Research Repository UCD

Title	The role of aggregation in estimating the effects of private school competition on student achievement
Authors(s)	Jepsen, Christopher
Publication date	2002-11
Publication information	Jepsen, Christopher. "The Role of Aggregation in Estimating the Effects of Private School Competition on Student Achievement" 52, no. 3 (November, 2002).
Publisher	Elsevier
Item record/more information	http://hdl.handle.net/10197/4436
Publisher's statement	This is the author's version of a work that was accepted for publication in Journal of Urban Economics. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Journal of Urban Economics (Volume 52, Issue 3, November 2002, Pages 477–500) DOI:10.1016/S0094-1190(02)00530-2 Elsevier Science (USA)
Publisher's version (DOI)	10.1016/S0094-1190(02)00530-2

Downloaded 2023-04-07T10:52:29Z

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)

© Some rights reserved. For more information

The Role of Aggregation in Estimating the Effects of Private School Competition on Student Achievement^{*}

Christopher Jepsen
Public Policy Institute of California
500 Washington Street, Suite 800
San Francisco, CA 94111
(415) 291-4479
fax: (415) 291-4428
jepsen@ppic.org

Keywords: education, private school competition, aggregation.

Suggested running head: aggregation and private school competition.

^{*} The views in this paper do not necessarily reflect those of the Public Policy Institute of California. I would like to thank Joseph Altonji, Rebecca Blank, Jan Brueckner, Tom Downes, Todd Elder, Lisa Jepsen, Dan Levy, Jens Ludwig, Bruce Meyer, Peter Meyer, Mark Montgomery, Dan Rosenbaum, Heather Rose, Paulo Santiago, Tim Sass, Jon Sonstelie, Christopher Taber, and two anonymous referees for their helpful suggestions. Pedro Cerdan provided excellent research assistance. I gratefully acknowledge the Inter-University Consortium for Political and Social Research (ICPSR) for providing access to several demographic data sets. Still, all errors are mine.

The Role of Aggregation in Estimating the Effects of Private School Competition on Student Achievement

Abstract: This paper investigates possible reasons for the disparity in results in the private school competition literature. In particular, the focus is on the data set, the grade range, and level of aggregation of the competition variable, and on the choice of OLS or IV estimation strategies. The results show that the size and significance of the competition variable depends on each of these attributes, although the grade range of the competition variable has a slightly smaller impact than the others. Private school competition does not have a consistently positive, significant effect on student achievement.

1. Introduction

One hotly contested reform to improve public high schools is expanded school choice opportunities. Proponents of this plan, beginning with Friedman [8], believe that school choice increases competition between schools. They argue that competition would drive inefficient schools out of business and that the quality of education, measured as student achievement, would improve. The level of competition among public schools is often difficult to measure, because it is affected by residency requirements and within-district choice. In contrast, private school enrollment provides a clear indicator of the competition provided by private schools. Several recent papers investigate the impact of private school competition on student achievement, but they fail to reach a consensus. The majority finds insignificant (but usually positive) effects of private school competition, but several others also find positive and significant effects.

The goal of this paper is to investigate whether the difference in findings across studies can be attributed to differences in the definition of the competition variable or the data set used. I extend the literature in four ways. First, using the conventional definition of competition, the percent of students attending private school, I investigate whether the effect of competition is sensitive to the area over which this percentage is measured. Specifically, estimates of competition at the county, PMSA, and MSA levels are compared.¹ Second, I explore whether the competition coefficient is sensitive to the grade range (elementary, secondary, or both) at which it is measured. Third, the analysis uses two student-level data sets, the National Longitudinal Survey of Youth 1979 (NLSY79) and the National Educational Longitudinal

¹ For most metropolitan areas, the PMSA and the MSA are the same. The exceptions are metropolitan areas that cover a large area and contain smaller metropolitan areas, such as the Bay Area (San Francisco, Oakland, and San Jose) and New York City. In 1990, there were 20 MSAs that contained more than one PMSA. Also, PMSAs were

Survey of 1988 (NELS88). Fourth, I compare the effect of competition when measured with Census data or with actual school enrollment data from the Department of Education.

The results show that the estimated effect of private school competition on student achievement depends on several attributes: the choice of the student-level data set (NLSY79 or NELS88), the dependent variable (i.e., the measure of student achievement), and the aggregation level of the competition variable. For example, in the NELS88 data the competition coefficient is positive and significant when defined at the county level but not at the PMSA or MSA levels. In addition, the impact of competition also depends on whether the model is estimated with OLS or IV. The grade range of the competition variable also affects the competition coefficient, but the range of estimates across grade ranges (all else constant) is smaller than that for other factors such as the level of aggregation. These findings suggest that much of the variation in the private school competition coefficients in the present literature stems from differences in outcome dataset and the level of aggregation of the competition measure.

2. Conceptual Framework

There is a large literature on the effects of private school competition on student achievement. Table 1 summarizes the results on this topic.² The studies in Table 1 focus on one measure of private school competition, market share. Specifically, competition is measured as the percentage of students who attend private school in a given area.

Most papers define the aggregation level (i.e., area) as the school district or county, but few discuss the importance of that choice. The exception is Hoxby [12], who argues that the

previously known as SMSAs, while MSAs were previously known as SCSAs. Metropolitan areas are referred to as either PMSAs or MSAs regardless of the time period.

² Belfield and Levin [2] present a more complete review of the literature on school competition, including the effects of private school competition.

appropriate aggregation level is the PMSA. She believes that more disaggregated levels, such as county or especially the school district, are problematic because household location is based in part on school quality. This household location choice confounds the effects of a competition variable with effects of unobservable school quality.

Measuring competition at the PMSA level is not without its problems. Because the largest metropolitan areas each contain multiple PMSAs, households in those areas choose the PMSA in which to reside. For example, Oakland and San Francisco are in different PMSAs, as are New York City and Jersey City, New Jersey. Because each metropolitan area (PMSA or MSA) covers a large and diverse geographic area, it is unlikely that the impact of competition is the same for each student in the area. In other words, aggregation bias is a potential concern.³ Therefore, the choice of the appropriate level of aggregation of the competition variable is unclear.

For studies using aggregate outcomes, the level of aggregation of the competition variable usually (but not always) is chosen in order to match the level of the dependent variable.⁴ One consequence of this choice is that differences in the estimated effects of competition could result either from the choice of aggregation level or from the choice of data sets. An example is the difference between the generally insignificant effects of district-level competition in Simon and Lovrich [21] using Washington data, and the positive, significant effects of county-level competition in Couch et al. [6] using North Carolina data.

In addition to the level of aggregation, the grade range of the competition variable is usually chosen to match the grade range of the outcome variable. Nearly all the studies in

³ Hanushek, Rivkin, and Taylor [11] show that the aggregation of school resources at the state or district level tends to increase the size of the school resources coefficients in student achievement regressions.

⁴ The two exceptions are Dee [7], who uses county-level competition data and district-level outcomes, and Sander [19], who uses district-level competition data and school-level outcome data.

