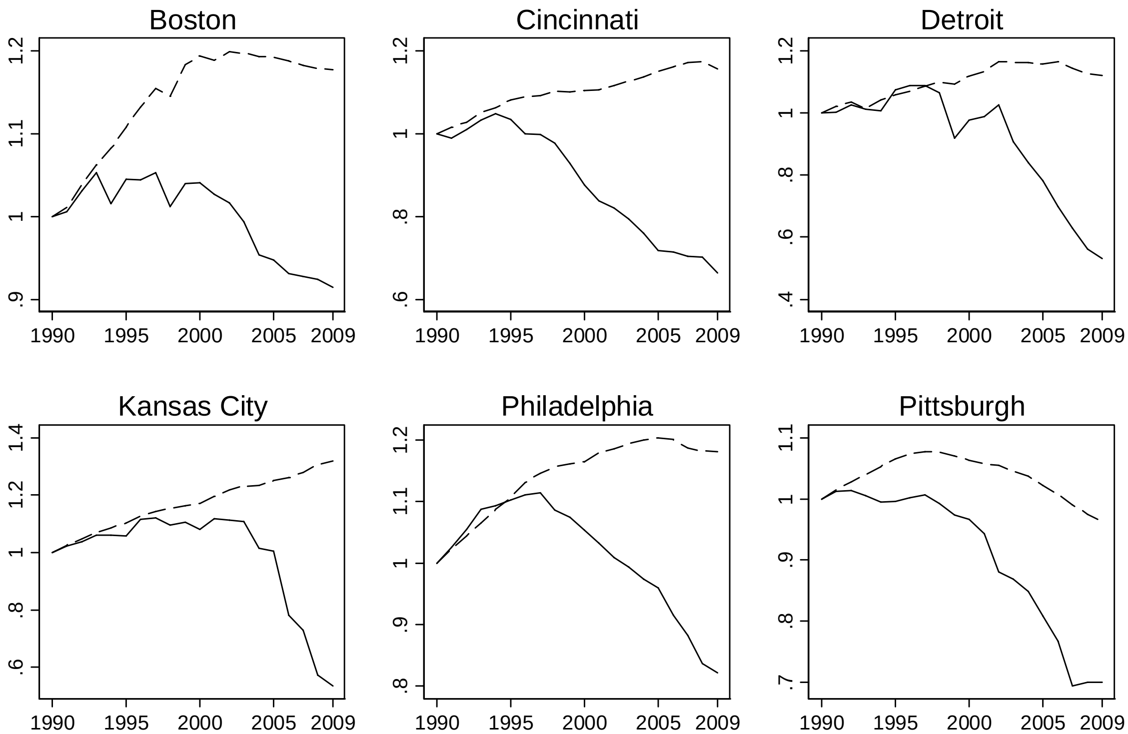
[13] I currently do not include data on the availability of vouchers or similar programs which may help offset the cost of private schools for some families, thereby increasing private schools' competitive advantage.

Together, the above factors have driven enrollment decline with disproportionate effects on urban core districts. Figure 2.4 shows the strikingly different paths of enrollment for the urban core public schools district versus suburban school districts in several of the large NEMW metropolitan areas. The graphs show enrollment relative to (own) 1990 levels with the dashed line for suburban enrollment and the solid line for urban core enrollment.

In my analysis, I consider the urban core district's "market share", i.e. its enrollment with respect to the total enrollment in the metropolitan area. By looking at changes in the urban core's market share instead of its enrollment, I abstract away from the changes in birth rates and migration of the NEMW described above in order to focus on migration to the suburbs and substitution to charter or private schools. I find that, where present, increasing charter school participation leads to urban core districts disproportionately losing students. With respect to overall enrollment losses in a metropolitan area, I find that in metropolitan areas with relatively large urban cores, where it is thus more costly (distance-wise) to move to the suburbs and commute to downtown jobs, the impact on the urban core district is mitigated. Also, in metropolitan areas where the urban core and suburbs are comparably wealthy or where the urban core is richer, the impact on the urban core district is mitigated.

In Section 2.2 I describe the formation and sources of my data sample and present descriptive statistics. In Section 2.3 I present my model specification and analytical results. Finally, in Section 2.4, I conclude.

Figure 2.4: Enrollment: Urban Core v. Suburbs



Enrollment relative to (own) 1990 levels, Dashed Line = Suburban Enrollment, Solid Line = Urban Core Enrollment

2.2 Data

My analysis focuses on enrollment patterns in 17 large metropolitan areas in the Northeast and Midwest U.S.[14] For each metropolitan area, the "education market" is defined as all elementary and secondary public (including charter) and private school districts located therein.[15] I omit districts designated solely as medical, special education, alternative education, or vocational education, since parental choice is limited in these cases. For public schools, I utilize a panel of data covering school years 1986-1987 to 2009-2010 from the National Center for Education Statistic's (NCES) Common Core of Data (CCD)[16]. The data includes fiscal and non-fiscal variables at the district and school level, such as student enrollment and characteristics, graduation rates, teacher and staff information, location, and Census 2000 population characteristics.[17] For private schools, I utilize a panel of data covering every other school year from 1989-90 to 2009-2010 from NCES's Private School Universe Survey.[18,19]

Table 2.3 presents descriptive statistics for the total population and students in the NEMW metropolitan areas. The populations range from 1 million in Rochester, New York to about 13 million in Chicago, with an average population of about 3.1 million. The percent of the population that lives in the urban core (Pop_i) ranges from 10% in Hartford to 38% in Milwaukee, averaging 20%. I define an education ratio, Edu_i, as the ratio of the percent of adults in the urban core population who earned at least a bachelor's degree to the percent of adults in the suburban population who earned at least a bachelor's degree. The education ratio is greater than 1 if urban core adults are more educated than suburban adults. The ratios range from 0.36 in Hartford (suburban adults are about three times as educated as urban adults) to 1.12 in Pittsburgh (urban adults are slightly more educated than suburban adults) with an average of 0.73. Similarly, I define an income ratio, Inc_i, as the ratio of per capita income in the urban core population to per capita income in the suburban population. The income

[14] New York City is omitted due to several major organizational restructurings of its schools.

[15] A district is considered to be in a metropolitan area if any part of it is in a component county. In the case of districts located partially within two adjacent metropolitan areas, the district is classified as in the one where more of its area is located. No district is considered to be in two different metropolitan areas. Because of this broad county based definition of metropolitan areas, I test more narrow definitions in my sensitivity analysis. In order to add better comparability across metropolitan areas in my main analysis, I exclude districts that are located more than 70 miles from the urban core.

[16] Data accessed online at http://nces.ed.gov/ccd/index.asp in 2012. The Common Core of Data (CCD) is a product of the U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

[17] State departments of education often provide more detailed data on school and student characteristics. However, such data is not comparable across states, an important requirement for this study.

[18] Data accessed online at http://nces.ed.gov/surveys/pss/index.asp in 2012. The Private School Universe Survey (PSS) is a product of the U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

[19] Since private school data is only available for every other year, I aggregate private school enrollment to the education market and state level and then estimate the missing years of enrollment by averaging the pre and post values.

ratio is greater than 1 if the urban core population has higher per capita income than the suburban population. The income ratio ranges from 0.48 in Hartford (the suburban population is twice as rich as the urban core population) to 0.89 in Pittsburgh, with an average of 0.70.

Table 2.3 Descriptive Statistics

	Mean	(St. Dev)	Minimum	Maximum
Census 2000				
Metro Population (thousands)	3063	(2894)	1044	12,898
Urban Core Population Percent (Pop_i)	0.20	(0.07)	0.10	0.38
Ratio: Bachelor's Degree + (Edu_i)	0.73	(0.24)	0.36	1.12
Ratio: Per Capita Income (Inc_i)	0.70	(0.11)	0.48	0.89
AY 2000-2001				
Market Enrollment (thousands)	510	(392)	193	1,726
Urban Core Share	0.17	(0.07)	0.09	0.33
Suburban Share	0.70	(0.09)	0.48	0.80
Private Share	0.13	(0.04)	0.08	0.20
Charter Share	0.01	(0.01)	0.00	0.03
Yearly Changes				
%Δ Total Enrollment	0.005	(0.014)	-0.029	0.092
Δ Urban Core Share	-0.002	(0.004)	-0.025	0.015
Δ Charter Share	0.001	(0.002)	-0.001	0.009
Δ Private Share	-0.002	(0.003)	-0.012	0.006

In the middle part of Table 2.3, I present school year 2000-2001 enrollment information which includes total education market enrollment and shares by type of enrollment. Variable definitions are as follows:

Total Enrollment	T = total enrollment in the education market
Urban Core Share	$U = \frac{\text{urban core enrollment}}{\text{total enrollment}}$
Private Share	$P = \frac{\text{private enrollment}}{\text{total enrollment}}$
Charter Share	$C = \frac{\text{charter enrollment}}{\text{total enrollment}}$
Suburban Share	$S = 1 - U - P - C$

Market enrollment (AY 2000-2001) ranged from 193,000 in Rochester, New York to 1.73 million in Chicago, with an average of 510,000. Urban core districts enrolled an average of 17% of students, ranging from 9% in St. Louis to 33% in Milwaukee. Suburban districts enrolled an average of 70% of students, private schools 13%, and charter schools 1%. In 2000, charter schools were present in 12 of the 17 NEMW metropolitan areas, with a maximum share of 3%.

In my analysis, I estimate the effect of percent change in total enrollment and level changes in charter and private share on the urban core district's enrollment share. The bottom portion of Table 2.3 reports the average values for these variables across all years and metropolitan areas. Total enrollment changed very little on average per year (about 0.5%), but ranged from a loss of 3% in Detroit from 2006 to 2007 to a gain of 9% in Minneapolis/St. Paul from 1992 to 1993.

The level changes in urban core share of enrollment were, on average, negative and small. The largest drop in urban core share occurred in Detroit from 1998 to 1999. In 1998 the urban core district had an enrollment of nearly 180,000 which comprised 21.5% of metropolitan area enrollment. From 1998 to 1999, enrollment decreased by about 24,500 students and the share decreased by 2.5 percentage points to 19.0%. At the same time, charter school enrollment grew by over 7000 students, increasing charter share by 0.9 percentage points. Suburban and private schools also lost students, but at a slower pace than the urban core, leading to a 0.2 percentage point increase in private school share and a 1.3 percentage point gain in suburban share. Detroit's urban core district's share decreased by more than 1 percentage point every year from 2003 to 2008 and, by 2009, the share was just 11.5%.

Charter shares grew, on average, by 0.1 percentage points per year, with only three cases of decreases in charter share (Milwaukee and Cincinnati in 2008 and Cincinnati in 2009). Detroit had the three largest gains in charter share, about 0.9 percentage points, in 1998, 1999, and 2005. In contrast, private shares decreased in nearly three-quarters of the cases. The average private share change was -0.2 percentage points per year, ranging from a 1.2 percentage point loss to a 0.6 percentage point gain.

2.3 Analytical Approach & Results

In the basic specification, I model the change in the urban core's share of total market enrollment $\left(\Delta U_{i,t} = U_{i,t} - U_{i,t-1}\right)$ as a linear function of the percent change in total enrollment $\left(\%\Delta T_{i,t} = \frac{T_{i,t} - T_{i,t-1}}{T_{i,t-1}}\right)$, the change in charter share $\left(\Delta C_{i,t} = C_{i,t} - C_{i,t-1}\right)$, and the change in private share $\left(\Delta P_{i,t} = P_{i,t} - P_{i,t-1}\right)$. To account for initial differences in the urban core's share, I include the 1990 level of urban core share $\left(U_{i,1990}\right)$ as a regressor. Finally, I include a constant (β_0) and a random error term $\left(\varepsilon_{i,t}\right)$.

