Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools[1]

Joseph G. Altonji[2]
Todd E. Elder[3]
Christopher R. Taber[2]

April 4, 2002

[1]We thank Timothy Donohue and Emiko Usui for excellent research assistance. We also received helpful comments from Glen Cain, Thomas DeLeire, Lars Hansen, James Heckman, Robert LaLonde, Jean-Marc Robin, George Jakubson and especially Tim Conley and Derek Neal as well as participants in seminars at the American Economic Association Winter meetings (January 2000), Boston College, Cornell University, CREST-INSEE, Duke University, IZA, Johns Hopkins University, Harvard University, MIT, Northwestern University, Princeton University, University of Chicago, University College London, University of Florida, University of Maryland, University of Missouri, University of Rochester, and University of Wisconsin at Madison. We are grateful for support from the National Science Foundation grant 9512009 (Altonji), the National Institute of Child Health and Development grant R01 HD36480-03 (Altonji and Taber), and the Institute for Policy Research, Northwestern University.

[2]Department of Economics and Institute for Policy Research, Northwestern University and NBER

[3]Department of Economics and Institute for Policy Research, Northwestern University

Abstract

In this paper we measure the effect of Catholic high school attendance on educational attainment and test scores. Because we do not have a good instrumental variable for Catholic school attendance, we develop new estimation methods based on the idea that the amount of selection on the observed explanatory variables in a model provides a guide to the amount of selection on the unobservables. We also propose an informal way to assess selectivity bias based on measuring the ratio of selection on unobservables to selection on observables that would be required if one is to attribute the entire effect of Catholic school attendance to selection bias. We use our methods to estimate the effect of attending a Catholic high school on a variety of outcomes. Our main conclusion is that Catholic high schools substantially increase the probability of graduating from high school and, more tentatively, college attendance. We find little evidence of an effect on test scores.

1 Introduction

Distinguishing between correlation and causality is the most difficult challenge faced by empirical researchers in the social sciences. Social scientists rarely are in a position to run a well controlled experiment. Consequently, they rely on a priori restrictions on the patterns of interaction among the variables that are observed or unobserved. These restrictions are typically in the form of exclusion restrictions or assumptions about the functional form of the model, the distribution of the unobserved variables, or dynamic interactions. Occasionally, the a priori restrictions are derived from a widely accepted theory or are supported by other studies that had access to a richer set of data. However, in most cases, doubt remains about the validity of the identifying assumptions and the inferences that are based on them.

The challenge of isolating causal effects is particularly difficult for the question addressed in our paper—"Do Catholic high schools provide better education than public schools?" This question is at the center of the debate in the United States over whether vouchers, charter schools, and other reforms that increase choice in education will improve the quality of education. It is also highly relevant to the search for ways to improve teaching and governance of public schools. Simple cross tabulations or multivariate regressions of outcomes such as test scores and post secondary educational attainment typically show a substantial positive effect of Catholic school attendance.[1] However, many prominent social scientists, such as Goldberger and Cain (1982), have argued that the positive effects of Catholic school attendance may be due to spurious correlations between Catholic school attendance and unobserved student and family characteristics. The argument begins with the observation that it costs parents time and money to send their children to private school. In the absence of experimental data, the challenge in addressing this potentially large bias is finding exogenous variation that affects school choice but not outcomes. Most student background characteristics that influence the Catholic school decision, such as income, attitudes, and education of the parents, are likely to influence outcomes independently of the school sector because they are likely to be related to other parental inputs. Characteristics of private and public schools that influence choice, such as tuition levels, student body characteristics, or

[1] The most influential examples are Coleman, Hoffer, and Kilgore (1982) and Coleman and Hoffer (1987). Other early examples of studies of Catholic schools and other private schools are Noell (1982), Goldberger and Cain (1982), and Alexander and Pallas (1985). Recent studies include Evans and Schwab (1995), Tyler (1994), Neal (1997), Grogger and Neal (2000), Figlio and Stone (2000), Sander (2000) and Jepsen (2000). Murnane (1984), Witte (1992), Chubb and Moe (1990), Cookson (1993) and Neal (1998) provide overviews of the discussion and references to the literature. Grogger and Neal provide citations to a small experimental literature, which for the most part has found positive effects of Catholic school.

school policies, are likely to be related to the effectiveness of the schools.

Several recent studies, including Evans and Schwab (1995), Neal (1997), Grogger and Neal (2000), Figlio and Stone (2000) and Altonji, Elder and Taber (1999) use various exclusion restrictions to estimate the Catholic school effect on a variety of outcomes. Evans and Schwab (1995) use religious affiliation as an exogenous source of variation in Catholic school attendance and confirm the large positive estimates of Catholic school effects on high school graduation and college attendance that they obtain when Catholic school attendance is treated as exogenous. However, as Evans and Schwab recognize and Murnane (1984) and Neal (1998) note, being Catholic could well be correlated with characteristics of the neighborhood and family that influence the effectiveness of schools. Another influential paper by Neal (1997) uses proxies for geographic proximity to Catholic schools as an exogenous source of variation in Catholic high school attendance. The basic assumption is that the location of Catholics and/or Catholic schools was determined by historical circumstances and is independent of unobservables that influence performance in schools. He finds evidence of a positive effect of Catholic high school attendance on high school and college graduation among students in urban areas, particularly in the case of nonwhites. In Altonji, Elder and Taber (1999, 2001), we employ a similar methodology using data on zip code of residence and the zip codes of all of the Catholic high schools in the country to compute a measure of distance from the nearest Catholic high school for our samples. We conclude that the use of location or location interacted with religion is not a good way to estimate Catholic school effects.[2] Grogger and Neal (2000) come to a similar conclusion.[3] Altonji, Elder, and Taber (2001) also find that Catholic religion has a strong association with graduation rates for students who attended public eighth grades even though such students rarely attend Catholic high school. This evidence and work by Ludwig (1997) raises serious doubts about the validity of Catholic religion as an instrument.

[2]We provide evidence based on links to observed variables and to eighth grade test scores that suggests that neither distance from Catholic high schools nor the interaction between distance and religion should be excluded from the outcome equations unless detailed controls for location are included. (This informal use of observables as a guide to correlation between the instrument and the unobservables led to the current paper.) Failure to control for these factors leads to negative biases in estimates of Catholic school effects. Unfortunately, including detailed geographic controls (such as 3 digit zip code) leads to very large standard errors. We also follow Neal (1997) and Evans and Schwab (1995) by using bivariate probit models to jointly estimate the Catholic School decision with the outcomes. We find that empirical identification comes largely from the functional form of the model rather than exclusion of the measure of distance from Catholic schools. Nonlinearities in the effects of student background rather than proxies for distance from Catholic schools seem to be the main source of identification.

[3]Grogger and Neal (2000) use NELS:88, the data set for the present study. Altonji, Elder and Taber (1999) analyze NELS:88, the National Longitudinal Survey of the High School Class of 1972, and NLSY79. Neal (1997) uses the NLSY79.

In this paper we develop new estimation strategies that may be helpful when strong prior information is unavailable regarding the exogeneity of either the variable of interest or instruments for that variable. We view this to be the situation in studies of Catholic school effects and in many other applications in economics and the other social sciences. We then use our strategies to assess the effectiveness of Catholic schools.

Our approach uses the degree of selection on observables as a guide to the degree of selection on the unobservables. Researchers often informally argue for the exogeneity of an explanatory variable or an instrumental variable by examining the relationship between the instrumental variable and a set of observed characteristics, or by assessing whether point estimates are sensitive to the inclusion of additional control variables.[4] We provide a formal analysis confirming the intuition that such evidence can be informative in some situations. More importantly, we provide a way to quantitatively assess the degree of omitted variables bias.[5]

Using our Catholic schools application, let the outcome Y be a function of the latent variable[6] Y^* which is determined as

$$\begin{aligned} Y^* &= \alpha CH + W'\Gamma \\ &= \alpha CH + X'\Gamma_X + \xi, \end{aligned}$$

where CH is an indicator for whether the student attends a Catholic high school, the parameter α is the effect of Catholic school attendance on Y^*, W is the vector of characteristics (observed and unobserved) that determine Y and Γ is the causal effect of W on Y^*. In the second part of the equation X is the vector of observed variables, Γ_X is the corresponding subvector of Γ, and the error component ξ is an index of the unobserved variables. Because it is extremely unlikely that the control variables X are all unrelated to ξ, we work with

[4] See for example, Currie and Duncan (1995), Engen et al (1996), Poterba et al (1994), Angrist and Evans (1988), Jacobsen et al. (1999), Bronars and Grogger (1994), Udry (1996),Cameron and Taber (2001), or Angrist and Krueger (1999). Wooldridge's (2000) undergraduate textbook contains a computer exercise (15.14) that instructs students to look for a relationship between an observable (IQ) and an instrumental variable (closeness to college).

[5] Two precursors to our study are Altonji's (1988) study of the importance of observed and unobserved family background and school characteristics in the school specific variance of educational outcomes and especially Murphy and Topel's (1990) study of the importance of selection on unobserved ability as an explanation for industry wage differentials.