Table 1 use the percentage of K-12 students. However, the preferred grade range is not obvious for outcomes at the high-school level or beyond. Education is a cumulative process, where learning in high school builds on learning in earlier grades. Consequently, the combined elementary and secondary school measure of private school enrollment is an important measure of the comprehensive effect of competition. As most private schools in the country serve elementary students, the competition variable at the elementary level has the potential to capture more completely the effects of competition than do measures at other grade ranges.

Another factor in the study of the private school competition is the choice of the dependent variable. The availability of aggregate-level data on student achievement is usually limited to state-specific standardized test scores. Consequently, each study that uses aggregate test scores contains data from only one state. Dropout rates are more uniform across states, so Dee [7] uses county- and district-level data for 18 states that have a common definition of high school dropout. In contrast, the national data sets with individual-level data were designed to collect student achievement data (including test scores and high school graduation) that are consistent across states.

A major concern in the competition literature is the potential endogeneity of the competition variable measured as the percentage of students attending private school. Specifically, there is the concern that parents choose private schooling *because* the local public schools are of low quality. Because school quality (past or present) is difficult to measure, private school competition and unobservable student achievement are likely correlated. Although half the studies in Table 1 ignore this endogeneity, the others use instrumental variable analysis with religious density variables as instruments for private school attendance. Even with such instruments, the use of individual-level data creates the concern that

unobservable student achievement is correlated for students living in the same area (county/PMSA/MSA). Hoxby [12] includes SMSA-level random effects when estimating OLS but not when estimating IV. McMillan [16] calculates robust standard errors in both OLS and IV estimation strategies.

Many attributes of the competition variable affect the estimated impact of private school competition, but the literature pays little attention to the choice of these attributes. Instead, the components appear to be chosen arbitrary or are the result of data limitations. In contrast, I combine data from several sources to construct multiple measures of competition for each of several factors: the aggregation level, the grade range, and the data set. The next section describes these data.

3. Data

The outcome variables in this study come from two longitudinal, individual-level data sets. One data set is the National Longitudinal Survey of Youth 1979 (NLSY79), a nationally representative sample of 12,686 people aged 14 to 21 in the spring of 1979.⁵ The data contain follow-up surveys and sample weights for each year until 1996.⁶ I use the weighted 1990 sample and create four dependent variables: the percentage score on the Armed Forces Qualification Test (AFQT), years of schooling, high school graduation (or GED receipt), and college attendance.

The second student-level data set is the National Educational Longitudinal Survey of 1988 (NELS88). This data set began in 1988 with a nationally representative sample of over 20,000 eighth-grade students. McMillan [16] uses the 1988 baseline data to explore the effect of competition on eighth-grade test scores. I use weighted data from the second and third

⁵ The NLSY79 data contain little school-level information, so no school-level data are included in the analysis.

follow-up surveys, conducted in 1992 and 1994, respectively, to create three dependent variables: the 12th grade mathematics test score, high school graduation (or GED receipt), and college attendance.⁷

Each individual-level data set is supplemented with demographic data from several sources, based on the county of residence for each respondent. County-based definitions of metropolitan areas allow the aggregation of counties into PMSAs and PMSAs into MSAs. Each county not in a multi-county PMSA is treated as its own PMSA, and each PMSA not in an MSA is treated as its own MSA. For example, the 1980 Census contains 158 SMSAs that cover 558 of the 3101 counties in the U.S. Using these definitions, I create three measures of each demographic variable: the county-level measure, the PMSA-level measure, and the MSA-level measure. For example, there are variables for the percentage of females in the county, the PMSA, and the MSA. For a student not in a metropolitan area, the three measures would be the same: the county percentage.

The 1980 and 1990 Surveys of Churches and Church Membership provide religious demographic data, including the Catholic population density variables used as instruments. The sources of non-religious demographic information are the 1980 and 1990 U.S. Censuses.

The Department of Education provides private school competition data in several data sets. The 1982 Census of Governments contains 1980 public school enrollment at the district level, whereas the 1980 Private Schools in America Survey contains school-level private school enrollment.⁸ The data source for 1990 public and private school enrollments is the National

⁶ I choose 1990, when each respondent was at least 24 years old, to facilitate direct comparisons with Hoxby [12], whose dependent variables from the NLSY79 are as of age 24.

⁷ In particular, all descriptive statistics and the high school graduation or college attendance models use the sampling weight for participation in the third follow-up survey. The 12th grade mathematics test score models use second follow-up participation weights.

⁸ While a substantial number of schools are not in the private school survey, the pattern of missing schools does not appear to be systematic in any way (U.S. Department of Education [23]).

Education Data Research Center (NEDRC).⁹ The public school enrollment data are from the 1989-1990 Common Core of Data, and the private school enrollment data are from the 1989-1990 Private Schools Survey. For each year, competition is defined as the percentage of students attending all types of private school for each of the following three grade ranges: K-8, 9-12, and combined K-12. An alternative definition of competition is the percentage of Catholic school students in each of the three grade ranges.

In addition to the Department of Education data, 1980 and 1990 Census data also contain competition variables. For the 1980 Census, I calculate percentage of students attending private school for each of the following three grade ranges: K-8, 9-12, and combined K-12. The 1990 Census data do not separate enrollment into K-8 and 9-12, so the only available grade range is the combined K-12 percentage. Census data on Catholic school attendance are not available for 1980 or 1990. Due to these limitations, the Department of Education data are preferred to the Census data.¹⁰

4. Econometric Approach

Equation (1) provides a linear representation of an education production function as in Hanushek [10]:¹¹

$$(1) \quad Y_{ij} = X_{ij}'\beta + Z_j'\gamma + P_j\alpha + \varepsilon_{ij}$$

In this equation, Y_{ij} is one of the observed dependent variables mentioned above, X_{ij} is a vector of individual characteristics, Z_j is a vector of Census demographic variables, P_j is a measure of

⁹ I would like to thank Mark Glander and his associates at the NEDRC for providing these data.

¹⁰ Another limitation of the Census data is that they use Census estimates rather than actual enrollment, and they include home-schooled students (Belfield and Levin [2]).

¹¹ While not reported, probit models are also estimated for all binary outcomes. The results are qualitatively similar to the OLS results presented here.

private-school competition, ε_{ij} is the unobservable term, i indexes individuals, and j indexes local demographic area (county, PMSA, or MSA, depending on the model).

Because unobservable student quality is likely correlated among students in the same geographic area, equation (1) is estimated with standard errors that allow for correlation of ε_{ij} for students in the same geographic area (as indexed by j). The use of these robust standard errors increases the standard errors by at least fifteen percent. This error structure does not specify the functional form of ε_{ij} , whereas a random-effects model (employed by Hoxby [12]) assumes that the unobserved component has two parts, a geographic part and an individual component ($\varepsilon_{ij} = \eta_j + \nu_{ij}$).¹² Finally, note that the geographic area (county, PMSA, or MSA) denoted by j serves two important functions. It identifies the level of measurement for the competition variable and the demographic variables, and it determines which observations have correlated ε_{ij} 's.