The basic model is:

$$\Delta U_{i,t} = \beta_0 + \beta_1 U_{i,1990} + \beta_2 \Delta C_{i,t} + \beta_3 \Delta P_{i,t} + \beta_4 \%\Delta T_{i,t} + \varepsilon_{i,t}, \qquad t = 1991 \text{ to } 2009 \qquad (1)$$

I extend the model to capture whether changes in total enrollment differentially affect the urban core share in metropolitan areas with a richer versus poorer urban core or a larger versus smaller urban core. I implement this by adding interaction terms between the percent change in total enrollment and (1) the income per capita ratio (Inc_i)[20] and (2) the urban core population percent (Pop_i).[21] The income ratio and population percent are based on Census 2000 data and are constant across years.

The extended model is:

$$\Delta U_{i,t} = \beta_0 + \beta_1 U_{i,1990} + \beta_2 \Delta C_{i,t} + \beta_3 \Delta P_{i,t} + \%\Delta T_{i,t}(\beta_4 + \beta_5 Inc_i + \beta_6 Pop_i) + \varepsilon_{i,t},$$
$$t = 1991 \text{ to } 2009 \qquad (2)$$

Due to the endogeneity of the changes in charter and private share[22], I employ an instrumental variables approach. As instruments, I use the change in the charter and private share in the remainder of the state(s) within which the metropolitan area is located $\left(\Delta C_{remainder,t}, \Delta P_{remainder,t}\right)$, the one period lags of these remainders, and an indicator variable for whether the remainder of the state has any charter school enrollment $\left(I_{remainder\ charter,t}\right)$.

---

[20] Because the correlation between the income ratio and the education ratio is high, I do not include both measures in the specification.

[21] Ideally, I would also include interactions of the change in charter share and the change in private share with Inc_i and Pop_i. However, as noted below, $\Delta C_{i,t}$ and $\Delta P_{i,t}$ are endogenous, so any interaction terms with them are also endogenous. I was unable to find enough appropriate instruments to include such terms in the regression without leading to under-identification and weak-identification problems.

[22] The endogeneity results because all of the school type shares are determined simultaneously: $U + C + P + S = 1$.

I implement the model using two-stage least squares. The results for specification (1) are in in the left half of Table 2.4. For reference, first stage results for the instrumental variables regressions are reported in Appendix Table C.1. The coefficient estimates for the percent change in total enrollment are positive in all cases, indicating that in metropolitan areas where total enrollment is growing, the urban core district gains share. Conversely, in metropolitan areas where total enrollment is declining, the urban core district loses share. The effects are statistically significant in specification (1) but not in specification (2). For the change in charter and private share, the coefficient estimates are statistically significant and negative in all cases. This indicates that as charter (private) share grows the urban core district loses share. Without accounting for the endogeneity of the private and charter shares, the OLS estimates over-state the impact of charter and private share changes.

Table 2.4

	Specification (1)			Specification (2)		
	OLS	OLS	IV	OLS	OLS	IV
$\%\Delta T$	0.104***	0.0786***	0.0835***	0.256	0.180	0.180
	(0.0206)	(0.0166)	(0.0205)	(0.159)	(0.134)	(0.144)
ΔC		-0.856***	-0.716***		-0.831***	-0.779***
		(0.107)	(0.247)		(0.100)	(0.225)
ΔP		-0.476***	-0.385*		-0.468***	-0.331
		(0.0826)	(0.212)		(0.0827)	(0.216)
U_{1990}	-0.0113***	-0.00741***	-0.00768***	-0.0129***	-0.00842***	-0.00809***
	(0.00304)	(0.00253)	(0.00277)	(0.00322)	(0.00272)	(0.00288)
$\%\Delta T * Pop_i$				0.370	0.204	0.304
				(0.291)	(0.260)	(0.289)
$\%\Delta T * Inc_i$				-0.321*	-0.200	-0.227
				(0.188)	(0.147)	(0.165)
Constant	-0.000932	-0.00111*	-0.00113*	-0.000637	-0.000946	-0.000842
	(0.000569)	(0.000564)	(0.000600)	(0.000601)	(0.000609)	(0.000635)
Observations	323	323	306	323	323	306
R-squared	0.153	0.370	0.369	0.176	0.378	0.378

Robust[23] standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

[23] Throughout, this indicates that the standard errors are robust to the presence of arbitrary heteroskedasticity.

Sensitivity Analysis

In my main analysis, I define the education market to include all (relevant) schools within the (census-defined) metropolitan area. However, this definition may overstate the scope of the actual education market that is relevant for families considering where to live or send their children to school. For instance, as seen in Table 2.5 below, in school year 2000, on average across the NEMW metropolitan areas, 11% of enrollment was in schools further than 30 miles from the urban core. To test the sensitivity of my analysis to my education market definition, I repeat the analysis for specification (1) at radii of 20 miles to 50 miles from the urban core district. The results, reported in Appendix C Table C.2, show that my estimates are robust to different education market definitions.

Table 2.5

	Within 10 miles	Within 20 miles	Within 30 miles	Within 40 miles	Within 50 miles	Within Entire City
Percent of Metropolitan Area Enrollment (in school year 2000)	44%	73%	89%	97%	99%	100%

2.4 Conclusions

Forthcoming

Chapter 3: Bounding the Impact of a Gifted Program On Student Retention using a Modified Regression Discontinuity Design[24]

Abstract

Student retention has increasingly become an important issue for large urban districts that have experienced large declines in student enrollments during the last decade. This paper examines whether gifted programs can help urban districts retain students with higher SES backgrounds. Gifted programs often employ IQ thresholds for admission, with those above the threshold being admitted. These types of admission rules are often mandated by state rules and create strong incentives to manipulate the IQ score of students to increase access to the program. We proposed two new tests that can be used to detect local manipulation of IQ scores. In the presence of local manipulation, the standard regression discontinuity estimator does not identify the local average treatment effect of the program. We show how to modify the approach to construct a lower bound for the effectiveness of the program. This lower bound can be estimated using a modified RD estimator. Our application uses a new and unique data set that is based on applications and admissions to a gifted program of an urban school district. Our point estimates suggest that there is a favorable effect on retention for students who are not eligible for free or subsidized lunch.

3.1 Introduction

Student retention has increasingly become an important issue for urban districts, as nearly half of large urban districts in cities classified as large or mid-size by the National Center for Education of Statistics lost students between school years 1999-2000 and 2009-2010.[25] Cities in the Midwest and East Coast have been hit especially hard. Urban districts in Buffalo, Cincinnati, Cleveland, Detroit, Kansas City, Milwaukee, Pittsburgh, and Philadelphia have lost thousands of students over the last several decades. State funding, which is allocated on a per-pupil basis, has shrunk dramatically, and the declining urban population has led to a lower local tax base. However, these districts have buildings, staffing, and pension systems designed for a much larger enrollment base. Downsizing an urban school district is not only costly, but can also impose serious disruptions for students whose schools are closed and need to transfer (Engberg et-al [2011]). Moreover, a decline in enrollment is often

[24] Co-authored with John Engberg, RAND Corporation, Dennis Epple, Carnegie Mellon University and NBER, Holger Sieg, University of Pennsylvania and NBER, Ron Zimmer, Vanderbilt University

[25] Information gathered from National Center for Educational Statistics Common Core data available at http://nces.ed.gov/ccd/bat/index.asp. Using these data, we examined school districts that were either in large or mid-size cities (as classified in the Common Core data) and had at least 10 schools in the 2009-10 school year. In total there were 260 such school districts.

accompanied by a decline in peer quality as high achieving students are more likely to opt out of public schools. As a consequence, declining enrollments often lead to declining achievement for the students that remain in the urban district, exacerbating already existing problems in many urban schools.

These pressures have caused urban districts to search for ways to help maintain enrollment numbers. Districts often look to specialized programs -- such as magnet and gifted programs -- to attract and retain students and households, especially middle class households that have many options in the educational market place. In searching for a school, families may consider not only the quality of facilities, curriculum, and instruction, but also the quality of educational opportunities and peers. Gifted programs create opportunities for students to be stimulated and challenged and have positive peer influences. Often, smaller suburban districts may not have the scale to offer such programs, and, as a result, gifted programs may be a mechanism for retaining strong students within an urban district. The purpose of this paper is to estimate the treatment effect of admittance into a gifted program on student retention using new and unique data from an anonymous urban district.

Evaluating the impact of gifted programs is challenging. Gifted programs are not randomly assigned to districts and may develop in districts as a function of unobservable family characteristics. Moreover, students are typically not randomly assigned to gifted programs. Rather students are admitted based upon IQ scores. The admission by IQ scores raises the possibility of using a regression discontinuity (RD) design, which was recently used by Bui, Craig, and Imberman (2011) to estimate the impact of an urban district's gifted program on test score outcomes.[26] In theory, using an RD design to estimate the retention effect of the gifted program should be straightforward. Since students are admitted into gifted programs based on an IQ threshold, we could just use the IQ score as the forcing variable for an RD design. However, IQ examinations are oral tests, which provides psychologists some discretion in scoring. A psychologist may "give the benefit of the doubt" in assessing performance of students who are near the threshold for admission to a gifted program. In our data, many test scores appear to be manipulated. Similar issues arise in other settings as concerns for transparency can lead to promulgation of the criteria for admission into programs. Knowing the criteria for admission, participants may undertake activities that alter the reported outcome on the variable that determines admission. For example, students who fall below the threshold on a test determining whether they will be subject to remediation may retake the test to attempt to obtain a score above the threshold (Calcagno & Long [2008]).

[26] The regression discontinuity design was first used by Thistlethwaite & Cambell (1960). Some well-known applications of the RD design in education include Angrist & Lavy (1999), van der Klaauw (2002), DiNardo & Lee (2004), and Jacob & Lefgren (2004). For a guide for implementing RD designs, see Imbens & Lemieux (2008).

In our case, theory suggests that manipulation of IQ scores should be local in nature. The costs of manipulation are too large for students that have IQ scores below a threshold since they would not benefit from the advanced curriculum and might create negative peer effects for more advanced students in the gifted program. Moreover, there is obviously no need to manipulate the test scores of students who have scores above the admission threshold. Local manipulation of IQ scores then gives rise to a non-standard error in variables problem.

We provide two tests for manipulation of IQ scores. These tests help us determine the range over which IQ scores are likely to be manipulated. Both tests exploit the local nature of the manipulation. The first test is based on a monotonicity property of the density of the IQ distribution near the admission threshold. Because IQ scores are standardized to a normal distribution, when the admission threshold is above the mean, the density of the IQ score is monotonically decreasing around the threshold. If there is no manipulation, we should see a higher fraction of students below versus above the threshold. With local manipulation, however, some students with true IQ scores below the threshold have reported IQ scores above the threshold. If the fraction of students whose scores are manipulated is sufficiently large, we will observe the opposite of our expectation, i.e., there will be more students with scores above versus below the threshold.[27]

The second test for manipulation is based on scores on achievement tests that are often given along with the IQ test. Achievement test performance is not directly referenced in the admission guidelines of the gifted program, and, more importantly, the scores do not appear to be manipulated. We can, therefore, use the observed IQ score to predict the achievement score. Local manipulation creates an error-in-variables problem over a bounded interval of values. Using a regression model to predict the achievement score, we can test for parameter stability over the manipulation range. We implement this stability test using a standard Chow test.