[6] We will focus on two special cases in this paper. The first is a continuous dependent variable in which $Y = Y^*$. The second is a binary variable in which $Y = 1(Y^* > 0)$.

$$Y^* = \alpha CH + X'\gamma + \varepsilon, \tag{1.1}$$

where γ is defined so that $\text{cov}(\varepsilon, X) = 0$. Consequently, γ captures both the direct effect of X on Y^*, Γ_X, as well as the relationship between X and the mean of ξ. Let CH^* be the latent variable that determines CH and let $CH = 1(CH^* > 0)$, where the indicator function $1(\cdot)$ is 1 when $CH^* > 0$ and 0 otherwise. Consider the linear projection of CH^* onto $X'\gamma$ and ε,

$$\text{Proj}(CH^*|X'\gamma, \varepsilon) = \phi_0 + \phi_{X'\gamma}X'\gamma + \phi_\varepsilon \varepsilon. \tag{1.2}$$

We formalize the idea that "selection on the unobservables is the same as selection on the observables" as

Condition 1

$$\phi_\varepsilon = \phi_{X'\gamma}.$$

We contrast this with the OLS condition,

Condition 2

$$\phi_\varepsilon = 0.$$

Roughly speaking, Condition 1 says that the part of Y^* that is related to the observables and the part related to the unobservables have the **same** relationship with CH^*. Condition 2 says that the part of Y related to the unobservables has **no** relationship with CH^*.

We show that Condition 1 requires three types of assumptions. The first is that the elements of X are chosen at random from W. The second is that the number of elements in X and W are large, so that none of the elements dominates the distribution of school choice CH or the latent variable Y^*. These two assumptions are enough to establish asymptotic equality of the coefficients of the projection of CH^* onto $X'\Gamma_X$ and ξ .[7] To establish Condition 1 we need an additional, very strong, assumption about the relationship between X and the unobservable elements that determine ξ.

While the assumptions that lead to Condition 1 are strong and unlikely to hold exactly, they are no less objectionable than the OLS assumptions leading to Condition 2:

[7] We take asymptotic approximations as the number of elements in W grows large.

$Cov(CH, \xi) = 0$ and $Cov(X, \xi) = 0$.[8] As we discuss in more detail in Section 3, major data sets with large samples and extensive questionnaires are not designed to address one relatively specific question, such as the effectiveness of Catholic schools, but rather to serve multiple purposes. Because there are a limited number of factors that we expect to matter for a particular outcome, know how to collect, and can afford to collect, many relevant variables are left out. This is reflected in the typically low explanatory power of social science models of individual behavior. Furthermore, in many applications, including ours, the endogenous variable is correlated with many of the elements of X. Given the constraints that shape the choice of X and the fact that many of the elements of X are systematically related to CH^*, it is unlikely that the many unobserved variables that determine ξ are unrelated to CH^*, which is basically what $Cov(X, \xi) = 0$ requires. Given that the X variables are intercorrelated, the assumption that $Cov(X, \xi) = 0$ is likely to be a poor approximation to reality even though it is made in virtually all empirical studies in the social sciences.

We prove that Condition 1 provides identifying information for α, although it does not always deliver point identification. More importantly, we argue that Conditions 1 and 2 represent extreme assumptions about the degree of selection on unobservables and the truth is probably somewhere in between, with

Condition 3

$$0 \le \phi_\varepsilon \le \phi_{X'\gamma}.$$

We prove that Condition 3 allows us to estimate a set of permissible values of α. **We view analysis based on Condition 1 and Condition 3 as a complement to the standard analysis based on Condition 2, not as a replacement for it.**

We also propose and justify a closely related but more informal way to use the relationship between the observables as a guide to endogeneity bias. It is related to the common practice of checking for a systematic relationship between CH and the mean of the elements of X. Let $E(.)$ and $Var(.)$ be expected value and variance operators. We compute estimates of $\frac{E(X'\gamma|CH=1)-E(X'\gamma|CH=0)}{Var(X'\gamma)}$, which is a measure of the degree to which the index of observables in the outcome equation varies with CH. We then ask how many times larger

[8] Technically these two assumptions are sufficient for Condition 2 but not necessary. One possibility is that $Cov(CH, X) = Cov(CH, \xi) = 0$ but this condition implies Condition 1 as well. The necessary condition for $\phi_\varepsilon = 0$ and for OLS to be unbiased is $Cov(CH, \varepsilon) \equiv Cov(CH, \xi - E(\xi|X))$
$= Cov(CH, \xi) - Cov(CH, X')Var(X)^{-1}Cov(X, \xi) = 0.$
The latter condition can hold if both $Cov(CH, \xi) = 0$ and $Cov(X, \xi) = 0$ happen to fail in a way that leads to a perfect cancellation of biases, or if $Cov(CH, \xi) = 0$, $Cov(X, \xi) \neq 0$, but $Cov(CH, X')Var(X)^{-1}Cov(X, \xi) = 0$.. Neither of these cases is very interesting.

the normalized shift in the index of the unobservables $\frac{E(\varepsilon|CH=1)-E(\varepsilon|CH=0)}{Var(\varepsilon)}$ would have to be to explain away the entire estimate of α. The null hypothesis that the single equation estimator of α is unbiased corresponds to the case in which $\frac{E(\varepsilon|CH=1)-E(\varepsilon|CH=0)}{Var(\varepsilon)}$ is 0, while the hypothesis that X is a randomly chosen subset of W implies that

$$(1.3) \qquad \frac{E(\varepsilon|CH=1)-E(\varepsilon|CH=0)}{Var(\varepsilon)} = \frac{E(X'\gamma|CH=1)-E(X'\gamma|CH=0)}{Var(X'\gamma)}.$$

If selection on unobservables must be several times stronger than selection on the observables to explain the entire estimate of α, then the case for a causal effect of Catholic school is strengthened. We provide similar estimation methods that can be used as complements to standard IV type estimators when an excluded variable (e.g., Catholic religion or proximity to a Catholic school in the Catholic schools literature) is used to identify a model, but there are concerns about whether it is exogenous.

In section 2 we set the stage for the development and application of our econometric methods by providing a standard multivariate analysis of the Catholic school effect using the National Educational Longitudinal Survey of 1988 (NELS:88). We present descriptive statistics on the relationship between Catholic school attendance and a broad range of observable measures of family background, eighth grade achievement, educational expectations, social behavior, and delinquency. The descriptive statistics show huge Catholic high school advantages in high school graduation and college attendance rates, and smaller ones in 12th grade test scores. However, the evidence across the wide range of observables, which have substantial explanatory power in our outcome equations, suggests fairly strong positive selection into Catholic schools. We also find that the link between observables and Catholic high school attendance is much weaker among children who attended Catholic eighth grade and that public school eighth graders almost never attend Catholic high school. These facts suggest that we can improve comparability of the "treatment" and control groups and avoid confounding the effect of attending Catholic high school with the effect of Catholic elementary school by focusing on the Catholic eighth grade sample. This is what we do, in contrast to most of the literature.

We present an initial set of regression and probit models containing detailed controls for student characteristics that are determined prior to high school. We find a small positive effect on 12th grade math scores and a zero effect on reading scores. However, our estimates of the effect of Catholic high school point to a very large positive effect of 0.15 on the probability of attending a 4 year college 2 years after high school and 0.08 on the high school graduation rate. The estimates are not very sensitive to the addition of a powerful set of controls, particularly in the case of the high school graduate rate. The insensitivity of the results to the controls and the "modest" association between the observables that determine the outcome and Catholic high school suggests that part of the educational attainment effect is real. However, the small positive effects on math test scores could easily be accounted for by positive selection on unobservables.

In sections 3 and 4 we develop and apply our methods for using the degree of selection on observables to provide better guidance about bias from selection on unobservables. Because high school outcomes depend on many variables that are determined after the decision to attend Catholic high school is made, selection on unobservables that affect outcomes is likely to be positive but weaker than selection on observables, with $0 \leq \phi_\varepsilon \leq \phi_{X'\gamma}$. Consequently, our estimates of a joint model of Catholic high school attendance and educational attainment subject to Condition 1 are likely to overstate selection and understate the Catholic school effect. Operationally, we use a bivariate probit model without exclusion restrictions as the functional form for the joint model even though the identification results are nonparametric. The estimate of the effect of Catholic school on high school graduation declines from the univariate estimate of about 0.08, which we view as an upper bound, to 0.07 when we impose equal selection, which we view as a lower bound, although sampling error widens this range. The estimate of the effect on college attendance declines from the univariate estimate of 0.15 to 0.07 or 0.02, depending on the details of the estimation method.

Using (1.3) we estimate that selection on unobservables would have to be between 2.78 and 4.29 times stronger than selection on the observables to explain away the entire Catholic school effect on high school graduation, which seems highly unlikely. It would have to be between 1.30 and 2.30 times stronger to explain away the entire college effect, which is also unlikely. However, more modest positive selection on the unobservables could explain away the entire Catholic school effect on math scores. We conclude that Catholic high school attendance substantially boosts high school graduation rates and, more tentatively, college attendance rates.

In section 5 we extend our analysis to a subsample of urban minorities, for whom we obtain larger univariate effects but also stronger evidence of selection. In section 6 we provide conclusions and an agenda for further research on the use of observables as a guide to selection bias.

2 A Preliminary Analysis of the Catholic School Effect

In section 2.1 we describe the data. In section 2.2 we present the sample means of outcomes, measures of family background, eighth grade achievement, social behavior, and delinquency as a way of assessing the potential importance of selection bias and to motivate the choice of sample. In section 2.3 we present probit and OLS regression estimates of the Catholic school effect. These serve as a benchmark for our subsequent analysis. In section 2.4 we present an analysis of the sensitivity of the Catholic high school effect to assumptions about the degree of selection on unobservables.