The main concern with competition variables in OLS regressions is that endogeneity problems often exist. Unobservable determinants of student achievement, such as local public school quality, are likely correlated with competition. One way to control for the effect of unobservable school quality is the use of instrumental variables (IV) estimation, as illustrated by the equations (2) and (3):¹³

$$(2) \quad P_j = Z_j' \lambda + C_j' \mu + \nu_j$$

$$(3) \quad Y_{ij} = X_{ij}' \phi + Z_j' \gamma + P_j' \alpha + \varepsilon_{ij}$$

¹² The estimation of a fixed-effects model for competition is not possible, as the fixed effects for area (county, PMSA, or MSA) would capture most, if not all, of the variation in the competition variable.

¹³ A probit model is also estimated when the outcome variable in equation (3) is a binary variable. As in the single-equation model, the results from this model (not reported) are qualitatively similar to those from linear IV estimation. See Murphy and Topel [17] for a more detailed explanation of two-stage estimation when the second stage is non-linear.

In these equations P_j , X_{ij} , Z_j , ε_{ij} , i and j are defined as in equation (1). Following the methods used in previous research of private schools, religious composition variables are used as instruments. Specifically, the six elements of C_j are the percentage of the area (county/PMSA/MSA) population that is Catholic, the population density of Catholics (i.e., Catholics per square mile), the density of Catholic churches, and the squares of these three variables.¹⁴ These variables provide good instruments because roughly half of all private schools students are in Catholic schools. The Catholic Church supports these schools, so areas with more Catholics, and specifically areas with concentrations of Catholics, are more likely to have more private schools. These areas then have higher percentages of private school students and lower distances to nearest private schools, all else equal.

The choice of the variables to include in C_j is crucial to the performance of the model. If the independent variables in equation (2), especially those in C_j , provide a poor fit, then the parameter estimates for equation (3) will be poorly estimated (see Bound, Jaeger and Baker [5]). Therefore, the following section includes an analysis of the fit of equation (2).

5. Results

Weighted descriptive statistics for the two data sets are presented in Appendix Tables 1 through 3. For a given measure of competition, the means are quite similar across data sets and levels of aggregation (county, PMSA, and MSA). The percent attending private school is slightly higher at the elementary level than at the secondary level, and the mean percent attending Catholic school for each grade range is roughly 60 percent of the mean for all private schools.

¹⁴ Note that these instruments are highly correlated with each other by definition, since they all depend on the number of Catholics (or Catholic churches) in the county/PMSA/MSA.

Tables 2 through 6 contain the regression results. All but Table 4, the first-stage results for the IV model, have a similar layout. The left third of each table contains results from specifications using county-level demographic variables and measures of competition, the middle third contains results from PMSA-level measures, whereas the right third contains results from MSA-level measures. Columns 1, 4, and 7 measure competition as the percentage of K-8 students (elementary); columns 2, 5, and 8 measure competition as the percentage of 9-12 students (secondary); columns 3, 6, and 9 measure competition at the percentage of K-12 students. All models include student-level controls for race, sex, parental education, religion, number of siblings, and frequency of religious attendance (NLSY79 only).¹⁵ In other words, each specification contains both student-level *and* area-level (county, PMSA, or MSA) controls for many demographic characteristics.

OLS Results

Table 2 shows that private school competition has insignificant effects on all outcomes in the NLSY79 data, except for the percent of grade 9-12 students at the MSA level in the college attendance model. County-level coefficients tend to be smaller than PMSA-level coefficients, which in turn tend to be smaller than MSA-level coefficients. The competition coefficients in Table 3 for the NELS88 data are always insignificant, and they have little variability by area or grade range of the competition variable.

Although not reported, the models are also estimated using two alternative definitions of private school competition. The first definition uses the Census data rather than school enrollment data, and the second definition is the percentage of children attending Catholic

¹⁵ Frequency of religious attendance is not available at baseline in the NELS88. The variable in the first follow-up measuring religious attendance is missing for many observations. Therefore, it is excluded from the analysis. NELS88 results that contain controls for religion attendance are similar to those reported in the tables.

school rather than all private schools. The results from these two measures are quite similar to the results in Tables 2 and 3.¹⁶

In summary, the effect of private competition is generally insignificant in Tables 2 and 3, although the coefficient varies across a number of dimensions such as area and the dependent variable. Hoxby [12], Arum [1], and McMillan [16] also find insignificant effects of competition on individual student test scores and years of schooling. With aggregate data, the results are mixed, although most of the competition coefficients are insignificant or negative.

IV First Stage Results

Because researchers have many reasons to doubt the exogeneity of the private school competition variable in student achievement models, these models are estimated using IV as in equations (2) and (3). Table 4 contains the results from equation (2), where the dependent variable is private school competition. The coefficients for the NLSY79 data are in the left half and those for the NELS88 data are in the right half. For simplicity, only the results where the dependent variable is the combined K-12 percent attending private school are reported. Results for other definitions of competition are similar to those reported here. Columns 1 and 4 use the county percentage, columns 2 and 5 use the PMSA percentage, and columns 3 and 6 use the MSA percentage. The instruments, the independent variables included in this equation but not in the outcome equation, are six measures of Catholic density: percentage of the county/PMSA/MSA population that is Catholic, the density of Catholics in the county/MSA, the density of Catholic churches in the county/PMSA/MSA, and the squares of these three variables.

¹⁶ Newmark [18] also estimates OLS models for Census-based and enrollment-based competition coefficients and finds little difference between the two sets of results.

In order to minimize the possibility of omitted-variable bias, the first-stage equations, as well as the OLS and second-stage student achievement equations, contain extensive demographic controls. For example, the religious homogeneity index and the percentage of the county population that regularly attends religious services (of any denomination) are independent variables, so that the coefficients on the instruments reflect Catholic density, rather than the effects of other religions. Other examples are the inclusion of population density, population, and area variables.

The adjusted R-squared values presented in Table 4 show that the model fits 80 to 90 percent of the variation in the percent attending private school. The joint F statistic for the significance of all six instruments is significant at one percent for all models, well above the lower bound suggested by Bound, Jaeger and Baker [5] for poorly fitted models. The general significance of the squared terms suggests that the relationship between these variables and private school competition is not linear.¹⁷ The coefficients in Table 4 for the instruments are more than twice as large as those reported in Hoxby [12] and Sander [19].¹⁸ The most powerful instrument in Table 4 is the Catholic percentage of the population, where the coefficient ranges from 0.25 to 0.60. This coefficient is highly significant at one percent for all definitions of competition, and the squared population is always significant and negative. Because all the instruments are measurements of Catholic religious influence, they are highly correlated with each other. Therefore, the other instruments – measures of Catholic density – are often insignificant and in some specifications are negative (with positive squared terms).

¹⁷ The intuition for a quadratic relationship between Catholic population/density and private school competition is as follows. As Catholic population density increases, the support for a Catholic school increases. However, at some point, the high percentage of Catholic residents influences the (secular) behavior of the public schools, somewhat reducing the demand for Catholic schooling. While this intuition is a simplification of the situation, it does suggest that a quadratic relationship may be more appropriate than a linear one.

IV Second Stage Results

The results from equation (3) are summarized in Tables 5 and 6 for NLSY79 and NELSS88 data, respectively. The dependent variable is one of the following measures of student achievement: test score, years of schooling, high school graduation, or college attendance. In both tables, the standard errors of the competition coefficients are nearly twice as large as in the specifications using OLS.