The second test for manipulation is based on additional achievement tests that are often available along with the overall IQ score. Some of these achievement tests are not directly referenced in the admission guidelines of the gifted program. More importantly, they do not appear to be manipulated. We can, therefore, use the observed IQ score to predict the additional achievement score. Local manipulation creates an error-in-variables problem over a bounded interval of values. Using a regression model to predict the achievement score, we can test for parameter stability over the manipulation range. We can implement this stability test using a standard Chow test.

[27] McCrary (2008) provides an alternative framework for testing for manipulation. He tests the null hypothesis of continuity of the underlying density function at the program cut-off points. In our application, the IQ test is measured in discrete increments which leads to our alternative tests.

One consequence of manipulation is that IQ scores at the admission threshold may not be valid instruments for program participation, which invalidates the key identifying assumption of the standard regression discontinuity design. However, we can exploit the local nature of the manipulation process and construct an estimator for a sharp lower bound of the relevant treatment effect. The key additional assumption that we need to invoke is that the mean outcome in the untreated state is monotonically declining in the IQ score. In our application, this assumption is natural and implies that parents with students that have higher IQ's are less likely to stay in the district in the absence of a gifted program. More able students are likely to have more alternatives, including being sought after by private schools. Under this assumption, we can construct a sharp lower bound of the treatment effect of the gifted program. The basic idea is to compare the mean outcome of students with the highest non-manipulated IQ score below the threshold with the outcome of students with the lowest non-manipulated IQ score above the threshold. Our approach draws on the pioneering work by Manski (1997) who first suggested exploiting monotone treatment responses to construct sharp bounds for treatment effects. We show that a similar idea can be used within a regression discontinuity design with a locally manipulated forcing variable.[28] Moreover, we show that we can implement this estimator using a modified RD estimator.

Our application focuses on a gifted program operated by a mid-sized urban school district that prefers not to be identified. We implement our estimation strategy for a sample of students tested for the gifted program while attending a district school in school years 2004-05 to 2007-08. We find large, significant point estimates which provide strong evidence for a favorable retention for higher income students.

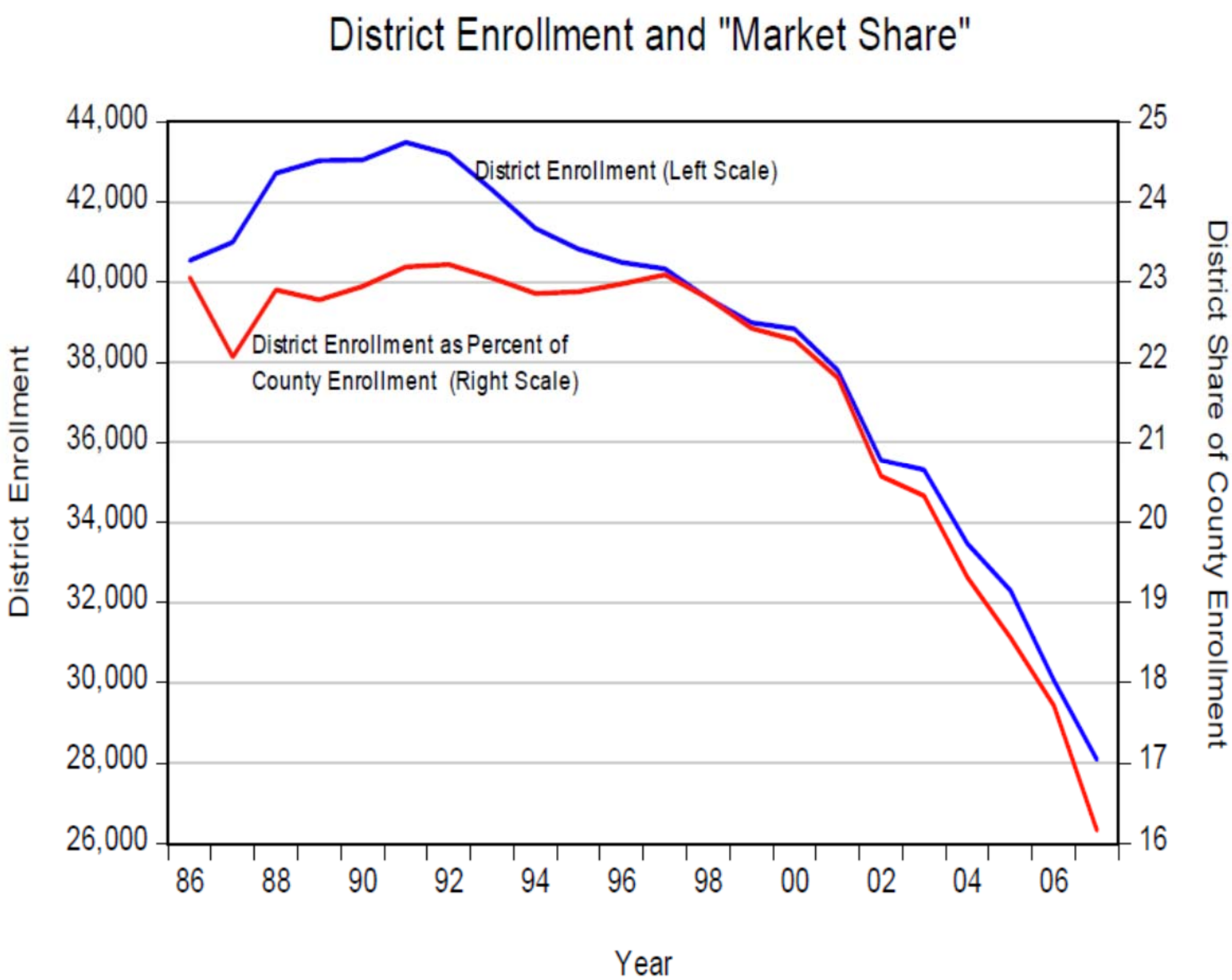
With the development of a bound analysis within an RD framework, this paper not only makes a contribution to the important policy question of whether gifted programs are a mechanism for retaining students, but also provides a new method that can be used in other educational contexts in which test scores, or other measures, are manipulated to gain entrance into a program. The rest of the paper is organized as follows. Section 3.2 provides information about our data set and describes the testing and admission procedures used in the gifted program studied in our application. Section 3.3 discusses estimation and inference in an RD design with local manipulation. Section 3.4 presents the empirical findings of our paper. Section 3.5 offers some conclusions.

[28] See also Manski & Pepper (2000) who extend this framework using monotone instrumental variables.

3.2 Institutional Background and Data

Student retention has become one of the key challenges faced by urban school districts. Figure 1 plots the market share of the urban district considered in this paper relative to the broader educational market (measured by all districts in the county.) The district was maintaining its student share during the 1990's when enrollment was rising in the market. When countywide enrollment began to decline in 1998, the district not only shared in the countywide decline in the student population, but experienced a further decline as more affluent households exited the city and moved up the school district income hierarchy ("voting with their feet"). The combination of these two effects resulted in the district bearing 75 percent of the countywide decline in public school enrollment.

Figure 1: Market Share of the Urban District



One promising tool to make an urban school district attractive to students and parents is to offer special education programs that cannot be provided by smaller districts. Gifted programs are one prominent example of such programs. Gifted and talented programs have a long history in the U.S., dating back to the late 19th century. However, gifted programs did not receive federal support until 1958 when the federal government established the National Defense Education Act. This act initiated federal support for specialized programs for math, science, and foreign languages (Bhatt [2009]). More recently, the federal government expanded its support to gifted programs through the Jacob Javits Gifted and Talented Educational Act in 1988 and the No Child Left Behind Act in 2002. Through these initiatives, gifted programs have gained popularity, especially in urban districts. For urban districts, these programs have the dual objective of engaging and challenging gifted students to reach advanced levels of achievement as well as attracting and retaining students who might otherwise leave for suburban or private schools. Despite receiving federal support, gifted programs are not mandated by the federal government. Individual states or districts decide if and how to use gifted programs, including how students are identified (Shaunessy [2003]).

Despite the fact that there are currently over 3 million students in gifted programs, these programs have generally been ignored by researchers (Bhatt [2011]). In reviewing the small literature up to that point, Vaughn, Feldhusen, and Asher (1991) conducted a meta-analysis of nine papers and found that participation in pull-out gifted programs led to improved achievement, critical thinking, and creativity, but student's self-concepts were not affected. However, these studies often had difficulties dealing with endogeneity issues associated with students self-selecting into the programs. More recently, Bui, Craig, and Imberman (2011), using an RD design to address the self-selection problem, examined the impact of a gifted program on test scores in an anonymous urban district and found that these programs did have a positive impact on language and science test scores after two years of participation, but did not have an effect in other subjects.[29]

The school district that we study in this paper operates a gifted program that is quite large in scope. Approximately 10 percent of the students in the district participate in some type of gifted education. Gifted students in grades 1 to 8 participate in a one-day-per-week pull out program at a designated location away from the student's home school. Students enroll in programs designed to enhance creative problem solving and leadership skills and are offered specially designed instruction in math, science, literature, and a variety of other

[29] A closely related literature considers school tracking, which sorts students into different tracks based upon ability. See, for example, Zimmer (2003), Figlio & Page (2002), Betts & Shkolnik (2000), Argys, Rees, & Brewer (1996), Hoffer (1992), and Kerckhoff (1986). However, this research has not examined the impact of gifted or tracking programs on retaining students.

fields. For high school students, gifted education is available within the school and involves the annual design of an individualized education program, full-time curricula, and a number of other enhancements.

The district adheres to state regulations concerning gifted students and services. The state regulations outline a multifaceted approach used to identify whether a student is gifted and whether gifted education is needed. A mentally gifted student is defined as someone with an IQ of at least 130 points or someone who shows outstanding intellectual and creative ability using other educational criteria. Further, to qualify for gifted services, the district must show that the student requires services or programs not available in regular education.

The state guidelines stress that IQ cannot be the only factor used in determining gifted ability. Specifically, low scores in memory or processing speed tests cannot be used alone to disqualify a student. Also, even if a student has an IQ below 130, she may be deemed gifted based on above grade level achievement on standardized tests, a superior rate of acquisition or retention of new academic content or skills, excellence in specific academic areas, or other factors that indicate superior functioning. Additionally, the guidelines specifically note that the gifted decision must account for any potential masking of gifted abilities because of disability, socio/cultural deprivation, gender or race bias, or English as a second language. Further, it is emphasized that the gifted decision may not be based on a single test or type of test. For limited English proficiency or students of racial-, linguistic-, or ethnic-minority background, it is specifically noted that an IQ score may not be used as the only measure to show low aptitude.

The evaluation process begins when a parent, teacher, administrator, or student requests a gifted evaluation. Once the student's parents are notified and give consent for the evaluation, a team consisting of parents, a certified school psychologist, teachers, and others familiar with the student's educational experience and performance or cultural background conducts the evaluation. The evaluation must include information on academic functioning, learning strengths, and educational needs. The information and findings from the evaluation of the student's educational needs and strengths is combined by the team into a written report. This report includes the team's recommendation as to whether the student is gifted and in need of specially designed instruction. Finally, the report is evaluated in a team meeting where the decision is made regarding the student's eligibility for gifted education.

The district adheres to the preceding guidelines for evaluating potential gifted students. As noted, one way to support a claim of giftedness is to show superior performance (above 130 points) on an intelligence test. In our district, every student considered for the gifted program is given some type of intelligence test. During the time-frame of our analysis (school years 2004-2005 to 2007-2008) district psychologists mainly used the Wechsler Intelligence Scale for Children, 4th edition (WISC4) test instrument. The WISC4 gives four index scores measuring verbal comprehension (VCI), perceptual reasoning (PRI), working memory (WMI), and processing speed (PSI). It also gives a "Full Scale IQ" (FSIQ) which combines the results from the four indexes and a "Generalized Ability Index" (GAI) which combines the results from the VCI and PRI. The FSIQ, indexes, and GAI are normed, by age, to be representative of the current population of children in the United State and have a mean of 100 and a standard deviation of 15. Thus, a score of 130 is two standard deviations above the mean.