2.1 Data

Our data set is NELS:88, a National Center for Education Statistics (NCES) survey which began in the Spring of 1988. The base year sample is a two stage stratified probability sample in which a set of schools containing eighth grades were chosen on the basis of school size and whether they were classified as private or public. In the second stage, as many as 26 eighth grade students from within a particular school were chosen based on race and gender. A total of 1032 schools contributed student data in the base year survey, resulting in 24,599 eighth graders participating. Subsamples of these individuals were reinterviewed in 1990, 1992, and 1994. The NCES only attempted to contact 20,062 base-year respondents in the first and second follow-ups, and only 14,041 in the 1994 survey. Additional observations are lost due to attrition.

The NELS:88 contains information on a wide variety of outcomes, including test scores and other measures of achievement, high school dropout and graduation status, and post-secondary education (in the 1994 survey only). Parent, student, and teacher surveys in the base year provide a rich set of information on family and individual background, as well as pre-high school achievement, behavior, and expectations of success in high school and beyond. Each student was also administered a series of cognitive tests in the 1988, 1990, and 1992 surveys to ascertain aptitude and achievement in math, science, reading, and history. We use standardized item response theory (IRT) test scores that account for the fact that the difficulty of the 10th and 12th grade tests taken by a student depends on the 8th grade scores. We use the 8th grade test scores as control variables and the 10th and 12th grade reading and math tests as outcome measures.

We also use high school graduation and college attendance as outcome measures. The high school graduation variable is equal to one if the respondent graduated high school by the date of the 1994 survey, and zero otherwise.[9] The "College attendance" indicator is one if the respondent was enrolled in a four-year college at the time of the 1994 survey and zero otherwise.[10] The indicator variable for Catholic high school attendance, CH, is one if the current or last school in which the respondent was enrolled was Catholic as of 1990 (two years after the eighth grade year) and zero otherwise.[11]

We estimate models using a full sample, a Catholic eighth grade sample, and various other subsamples. We always exclude approximately 400 respondents who attended non-Catholic private high schools or private, non-Catholic eighth grades. Observations with missing values of key eighth grade or geographic control variables (such as distance from the nearest Catholic high school) were dropped. Sample sizes vary across dependent variables because of data availability and are presented in the tables. The sampling probabilities for the NELS:88 followups depend on choice of private high school and the dropout decision, so sample weights are used to avoid bias from a choice based sample. Unless noted otherwise, the results reported in the paper are weighted.[12] Details regarding construction of variables

[9] We obtain similar results using a "drop out" dummy variable which equals one if a student dropped out of high school by 1992, or if the student dropped out of high school by 1990 and was not reinterviewed in 1992 or 1994, zero otherwise. This variable catches dropouts who left the survey by 1990 and were either dropped from the sample or were nonrespondents.

[10] Our major findings are robust to whether or not college attendance is limited to 4-year universities, full-time versus part-time, or enrolled in college "at some time since high school" or at the survey date.

[11] A student who started in a Catholic high school and transferred to a public school prior to the tenth grade survey would be coded as attending a public high school ($CH = 0$). If such transfers are frequently motivated by discipline problems, poor performance, or alienation from school, then misclassification of the transfers as public high school students could lead to upward bias in estimates of the effect of CH on educational attainment. We investigated this issue using an 8th grade question about whether the student expected to attend Catholic high school and information about whether the student had changed high schools prior to the 10th grade survey. Among Catholic school 8th graders for whom we have the relevant data, 832 of 889 kids (94%) who reported that they expected to attend Catholic high school actually attended Catholic high school. Among the remaining 57, only 12 students had transferred at least once and of these only 3 failed to graduate high school. Furthermore, it is quite possible that 1 or 2 of these students never started Catholic high school, perhaps because of a family move. We conclude that any bias from misclassification of students is small.

[12] In the initial sample, private schools and schools with a minority enrollment of over 19 percent were oversampled. The probability of sampling in the first and second follow-ups is smaller for high schools attended by fewer than 10 students from the NELS:88 base year sample, and the sample weights are inversely related to the number of sample members in the high school. This is likely to lead to undersampling of students who attend private high schools. In contrast, the third follow-up sample design oversamples those who attended private high schools. Furthermore, the sampling probability depends on whether the student was believed to have dropped out of high school. Because the sample probabilities depend on an endogenous right-hand side variable and the school attainment variables, it is necessary to weight the analysis to obtain consistent parameter estimates. We use the first follow-up panel weights for the analysis of 10th grade test scores, the second follow-up panel weights for the analysis of 12th grade scores, and the third follow-up cross section weights for the analysis of high school graduation and college attendance. The results are somewhat sensitive to the use of sample weights, although our main findings are robust to weighting. Given the sampling scheme the weighted estimates are clearly preferred.

and the composition of the sample are provided in Appendix B.

2.2 Characteristics of Catholic High School and Public High School Students by Eighth Grade Sector

In Table 1 we report the weighted means by high school sector of a set of family background characteristics, student characteristics, eighth grade outcomes, and high school outcomes. We report results separately for students who attended Catholic eighth grades (the "C8" sample) and for the full sample. The "outcomes" category displays by high school sector the college attendance rate, high school graduation rate, and 10th and 12th grade math and reading test scores for students from the NELS:88 sample.[13] Looking at the full sample, the graduation and college attendance rates differ enormously between the two sectors. Catholic high school students are far less likely to drop out of high school than their public school counterparts (0.02 versus 0.15), and are almost twice as likely to be enrolled in a four year college in 1994 (0.59 versus 0.31). Differences in twelfth grade test scores are more modest but still substantial—about 0.4 of a standard deviation higher for Catholic high school students. In the C8 sample the gap in the dropout rate is also very large (0.02 versus 0.12) as is the gap in the college attendance rate (0.61 versus 0.38). In contrast, the gap in the 12th grade math score is only about 0.25 standard deviations. Table 2 shows that the gaps in school attainment and test scores are even more dramatic for minority students in urban schools.

Tables 1 and 2 also show that the means of favorable family background measures, 8th grade test scores and grades, and positive behavior measures in eighth grade are substantially higher for the students who attend Catholic high schools. The large discrepancies for many of the variables raise the possibility that part or even all of the gap in outcomes may be a reflection of who attends Catholic high school. However, the gap is much lower for most variables in the case of Catholic eighth graders. For example, the gap in log family income is 0.49 for the full sample but only 0.19 for the C8 sample. The high school sector gap in measures of the parents' educational expectations for the child is more favorable to the students who attend Catholic high school in the full sample than in the eighth grade sample, and the difference in the student's expected years of schooling is 0.72 in the full sample but only 0.40 in the C8 sample.[14] The high school sector differential in father's

[13] In Table 1 and Table 2 the outcome variables are weighted with the same weights used in the regression analysis, as described in the previous section. All other variables are weighted using second follow-up panel weights.

[14] Appendix B and the footnotes to Table 3 provide the complete list of variables used in our multivariate models. Many are excluded from Tables 1 and 2 to keep them manageable. The expectations variables in

education is about one year in both samples, but for mother's education it is 0.75 for the full sample and 0.54 for the C8 sample. The discrepancy in the fraction of students who repeated a grade in grades 4-8 is -0.05 in the full sample and only -0.01 in the C8 sample, and the gap in the fraction of students who are frequently disruptive is -0.05 in the full sample and 0 in the C8 sample. Both of these variables are powerful predictors of high school graduation. Finally, the gap in the 8th grade reading and math scores are 3.86 and 3.44, respectively, in the full sample, but only 1.47 and 1.09, respectively, in the C8 sample.

These results hold for most of the other variables in Table 1. Specifically, differences by high school sector among the family background characteristics and eighth grade outcomes are much smaller for Catholic eighth graders than for public eighth graders. This pattern is consistent with the presumption that since the parents of 8th graders from Catholic schools have already chosen to avoid public school at the primary level, other, arguably more idiosyncratic factors, are likely to drive selection into Catholic high schools from Catholic eighth grade. Intuitively, it seems likely that these factors could lead to less selection bias than in the full sample, although the overwhelming evidence based on a very broad set of 8th grade observables is that selection bias is positive in both samples. These considerations, the desire to avoid confounding the Catholic high school effect with the effect of Catholic elementary school, and a concern about selection bias that arises from the fact that only 0.3% of public school eighth graders in our effective sample go to Catholic high school lead us to focus on the sample of Catholic eighth graders in most of our analysis.[15]

2.3 Estimates of the Effect of Catholic High Schools

In this section of the paper we present regression and univariate probit estimates of the effects of Catholic high school attendance on a set of outcomes. For reasons discussed above, we focus on the subsample who attended Catholic eighth grade, although we also present results for the full sample.

In the top panel of Table 3 we report the coefficient on CH in univariate probit, OLS,

Tables 1 and 2 are excluded from our outcome models because if Catholic school has an effect on outcomes, this may be influence expectations.