In the NLSY79 results in Table 5, I find that the size and significance of the competition coefficients depend on the area and the grade range of the competition variable, as well as on the outcome variable. For AFQT scores, two of the nine coefficients are significant: grades 9-12 at the PMSA level and grades K-12 at the MSA level. All the MSA-level competition variables are positive and significant in the years of school specifications, where a ten-percentage point increase in the MSA private school percentage corresponds with an increase in the years of school of 0.27 to 0.30 years. For college attendance, the grade 9-12 competition variables at the county and PMSA levels have positive, significant effects (of 5 percentage points for a ten-percentage point increase in competition), but all other coefficients are insignificant. As Hoxby [12] finds, the results in Table 6 show that grade 9-12 competition has larger effects than competition at grades K-8.

The results for the NELSS88 data in Table 6 are markedly different, as the private school competition coefficients are usually insignificant and negative. The only significant effects of competition are the positive impacts of county-level competition in grades K-8 and K-12 on college attendance. At the same time, all of the estimates are imprecisely estimated, so that the

¹⁸ Although the instruments for Catholic population and density in the first-stage models estimated by Dee [7] differ slightly from the models estimated here, he also finds large, highly significant effects of the instruments in his first-stage models.

90 percent confidence intervals contain substantial competitive effects, even for the MSA-level specifications.

Models with alternative definitions of private school competition are estimated with IV. As with the OLS estimates, the IV results for these alternative specifications are similar to the results in Tables 5 and 6. The only notable difference is for the Census-based competition variables in the NLSY79 data, where the county-level competition measures are often significant for AFQT test scores and college attendance, but the PMSA- and MSA-level coefficients are generally insignificant.

6. Discussion

Among the most surprising results is the discrepancy between my findings for the NLSY79 and those in Hoxby [12]. In particular, Hoxby's [12] main finding is that the PMSA percentage of students in Catholic high schools has a positive effect on years of schooling when estimated with IV. In contrast, all the measures of competition at the PMSA level in Table 5, along with the unreported results using Census data or the Catholic percentage of students, have insignificant coefficients for years of schooling. Given that both studies use NLSY79 data with the same measure of competition, this difference is noteworthy. The primary cause of the difference is the treatment of the standard errors. Estimating a model that allows for correlation among respondents in the same area increases the standard errors by at least 15 percent, and Hoxby [12] does not allow for such correlation in her IV models. Re-estimating the model in Table 5 without allowing for within-area correlation leads to significant, positive effects of PMSA-level competition on years of schooling.¹⁹

¹⁹ When I define years of schooling by age 24 (as Hoxby [12] does), rather as years of schooling by 1990, the significance of the private school competition coefficient increases only slightly.

This divergence in results illustrates how the methodological differences also affect the coefficient on private school competition. Other observable differences across the private school competition literature are: the choice of the outcome data set, the measure of student achievement (test scores, years of schooling, etc.), the econometric approach (IV versus OLS), and the aggregation level and grade range of the competition variable. The results in the tables demonstrate that the competition coefficient is determined in large part by the treatment of these factors.

Perhaps the most important implication of the results is that the level of aggregation of the competition variable is an important choice made by researchers. At first glance, the pattern of results in the competition literature suggests that researchers are more likely to find an effect of competition as the level of aggregation increases, as illustrated by Table 1. The studies using district-level competition data generally find insignificant effects of competition (Sander [19], Simon and Lovrich [21], and McMillan [16]).²⁰ At the county level, the results are mixed: McMillan [16], Newmark [18], and Geller et al. [9] find insignificant effects, but Dee [7] and Couch et al. [6] find positive and significant effects.²¹ At the PMSA level, Hoxby [12] finds positive and significant effects (using IV). The findings in Tables 5 and 6 illustrate that this pattern does not hold when the level of aggregation changes but all other components of the competition variable are held constant.

While the aggregation level of the competition variable has a substantial impact on the competition coefficient, the grade range at which competition is measured has a more modest effect. In some cases, the competition coefficient at grades 9-12 differs from the K-8 and K-12 competition coefficients, but the changes are less dramatic than corresponding changes in other

²⁰ Similarly, Jepsen [14] finds insignificant effects of competition defined at the zip-code level.

attributes of the competition variable, such as the level of aggregation or the choice of dependent variable.

The choice of data set for the dependent variable is an extremely important determinant of the effect of the competition variable, as well as of the inferences that can be drawn from the results. Competition has little effect on outcomes in the NELS88 data in IV specifications but has some positive effects in the NLSY79 data. These two data sets differ along a number of dimensions, each of which could produce differences in the effects of competition. The most obvious difference is timing. In addition to being administered a decade apart, the data sets also differ in the age ranges of respondents. The NLSY79 covers respondents well into their adult lives, while the most recent wave of the NELS88 occurs when most respondents are aged 20. The effect of competition on student achievement may depend both on the year and on the age level of the respondent.

A related difference is the goal of the data sets. The NLSY79 was administered by the Bureau of Labor Statistics to study labor issues, while the NELS88 is from the Department of Education. This difference in goals also produces differences in sampling techniques and therefore sample populations. Finally, the NLSY79 and the NELS88 data sets also differ in terms of geographic area covered, as each data set only covers a subset of the counties (and therefore PMSAs and MSAs) in the country. The purpose of the sample weights is to make both data sets nationally representative.²² Although the implications of these differences for a study of private school competition are unclear, the resulting discrepancy in coefficients

²¹ The results for Florida, where counties and districts are identical, are also mixed (Smith and Meier [22], Maranto et al. [15], and Sass [20]).

²² I would like to thank an anonymous referee for making this observation. It is not possible to discern whether geographic differences explain the difference in results between these two data sets. Restricting each data set to the set of counties in both the NLSY79 and the NELS88 reduces the size of each data set by roughly one-third. Competition coefficients from this “common” sample are quite noisy; no distinct patterns exist either within or between data sets.

by data set provides further proof against making inferences out of sample for any data set. For example, the results in the tables provide little information on the effect of competition for today's students.

Of similar importance is the choice of dependent variable, the measure of student achievement. Again, when all else is held constant (such as the level of aggregation, the grade range, and the outcome data set), a given measure of private school competition has a positive and significant impact on one outcome but an insignificant effect on other outcomes. An example is that competition at the MSA level is positive and significant for years of schooling but generally insignificant for other outcomes, as reported in the right third of Table 5.

Another potential influence on the effects of competition is the data source of the competition variable. The two main sources of data on public and private school enrollment are the Department of Education (national and state) and the Census. Competition coefficients based on Census data are less precisely estimated than those based on enrollment data. However, in models estimated with IV, the choice of data set for the competition variable has little impact on the results. Thus, the choice of the data set for the dependent variable is more critical than the choice of the data set for the competition variable.

The choice between OLS and IV estimation is another crucial determinant of the estimated effects of the competition variable. Most authors generally find insignificant effects of competition when using OLS, but the results when using IV differ greatly. The results in the tables are no exception. In the NLSY79 data, almost every measure of competition is insignificant when using OLS, but several of the coefficients using IV are significant.