In the district, each student takes an intelligence test and is then categorized as meeting the IQ criteria if the FSIQ or GAI is 130 or above.[30] Students with a FSIQ or GAI of 125 to 129 or a VCI or PRI of 130 or above do not meet the IQ criteria but do qualify for special further consideration through a portfolio evaluation. Students who score below these cutoffs may still be considered gifted based on a further review of other factors. In practice, the probability of a regular lunch student being admitted into the gifted program increases most at the portfolio cutoff of 125 points. (See Figure 3.2 in Section 3.4.)

To investigate the impact of the gifted program on student retention, we start with a set of 1726 students first tested for the gifted program in school years 2004-05 to 2007-08 while attending a district public school.[31] Of these students, we keep 806 who do not receive subsidized lunch[32]. Students eligible for subsidized lunch face different IQ thresholds to be considered for admission into the gifted program. Moreover, retention effects are likely to be more important for higher income households since they have access to more educational options for their children. We exclude 141 students who have more than one set of test results for gifted consideration reported at any point up until summer 2009. Dropping observations with multiple tests removes potential manipulation of the type uncovered by Calcagno and Long (2008). Next, we drop 21 students not tested by a district psychologist, since a private psychologist hired by a student's parent may have incentives to inflate

[30] Note that the GAI excludes processing speed and working memory sub-tests. Hence, the GAI offers a way to address the state requirement pertaining to not excluding students based solely on low scores in memory and processing speed.
[31] Note that this excludes students attending a private school when tested. This is in order to isolate retention effects of the program. While attraction effects are also of interest, they are beyond the scope of this paper.
[32] Throughout our analysis, a designation of subsidized lunch means that the student was tagged as receiving subsidized lunch at some point in the district's data from school years 1999 to 2009. Thus, it is a constant variable by student (as are race and sex).

scores and scores would likely only be reported to the district if the score is above the admission thresholds. Additionally, parents who hire a private psychologist may differ on unobservables. We also omit 26 students with nonverbal test scores. Recall that nonverbal tests are given to students who are culturally or linguistically diverse. Thus, these students likely differ on unobservables. Finally, for data comparability, we exclude 9 students with invalid or missing data and 69 students not administered the WISC4 test instrument. This leaves us with our final sample of 540 students.

In addition to IQ and achievement test results for students being evaluated for gifted admission, we have longitudinal student level data for all district students in school years 2004-05 to 2007-08. The data includes information about race, gender, standardized test scores, subsidized lunch status, school attended, home census tract, etc. Table 3.1 reports descriptive statistics for the state (data column 1), the district (data column 2), and students in our sample (data columns 3 and 4).[33] Comparing the district to the state, we find that the district has a similar proportion of males and a larger proportion of African American students than the state. Also, district students, on average, live in neighborhoods with lower median income and fewer college educated adults and perform below the state average on a 5th grade state-wide standardized test in math and reading.

Next, comparing students in our sample to all district students, we see that tested students have a similar proportion of males but a smaller proportion of African American students than the district. Also, tested students come from relatively richer and more educated neighborhoods and score higher on the standardized tests. Admitted students come from even richer and more educated neighborhoods and score even higher on both math and reading. These demographic differences are partly due to the fact that our sample excludes students who receive subsidized lunch.

Considering just the students in our sample, we see that tested and admitted students, unsurprisingly, have a higher average IQ than students who were tested but not admitted. Finally, Table 3.1 shows that the unconditional mean of one and two year retention for admitted students is higher than that for students tested but not admitted. This suggests that admittance into the gifted program may have some positive impact on retention.

[33] State averages for race, gender, and test scores are for school year 2005-06. State averages for income and college are from Census 2000. District averages are for students in kindergarten through 12th grade enrolled in the district at some time between 2004 and 2007. Standard deviations are in parentheses.

Table 3.1: Descriptive Statistics

	District (K-12)	Tested & Admitted	Tested & Not Admitted
Male	0.506	0.508	0.509
African American	0.559	0.078	0.174
Income	28,868 (11,153)	43,555 (17,244)	37,231 (11,946)
College	.185 (0.157)	.442 (0.234)	.282 (0.208)
Math	1308 (221)	1732 (178)	1513 (154)
Reading	1246 (217)	1585 (142)	1446 (154)
Offenses	.877 (1.884)	.02 (0.140)	.131 (0.410)
IQ		131.6 (7.630)	112.0 (8.219)
1 year retention		.918 (0.274)	.844 (0.364)
2 year retention		.837 (0.370)	.737 (0.441)
Count	47,506	319	224

Standard deviations in parentheses.

3.3 Estimation and Inference in an RD Design with Local Manipulation

3.3.1 The Fuzzy Regression Discontinuity Design

We adopt standard notation in the program evaluation literature and consider a model with two potential outcomes.[34] Let D denote an indicator variable that is equal to one if a person receives treatment and zero otherwise. In our application, treatment is admittance into the gifted program. Let Y_1 denote the outcome with treatment and Y_0 the outcome without treatment, where the outcome of interest is retention in the district. The researcher observes:

$$Y = DY_1 + (1 - D)Y_0 \quad (1)$$

The difference from receiving the treatment is defined as

$$\Delta = Y_1 - Y_0 \quad (2)$$

and note that this gain is unobserved for every single person in the sample. In terms of the treatment effect, the model can be written as

$$Y = Y_0 + D\Delta \quad (3)$$

[34] See Quandt (1972), Rubin (1974, 1978), and Heckman (1978, 1979).

We focus on the "fuzzy" regression discontinuity design in which the probability of receiving the treatment changes discontinuously at certain points in the support of a forcing variable. In our application the forcing variable is ability measured by an IQ score. Let Z denote the observed IQ score. Under the fuzzy design D is a random variable given Z. The propensity score is defined as:

$$E(D|Z) = \Pr(D = 1|Z) \quad (4)$$

The RD design is then based on three assumptions. The first assumption formalizes the notion that the probability of admission into the program is discontinuous at a known point z_0

Assumption 1: $Pr(D = 1|Z)$ *is known to be discontinuous at* z_0.

Following Hahn, Todd, and van der Klaauw (2001), identification of the treatment effect can be established using the following argument. Let $e > 0$ denote an arbitrarily small positive number. Then

$$E[Y|Z = z_0 + e] - E[Y|Z = z_0 - e] = E[\Delta D|Z = z_0 + e] - E[\Delta D|Z = z_0 - e] + E[Y_0|Z = z_0 + e] - E[Y_0|Z = z_0 - e] \quad (5)$$

The second assumption of the RD design guarantees that the second term in equation (5) vanishes:

Assumption 2: $E[Y_0|Z]$ *is continuous at* z_0.

Hence

$$\lim_{e \to 0} E[Y_0|Z = z_0 + e] - E[Y_0|Z = z_0 - e] = 0 \quad (6)$$

The third assumption of the RD design is:

Assumption 3: D *is conditionally independent of* Δ *near* z_0.

This assumption implies that the first term in equation (5) can be written as:

$$E[\Delta D|Z = z_0 + e] - E[\Delta D|Z = z_0 - e] = E[\Delta|Z = z_0 + e]\, E[D|Z = z_0 + e] - E[\Delta|Z = z_0 - e]\, E[D|Z = z_0 - e] \quad (7)$$

Assumptions 1-3 together imply that the treatment effect is identified from the following equation:

$$E[\Delta|Z = z_0] = \lim_{e \to 0+} \frac{E[Y|Z = z_0 + e] - E[Y|Z = z_0 - e]}{E[D|Z = z_0 + e] - E[D|Z = z_0 - e]} \quad (8)$$

3.3.2 Local Manipulation of the IQ Score

If the observed IQ score Z is manipulated, the true IQ score must be treated as a latent variable. We observe instead a manipulated score denoted by $\tilde{Z}$. We make the following assumption regarding the extent of the manipulation:

Assumption 4

1. The IQ is not manipulated for values that are sufficiently far away from the cut-off point z_0, i.e., there exists a known value $\underline{z}$ such that if $Z_i < \underline{z}$ then $\tilde{Z}_i = Z_i$.

2. Scores above z_0 are not manipulates, i.e. $Z_i \geq z_0$ implies $\tilde{Z}_i = Z_i$.

Assumption 4 implies that manipulation is local in nature. Only IQ scores between $\underline{z}$ and z_0 are potentially manipulated. This assumption is plausible since the district psychologists have no incentives to manipulate scores above the admission threshold. Moreover, the cost of admitting unqualified students to the program is increasing in the distance from the threshold. The detrimental effects to the program through lower peer quality are one example. This implies that students with sufficiently low scores should not be admitted into the program. In that case, district psychologists have no incentive to manipulate the scores of students below a given lower bound denoted by $\underline{z}$.[35]

However, not all IQ scores in the relevant range are manipulated. Let M denote a random variable which is equal to one if the observation is subject to manipulation and zero otherwise. We assume that district psychologists engage in selective manipulation:

Assumption 5:

1. $M = 1$ with probability P^M. In that case

$$\tilde{Z} = z_0 + E \quad (9)$$

where E is an individual specific manipulation error.

[35] It is fairly straightforward to derive this result in a simple theoretical model in which a decision maker optimally manipulates scores to increase access to the gifted program.

2. Moreover there exists a known value $\bar{z}$ *such that* $z_0 \leq \tilde{Z} \leq \bar{z}$.

Assumption 4 and 5 imply that observations outside the range [$\underline{z}$,$\bar{z}$] are not manipulated. Note that it is straightforward to allow P^M to be a function of observables and unobservables. Strategic manipulation arises if P^M is a function of U, i.e. if $P^M = P^M(Z, U)$. If P^M is an increasing function of U, then students that are more likely to be retained are receiving favorable treatment and vice versa.

3.3.3 Bounding the Treatment Effect

With local manipulation of the IQ score the treatment effect is no longer point identified under the assumptions made above. Here we construct a lower bound of the treatment effect based on a monotonicity assumption of the treatment effect and exploiting the fact that manipulation is only local in nature.

Assumption 6: *For values in the interval* [$\underline{z}$,$\bar{z}$], *we assume that* $E[Y_0|Z]$ *is monotonically decreasing in* Z.

This type of monotonicity assumption was first used by Manski (1997) who showed how to exploit monotone treatment responses to construct sharp bounds for treatment effects. Here we show that a similar idea can be used within a regression discontinuity design with locally manipulated forcing variables to construct sharp bounds for treatment effects.

A slightly weaker assumption is that $E[Y_0|Z = \underline{z}] > E[Y_0|Z = \bar{z}]$. In our application, this assumption means that parents with students who have higher IQ's are less likely to stay in the district in the absence of a gifted program. As we noted earlier, this is a natural assumption because more intelligent students will have more attractive outside alternatives.

Finally, we assume that the expected marginal treatment effect is constant over the interval [$\underline{z}$,$\bar{z}$]:

Assumption 7: $E[\Delta|\tilde{Z} = \underline{z}] = E[\Delta|\tilde{Z} = z_0] = E[\Delta|\tilde{Z} = \bar{z}]$.