[15] This is an unweighted percentage. The weighted percentage is 0.8%. We have made similar calculations based on the sample of 16,070 individuals for whom information on sector of eighth-grade and sector of 10th grade is available. The corresponding estimate of the percentage of the eighth graders from public schools who attend Catholic high schools is 0.3%. If one restricts the analysis to individuals whose parents are Catholic, only 0.7% of students who attended public eighth-grade attend a Catholic high school. The unweighted and weighted estimates of the percentage of Catholic high school 10th graders who attended Catholic eighth-grade are 95.2 percent and 84.7 percent.

and school fixed effects models for high school graduation.[16] The difference in means for the C8 sample is 0.105 when no controls are included, as shown by the marginal effect in the probit with no controls (column 5). When we add the first set of controls, the coefficient falls to 0.88 (0.25). The associated average marginal effect on the graduation rate is 0.084, which is very large given that the graduation rate is 0.947 among students from Catholic eighth grades. In the C8 sample the family background and geographic controls explain only a small fraction of the raw difference in the graduation rate. The point estimate of the marginal effect of CH declines slightly to 0.081 when we add eighth grade test scores in column 7, and increases to 0.088 when we add a large set of eighth grade measures of attendance, attitudes toward school, academic track in eighth grade, achievement, and behavioral problems. The stability of the Catholic school effect is remarkable, especially given the fact that the control variables in column 8 are quite powerful. One can see that the Psuedo R^2 of the regression model rises from 0.11 to 0.35 as we add the first set of controls, and all the way to 0.58 when we add the full set of controls. These covariates are powerful predictors of dropout behavior but lead to only a small change in the estimated Catholic schooling effect.

The second row in Table 3 is based on linear probability models of high school graduation. The coefficient on CH varies from 0.105 to 0.080 and closely agrees with the probit estimates. Row 3 of columns 5-8 adds eighth grade fixed effects to the specifications reported in row 2.[17] The fixed effects estimate is 0.143 for the basic specification and 0.102 when the full set of controls is included.[18]

In Table 3 we also report estimates of the effect of Catholic high school attendance on the probability that a student is enrolled in a 4 year college at the time of the 3rd follow-up survey in 1994, 2 years after most students graduate from high school. For the basic specification (column 6) the probit estimate implies that CH raises the college enrollment probability by 0.154, which compares to a raw difference of 0.24 (column 5). This estimate falls to 0.149 when we add detailed controls to the model. Once again the Psuedo R^2 rises substantially as we add more control variables. Linear probability models yield similar estimates.

In Table 4 we report estimates of the effect of CH on 10th and 12th grade reading and

[16] Huber-White standard errors are reported throughout the paper. The standard errors account for the use of weights and, with the exception of Table 7 and 8, they account for correlation among students from the same eighth grade.

[17] That is, it includes separate intercepts for each eighth grade.

[18] We report fixed effects results to show that factors that vary across Catholic elementary schools (such as public high school quality) do not drive the large positive estimates of the Catholic high school effect. Bias from individual heterogeneity could well be more severe in the within-school analysis.

math scores. In contrast to the above findings, we obtain modest **negative** estimates of the effects of Catholic high schools on 10th grade reading scores. In the simplest specification for the Catholic eighth grade subsample, we obtain a coefficient of -1.07 (0.97), which rises to -0.87 (0.77) when the full set of controls and eighth grade fixed effects are added. We obtain a small but statistically insignificant coefficient of -0.32 (1.01) in the case of math, but this estimate declines to essentially 0 when we add detailed controls.

In the bottom panel of Table 4 we report estimates of the effects on 12th grade reading and math scores. For the Catholic eighth grade sample with the full set of controls we obtain a small positive effect of 1.14 (0.46) on the math score and 0.33 (0.62) on the reading score. As Grogger and Neal (2000) emphasize, a positive effect of Catholic schools on the high school graduation rate might lead to a downward bias in the Catholic high school coefficient in the 12th grade test equations given that dropouts have lower test scores and that dropouts have a lower probability of taking the 12th grade test. However, the issue appears to be of only minor importance.[19]

To facilitate comparison with other studies, we also present estimates for the combined sample of students from Catholic and public eighth grades. For this sample the effect of CH on high school graduation is reduced from 0.123 to 0.052 after we add the full set of controls (Table 3, columns 1-4). It is interesting to note, however, that the OLS estimate is only 0.023 once the full set of controls are added and differs substantially from the probit estimate of the average marginal effect. This reflects problems with the linear probability specification when the outcomes is relatively rare and the explanatory variables are powerful and vary widely in the sample. The college attendance results largely mirror the high school graduation results. The probit estimate of the effect of Catholic school attendance is 0.074 once the full set of controls are included, which is substantial relative to the mean college attendance probability of 0.28.

[19]We deal with this issue by filling in missing data for both high school graduates and dropouts using predicted values from a regression of the 12th grade score on the full set of controls in the outcome regression, plus the Catholic high school dummy and the 10th grade test scores and a dummy variable for whether the individual graduated from high school (high school graduation has a small and statistically insignificant coefficient). Using the new dependent variable and sample the estimated effect of Catholic high schools for 12th grade math and reading are 1.20 and 0.58 respectively. We obtain 1.20 and 0.56, respectively, when we use an alternative imputation in which we adjust for differences in unobservables using the assumption that the difference between dropouts with and without 12th grade test scores in the mean/variance of the regression residual from the test score prediction regression is the same as the difference in the mean/variance of the predicted values of the tests. The R^2 of the prediction equations are 0.70 for reading and 0.86 for math. The estimates of the reliability of the math test reported in the NELS:88 documentation, while probably downward biased, are in the 0.87 to 0.90 range. Consequently, a substantial part of the test score residual probably reflects random variation in test performance and is unrelated to achievement levels. For this reason selection on unobservables in the availability of test data is probably less strong than selection on the predicted portion of the test scores.

Note that the choice of controls make a much larger difference in the full sample than in the Catholic eighth grade sample. We do not fully understand this pattern. However, conditioning on eighth grade variables is tricky in the full sample. The problem is that a substantial number of variables are supplied by schools and teachers and may reflect school specific standards. For example, the standards for being a "trouble maker" may differ substantially between Catholic and public eight grades. As a result, in order to draw inferences for the full sample we would want to control for type of eighth grade and interact the covariates with this variable. However, since virtually all of the Catholic high school students come from Catholic eighth grade, this essentially amounts to using the Catholic eighth grade sample to identify the Catholic school effect. It is thus hard to justify why one should be interested in the full sample.

Once detailed controls for eighth-grade outcomes are included, the estimates of the effect of Catholic high schools on 10th grade math and reading scores are essentially 0, and the estimates of the effects on 12th grade reading and math are only 1.14 and 0.92, respectively. Again, there is little evidence that Catholic high schools increase achievement by 10th grade, in accordance with the findings based on the Catholic eighth grade subsample. In contrast, the 12th grade math and reading score results indicate a small but statistically significant positive effect. Given the high degree of selection into Catholic high school in the full sample on the basis of observable traits, these estimates may reflect the effects of unobserved differences between public and Catholic high school students rather than actual effects on test scores, and should be interpreted with caution.

2.4 A Sensitivity Analysis

Based on observables, there is not that much evidence of selection in the $C8$ subsample. However, it is possible that a small amount of selection on unobservables could explain the whole Catholic school effect. We now explore this possibility by examining the sensitivity of the estimates of the Catholic high school effect to the correlation between the unobserved factors that determine CH and the various outcomes Y. We display estimates of the Catholic school effect for a range of values of the correlation between the unobserved determinants of school choice and the outcome.

Consider the bivariate probit model

$$
\begin{aligned}
(2.1) \qquad CH &= 1(X'\beta + u > 0) \\
(2.2) \qquad Y &= 1(X'\gamma + \alpha CH + \varepsilon > 0) \\
(2.3) \qquad \begin{bmatrix} u \\ \varepsilon \end{bmatrix} &\sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right).
\end{aligned}
$$

While this model is formally identified without an exclusion restriction, semiparametric identification requires such an excluded variable. Furthermore, empirical researchers are highly skeptical of results from this model in the absence of an exclusion restriction. Our thought exercise in this section is to treat this model as if it were underidentified by one parameter. In particular, we act as if ρ is not identified.[20]

In Table 7 we display estimates of Catholic schooling effects that correspond to various assumptions about ρ, the correlation between the error components in the equation for CH and Y.[21] We report results for high school graduation in the top panel and college attendance in the bottom panel, and include both probit coefficients and average marginal effects on the outcome probabilities (in brackets). We include family background, eighth grade tests, and other eighth grade measures. However, because of convergence problems in estimating the bivariate probit models we eliminated the dummy variables for household composition (but not marital status of parents), urbanicity, region, and indicators for "student rarely completes homework", "student performs below ability", "student inattentive in class", "a limited English proficiency index", and "parents contacted about behavior" from the set of controls. We vary ρ from 0 (the univariate probit case that we have already considered above) to 0.5 by estimating bivariate probit models constraining ρ to the specified value. For the full sample, the raw difference in the high school graduation probability is 0.13. When $\rho = 0$ the estimated effect is 0.058, and the figure declines to 0.037 when $\rho = 0.1$ and to 0.011 if $\rho = 0.2$. Given the strong relationship between the observables that determine high school graduation and Catholic school attendance in the full sample, the evidence for a strong Catholic school effect is considerably weaker than suggested by the estimates that take Catholic school attendance as exogenous.

For our preferred sample of Catholic 8th graders, the results are less sensitive to ρ, presumably because the pseudo R^2 is higher. The effect on high school graduation is 0.078 when $\rho = 0$,which is slightly below the estimate we obtain with the full set of controls in

[20]We use the bivariate probit because it is convenient. An alternative would be to treat ε and u nonparametrically subject to the normalization $var(\varepsilon) = var(u) = 1$ and the restriction $corr(\varepsilon, u) = \rho$.