7. Conclusion

Several components of the private school competition variable affect the size and significance of the estimated effect on student achievement. The choice of the outcome data set and the level of aggregation of the competition variable have major effects on the size and significance of the competition effect, whereas the grade range and the choice of competition data sets have smaller impacts. The choice between OLS and IV is also important. This pattern is likely responsible for much of the variation in the private school competition literature.

This study, as well as the literature more generally, has limitations that should be investigated in future research. Private school enrollment is only one component of competition, although it is by far the dominant measure in the literature. Two alternative measures of competition are the distance to the nearest private school (Jepsen [14]) and the percentage of private schools (Geller et al. [9]). Both studies find insignificant effects for these measures of competition for both OLS and IV.

The private school competition literature essentially ignores competition from public schools. A thorough analysis of school competition should include controls for both public *and* private school competition, with consideration of the likely endogeneity of both types of competition. Borland and Howsen [4] use one variable to control for competition from both private and public schools, and they find marginally significant, positive effects of competition on student achievement.²³ Hoxby's [12] analysis of private school competition controls for public school competition but not its endogeneity - a major theme in Hoxby [13].

The reliance in the competition literature on religious density measures as instruments is also worth further investigation, especially given the vast differences between competition

estimates from OLS and those from IV. This issue is not explored in this study because the choice of instruments is relatively constant in this literature, and the focus here is on explaining differences in the literature. However, McMillan [16] raises valid concerns about the use of these instruments and explores alternative instruments.

Even with these limitations, the results here provide some information in response to the main question in this literature: how does private school competition affect student achievement? The lack of consistent results in the tables and in the previous literature, along with the sensitivity of the results to such factors as the choice of data set and level of aggregation, suggests that private school competition as measured by the percentage of students attending private school does not have a clear, positive impact on student achievement. At the same time, the results in tables have large standard errors, and therefore confidence intervals contain sizable positive and, in some cases, negative competition coefficients. Therefore, much work remains to be done before policymakers can reach any conclusions about the effects of private school competition on student achievement.

²³ Borland and Howsen [3] use the same measure of competition but do not control for its endogeneity.

References

1. R. Arum, Do private schools force public schools to compete? *American Sociological Review*, 61, 29-46 (1996).
2. C.R. Belfield and H.M. Levin, The effects of competition on educational outcomes: A review of the U.S. Evidence, National Center for the Study of Privatization in Education Occasional Paper No. 35 (2001).
3. M.V. Borland and R.M. Howsen, Student academic achievement and the degree of market concentration in education, *Economics of Education Review*, 11, 31-39 (1992).
4. M.V. Borland and R.M. Howsen, Manipulable variables of policy importance: The case of education, *Education Economics*, 8, 241-248 (2000).
5. J. Bound, R.M. Jaeger, and D. Baker, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, *Journal of the American Statistical Association*, 90, 443-450 (1995).
6. J.F. Couch, W.F. Shughart II, and A.L. Williams, Private school enrollment and public school performance, *Public Choice*, 76, 301-312 (1993).
7. T.S. Dee, Competition and the quality of public schools, *Economics of Education Review*, 17, 419-427 (1998).
8. M. Friedman, "Capitalism and Freedom," University of Chicago Press, Chicago (1955).
9. C.R. Geller, D.L. Sjoquist, and M.B. Walker, The effects of private school competition on public school performance, National Center for the Study of Privatization in Education Occasional Paper No. 15 (2001).
10. E.A. Hanushek, The economics of schooling: Production and efficiency in public schools, *Journal of Economic Literature*, 24, 1141-1177 (1986).
11. E.A. Hanushek, S.G. Rivkin, and L.L. Taylor, Aggregation and the estimated effects of school resources, *Review of Economics and Statistics*, 78, 611-27 (1996).
12. C.M. Hoxby, Do private schools provide competition for public schools? National Bureau of Economic Research Working Paper 4978 (1994).
13. C.M. Hoxby, Does competition among public schools benefit students and tax payers? *American Economic Review*, 90, 1209-1238 (2000).
14. C. Jepsen, "The Private Schooling Market and Its Effects on Student Achievement," Ph.D. Dissertation, Northwestern University (2000).

15. R. Maranto, S. Millman, and S. Stevens, Does private school competition harm public schools? Revisiting Smith and Meier's *The Case Against School Choice*, *Political Research Quarterly*, 53, 177-192 (2000).
16. R. McMillan, The identification of competitive effects using cross-sectional data: An empirical analysis of public school performance, unpublished manuscript, University of Toronto (2001).
17. K.M. Murphy and R.H. Topel, Estimation and inference in two-step econometric models, *Journal of Business and Economic Statistics*, 3, 370-379 (1985).
18. C.M. Newmark, Another look at whether private schools influence public school quality, *Public Choice*, 82, 365-373 (1995).
19. W. Sander, Private schools and public school achievement, *Journal of Human Resources*, 34, 697-709 (1999).
20. T.R. Sass, The effects of private school competition on public school performance: Does the type of competitor matter? unpublished manuscript, Florida State University (2001).
21. C. A. Simon and N.P. Lovrich, Jr., Private school performance and public school performance: Assessing the effects of competition upon public school student achievement in Washington state, *Policy Studies Journal*, 24, 666-675 (1996).
22. K.B. Smith and K.J. Meier, Public choice in education: Markets and the demand for quality education, *Political Research Quarterly*, 48, 461-478 (1995).
23. U.S. Department of Education, National Center for Education Statistics, Data base documentation for Universe of Private Schools, 1976-80, Washington, D.C. (1980).

Table 1: Summary of Literature on Private School Competition

Author	Time Period	Location / Survey	Competition Variable			Dependent Variable	Results	
			Area ¹	Grade	Source ³		OLS	IV
Aggregate Data								
Simon and Lovrich [21]	1990	Washington	District	K-12	State	4th, 8th grade test scores	0	N/A
Sander [19]	1996	Illinois	District	K-12	U.S. DOE	Test scores, graduation rates	0	0
Couch et al. [6]	1989	North Carolina	County	K-12	State	8th-12th grade test scores	+	N/A
Newmark [18]	1989	North Carolina	County	K-12	State	8th-12th grade test scores	0	N/A
Smith and Meier [22]	1988-1990	Florida	County	K-12	Not given	10th grade test scores	- / 0	N/A
Maranto et al. [15]	1988-1990	Florida	County	K-12	Not given	5th, 8th, 10th grade test scores	- / 0	N/A
Dee [7]	1994	18 States	County	K-12	U.S. DOE	High school graduation rates	+ / 0	+
Geller et al. [9]	1988-1990	Georgia	County	K-12	State	3rd, 10th grade test scores	- / 0	- / 0
Sass [20]	1998-2000	Florida	County	Varies ²	State	4th, 5th, 8th, 10th grade test scores	+ / 0	+ / 0
Individual Data								
Hoxby [12]	1979-1990	NLSY79	PMSA	9-12	U.S. DOE	Educational attainment, test scores, wages	0	+
Arum [1]	1980-1982	High School and Beyond	State	K-12	U.S. DOE	12th grade test scores	+ / 0	N/A
McMillan [16]	1988	NELS88	District, County	5-8	U.S. DOE	8th grade test scores	- / 0	- / 0

Notes: The results summarized in the table are for the authors' preferred specifications. Borland and Howsen [3], [4] are excluded from the tables because their measure of competition includes competition from public schools.