We then obtain the following result:

Proposition 1: A sharp lower bound for the treatment effect of the program is given by:

$$LB = \frac{E[Y|\tilde{Z} = \bar{z}] - E[Y|\tilde{Z} = \underline{z}]}{Pr\{D = 1|\tilde{Z} = \bar{z}\} - Pr\{D = 1|\tilde{Z} = \underline{z}\}} \qquad (10)$$

Proof: Assumptions 4 and 5 imply that $Z\ (= \tilde{Z})$ is observed without error outside the manipulation range $[\underline{z}, \bar{z}]$. Thus all the conditional expectations in equation (10) are identified. Assumption 1 implies that the denominator of the right hand side of equation (10) is greater than zero. Moreover, we have

$$E[Y|Z = \bar{z}] - E[Y|Z = \underline{z}] = E[\Delta D|Z = \bar{z}] - E[\Delta D|Z = \underline{z}] + E[Y_0|Z = \bar{z}] - E[Y_0|Z = \underline{z}] \qquad (11)$$

Assumption 6 implies that

$$E[Y|Z = \bar{z}] - E[Y|Z = \underline{z}] < 0 \qquad (12)$$

Assumptions 3[36] and 7 imply that

$$E[\Delta D|Z = \bar{z}] - E[\Delta D|Z = \underline{z}] = E[\Delta|Z = z_0]\ (\Pr\{D = 1|Z = \bar{z}\} - \Pr\{D = 1|Z = \underline{z}\}) \qquad (13)$$

Combining these results, we have:

$$\frac{E[Y|Z = \bar{z}] - E[Y|Z = \underline{z}]}{Pr\{D = 1|Z = \bar{z}\} - Pr\{D = 1|Z = \underline{z}\}} < E[\Delta|Z = z_0] \qquad (14)$$

We thus have a lower bound for the treatment effect. Q.E.D.

3.3.4 Detecting Local Manipulation

Here we discuss two tests for local manipulation of the IQ score. The first test is based on a monotonicity assumption of the density of the IQ distribution.

Assumption 8: *We assume that the IQ threshold, z_0, that determines access to the programs is sufficiently high that the density of the true IQ score is monotonically decreasing around the cut-off point.*

[36] We need the slightly stronger assumption of conditional independence for all values in $[\underline{z}, \bar{z}]$, not just those near z_0.

The monotonicity assumption is realistic for the whole underlying population. However, the students that take the IQ test are not a random sample of the underlying population. If parents and teachers can accurately predict who will pass the admission test, one may expect that this assumption could be violated in the selected sample of test takers.

In practice, IQ scores are reported in discrete intervals. Therefore, to construct a test for local manipulation it is useful to treat Z as a discrete random variable. The proportion of students with scores that are within a bandwidth of h below the cut-off is denoted by $\Pr\{z_0 - h \leq Z < z_0\}$.

Similarly, the proportion of students in the same length interval above the threshold is $\Pr\{z_0 \leq Z < z_0 + h\}$. In the absence of manipulation, $\Pr\{z_0 - h \leq Z < z_0\} > \Pr\{z_0 \leq Z < z_0 + h\}$. We test to see if this holds.

The second test exploits the availability of a non-manipulated achievement test score denoted by X. By definition, we have

$$X = E[X|Z] + V \tag{15}$$

Notice that $E[X|Z]$ is not known for values in the range *[$\underline{z}$,$\bar{z}$]* since IQ scores are potentially manipulated in that range. We therefore need to impose a functional form assumption:

Assumption 9: *The functional form of $E[X|Z]$ is known up to a parameter value that can be consistently estimated based on the non-manipulated subsample.*

For example, if $E[X|Z] = \alpha_0 + \alpha_1 Z$, we can consistently estimate the parameters of this function based on the non-manipulated sample using OLS. In the absence of manipulation a regression of Xon $\tilde{Z}$ should give a consistent estimator of the parameters of the regression model whether or not the observations in the range *[$\underline{z}$,$\bar{z}$]* are used.

If manipulation occurs, however, a regression of X on $\tilde{Z}$ using the full sample will not yield a consistent estimator due to the error-in-variables problem that is created by manipulation. Regressions using observations below $\underline{z}$ and/or above $\bar{z}$ will give consistent estimators in the presence of manipulation. This suggests a strategy for investigating manipulation. In particular, we conduct Chow tests to assess whether the parameters estimated using observations in the interval *[$\underline{z}$,$\bar{z}$]* differ significantly from parameters estimated using observations above and below that interval. In addition, Assumption 9 can be tested by investigating stability between observations below and observations above the interval $[\underline{z},\bar{z}]$.[37]

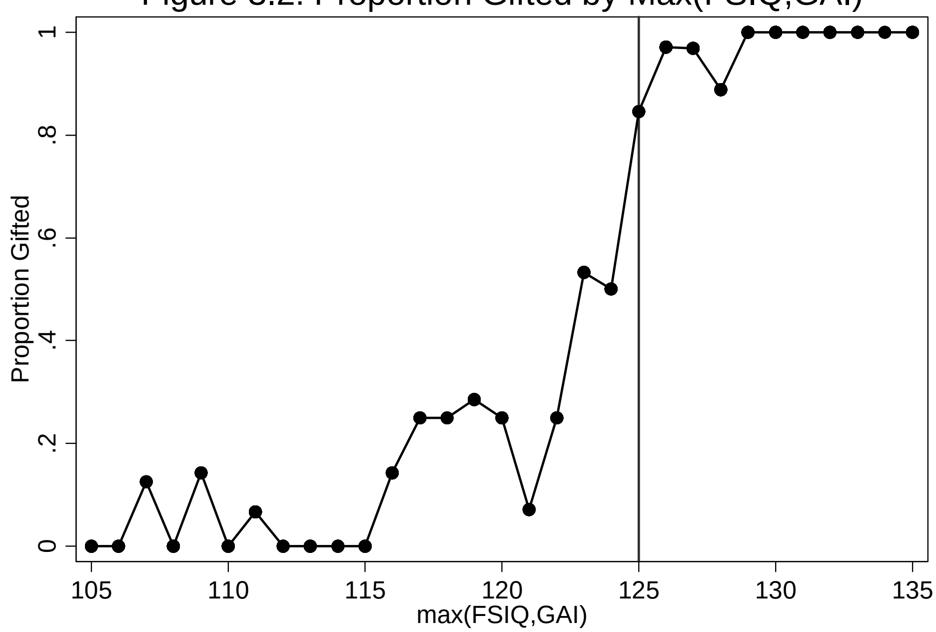
An alternative and potentially more robust way to implement this test is to (i) estimate the parametric model for $E[X|Z]$ on the range $(-\infty, \underline{z})$; (ii) test whether this model holds on $(\bar{z},\infty)$, and (iii) test whether it holds on $[\underline{z},\bar{z}]$. If the test is not rejected on $(\bar{z},\infty)$, this can be viewed as a good indication that the model is valid for the full support.

3.4 Empirical Results

3.4.1 Manipulation of IQ Scores

Since the district's regulations are in terms of both FSIQ and GAI, a natural starting point for the analysis is to consider the maximum of these two scores as the forcing variable that determines access to the gifted program. Figure 3.2 shows the proportion of students who are gifted as a function of the maximum of the FSIQ and GAI score.

Figure 3.2: Proportion Gifted by Max(FSIQ,GAI)

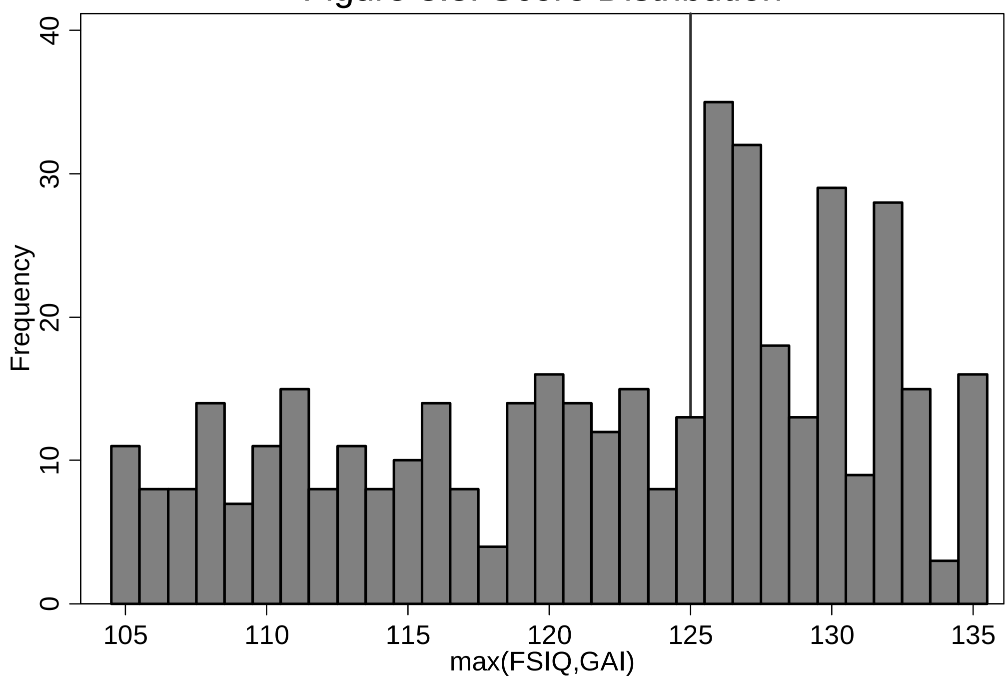


[37] We have assumed that $\underline{z}$ and $\bar{z}$ are known to the econometrician or at least can be bounded from above and below. In most applications that assumption will be valid. In any empirical application, it is useful to conduct some robustness analyses. We discuss these checks in the next section.

We find that higher scores generally correspond to a higher proportion of students admitted into the gifted program. There is the largest jump in proportion gifted at the cut-off 125.[38]

For the fuzzy RDD approach to be valid, the distributions of any covariates, including the forcing variable, should not show a discontinuity at the cut-off. Here we encounter a puzzling feature of the distribution of the main forcing variable. The maximum of FSIQ and GAI does not exhibit a smooth frequency distribution. Figure 3.3 plots a histogram of the distribution. We find that the distribution is heavily skewed to the right around the cut-off point of 125. This finding is robust to a number of sensitivity checks. For example, the patterns are not driven by any one administering psychologist or by testing in one particular grade or school year.[39]

Figure 3.3: Score Distribution



Next, we implement the two tests discussed above to provide more formal evidence of manipulation of the forcing variable.

[38] Note that there are some students who are admitted into the program without meeting the IQ requirements. This is consistent with the requirement that students not be rejected solely on failure to meet the IQ thresholds.
[39] The protocol for converting raw scores into IQ scores creates spikes at certain values of the distribution, as evidenced by spikes at 132 and 135. This phenomenon may also be a contributor to spikes earlier in the distribution, but this phenomenon alone cannot account for the evidence of manipulation documented below.

The first test exploits a monotonicity assumption of the non-manipulated IQ distribution. Recall that this means that in the absence of manipulation we should see $\Pr\{z_0 - h \le Z < z_0\} > \Pr\{z_0 \le Z < z_0 + h\}$. We implement the first test for a variety of different bandwidth parameters. We also consider two samples: the full sample and a subsample of students for whom we also have additional achievement test scores, namely from the Wechsler Individual Achievement Test (WIAT). Note that we can only implement the second manipulation test for the WIAT subsample. The empirical results associated with the monotonicity test are summarized in Table 3.2.

Table 3.2: Monotonicity Test

	Full Sample		WIAT subsample	
BW	Proportion Below Cutoff	Observations in Interval	Proportion Below Cutoff	Observations in Interval
2	0.31	70	0.27	44
3	0.30	114	0.28	76
4	0.33	146	0.32	104
5	0.37	175	0.36	127
6	0.36	218	0.36	163

Table 3.2 provides strong evidence to suggest that IQ scores are locally manipulated. The null hypothesis that the proportion of observations below the cut-off is at least 50% is rejected at the 1% level at all bandwidths for the full sample and for the WIAT subsample. That is, we find that, contrary to the expected relation, the proportion of students in the interval at or above the IQ cut-off of 125 is significantly higher than the proportion below the cut-off. For instance, at a bandwidth of 6, 36% of all students in the interval are below the 125 cutoff while the remaining 64% are at or above the cutoff.