[21]See Rosenbaum (1995) for examples of this type of sensitivity analysis.

Table 3. It declines to 0.038 when $\rho = 0.3$ and is still positive when $\rho = 0.5$. Thus, for the Catholic 8th grade sample, the correlation between the unobservable components of Catholic school attendance and high school graduation would have to be greater than 0.5 to explain the estimated effect under the null of no "true" Catholic high school effect.

In the bottom panel of Table 7 we present the results for college attendance. For the full sample, the results are very similar to the high school graduation results. The evidence for a positive effect of CH on college attendance is stronger in the Catholic 8th grade sample than in the full sample, with the effect remaining positive until ρ is about 0.3. However, in this sample the strongest evidence is for a positive effect of CH on high school graduation.

The problem with this type of analysis is that, without prior knowledge, it is hard to judge the magnitude of ρ. We will show in section 3 that the assumption that "selection on the unobservables is similar to selection on the observables" can help solve this problem.

Summarizing the results to this point, our preferred estimates, which are based on the Catholic eighth grade sample, suggest a strong positive effect of CH on high school graduation and college attendance. For this subsample, the relationship between Catholic high school attendance and other observables seems weak and the estimates are not very sensitive to the addition of a powerful set of control variables, especially in the high school graduation case. Finally, in Table 7 we show that in the $C8$ sample the degree of selection on unobservables must be quite high to explain the full Catholic high school effect. This is where the typical analysis of bias due to selection on unobservables based on patterns in the observables would end. We would conclude that part of the Catholic school effect on educational attainment is real, but could not go much beyond such a statement. In the remainder of the paper, we formalize the idea of using the degree of selection on the observables as a guide to bias from selection on unobservables and provide ways of formally incorporating such information into the sensitivity analysis. We then apply our methods to study the effect of CH. Readers who are primarily interested in the empirical results may wish to skip to Section 4.

3 Selection Bias and the Link Between the Observed and Unobserved Determinants of School Choice and Outcomes

In the empirical work to this point we have casually argued that "if the unobservables are anything like the observables," there seems to be a substantial effect of Catholic high school on schooling outcomes. As mentioned above many other papers also use the relationship between an endogenous variable or an instrumental variable and the observables to make inferences about the relationship between these variables and the unobservables. The main goal of this section is to develop a theoretical foundation for this practice and to provide a way to quantitatively assess the importance of the bias from the unobservables. In particular we show that modeling how the set of observed variables is determined can yield conditions that are useful for identification or the construction of bounds of treatment effects.

In section 3.1 we present a formal model of observables and unobservables, which we motivate with a discussion of the practical considerations that determine the content of data sets. The purpose of the model is to make clear the type of assumptions that are likely to yield Condition 1 and to aid interpretation of the parameters $\phi_{X\gamma}$ and ϕ_e in (1.2). It also justifies the particular form of Condition 1 as a way to represent the idea of equality of selection on observables and unobservables.

In section 3.2 we point out that a structural model of school choice, in which the odds of attending a Catholic school depend directly on the outcome, can also lead to Condition 1. In section 3.3 we consider the implications of Condition 1 for point identification of the treatment effect, although in our application we treat it as a bound on the amount of selection. In Section 3.4 we consider the case of continuous dependent variables and the case in which a possibly invalid instrumental variable is available. In Appendix A.8 we provide a preliminary discussion of how one may extend the methodology to allow for heterogeneous treatment effects. In section 3.5 we discuss the relevance of our analysis for studies of the Catholic school effect. We argue that the truth is somewhere between Conditions 1 and 2, with $0 \leq \phi_\varepsilon \leq \phi_{X'\gamma}$ and then show how one can use the inequality to construct bounds on α.

3.1 A Model of Observed and Unobserved Variables

Our outcome variable is some function of the latent variable Y^*. In some cases we may be interested in a binary variable such as graduating from high school (GHS) in which the outcome may be $GHS = 1(Y^* \geq 0)$. In others the continuous variable Y^* itself may be the variable of interest (such as test scores in the analysis above). Let W be the full set of variables that determine Y^* according to

$$Y^* = \alpha CH + W'\Gamma, \tag{3.1}$$

where Γ is a conformable coefficient vector. We assume that Γ is random, but is drawn once and is identical for everyone in the population. However, W and CH are random variables that vary across members of the population, so that each individual obtains an independent draw of W and CH but common values of Γ and α.

Assume that some of the elements of W are observable to the econometrician and others are not (or that the econometrician does not know that some of the observed variables belong in the model for Y^*). Following the notation above, denote the observable portion of W as X and the corresponding elements of Γ as Γ_X so that

$$Y^* = \alpha CH + X'\Gamma_X + \xi, \tag{3.2}$$

where ξ is unobserved. That is, for each potential covariate, W_j, let S_j be a dummy variable indicating whether W_j is observable. Then

$$X'\Gamma_X = \sum_{j=1}^{K} S_j W_j \Gamma_j, \quad \xi = \sum_{j=1}^{K} (1 - S_j) W_j \Gamma_j. \tag{3.3}$$

Like Γ_j, S_j does not vary across the population.

Define the projection of the latent variable CH^* onto $X'\Gamma_X$ and ξ to be

$$\mathrm{Proj}(CH^* | X'\Gamma_X, \xi) = \phi_0 + \phi_{X'\Gamma_X} X'\Gamma_X + \phi_\xi \xi. \tag{3.4}$$

We will make use of a condition similar to condition 1,

$$\phi_{X'\Gamma_X} = \phi_\xi \tag{3.5}$$

We do not know of a formal discussion of how variables are chosen for inclusion in data sets. Here we make a few general comments that apply to many social science data sets, including NELS:88. First, most large scale data sets such as NLSY, NELS:88, the PSID, and the German Socioeconomic Panel are collected to address many questions. Data set content is a compromise among the interests of multiple research, policy making, and funding constituencies. Burden on the respondents, budget, and access to administrative data sources serve as constraints. Obviously, content is also shaped by what is known about the factors that really matter for particular outcomes and by variation in the feasibility of collecting useful information on particular topics. Explanatory variables that influence a large set of important outcomes (such as family income, race, education, gender, or geographical information) are more likely to be collected. Major data sets with large samples and extensive questionnaires are designed to serve multiple purposes rather than to address one relatively specific question, such as the effectiveness of Catholic schools. As a result of the limits on the number of the factors that we know matter and that we know how to collect and can afford to collect, many elements of W are left out. This is reflected in the relatively low explanatory power of social science models of individual behavior. Furthermore, in many applications, including ours, the endogenous variable is correlated with many of the elements of X.

These considerations suggest that Condition 2, which underlies single equation methods in econometrics, will rarely hold in practice. The optimal survey design for estimation of α would be to assign the highest priority to variables that are important determinants of *both* CH^* and Y^* when choosing S. (It would also be to collect potential instrumental variables that determine CH^* but not Y^*.) However, many factors that influence Y^* and are correlated with CH^* and/or X are left out.

The other extreme is that the constraints on data collection are sufficiently severe that it is better to think of the elements of X as a more or less random subset of the elements of W rather than a set that has been systematically chosen to eliminate bias. Indeed, a natural way to formalize the idea that "selection on the observables is the same as selection on the unobservables" is to treat observables and unobservables symmetrically by assuming that the observables are a random subset of a large number of underlying variables. In our notation this amounts to assuming that S_j is an *iid* binary random variable which is equal to one with probability P_S. The outcome of S_j determines whether covariate W_j is observed. Of course, there are other ways to capture the idea of equality of selection on observables and unobservables. For example, P_S may vary across types of variables but have no systematic relationship with the values of Γ_j relative to the influence of the variables on CH^*. To the extent that the data set was designed for the study of the effect of CH^* on Y^*, one might expect $\phi_{X'\Gamma_X} > \phi_\xi$ in equation (3.4). For this and other reasons discussed in section 3.5, we focus on $\phi_{X'\gamma} > \phi_\varepsilon > 0$ as the basis for our empirical work.

We are now ready to consider the implications of random selection from W. It is easy to show that (3.5) holds on average over draws of the vector $\{S_1....S_K\}$.[22] This result in itself is not useful in practice because we only observe one draw of the sequence of S_j and Γ_j. To justify the condition we now show that as the number of covariates W gets large, the condition $\phi_{X'\Gamma_X} = \phi_\xi$ will become approximately true for a given draw of S_j and Γ_j.

[22]Define ϕ_c and ω such that

$$\begin{aligned} \text{Proj}\,(CH^* \mid W'\Gamma) &= \phi_0 + \phi_c W'\Gamma \\ \omega &= CH^* - \phi_0 - \phi_c W'\Gamma. \end{aligned} \tag{3.6}$$

That is, equality of selection on observables and unobservables will hold.