¹ Counties and school districts in Florida are the same. The area of the dependent variable is the school level in Sander [19] and the district level in Dee [7].

² The grade range matches the dependent variable: elementary percentages were used for 4th and 5th grade test scores; middle school percentages for 8th grade test scores; and secondary percentages for 10th grade test scores.

³ State = the location state's Department of Education; U.S. DOE = U. S. Department of Education

Table 2: OLS Coefficients for Percent Attending Private School in NLSY79 Data
Each Coefficient Represents a Separate Regression

Area Competition Grade Level	County			PMSA			MSA		
	K-8	9-12	K-12	K-8	9-12	K-12	K-8	9-12	K-12
	1	2	3	4	5	6	7	8	9
Dependent Variable									
AFQT Test (N=9,096)	-0.0435 (0.0751)	-0.0452 (0.0668)	-0.0526 (0.0778)	0.0970 (0.0922)	0.1088 (0.0803)	0.1096 (0.0947)	0.1429 (0.0986)	0.1197 (0.0865)	0.1457 (0.1013)
Years of Schooling (N=9,447)	-0.0011 (0.0068)	-0.0040 (0.0059)	-0.0027 (0.0069)	0.0005 (0.0086)	0.0011 (0.0073)	0.0006 (0.0086)	0.0084 (0.0092)	0.0089 (0.0075)	0.0092 (0.0092)
High School Graduation (N=9,472)	0.0002 (0.0008)	-0.0003 (0.0007)	-0.00003 (0.0008)	0.0002 (0.0011)	0.0005 (0.0008)	0.0003 (0.0010)	0.0009 (0.0011)	0.0006 (0.0008)	0.0008 (0.0011)
College Attendance (N=9,447)	0.0004 (0.0014)	0.0007 (0.0012)	0.0006 (0.0014)	0.0008 (0.0017)	0.0023 (0.0014)	0.0015 (0.0017)	0.0020 (0.0018)	0.0035*** (0.0015)	0.0027 (0.0018)

Notes: Robust standard errors that allow within county/PMSA/MSA correlation are in parentheses. All models contain the following individual-level variables: race, gender, parental education, religion, number of siblings, and frequency of religious attendance. All models also include the following area-level demographic variables: percent attending religious services, religious homogeneity index, various racial demographic variables, percent high school graduate, percent college graduate, percent male, percent urban, area, area squared, population density, percent female headed households, percent of households on public assistance, percent low-income households, per-capita income, median income, percent of households with incomes under \$20,000 and percent of households with incomes over \$40,000. State dummies are included. All models include sample weights. *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 3: OLS Coefficients for Percent Attending Private School in NELS88 Data
Each Coefficient Represents a Separate Regression

Area Competition Grade Level	County			PMSA			MSA		
	K-8	9-12	K-12	K-8	9-12	K-12	K-8	9-12	K-12
	1	2	3	4	5	6	7	8	9
Dependent Variable									
12th Grade Math Test (N=12,573)	-0.0230 (0.0496)	-0.0680 (0.0472)	-0.0452 (0.0540)	-0.0587 (0.0571)	-0.0533 (0.0589)	-0.0685 (0.0628)	-0.0521 (0.0544)	-0.0626 (0.0569)	-0.0673 (0.0597)
High School Graduation (N=13,243)	0.0003 (0.0012)	-0.0006 (0.0010)	-0.00002 (0.0012)	0.0007 (0.0014)	0.0002 (0.0013)	0.0006 (0.0015)	0.0004 (0.0013)	0.0002 (0.0014)	0.0004 (0.0015)
College Attendance (N=13,252)	0.0024 (0.0015)	-0.0001 (0.0013)	0.0018 (0.0015)	0.0015 (0.0020)	0.0014 (0.0020)	0.0016 (0.0022)	0.0022 (0.0018)	0.0027 (0.0019)	0.0026 (0.0019)

Notes: Robust standard errors that allow within county/PMSA/MSA correlation are in parentheses. All models contain the following individual-level variables: race, gender, parental education, religion, and number of siblings. All models also include the following area-level demographic variables: percent attending religious services, religious homogeneity index, various racial demographic variables, percent high school graduate, percent college graduate, percent male, percent urban, area, area squared, population density, percent female headed households, percent of households on public assistance, percent low-income households, per-capita income, median income, percent of households with incomes under \$20,000 and percent of households with incomes over \$40,000. State dummies are included. All models include sample weights. All models include sample weights. *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 4: First Stage Results from NLSY79 and NELS88
Dependent Variable is Percent of K-12 Students Attending Private School in Geographic Area

Geographic Area	NLSY79 Data			NELS88 Data		
	County 1	PMSA 2	MSA 3	County 4	PMSA 5	MSA 6
Catholic Percent of Population	0.5941*** (0.0780)	0.5843*** (0.0770)	0.5451*** (0.0789)	0.3264*** (0.0484)	0.2959*** (0.0515)	0.2432*** (0.0493)
Catholic Pct of Pop Sq	-0.0036*** (0.0007)	-0.0037*** (0.0006)	-0.0031*** (0.0007)	-0.0023*** (0.0006)	-0.0029*** (0.0007)	-0.0020*** (0.0006)
Catholic Population Density	0.0009 (0.0022)	-0.0090 (0.0077)	0.0003 (0.0115)	0.0049 (0.0050)	-0.0283*** (0.0099)	-0.0389 (0.0246)
Catholic Pop Density Sq (1000s)	0.00001 (0.0001)	0.000002 (0.000005)	0.01709 (0.0113)	-0.0003 (0.0007)	0.0266** (0.0130)	-0.0007 (0.0995)
Catholic Church Density	-5.32 (10.94)	16.48 (33.98)	63.79 (54.11)	42.09** (17.29)	345*** (57.42)	437*** (117.23)
Catholic Church Density Squared	-0.07 (1.45)	-50.0 (125)	-617** (308)	-9.06 (18.53)	-1458** (585)	-3006* (1810)
Percent Church Adherents	-0.1057*** (0.0372)	-0.1397*** (0.0408)	-0.1580*** (0.0363)	0.0056 (0.0179)	-0.0036 (0.0172)	-0.0058 (0.0164)
Religious Homogeneity Index	-0.0184 (0.0262)	-0.0005 (0.0257)	0.0174 (0.0246)	-0.0415* (0.0228)	-0.0590*** (0.0204)	-0.0616*** (0.0198)
Percent Black	0.1550** (0.0738)	0.34971*** (0.0747)	0.3327*** (0.0741)	0.0052 (0.0519)	0.1993*** (0.0468)	0.1706*** (0.0476)
Percent Hispanic	-0.1145*** (0.0378)	-0.0992** (0.0410)	-0.0926** (0.0370)	-0.0698* (0.0306)	0.0271 (0.0269)	-0.0046 (0.0281)
Percent High School Graduates	-0.0165 (0.0788)	-0.0616 (0.0675)	-0.0259 (0.0692)	-0.1103** (0.0550)	-0.1650*** (0.0560)	-0.1425** (0.0585)
Percent College Graduates	0.2268 (0.1492)	0.1761 (0.1248)	0.0744 (0.1572)	0.0538 (0.0706)	0.0386 (0.0731)	0.0338 (0.0807)
Percent Urban	0.0509** (0.0209)	0.0616*** (0.0224)	0.0796*** (0.0236)	0.0512*** (0.0128)	0.0476*** (0.0111)	0.0548*** (0.0126)
Pct Female Headed HH	-0.2227 (0.2273)	-0.1269 (0.2575)	-0.0976 (0.2476)	0.1303 (0.1139)	0.2539** (0.1212)	0.1880 (0.1224)
Observations	9,447	9,447	9,447	13,243	13,243	13,243
R-squared	0.79	0.84	0.86	0.83	0.88	0.89