To implement the second test for manipulation, we need a non-manipulated achievement test score. Natural candidates are the WIAT scores for numerical operations and word reading. These scores are not used in the gifted admission decision and thus are unlikely to be manipulated. Using the WIAT subsample, we implement Chow tests for IQ manipulation. Table 3.3 summarizes the main findings of the Chow tests using the two WIAT scores as dependent variables.

Test 1 in Table 3.3 tests Assumption 9 by investigating the stability of parameters across observations below (Group A) and observations above (Group C) the manipulation range. Consistent with Assumption 9, we find no evidence to reject the hypothesis of parameter stability across these two intervals outside the manipulation range. In particular, the tests of parameter stability for the two WIAT scores yield p-values of 0.64 and 0.55. In

Test 2, we compare parameter stability for observations outside the manipulation range (Groups A and C) to observations in the manipulation range (Group B). For Numeric Operations we find a p-value of .04, which provides relatively strong evidence of manipulation. We do not find evidence for manipulation when we use Word Reading as the dependent variable for Test 2. Test 3 compares observations below the manipulation range (Group A) to observations in the manipulation range (Group B). Here the findings echo those from Test 2. We find evidence of manipulation using Numerical Operations but not using Word Reading. Overall, the two testing strategies provide relatively strong evidence of manipulation. Under these circumstances, the standard Fuzzy Regression Discontinuity approach would clearly be problematic.

Table 3.3: Chow Test

	Test 1		Test 2		Test 3	
	Group A vs Group C		Groups A&C vs Group B		Group A vs Group B	
	Numeric Operations	Word Reading	Numeric Operations	Word Reading	Numeric Operations	Word Reading
IQ	0.518**	0.680***	0.684***	0.539***	0.518**	0.680***
2nd Group (dummy)	-7.297	30.18	-75.70**	-24.3	-93.98**	-8.66
2nd Group * IQ	0.0876	-0.248	0.589**	0.198	0.755**	0.0573
Constant	48.14**	33.75**	29.86***	49.39***	48.14**	33.75**
Observations	232	232	382	382	279	279
R-squared	0.375	0.39	0.275	0.313	0.162	0.271
Joint Test:						
F(2, n-4)	0.45	0.59	3.19	0.52	2.76	0.4
P-value	0.6385	0.5533	0.0422	0.5941	0.0652	0.6683

Robust standard errors, *** p<0.01, ** p<0.05, * p<0.1

Group A: Observations below manipulation range: [100,119]

Group B: Observations in the manipulation range: [120,130]

Group C: Observations above the manipulation range: [131,150]

Test 1: Chow test of Group A vs Group C; 2nd Group is Group C

Test 2: Chow test of combined Groups A&C vs Group B; 2nd Group is Group B

Test 3: Chow test of Group A vs Group B; 2nd Group is Group B

3.4.2 Retention Effects

Given that evidence above of local manipulation of IQ scores, the standard RD estimator is likely to be biased. Therefore, we implement our modified RD estimator which provides a sharp lower bound of the treatment effect. We can estimate this lower bound using the Imbens (2007) approach for standard RD estimators. In our application, a conservative estimate for the manipulation range is $\underline{z}$=120 and $\bar{z}$=130.[40]

We can then write the outcome equation as:

$$Y_i = \beta_0 + \beta_1 1\{Z_i < 120\}(120 - Z_i) + \beta_2 1\{Z_i > 130\}(Z_i - 130) + \Delta D_i + v_i \quad (16)$$

where Z_i is the observed IQ score. For a bandwidth of five, for instance, we only use data in the intervals [115,119] and [131,135] to estimate the model. The new instrument is $1\{Z_i > 130\}$.

We implement this estimator for a variety of different bandwidths and use both local linear and local constant estimators. Table 3.4 reports the first stage of the 2SLS estimator that implements our lower bound.

Table 3.4: First Stage of the Modified RDD

Full Sample						
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Difference	0.775***	0.82***	0.842***	0.661***	0.624***	0.626***
St. Err	(0.067)	(0.055)	(0.049)	(0.024)	(0.017)	(0.014)
Obs.	95	121	140	95	121	140
WIAT Subsample						
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Difference	0.742***	0.805***	0.83***	0.636***	0.588***	0.591***
St. Err	(0.080)	(0.063)	(0.055)	(0.030)	(0.021)	(0.017)
Obs.	73	95	111	73	95	111

The first stage results show that the proportion of students admitted into the gifted program is larger above the IQ cut-off. These results are statistically significant at various bandwidths in both the constant and linear specifications. For instance, in the full sample, there are 40 students with IQ's 116, 117, 118, or 119 (bandwidth 4 below the manipulation range) and 55 students with IQ's 131, 132, 133, or 134 (bandwidth 4 above the

[40] Robustness analysis regarding these cut-off points is discussed below.

manipulation range). Of those below the cut-off, 22.5% are admitted while 100% of those above are admitted. Thus, the difference in the constant specification is 77.5%.

Next, we investigate whether admittance into the gifted program helps the district retain students. We use one and two year retention (whether a student is in a district school one year or two years after being tested) as the outcome variables.[41] The main results are summarized in Table 3.5. Again we conduct a variety of robustness checks and implement the estimator on two different samples.

Table 3.5: Second Stage of the Modified RDD

	Full Sample: One Year Retention					
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Gifted	0.114	0.112	0.068	-0.175	-0.124	0.062
St. Error	(0.075)	(0.061)	(0.056)	(0.115)	(0.089)	(0.098)
p-value	0.129	0.068	0.226	0.128	0.166	0.53
Obs.	95	121	140	95	121	140
	Full Sample: Two Year Retention					
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Gifted	0.279**	0.228**	0.162*	0.447	0.448	0.575**
St. Error	(0.113)	(0.093)	(0.085)	(0.316)	(0.283)	(0.271)
p-value	0.013	0.014	0.058	0.158	0.113	0.034
Obs.	95	121	140	95	121	140
	WIAT Subsample: One Year Retention					
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Gifted	0.066	0.075	0.027	-0.108	-0.117	0.062
St. Error	(0.085)	(0.066)	(0.061)	(0.124)	(0.106)	(0.122)
p-value	0.435	0.257	0.656	0.382	0.271	0.609
Obs.	73	95	111	73	95	111
	WIAT Subsample: Two Year Retention					
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Gifted	0.318**	0.233**	0.152	0.681*	0.658	0.762**
St. Error	(0.139)	(0.109)	(0.10)	(0.387)	(0.350)	(0.338)
p-value	0.022	0.033	0.127	0.078	0.06	0.024
Obs.	73	95	111	73	95	111

[41] A student is considered in the district two years after testing if she graduated from a district school one year after testing.

Table 3.5 provides no clear evidence that the gifted program increases one year retention. The point estimates are positive in the majority of the reported cases, but standard errors are large and the overall effects are insignificant. This result for one year retention is not too surprising since it would likely take some time for families of non-admitted students to respond and find suitable alternatives. Considering two year retention, though, we find positive and statistically significant results. For the constant model, the treatment effect is approximately 25 percentage points. The point estimates are even larger for the local linear specification which is the preferred method for implementing RDD estimators.

Finally, we conducted sensitivity analysis with respect to the lower limit of the manipulation range. In the previous analysis, we used 120 as the lower limit. In Table 3.6 we show results using 119 as the lower limit for the range. The results between the two analyses are quite similar. If anything, the main findings are stronger using 119 as the lower bound since one year retention effects are then significant for the constant specification. We also implemented the estimator using 121 as the lower limit of the manipulation range and again we find the same results. We thus conclude that our main empirical findings are robust to the choice of the manipulation range.

Summarizing the empirical findings, the relatively large positive point estimates for two year retention suggest that there is a favorable effect on retention for students in our sample, i.e., those not on subsidized lunch. These effects are statistically significant in most cases.

Table 3.6: Robustness Analysis: Second Stage of the Modified RDD

	Full Sample: Lower Limit 119: One Year Retention					
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Gifted	0.151*	0.126*	0.131**	0.139	0.248	0.161
St. Error	(0.079)	(0.065)	(0.062)	(0.233)	(0.20)	(0.195)
p-value	0.055	0.051	0.035	0.551	0.215	0.407
Obs.	91	114	137	91	114	137
	Full Sample: Lower Limit 119: Two Year Retention					
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Gifted	0.228**	0.188**	0.147*	0.84	0.703*	0.578*
St. Error	(0.104)	(0.090)	(0.080)	(0.589)	(0.417)	(0.305)
p-value	0.028	0.036	0.067	0.154	0.092	0.058
Obs.	91	114	137	91	114	137
	WIAT subsample: Lower Limit 119: One Year Retention					
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Gifted	0.111	0.091	0.106	0.304	0.314	0.11
St. Error	(0.088)	(0.070)	(0.070)	(0.416)	(0.307)	(0.262)
p-value	0.204	0.192	0.128	0.466	0.306	0.674
Obs.	70	88	108	70	88	108
	WIAT subsample: Lower Limit 119: Two Year Retention					
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Gifted	0.236*	0.186*	0.136	1.808	1.251*	0.837*
St. Error	(0.122)	(0.105)	(0.093)	(1.159)	(0.684)	(0.429)
p-value	0.053	0.076	0.143	0.119	0.067	0.051
Obs.	70	88	108	70	88	108

3.5 Conclusions

Student retention has increasingly become an important issue for urban districts that have experienced large declines in student enrollment during the past decade. This paper examines whether a gifted program can help an urban district retain talented students. We focus on students from higher income families (i.e., those not on subsidized lunch) since such students and families are the most mobile yet the most important for districts to keep in order to maintain the tax base.

Many urban districts operate gifted programs that are large in scope and can be expensive to maintain, making the effectiveness of retaining students an important policy question. Gifted programs often employ IQ thresholds for admission, with those above the threshold being admitted. These types of admission criteria are often mandated by state rules. Unfortunately, such rules create strong incentives to manipulate students' IQ scores in order to increase access to the program. We have proposed two tests that can be used to detect local manipulation of IQ scores. Our tests show that manipulation of IQ scores is present in our data. One consequence of manipulation is that the standard regression discontinuity estimator does not identify the local average treatment effect of the gifted program.

In this paper, we show how to modify the standard RD approach to construct a lower bound for the effectiveness of the program. This lower bound can be estimated using a modified RD estimator. This estimator may prove fruitful for studying retention in other RD applications - especially in education settings since test scores are often used for admission into schools or specialized programs and in awarding scholarships. In these settings, as in ours, manipulation may be "local," i.e., the scope for successful manipulation will be greatest in the vicinity of the threshold. Such settings offer the potential for application of the bounding approach developed in this paper.

Our point estimates suggest that there is a favorable effect on retention for more affluent students, i.e., students who are not eligible for subsidized lunch. These results are encouraging and suggest that parents value gifted programs and are more likely to keep their children in urban schools if the children are recognized as gifted and obtain special instruction.

BIBLIOGRAPHY

Angrist, J. & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics*, 114, 533-75.

Baum-Snow, N. (2007) “Did Highways Cause Suburbanization?” *Quarterly Journal of Economics*. 122(2), pp.775-805.

Betts, J. & Shkolnik, J. (2000). Key Difficulties in Identifying the Effects of Ability Grouping on Student Achievement. *Economics of Education Review*, 19:1, 21-6.