We define Y_K^*, CH_K, and CH_K^* as outcomes for a sequence of models where there are K factors that determine Y_K^*.[23] A natural part of the thought experiment in which K varies across models is the idea that the importance of each individual factor declines with K. That is,

$$Y_K^* = \alpha CH_K + \sum_{j=1}^{K} W_j^K \Gamma_j^K, \tag{3.8}$$

where W_j^K and/or Γ_j^K depend on K. To insure that Y_K^* is well behaved as K gets large, we specify that the net effect of the change in scale of W_j^K and/or Γ_j^K on the scale of $W_j^K\Gamma_j^K$ is inversely proportional to $\sqrt{K}$, which means that the above equation may be rewritten as

$$Y_K^* = \alpha CH_K + \frac{1}{\sqrt{K}} \sum_{j=1}^{K} W_j \Gamma_j. \tag{3.9}$$

We restrict $W_j\Gamma_j$ in this sequence to be stationary so that no particular covariate will be any more important ex-ante than others. This embodies the idea that a large number of factors are important in determining outcomes in social science data and that none dominate. Without loss of generality we normalize the model by assuming that $E(W_j\Gamma_j) = 0$.

We now show that under certain assumptions (3.5) will hold as the number of elements of W gets large. Note that our asymptotic analysis is nonstandard. First, we are allowing the number of underlying factors, K, to get large. Second, the random variable W_j is different in a sense than the random variables Γ_j and S_j. For each j we draw one observation on

Notice that

$$\begin{aligned} E(X'\Gamma_X\omega) &= E\left(\sum_{j=1}^{K} S_j W_j \Gamma_j \omega\right) \\ &= P_S E\left(\omega \sum_{j=1}^{K} W_j \Gamma_j\right) \\ &= P_S E(\omega W'\Gamma) = 0. \end{aligned} \tag{3.7}$$

Similar logic yields $E(\xi\omega) = 0$. Consequently, since

$$\begin{aligned} CH^* &= \phi_0 + \phi_c W'\Gamma + \omega \\ &= \phi_0 + \phi_c X'\Gamma_X + \phi_c \xi + \omega \end{aligned}$$

and since $X'\Gamma_X$ and ξ are orthogonal to ω, $\phi_{X'\gamma} = \phi_\varepsilon = \phi_c$.

[23] The "local to unity" literature in time series econometrics" (discussed in Stock, 1994) and the "weak instruments" literatures (Staiger and Stock, 1997) are other examples in econometrics in which the asymptotic approximation is taken over a sequence of models, which in the case of those literatures, depend on sample size.

Γ_j and S_j which is the same for every person in the population; however, each individual will draw his own W_j. Consider the projection of CH_K^* on the observable portion of Y_K^*, given by $\frac{1}{\sqrt{K}}\sum_{j=1}^{K} S_j W_j \Gamma_j$, and the unobservable portion, given by $\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(1-S_j)W_j\Gamma_j$. This projection is meant to be the population projection (i.e., for a very large number of persons) but with K fixed. That is, this projection conditions on a particular realization of Γ_j and S_j, $j = 1...K$. Theorem 1 states that the projection coefficients on $\frac{1}{\sqrt{K}}\sum_{j=1}^{K} S_j W_j \Gamma_j$ and $\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(1-S_j)W_j\Gamma_j$ approach each other with probability one as K gets large.

Theorem 1 *Assume that (1) W_j and Γ_j are independent nondegenerate, stationary, ergodic processes that satisfy the conditions for White's (1984) Central Limit Theorem 5.15, (2) $E(W_j\Gamma_j) = E(CH_K^*) = 0$, and (3) S_j is independent and identically distributed with $0 < Pr(S_j = 1) < 1$.*

Let

$$V_j \equiv \left\{ \underset{K\to\infty}{plim} \sqrt{K} E\left(CH_K^* W_j \mid \Gamma_1, ...\Gamma_K\right)\right\}.$$

Assume that for each j, $E(V_j) < \infty$, the sequence $\{\Gamma_j V_j\}$ satisfies the mixing conditions specified in McLeish's (1975) law of large numbers, and

$$\underset{K\to\infty}{plim} \sup_j \left|\Gamma_j \left(V_j - \sqrt{K} E\left(CH_K^* W_j \mid \Gamma_1, ...\Gamma_K\right)\right)\right| = 0.$$

Define $\phi_{X'\Gamma_X,K}$ and $\phi_{\xi,K}$ such that conditional on $S_1, ..., S_K, \Gamma_1, ..., \Gamma_K$,

$$Proj\left(CH_K^* \mid \frac{1}{\sqrt{K}}\sum_{j=1}^{K} S_j W_j \Gamma_j, \frac{1}{\sqrt{K}}\sum_{j=1}^{K}(1-S_j)W_j\Gamma_j\right)$$
$$= \phi_{X'\Gamma_X,K}\frac{1}{\sqrt{K}}\sum_{j=1}^{K} S_j W_j \Gamma_j + \phi_{\xi,K}\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(1-S_j)W_j\Gamma_j.$$

Then as K gets large, $\left(\phi_{X'\Gamma_X,K} - \phi_{\xi,K}\right)$ converges in probability to zero.

(Proof in Appendix A.1)

The assumptions of the theorem concerning V_j say that the relationship between CH_K^* and W_j is well behaved as K gets large. To motivate these assumptions suppose that CH_K^* were treated symmetrically with $(Y_K^* - \alpha CH_K)$ so that

$$CH_K^* = \frac{1}{\sqrt{K}}\sum_{j=1}^{K} W_j\beta_j + \eta_K, \tag{3.10}$$

where $E(W_j\beta_j) = 0$ and the sequence $\{W_j\beta_j\}$ is stationary. Under standard mixing assumptions about W_j and $\{\Gamma_j, \beta_j\}$ this will satisfy the conditions in the theorem.[24] Note that the theorem does not require any assumptions about the form of CH, CH^* or the relationship between them. In particular, it applies to the special cases $CH = CH^*$ both continuous, $CH = CH^*$ both binary, and CH binary/ CH^* discrete. We make use of all three of these cases below.

3.2 Correlation Between X and ξ

Mean independence of ξ and X is maintained in virtually all studies of selection problems, because without it, α is not identified even if one has a valid exclusion restriction.[25] Similarly, (3.5) is not operational unless $E(\xi|X) = 0$ because Γ_X is not identified. Our discussion of how the observables are arrived at makes clear that it is hard to justify in most settings, including ours. If the observables are correlated with one another, as in most applications, then the observed and unobserved determinants of Y^* are also likely to be correlated.

Assume that the conditional expectation is linear. Following the notation above, define γ and ε to be the slope vector and error term of the "reduced form"

$$E\left(Y^* - \alpha CH \mid X\right) \equiv X'\gamma \tag{3.11}$$

$$Y^* - E\left(Y^* - \alpha CH \mid X\right) \equiv \varepsilon. \tag{3.12}$$

We consider the case in which CH^* is linear as defined in (3.10) and consider our data generation process defined above. In appendix A.2 we provide a sufficient condition for the

[24] To see this note that

$$\begin{aligned} V_j &= \operatorname*{plim}_{K\to\infty} \sqrt{K} E\left(CH_K^* W_j \mid \Gamma_1, ...\Gamma_K\right) \\ &= \operatorname*{plim}_{K\to\infty} E\left(\sum_{j_2=1}^{K} W_j W_{j_2}\beta_{j_2} \mid \Gamma_1, ...\Gamma_K\right) \\ &= E\left(\sum_{j_2=1}^{\infty} W_j W_{j_2}\beta_{j_2} \mid \Gamma_1, \Gamma_2, ...\right). \end{aligned}$$

Conditional on a sequence of Γ_j this will be finite as long as the serial dependence between $W_j\Gamma_j$ and $W_j\beta_j$ falls fast enough. Since Γ_j is random, V_j is also random and well behaved as K gets large. It is then straightforward to provide conditions about $\{\Gamma_j, \beta_j\}$ under which V_j will satisfy the conditions for the law of large numbers.

[25] The exception is when the instrument is uncorrelated with X as well as ξ, as when the instrument is randomly assigned in an experimental setting.

coefficients of the projection of CH^* on $X'\gamma$ and ε to be equal and thus satisfy Condition 1. The sufficient conditions are the conditions of Theorem 1 and

$$\frac{\sum_{\ell=-\infty}^{\infty} E\left(W_j W_{j-\ell}\right) E\left(\beta_j \Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(W_j W_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right)} = \frac{\sum_{\ell=-\infty}^{\infty} E(\tilde{W}_j \tilde{W}_{j-\ell}) E\left(\beta_j \Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E(\tilde{W}_j \tilde{W}_{j-\ell}) E\left(\Gamma_j \Gamma_{j-\ell}\right)}, \tag{3.13}$$

where $\tilde{W}_j$ is the component of W_j that is orthogonal to X. Roughly speaking (3.13) says that the regression of CH^* on $Y^* - \alpha C$ is equal to the regression of the part of CH^* that is orthogonal to X on the corresponding part of $Y^* - \alpha C$. One can show that this condition holds under the standard assumption $E(\xi \mid X) = 0$, in which case γ and ε are identical to Γ_X and ξ, respectively. However, $E(\xi \mid X) = 0$ is not necessary for (3.13). For example, in appendix A.2 we show that (3.13) will also hold if $E\left(\beta_j \Gamma_{j-\ell}\right)$ is proportional to $E\left(\Gamma_j \Gamma_{j-\ell}\right)$ regardless of the correlations among the W_j.

In a Monte Carlo analysis not reported, we did not find large biases even when the unobservables were correlated with the observables in the original data generating process, which provides additional reassurance. In fact we often found the coefficients on $X'\gamma$ and ε to be closer to each other than the coefficients on $X'\Gamma_X$ and ξ. Exploring the consequences of the relationship between X and ξ deserves a high research priority. It is important for our case in which one is treats observables similar to unobservables, but it is also important for the more standard case in which the treatment effect or an instrumental variable are assumed to be uncorrelated with the unobservables. We believe the data generation framework we have developed will prove useful in that regard. However, given our Monte Carlo evidence and the strength of our empirical results below we do not think the link between X and ξ drives our results. Given the amount of material in this paper we leave a full analysis of this difficult issue to further research.