Notes: Robust standard errors that allow within county/PMSA/MSA correlation are in parentheses. All models contain the following individual-level variables: race, gender, parental education, religion, number of siblings, and frequency of religious attendance (NLSY79 only). In addition to the variables shown, all models also include the following area-level demographic variables: various racial demographic variables, percent male, area, area squared, population density, percent of households on public assistance, percent low-income households, per-capita income, median income, percent of households with incomes under \$20,000 and percent of households with incomes over \$40,000. State dummies are included. All models include sample weights. *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 5: IV Coefficients for Percent Attending Private School in NLSY79 Data
Each Coefficient Represents a Separate Regression

Area Competition Grade Level	County			PMSA			MSA		
	K-8	9-12	K-12	K-8	9-12	K-12	K-8	9-12	K-12
	1	2	3	4	5	6	7	8	9
Dependent Variable									
AFQT Test (N=9,096)	0.2079 (0.1544)	0.2367 (0.1649)	0.2166 (0.1570)	0.2260 (0.1605)	0.3785** (0.1853)	0.2699 (0.1673)	0.2496 (0.1541)	0.2587 (0.1704)	0.2626* (0.1583)
Years of Schooling (N=9,447)	0.0185 (0.0150)	0.0195 (0.0161)	0.0189 (0.0153)	0.0143 (0.0154)	0.0272 (0.0175)	0.0179 (0.0160)	0.0270* (0.0147)	0.0296* (0.0158)	0.0287* (0.0148)
High School Graduation (N=9,472)	0.0021 (0.0019)	0.0022 (0.0021)	0.0021 (0.0019)	0.0016 (0.0019)	0.0025 (0.0022)	0.0018 (0.0020)	0.0023 (0.0019)	0.0016 (0.0019)	0.0021 (0.0019)
College Attendance (N=9,447)	0.0042 (0.0028)	0.0051* (0.0030)	0.0045 (0.0029)	0.0028 (0.0030)	0.0059* (0.0035)	0.0036 (0.0031)	0.0043 (0.0029)	0.0049 (0.0031)	0.0047 (0.0029)

Notes: Robust standard errors that allow within county/PMSA/MSA correlation are in parentheses. All models contain the following individual-level variables: race, gender, parental education, religion, number of siblings, and frequency of religious attendance. All models also include the following area-level demographic variables: percent attending religious services, religious homogeneity index, various racial demographic variables, percent high school graduate, percent college graduate, percent male, percent urban, area, area squared, population density, percent female headed households, percent of households on public assistance, percent low-income households, per-capita income, median income, percent of households with incomes under \$20,000 and percent of households with incomes over \$40,000. State dummies are included. All models include sample weights. *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Table 6: IV Coefficients for Percent Attending Private School in NELS88 Data
Each Coefficient Represents a Separate Regression

Area Competition Grade Level	County			PMSA			MSA		
	K-8	9-12	K-12	K-8	9-12	K-12	K-8	9-12	K-12
	1	2	3	4	5	6	7	8	9
Dependent Variable									
12th Grade Math Test (N=12,573)	-0.0548 (0.0897)	-0.0906 (0.1145)	-0.0684 (0.0974)	-0.0958 (0.0844)	-0.0966 (0.1189)	-0.1026 (0.0937)	-0.0799 (0.0906)	-0.1288 (0.1350)	-0.0935 (0.1022)
High School Graduation (N=13,243)	-0.0008 (0.0025)	-0.0001 (0.0028)	-0.0006 (0.0026)	-0.0024 (0.0024)	-0.0021 (0.0034)	-0.0025 (0.0027)	-0.0028 (0.0022)	-0.0041 (0.0034)	-0.0032 (0.0025)
College Attendance (N=13,252)	0.0059** (0.0027)	0.0044 (0.0027)	0.0059** (0.0028)	0.0045 (0.0030)	0.0040 (0.0042)	0.0047 (0.0034)	0.0017 (0.0031)	-0.0006 (0.0046)	0.0014 (0.0035)

Notes: Robust standard errors that allow within county/PMSA/MSA correlation are in parentheses. All models contain the following individual-level variables: race, gender, parental education, religion, and number of siblings. All models also include the following area-level demographic variables: percent attending religious services, religious homogeneity index, various racial demographic variables, percent high school graduate, percent college graduate, percent male, percent urban, area, area squared, population density, percent female headed households, percent of households on public assistance, percent low-income households, per-capita income, median income, percent of households with incomes under \$20,000 and percent of households with incomes over \$40,000. State dummies are included. All models include sample weights. *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

App. Table 1: Weighted Descriptive Statistics for Individual Level NLSY79 and NELS88 Data

NLSY79 Data	Obs	Mean	Std Dev	Min	Max
AFQT Test Score	9,075	49.01	28.47	1	99
Years of Schooling	9,447	13.12	2.48	0	20
High School Grad / GED	9,447	0.89	0.31	0	1
College Attendance	9,447	0.45	0.50	0	1
Female	9,447	0.50	0.50	0	1
White/Other Race	9,447	0.81	0.39	0	1
Black	9,447	0.14	0.34	0	1
Hispanic	9,447	0.06	0.23	0	1
Parental Education	9,447	12.56	3.03	0	20
Number of Siblings	9,447	3.31	2.27	0	22
Catholic Religion	9,447	0.32	0.47	0	1
Lutheran Religion	9,447	0.08	0.26	0	1
Methodist Religion	9,447	0.09	0.29	0	1
Baptist Religion	9,447	0.24	0.43	0	1
Episcopalian Religion	9,447	0.02	0.14	0	1
Presbyterian Religion	9,447	0.04	0.19	0	1
Jewish Religion	9,447	0.01	0.12	0	1
Other Christian Religion	9,447	0.16	0.36	0	1
Other Religion	9,447	0.002	0.05	0	1
No Religion	9,447	0.04	0.20	0	1
Freq. of Religious Attendance	9,447	3.20	1.71	1	6
NELS88 Data	Obs	Mean	Std Dev	Min	Max
Grade 12 Math Test	10,009	48.07	14.45	17	78
Grade 12 Reading Test	10,012	33.06	10.11	10	51
High School Graduation	13,243	0.83	0.37	0	1
College Attendance	13,243	0.65	0.48	0	1
Female	13,243	0.50	0.50	0	1
White/Other Race	13,243	0.77	0.42	0	1
Black	13,243	0.13	0.33	0	1
Hispanic	13,243	0.10	0.30	0	1
Parental Education	13,243	14.11	2.43	10	20
Number of Siblings	13,243	2.27	1.57	0	6
Catholic Religion	13,243	0.29	0.45	0	1
Lutheran Religion	13,243	0.06	0.24	0	1
Methodist Religion	13,243	0.09	0.29	0	1
Baptist Religion	13,243	0.23	0.42	0	1
Episcopalian Religion	13,243	0.02	0.13	0	1
Presbyterian Religion	13,243	0.04	0.19	0	1
Jewish Religion	13,243	0.02	0.15	0	1
Other Christian Religion	13,243	0.16	0.37	0	1
Other Religion	13,243	0.05	0.22	0	1
No Religion	13,243	0.03	0.18	0	1