Bhatt, R. (2009). The Impacts of Gifted and Talented Education. Andrew Young School of Policy Studies Research Paper Series No. 09-11.

Bhatt, R. (2011). A Review of Gifted and Talented Education in the United States. *Education, Finance, and Policy*, 6:4, 557-582

Booker, K., Gilpatric, S.M., Gronberg, T., Jansen, D. (2007). The impact of charter school attendance on student performance. *Journal of Public Economics* 91, 849–876.

Boustain, L, and Schertzer, A (2010). “Demography and Population Loss from Central Cities, 1950 to 2000”, *NBER Working Paper Series*. Working Paper 16435.

Buddin, Richard (2012). How Are Charter Schools Affecting Public and Private School Enrollments?

Bui, S., Craig, S. & S. Imberman (2011). Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Programs on Achievement. NBER Working Paper 17089.

Calcagno, J. & Long, B. (2008). The impact of postsecondary remediation using a regression discontinuity approach: addressing endogenous sorting and noncompliance. CCRC Working Paper.

Card, D, Mas, A, and Rothstein, J. (2008). “Tipping and the Dynamics of Segregation”, *The Quarterly Journal of Economics*. 123(1), pp.177-218.

Center for Education Reform (2013). Charter School Law Rankings and Scorecard in Charter School Law and Rankings Across the States.

Cullen, J. B., & Levitt, S. D. (1999). Crime, urban flight, and the consequences for cities. *Review of economics and statistics*, *81*(2), 159-169.

de la Torre, M., & Gwynne, J. (2009). When schools close: Effects on displaced students in Chicago Public Schools. Chicago: Consortium on Chicago School Research, Chicago.

DiNardo, J. & Lee, D. (2004). Economic impacts of new unionization on private sector employers: 1984-2001. *Quarterly Journal of Economics*, 119 (4), 1383-1441.

Engberg, J., Gill, B., Zamarro, G. & Zimmer, R. (2011). Closing schools in a shrinking district: Do student outcomes depend on which schools are closed? *Journal of Urban Economics*.

Figlio, D. & Page, M. (2002). School choice and the distributional effects of ability tracking: Does separation increase inequality? *Journal of Urban Economics*, 51, 497-514.

Hahn, J., Todd, P., & van der Klaauw, W. (2001). Identification and estimation of treatment effects with regression discontinuity design. *Econometrica*, 69 (1), 201-09.

Hanushek, E.A., Kain, J.F., Rivkin, S.G. (2004). Disruption versus Tiebout improvement: the costs and benefits of switching schools. *Journal of Public Economics* 88, 1721–1746.

Heckman, J. (1978). Dummy Endogenous Variables in a Simultaneous Equation System. *Econometrica*, 46, 931-960.

Heckman, J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47 (1), 153-161.

Hoffer, T. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis*, 14 (3), 205-27.

Imbens, G. (2007). Regression discontinuity design. NBER Lecture Notes: What's New in Econometrics?

Imbens, G. & T. Lemieux (2008). Regression Discontinuity Designs: A Guide to Practice. *Journal of Labor Economics*, 142, 615-35.

Jacob, B. & Lefgren, L. (2004). The Impact of Teacher Training on Student Achievement: Quasi-experimental Evidence from School Reform Efforts in Chicago. *Journal of Human Resources*, 39, 50-79.

Kirshner, B., Gaertner, M., Pozzoboni, K. (2009). Tracing transitions: a multiple methods analysis of the impact of school closures on displaced students. *Educational Evaluation and Policy Analysis,* 32, 407–429.

Kerckhoff, A. (1986). Effects of Ability Grouping in British Secondary Schools. *American Sociological Review*, 51, 842-58.

Manski, C. (1997). Monotone treatment response. *Econometrica*, 65, 1311-1334.

Manski, C. & Pepper, J. (2000). Monotone instrumental variables: with an application to the returns to schooling.. *Econometrica*, 68, 997-1010.

McCrary, J. (2008). Manipulation of the forcing variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142:2, 698–714.

Ozek, U. (2009) The Effects of Open Enrollment on School Choice and Student Outcomes. The Urban Institute, National Center for the Analysis of Longitudinal Data in Education Research, Washington.

Quandt, R. (1972). A new approach to estimating switching regression models. *Journal of the American Statistical Association*, 67, 306-310.

Rubin, D. (1974). Estimating causal e_ects of treatment in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701.

Rubin, D. (1978). Bayesian inference for causal effects. *Annals of Statistics*, 6, 34-58.

Shaunessy, E. (2003). State policies regarding gifted education. Gifted Child Today, 26, 16-21.

Sunderman, G.L., & Payne, A. (2009). Does Closing Schools Cause Educational Harm? A Review of the Research. Information Brief. Mid-Atlantic Equity Center, Arlington, VA.

Thistlethwaite, D. & D.Cambell (1960). Regression-discontinuity analysis: an alternative to the ex post factor experiment. *Journal of Educational Psychology*, 51, 309-17.

Tiebout, C. (1956), "A Pure Theory of Local Expenditures", *Journal of Political Economy,* 64 (5): 416–424.

van der Klaauw, W. (2002). Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression Discontinuity Approach. *International Economic Review*, 43 (4), 1249-87.

Vaughn, V. L., Feldhusen, J. F., & Asher, J. W. (1991). Meta-analysis and review of research on pull-out programs in gifted education. *Gifted Child Quarterly*, 35 (2), 92-98.

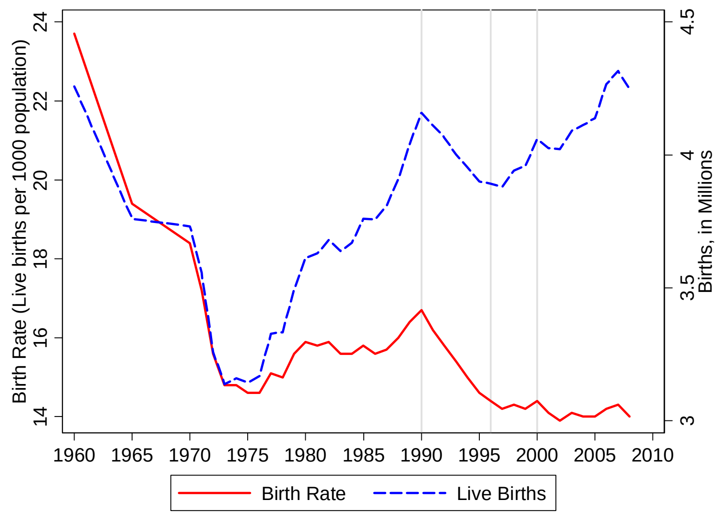
Xu, Z., Hannaway, J., D'Souza, S. (2009). Student transience in North Carolina: the effect of mobility on student outcomes using longitudinal data. National Center for Analysis of Longitudinal Data in Education Research, Working Paper 22.

Young, V.M., Humphrey, D.C., Wang, H., Bosetti, K.R., Cassidy, L., Wechsler, M.E., Rivera, M., Murray, S., Schanzenbach, D.W. (2009). Renaissance Schools Fund-Supported Schools: Early Outcomes, Challenges, and Oportunities. SRI International, Menlo Park, CA.

Zimmer, R. (2003). A new twist in the tracking debate. *Economics of Education Review*, 22, 307-15.

Appendix A: The Echo Boom

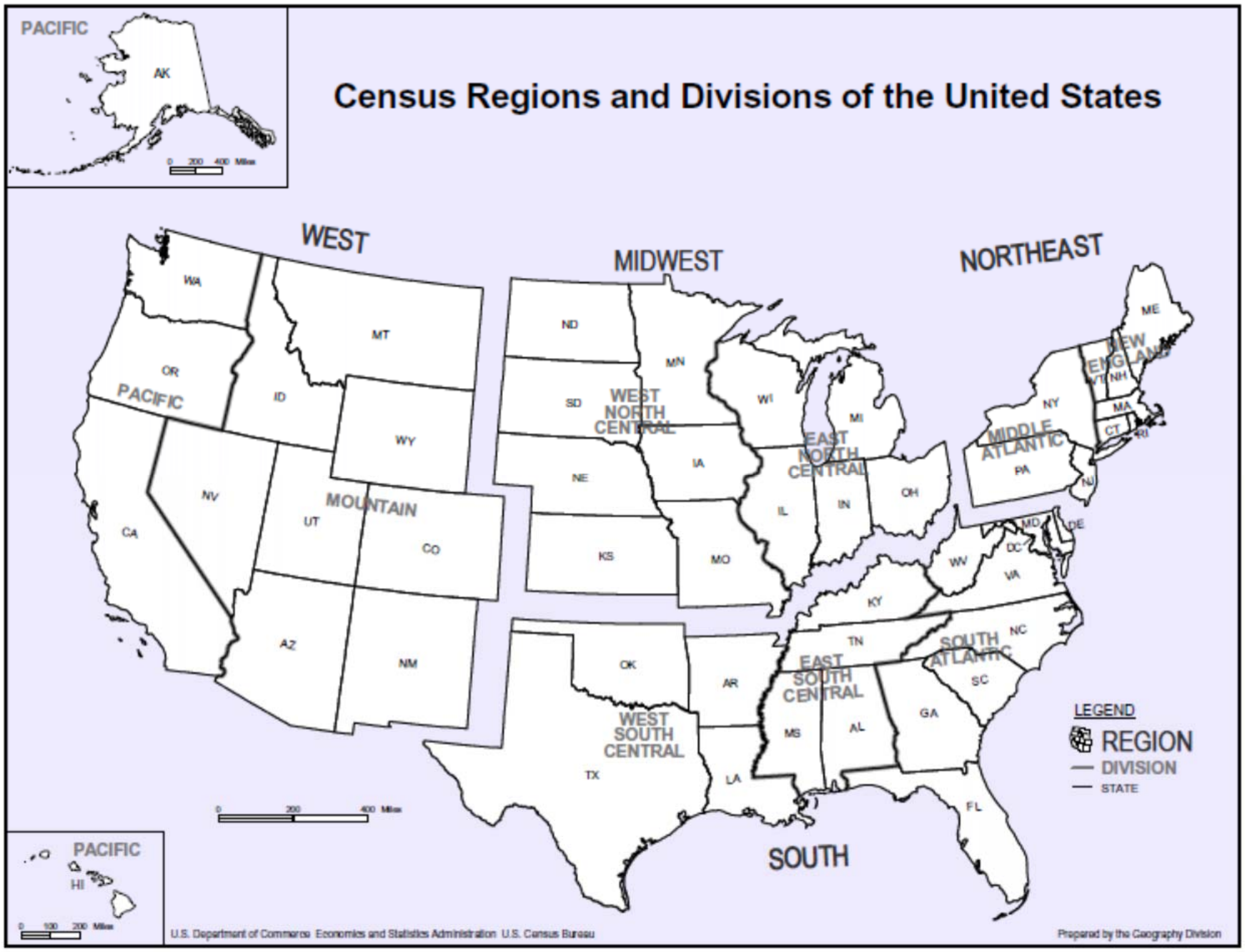
During the post-World War II “Baby Boom” in 1946 to 1960, the birth rate in the United States hovered around 24 to 26 live births per 1000 population.[42] The birth rate then fell to a low of 14.6 in the mid-1970s before climbing back up during the “Echo Boom” of the late 1970s and 1980s. While not nearly as dramatic as the post-WWII boom, the Echo Boom saw an increase in the birth rate to 16.7 at its peak in 1990. In the 1990s and 2000s, the birth rate steadily declined and fell below pre-Echo levels, stabilizing around 14 in the 2000s and falling to 13.5 in 2009.[43] The total number of live births (not normed for population) follows a similar pattern, increasing until 1990 and then decreasing until 1995. Then, instead of leveling off, the number of births began to rebound and rose somewhat steadily through 2009. These trends are shown in Figure X below. The solid red line shows birth rates (scale on left-hand y-axis) and the blue dashed line shows the total number of live births (scale on right-hand y-axis).