3.3 Structural Models of School Choice and Condition 1

A conventional path to identification of causal effects in the presence of endogeneous variables is through the use of an economic model as a source of informed restrictions. Here we digress briefly to show that this kind of approach can also deliver restrictions like Condition 1. Suppose that Catholic school attendance depends on X and ε only through Y^*. In addition, CH^* may depend on some additional unobserved variables that are unrelated to X and ε. In this case, the equation for CH^* would take the form

$$CH^* = a_1\left(Y^* - \alpha CH\right) + \varsigma, \tag{3.14}$$

where ς is uncorrelated with X and ε. Combining (3.14) with (1.1) one obtains

$$CH^* = \phi_c X'\gamma + \phi_c \varepsilon + \varsigma, \tag{3.15}$$

where $\phi_c = a_1$, and Condition 1 is satisfied. However, (3.14) is much stronger than our assumptions justifying Condition 1 because it implies CH^* is linear with coefficients $\beta_j = \phi_c \gamma_j$ for *every* j.

The above model might be a plausible approximation to the decision making process of the schools, parents, and children in situations in which schools are oversubscribed and select students to maximize outcomes such as achievement or college attendance. Many Catholic high schools give admissions tests and base decisions in part on the results, so the criterion of the high schools is partly related to 10th grade or 12th grade test performance. But particular elements of X may influence CH^* quite differently from the way in which they influence secondary school outcomes.[26] Our point is simply to establish that structural models of school choice and outcomes may also lead to Condition 1. Our model of the data generation process is sufficient for Condition 1, but it is not necessary.

3.4 Identification Based on Condition 1

We are now ready to discuss identification of α. Model (3.1) above is linear; however, in studying identification we want to isolate the contribution of Condition 1 from the role of linearity or large sample properties (e.g., normality of ε is implied by our model as the number of factors gets large). We want to rely only on Condition 1 rather than all of the implications of the model.[27] We also wish to avoid some of the complications that arise in studying nonparametric identification of discrete choice models. Consequently, we study identification of α using the familiar "treatment effect" model without exclusion restrictions:

[26] For example, the relative effects of specific variables such as religion, race, parental education, and the ability and motivation of the child on sector choice and outcomes may be different. Allowing the effects of a subset of the observed variables to enter freely into (3.14) may not be sufficient and one would require a priori information about which variables to enter. The implicit restrictions on the unobservables embodied in (3.14) also pose a problem, since whether or not a student graduates from high school or attends college will be influenced by many factors that are determined after the child decides whether to attend a Catholic high school.

[27] The use of a subset of restrictions implied by a model for identification is common in applied work. For example, as long as the probability of going to a Catholic school is nonlinear, linearity of g in (3.16) below is sufficient for identification of α and one does not need an exclusion restriction. The propensity score could be used as an instrument. We are taking a similar approach here in that we do not want identification to come from the linearity assumption, but rather from the relationship between observables and unobservables.

(3.16) $$CH = 1(CH^* \geq 0)$$

(3.17) $$Y^* = \alpha CH + g(X) + \varepsilon$$

(3.18) $$E(\varepsilon \mid X) = 0.$$

The econometrician observes (X, CH, Y), but not ε or the latent variable CH^*.[28] $g(X)$corresponds to $X'\gamma$and is defied so that $E(\varepsilon \mid X) = 0$.

It is well known that (3.16)-(3.18) is not nonparametrically identified without an exclusion restriction. We are essentially one parameter (or one equation) short of identification. This result suggests that one more restriction on this set of equations may deliver identification of α.

In the notation of (3.16)-(3.18), the analog to Condition 1 is

Condition 1-NP

$$Proj(CH^* | g(X), \varepsilon) = \phi_0 + \phi_{g(X)} g(X) + \phi_\varepsilon \varepsilon \ ; \ \phi_{g(X)} = \phi_\varepsilon \ .$$

This is equivalent to

(3.19) $$\frac{cov(CH^*, g(X))}{var(g(X))} = \frac{cov(CH^*, \varepsilon)}{var(\varepsilon)}.$$

The next theorem says that Condition 1-NP sometimes delivers point identification and always restricts the model so that the solutions α^* for α are the roots of a cubic.

Theorem 2 *In the selection model (3.16)-(3.18) let α be the true value of the treatment effect. Under Condition 1-NP, from the data we can identify a set $\mathcal{A}$ of which α is a member. The elements α^* of $\mathcal{A}$ are the roots of an identified cubic equation.*

(Proof in Appendix A.3)

The value $\alpha^* = \alpha$ is one root of the cubic. Except for pathological cases, there will be either no other real roots, or two others.[29]

In the Appendix A.3 we discuss why Condition 1 does not always yield point identification. In our application we do not obtain multiple roots. In any case, in most applications attention will focus on construction of bounds based on Theorem 4 below rather than on point identification.

[28] At this point we abstract from most of the recent literature on program evaluation by assuming that α does not vary across individuals. Allowing for heterogeneity in this parameter adds a number of additional issues even in the presence of an exclusion restriction (see e.g. Cameron and Heckman (1998), Heckman and Robb (1985), Heckman (1990), Imbens and Angrist (1994), and Manski (1989,1994)). We consider heterogeneity in α in section A.8.

[29] If all three coefficients of the cubic are 0, there are infinitely many solutions. If the cubic is tangent to 0, there can be two roots. While both of these cases are possible, they are very special.

3.5 Extensions

3.5.1 Continuous Endogenous Variables

The discussion in the previous subsection focused on a model such as Catholic schooling in which CH is binary and the restriction applies to the underlying latent variable. However, as we have already noted, the link between CH and CH^* in Theorem 1 is not restricted to $CH = 1(CH^* > 0)$. Many potential applications of the idea involve continuous endogenous variables. We maintain the model

$$Y^* = \alpha CH + g(X) + \varepsilon \tag{3.20}$$

but no longer require that CH be binary. Instead assume that

$$CH = CH^* \tag{3.21}$$

and define

$$b(X) = E(CH \mid X) \tag{3.22}$$

$$u = CH - b(X). \tag{3.23}$$

In this case we obtain a stronger identification result using Condition 1-NP and one additional assumption:

Theorem 3 *In model (3.20)-(3.23) assume Condition 1-NP and that*

$$\frac{var(u)}{var(b(X))} \neq \frac{var(\varepsilon)}{var(g(X))}.$$

Then one can identify the set $\mathcal{A}$ which consists of two values, the true α and $\alpha + \frac{var(\varepsilon)}{cov(u,\varepsilon)}$.

(Proof in Appendix A.4)

Although there are two roots, this result is useful. When an applied researcher is worried about the bias in a regression type estimator, he or she often has a strong prior about the sign of the bias, which is the sign of $cov(u, \varepsilon)$. Imposing an assumption about the sign of $cov(u, \varepsilon)$ on the data delivers point identification; if one imposes that $cov(u, \varepsilon)$ is positive (negative), then the smaller (larger) of the two elements in $\mathcal{A}$ is the true value. We stress again, however, that in most applications attention will focus on construction of bounds based on Theorem 4 below rather than point identification.

3.5.2 Using an Invalid Instrumental Variable

The results above extend to the important case in which the researcher works with an instrumental variable Z but suspects that it may be correlated with the error term in the outcome equation. For simplicity we focus on the linear case and maintain our notation

$$Y^* = \alpha CH + X'\gamma + \varepsilon,$$

where X is observable but ε is not. CH is a binary variable in our case but could also be continuous. γ is defined so that X is uncorrelated with ε, but CH is potentially endogenous and thus correlated with ε. We assume that the instrument Z does not influence Y directly, but is correlated with CH. However, Z is not necessarily a valid instrument because it might be correlated with ε. We extend the idea of using the data generation process as an aid to identification by showing that if the relationship between $X'\gamma$ and Z is similar to the relationship between ε and Z, then we can sometimes obtain identification.

Define β and π such that

$$\text{Proj}\,(Z \mid X) = X'\pi, \tag{3.24}$$

$$\text{Proj}\,(CH \mid X, Z) = X'\beta + \lambda Z, \tag{3.25}$$

and define v as the residual component of Z, so that

$$Z = X'\pi + v. \tag{3.26}$$

Consider running two stage least squares. The coefficient on the endogenous variable in this regression converges to

$$\widehat{\alpha} = \alpha + \frac{cov\,(v, \varepsilon)}{\lambda var(v)}.$$

If Z were a valid instrument, v would be uncorrelated with ε and $\widehat{\alpha}$ would equal α. The assumption $\frac{cov(v,\varepsilon)}{\lambda var(v)} = 0$ is equivalent to assuming that

Condition 2-IV

$$Proj(Z|X'\gamma, \varepsilon) = \phi_0 + \phi_{Z,X'\gamma} X'\gamma + \phi_{Z,\varepsilon}\varepsilon;\ \ \phi_{Z,\varepsilon} = 0,$$

where $\phi_{Z,X'\gamma}$ and $\phi_{Z,\varepsilon}$ are defined to be the coefficients of the projection. However, without an unusually strong form of a priori information, it is hard to argue that Z is orthogonal to the index ε of unobservable determinants of Y^* if the relationships of Z and

Y^* to a broad set of observable determinants of Y^* are similar. A large absolute value of $\phi_{Z,X'\gamma}$would imply such a similarity.