Appendix Table 2: Weighted Descriptive Statistics for Area Level NLSY79 Data

Area Level Variables	County		PMSA		MSA	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
K-8 Percent Private School	11.65	7.60	11.48	6.97	11.78	6.95
9-12 Percent Private School	9.53	7.71	9.38	6.71	9.56	6.62
K-12 Percent Private School	11.05	7.42	10.88	6.71	11.15	6.68
K-8 Pct Private School (Census)	10.34	5.55	10.28	5.09	10.49	5.08
9-12 Pct Private School (Census)	7.75	4.91	7.71	4.35	7.82	4.25
K-12 Pct Private School (Census)	9.52	5.22	9.46	4.73	9.64	4.69
K-8 Percent Catholic School	7.59	6.69	7.55	6.14	7.78	6.18
9-12 Percent Catholic School	6.28	6.50	6.24	5.62	6.39	5.57
K-12 Percent Catholic School	7.22	6.48	7.17	5.88	7.38	5.90
Catholic Percent of Population	21.73	16.58	21.74	16.10	21.96	16.07
Catholic Population Density	531	1487	246	622	207	346
Catholic Church Density	0.14	0.35	0.06	0.12	0.05	0.07
Percent Church Adherents	51.16	14.69	50.79	14.18	50.85	14.05
Religious Homogeneity Index	34.29	16.24	34.01	15.40	34.49	15.42
Racial Homogeneity Index	76.56	16.81	75.90	15.00	75.23	14.75
Percent White	83.40	14.32	83.99	11.34	83.68	10.95
Percent Black	12.48	13.59	11.85	10.60	11.94	10.38
Percent Hispanic	5.89	10.54	5.92	10.44	6.07	10.16
Percent Asian/Pacific Islander	1.15	2.14	1.17	2.01	1.27	2.04
Percent Native American	0.60	2.03	0.60	2.03	0.60	2.03
Percent High School Graduates	65.51	10.06	65.73	9.27	65.74	9.14
Percent College Graduates	15.26	5.65	15.36	5.00	15.53	4.86
Percent Male	48.52	1.23	48.54	1.02	48.52	1.01
Percent Urban	72.27	28.34	71.87	25.46	72.41	25.71
Area	1,125	2,060	2,254	2,882	3,764	6,058
Population (000s)	714	1,345	1,483	2,213	2,953	4,805
Population Density	1,664	4,775	730	1,347	630	846
Percent Female Headed HH	14.12	5.24	13.78	3.56	13.90	3.40
Percent Receiving Public Asst.	4.38	2.60	4.22	1.77	4.26	1.68
Percent Poor	9.82	4.80	9.57	4.06	9.62	3.91
Per Capita Income (000s)	10.44	2.19	10.45	2.00	10.51	1.97
Median Income (000s)	17.00	3.61	17.14	3.15	17.12	2.96
Percent HH w/ income <20,000	58.95	10.80	58.66	9.55	58.79	9.02
Percent HH w/ income 20,000-40,000	32.30	7.18	32.50	6.26	32.38	5.96
Percent HH w/ income >40,000	8.75	4.53	8.84	4.08	8.83	3.80

Appendix Table 3: Weighted Descriptive Statistics for Area Level NELS88 Data

Area Level Variables	County		PMSA		MSA	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
K-8 Percent Private School	9.93	7.17	10.07	6.66	10.04	6.39
9-12 Percent Private School	7.78	7.44	7.97	6.57	7.96	6.27
K-12 Percent Private School	9.34	7.01	9.49	6.45	9.46	6.18
K-12 Pct Private School (Census)	9.42	5.80	9.50	5.40	9.49	5.20
K-8 Percent Catholic School	5.95	5.73	6.02	5.40	6.01	5.26
9-12 Percent Catholic School	4.82	5.92	4.94	5.31	4.96	5.11
K-12 Percent Catholic School	5.64	5.58	5.72	5.22	5.72	5.08
Catholic Percent of Population	21.36	17.57	21.48	16.95	21.60	16.85
Catholic Population Density	195.88	711.75	96.79	213.27	65.33	106.36
Catholic Church Density	0.04	0.12	0.02	0.03	0.02	0.02
Percent Church Adherents	57.05	15.28	56.96	14.45	57.07	14.22
Religious Homogeneity Index	30.39	15.51	30.07	15.10	30.15	14.88
Racial Homogeneity Index	69.57	19.93	67.68	19.21	67.04	19.06
Percent White	76.84	20.50	76.26	18.94	76.28	17.81
Percent Black	11.22	12.75	11.58	10.74	11.64	10.45
Percent Hispanic	8.75	15.11	8.93	15.04	8.86	14.44
Percent Asian/Pacific Islander	2.32	4.51	2.35	4.45	2.34	4.23
Percent Native American	0.79	3.56	0.79	3.56	0.79	3.56
Percent Other Ethnicity	0.09	0.14	0.09	0.11	0.09	0.11
Percent High School Graduates	74.37	8.98	74.16	8.18	74.18	7.97
Percent College Graduates	19.25	7.85	19.20	6.83	19.18	6.44
Percent Male	48.65	1.20	48.64	1.11	48.64	1.08
Percent Urban	71.43	28.91	71.75	26.58	71.94	26.57
Area	3,222	5,442	5,955	7,267	11,449	17,504
Population (000s)	814	1,645	1,510	2,227	3,172	5,205
Population Density	596	1,895	291	514	206	237
Percent Female Headed HH	15.79	5.59	15.92	4.11	15.93	3.69
Percent Receiving Public Asst.	7.67	3.75	7.78	3.19	7.81	2.98
Percent Poor	13.68	6.99	13.85	6.30	13.85	6.06
Per Capita Income (000s)	13.93	3.73	13.87	3.32	13.86	3.11
Median Income (000s)	29.94	8.22	29.66	7.16	29.69	6.81
Percent HH w/ income <20,000	34.31	10.87	34.68	9.75	34.66	9.49
Percent HH w/ income 20,000-40,000	31.45	3.96	31.52	3.60	31.50	3.51
Percent HH w/ income >40,000	34.24	12.41	33.80	11.01	33.83	10.62