[42] Vital Statistics of the United States, 1993. Volume I, Natality, Table 1-1 "Live births, birth rates, and fertility rates, by race: United States, 1909-2003."

[43] Centers for Disease Control and Prevention. National Center for Health Statistics. VitalStats. http://www.cdc.gov/nchs/vitalstats.htm. [4/24/2012].

Appendix B: Census Regions and Divisions of the United States



Source: https://www.census.gov/geo/www/us_regdiv.pdf

Region 1 (Northeast)

- Division 1 (New England) Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut
- Division 2 (Mid-Atlantic) New York, Pennsylvania, New Jersey

Region 2 (Midwest)

- Division 3 (East North Central) Wisconsin, Michigan, Illinois, Indiana, Ohio
- Division 4 (West North Central) Missouri, North Dakota, South Dakota, Nebraska, Kansas, Minnesota, Iowa

Region 3 (South)

- Division 5 (South Atlantic) Delaware, Maryland, District of Columbia, Virginia, West Virginia, North Carolina, South Carolina, Georgia, Florida
- Division 6 (East South Central) Kentucky, Tennessee, Mississippi, Alabama
- Division 7 (West South Central) Oklahoma, Texas, Arkansas, Louisiana

Region 4 (West)

- Division 8 (Mountain) Idaho, Montana, Wyoming, Nevada, Utah, Colorado, Arizona, New Mexico
- Division 9 (Pacific) Alaska, Washington, Oregon, California, Hawaii

Appendix C

Table C.1 presents the first stage estimates for the instrumental variables estimations of specifications (1) and (2).

Table C.1: First Stage Estimates

	Specification (1)		Specification (2)	
	ΔC	ΔP	ΔC	ΔP
ΔC_{rem}	0.206	0.162	0.192	0.163
	(0.151)	(0.140)	(0.147)	(0.139)
ΔC_{rem}, lag 1	0.477***	-0.164	0.542***	-0.160
	(0.151)	(0.133)	(0.148)	(0.134)
$I_{rem\ ch}$	0.000800***	-0.000732*	0.000817***	-0.000675*
	(0.000198)	(0.000397)	(0.000199)	(0.000378)
ΔP_{rem}	0.0395	0.562***	0.00231	0.558***
	(0.0592)	(0.0839)	(0.0604)	(0.0846)
ΔP_{rem}, lag 1	-0.0723	-0.0107	-0.0632	-0.00939
	(0.0560)	(0.0916)	(0.0565)	(0.0899)
$\%\Delta T$	-0.0128*	-0.00858	-0.105**	0.0674
	(0.00704)	(0.0184)	(0.0493)	(0.0960)
$\%\Delta T * Pop$			-0.327***	-0.185
			(0.0936)	(0.176)
$\%\Delta T * Inc$			0.226***	-0.0537
			(0.0636)	(0.120)
U_{1990}	0.00561***	-0.000845	0.00656***	-0.000362
	(0.00131)	(0.00213)	(0.00139)	(0.00222)
Constant	-0.000822***	-0.000379	-0.00111***	-0.000510
	(0.000228)	(0.000423)	(0.000249)	(0.000434)
Observations	306	306	306	306
R-squared	0.361	0.180	0.404	0.184

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table C.2 presents instrumental variables estimations of specifications (1) at various radius definitions of the education market.

Table C.2: Specification (1), IV estimation at various radii

	Radius=20	Radius=30	Radius=40	Radius=50	Radius=60
ΔC	-0.753***	-0.695***	-0.699***	-0.716***	-0.716***
	(0.240)	(0.245)	(0.249)	(0.248)	(0.247)
ΔP	-0.361*	-0.357	-0.400*	-0.390*	-0.385*
	(0.217)	(0.218)	(0.216)	(0.214)	(0.212)
$\%\Delta T$	0.113***	0.0989***	0.0923***	0.0850***	0.0835***
	(0.0254)	(0.0231)	(0.0206)	(0.0206)	(0.0205)
U_{1990}	-0.00244	-0.00430	-0.00739***	-0.00773***	-0.00768***
	(0.00212)	(0.00266)	(0.00273)	(0.00275)	(0.00277)
Constant	-0.00216***	-0.00174**	-0.00126**	-0.00114*	-0.00113*
	(0.000733)	(0.000675)	(0.000617)	(0.000605)	(0.000600)
Observations	306	306	306	306	306
R-squared	0.405	0.383	0.376	0.368	0.369

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table C.3: Analysis restricted to schools years 2000 to 2009

	Specification (1)			Specification (2)		
	OLS	OLS	IV	OLS	OLS	IV
$\%\Delta T$	0.0715**	0.0711**	0.0369	0.545**	0.361	0.446*
	(0.0322)	(0.0302)	(0.0271)	(0.272)	(0.270)	(0.270)
ΔC		-0.772***	-0.807*		-0.736***	-0.879**
		(0.118)	(0.420)		(0.117)	(0.382)
ΔP		-0.464***	-0.484		-0.430***	-0.424
		(0.130)	(0.425)		(0.131)	(0.418)
U_{1990}	-0.0153***	-0.00883**	-0.00853	-0.0119**	-0.00682	-0.00678
	(0.00455)	(0.00383)	(0.00529)	(0.00483)	(0.00434)	(0.00503)
$\%\Delta T * Pop_i$				0.0172	0.0711	-0.372
				(0.622)	(0.620)	(0.468)
$\%\Delta T * Inc_i$				-0.703**	-0.449*	-0.488
				(0.277)	(0.265)	(0.304)
Constant	-0.000571	-0.00114	-0.00143	-0.00115	-0.00149	-0.00134
	(0.000887)	(0.000950)	(0.00158)	(0.000948)	(0.000995)	(0.00157)
Observations	170	170	153	170	170	153
R-squared	0.084	0.279	0.292	0.127	0.297	0.308

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table C.4: IV estimation using a smaller set of instruments, first and second stage estimates

	Specification (1)			Specification (2)		
	First Stage		Second Stage	First Stage		Second Stage
	ΔC	ΔP		ΔC	ΔP	
ΔC_{rem}	0.563***	0.00203		0.591***	-0.00116	
	(0.114)	(0.117)		(0.110)	(0.120)	
$I_{rem\ ch}$	0.000867***	-0.000972**		0.000908***	-0.000930**	
	(0.000193)	(0.000426)		(0.000198)	(0.000400)	
ΔP_{rem}	-0.0141	0.533***		-0.0352	0.540***	
	(0.0479)	(0.0982)		(0.0484)	(0.101)	
$\%\Delta T$	-0.0160**	1.74e-05	0.0884***	-0.0829	0.0711	0.199
	(0.00729)	(0.0185)	(0.0205)	(0.0506)	(0.102)	(0.136)
U_{1990}	0.00496***	-0.000219	-0.00878***	0.00614***	-3.23e-05	-0.00963***
	(0.00127)	(0.00220)	(0.00250)	(0.00139)	(0.00224)	(0.00267)
ΔC			-0.569**			-0.604**
			(0.272)			(0.253)
ΔP			-0.368*			-0.326
			(0.209)			(0.208)
$\%\Delta T * Pop_i$				-0.285***	-0.0748	0.252
				(0.0840)	(0.199)	(0.263)
$\%\Delta T * Inc_i$				0.179***	-0.0791	-0.233
				(0.0634)	(0.131)	(0.154)
Constant	-0.000644***	-0.000204	-0.00113*	-0.000953***	-0.000250	-0.000838
	(0.000222)	(0.000443)	(0.000598)	(0.000244)	(0.000435)	(0.000627)
Observations	323	323	323	323	323	323
R-squared	0.331	0.184	0.350	0.365	0.187	0.362

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix D: The WIAT Subsample

Table D.1 compares students with and without WIAT scores. Overall there are small differences between the two samples. We find that we are primarily missing WIAT scores for students in first and second grade. Moreover, the sample without WIAT scores has significantly fewer African American students. Two year retention is also higher for the subsample without WIAT scores. Our analysis of treatment effects shows that the results are qualitatively the same across samples.

Table D.1: Comparing student with and without WIAT scores

	Have WIAT	No WIAT	Two-sided p-value
Year	2005.63	2005.53	0.36
Grade	*3.67*	*2.96*	*0.01*
Male	0.50	0.54	0.42
African American	*0.14*	*0.07*	*0.03*
College	0.38	0.38	0.86
Income	41048.97	40657.46	0.80
Math	1633.33	1602.25	0.29
Reading	1515.81	1520.10	0.86
FSIQ	119.72	120.66	0.40
GAI	122.34	122.68	0.79
Max(FSIQ, GAI)	123.47	123.94	0.70
Gifted	0.58	0.62	0.36
1 year retention	0.88	0.91	0.34
2 year retention	*0.78*	*0.86*	*0.05*
Count	395	145	

Appendix E: Standard RDD Estimates

Table E.1 reports the first stage of the standard RDD estimator. We find that there exists a discontinuity at the admission threshold of 125. The estimates are significant in all specifications and have the correct sign in 11 of the 12 estimated models.

Table E.1: First Stage of the **Standard** RDD

Full Sample						
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Difference	0.626***	0.649***	0.662***	0.193***	0.343***	0.43***
St. Err	(0.072)	(0.061)	(0.055)	(0.026)	(0.019)	(0.015)
Obs.	146	175	218	146	175	218
WIAT Subsample						
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Difference	0.612***	0.643***	0.654***	- 0.129***	0.146***	0.33***
St. Err	(0.088)	(0.073)	(0.064)	(0.038)	(0.028)	(0.022)
Obs.	104	127	163	104	127	163

Table E.2 reports the results for the second stage of the standard RD estimator. Here the results are less promising. We find small positive, but insignificant, retention effects when we use the local constant estimator. That finding is robust across different bandwidths. When we use the local linear speci_cation, the estimates are very sensitive to the specification of the bandwidth. This is not surprising since local manipulation implies that this estimator is not consistent.

Table E.2: Second Stage of the **Standard** RDD

	Full Sample: One Year Retention					
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Gifted	0.103	0.092	0.034	-0.547	-0.039	0.186
St. Error	(0.092)	(0.077)	(0.066)	(1.102)	(0.331)	(0.209)
p-value	0.264	0.234	0.605	0.619	0.907	0.373
Obs.	146	175	218	146	175	218
	Full Sample: Two Year Retention					
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Gifted	0.072	0.039	0.051	-1.32	-0.235	-0.161
St. Error	(0.109)	(0.092)	(0.083)	(1.837)	(0.431)	(0.294)
p-value	0.511	0.671	0.542	0.473	0.586	0.584
Obs.	146	175	218	146	175	218
	WIAT Subsample: One Year Retention					
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Gifted	0.159	0.121	0.049	2.359	-0.468	0.26
St. Error	(0.121)	(0.093)	(0.077)	(4.880)	(1.068)	(0.355)
p-value	0.19	0.193	0.522	0.629	0.661	0.464
Obs.	104	127	163	104	127	163
	WIAT Subsample: Two Year Retention					
	Constant			Linear		
	BW4	BW5	BW6	BW4	BW5	BW6
Gifted	0.12	0.055	0.072	3.422	-0.847	-0.304
St. Error	(0.139)	(0.110)	(0.099)	(6.717)	(1.802)	(0.510)
p-value	0.388	0.619	0.466	0.611	0.638	0.55
Obs.	104	127	163	104	127	163