An alternative is the type of data set generation process that leads to Condition 1. One can apply Theorem 1 directly replacing CH^* with Z, which yields a condition analogous to (3.5). Using an argument similar to that in Section 3.2, one obtains a condition analogous to Condition 1.

Condition 1-IV $Proj(Z \mid X'\gamma, \varepsilon) = \phi_0 + \phi_{Z,X'\gamma}X'\gamma + \phi_{Z,\varepsilon}\varepsilon$; $\phi_{Z,X'\gamma} = \phi_{Z,\varepsilon}$.

This is equivalent to assuming that

$$\frac{cov(X'\pi, X'\gamma)}{var(X'\gamma)} = \frac{cov(v, \varepsilon)}{var(\varepsilon)}.$$

Many of the points made above about Condition 1 and 2 apply to Condition 1-IV and Condition 2-IV. In particular, it is hard to argue that $\phi_{Z,\varepsilon} = 0$ if the relationships of Z and a broad set of observable determinants of Y^* are similar and if that leads to a large absolute value of $\phi_{Z,X'\gamma}$.

Condition 1-IV restricts the solutions α^* to be the solutions of a cubic equation, one of which is α. This means that typically there are either three solutions (i.e. three values of α^* that we can not distinguish between) or there is a unique solution that equals α. The details are in Appendix A.5. In practice, one would use IV estimation subject to Condition 1-IV as part of a sensitivity analysis rather than for point identification. The analysis of bounds in Section 3.5 can be extended to the IV case.

3.6 Bounding the Catholic School Effect using Condition 3

We have data on a broad set of family background measures, teacher evaluations, test scores, grades, and behavioral outcomes in eighth grade, as well as measures of proximity to a Catholic high school. These measures have substantial explanatory power for the outcomes that we examine, and a large number of the variables play a role, particularly in the case of high school graduation and college attendance. Once we restrict the sample to Catholic eighth graders and condition on Catholic religion and distance from a Catholic high school, a broad set of variables make minor contributions to the probability of Catholic high school attendance. The relatively large number and wide variety of observables that enter into our problem suggests that the observables may provide a useful guide to the unobservables.

However, the "random selection of observables" model that leads to Condition 1 is not to be taken literally. There are in fact strong reasons to expect the relationship between the unobservables to be weaker than the relationship between the observables. First, random selection of observables is an extreme assumption. In reality, X has been selected with an eye toward reducing bias in single equation estimates rather than at random. For example, we control for race and ethnicity, which are strongly related to both Catholic school attendance and education attainment. We also include parental background measures that figure prominently in discussions of selection bias as well as for detailed eighth grade achievement and behavior measures. Second, in the case of the 12th grade test scores, ε will also reflect variability in test performance on a particular day, which presumably has nothing to do with the decision to start Catholic high school. Finally, the most important reason is that shocks that occur after eighth grade are excluded from X. These will influence high school outcomes but not the probability of starting a Catholic high school. To see this rewrite ε as $\varepsilon = \varepsilon_1 + \varepsilon_2$, where ε_1 includes factors determined prior to high school and ε_2 is the independent innovation in the error term that is determined during high school. Since CH^* is determined in eighth grade, we can impose our data generation condition on the variables determined prior to high school, in which case

$$\phi_{g(X)} = \frac{cov(CH^*, g(X))}{var(g(X))} = \frac{cov(CH^*, \varepsilon_1)}{var(\varepsilon_1)}. \tag{3.27}$$

Assume without loss of generality that $cov(CH^*, g(X)) \geq 0$ as is true in our data. Since $var(\varepsilon) > var(\varepsilon_1)$ and $cov(CH^*, \varepsilon) = cov(CH^*, \varepsilon_1)$ then

$$\phi_{\varepsilon} = \frac{cov(CH^*, \varepsilon)}{var(\varepsilon)} \leq \frac{cov(CH^*, \varepsilon_1)}{var(\varepsilon_1)} = \phi_{g(X)}.$$

Since $cov(CH^*, \varepsilon_1) \geq 0$ and $\phi_{\varepsilon} \geq 0$, Condition 1 is replaced by Condition 3. With this condition, we are able to identify bounds on α, as stated in the following theorem.

Theorem 4 *For any value α^* there is a unique g and ε consistent with the selection model (3.16)-(3.18). Define g_{α^*} and ε_{α^*} as these objects. Assuming without loss of generality that $cov(CH^*, g_{\alpha^*}(X)) > 0$, we can identify the set*

$$\mathcal{A} = \left\{ \alpha^* \in \Re : 0 \leq \frac{cov(CH^*, \varepsilon_{\alpha^*})}{var(\varepsilon_{\alpha^*}))} \leq \frac{cov(CH^*, g_{\alpha^*}(X))}{var(g_{\alpha^*}(X))} \right\}.$$

Under Condition 3 the true value α is a member of this set.

(Proof in Appendix A.6)

Treating the restriction as a bound actually simplifies the identification procedure. We simply identify the set of values of α that are consistent with (3.27). The possibility that more than one value of α^* solves (3.27) exactly plays no special role. The identification proof is constructive and suggests a manner for testing hypotheses about α (or constructing confidence intervals). Following the logic of the proof, for any potential value α_0 we can construct g_{α_0} and ε_{α_0} and then test whether the restriction holds for those values.

For our data and empirical specification we find that the upper bound on α occurs when one assumes that $\frac{cov(CH^*,\varepsilon)}{var(\varepsilon)} = 0$ and the lower bound occurs when one assumes that $\frac{cov(CH^*,g(X))}{var(g(X))} = \frac{cov(CH^*,\varepsilon)}{var(\varepsilon)}$. Thus, in the empirical work below, we interpret estimates of α that impose Condition 1 as a lower bound for α and single equation estimates with CH treated as exogenous (which impose Condition 2) as an upper bound. This simplifies the analysis substantially. If the lower bound estimates point to a substantial Catholic school effect, we interpret this as strong evidence in favor of such an effect. As it turns out, for some outcomes and samples, such as high school graduation, the single equation estimates are so large relative to the degree of selection on the observables that the lower bound estimate is still substantial. In other cases, even an amount of selection on the unobservables that is small relative to the selection on the observables is sufficient to eliminate the entire Catholic School effect.

4 Estimates of the Catholic School Effect Using Selection on the Observables to Assess Selection Bias

4.1 Using the indices of Observables in the School Choice and Outcome Equations to Bound ρ.

We now return to the bivariate probit model given by (2.1), (2.2), and (2.3) and use Theorem 4 to bound α. We argued above that Condition 1 represents an extreme assumption. The true amount of selection is in accordance with Condition 3—somewhere between independence (Condition 2) and Condition 1. In practice we have found the model to be monotonic so the highest value of the treatment effect α occurs at $\rho = 0$ while the minimum value occurs when Condition 1 is binding. Consequently, we focus on estimating the model while imposing Condition 1 and interpret the result as a lower bound on α. In the bivariate probit case, Condition 1 may be re-written[30] as

[30] Keep in mind that in the binary probit the variances of ε and u are normalized to 1.

$$\rho = \frac{Cov(X'\beta, X'\gamma)}{Var(X'\gamma)}. \tag{4.1}$$

In the top panel of Table 8, we present estimates that use the Catholic eighth grade sample directly and maximize the likelihood subject to (4.1). The estimate of ρ is 0.24. The estimate of α is 0.59 (0.33), which implies an effect of 0.07 on the probability of high school graduation. Consequently, even with the extreme assumption of equality of selection on observables and unobservables imposed, there is evidence of a large positive effect of attending Catholic high school on high school graduation.

The results for college attendance follow a similar pattern. The regression relationship between the indices of observables that determine CH and college attendance is sufficiently strong that imposing the restriction leads to a reduction in the estimated effect of Catholic schooling. The point estimate of 0.07 is substantial, although it is not statistically significant.

To improve precision of the estimates of α and as a check on the robustness of the results, we also try an alternative method that uses information contained in the public 8th grade sample. We partition X and γ into the subvectors $\{X_1, X_2, ..., X_G\}$ and $\{\gamma_1, \gamma_2, ..., \gamma_G\}$ consisting of variables and parameters that fall into similar categories. In practice, G is 6. We estimate γ on the public 8th grade sample on the grounds that very few such students go to Catholic school, and so selectivity will not influence the estimates of γ even though the mean of the error term may be different for this sample. We assume that the values of γ are the same for students from Catholic and public 8th grades, up to a proportionality factor for each subvector. Note that the univariate models reported above for the full sample implicitly assume that γ does not depend on the sector of the 8th grade. We are relaxing that assumption to some extent. The results using the second estimation method are reported in the middle panel of Table 8. In the case of high school graduation, ρ is only 0.09 and the estimate of the effect on the graduate probability is 0.09. However, the college effect is only 0.02. The restrictions on γ pass with a p-value of .12 in the high school graduation case, but fail with a p-value of .03 in the college attendance case, so perhaps the method 2 results for college attendance should be discounted. Details are in Table 8 note 4.[31]

[31] For completeness, Table 8 also presents estimates of α and ρ from an unrestricted bivariate probit on the Catholic school sample. The estimates α and ρ for high school graduation are quite close to the restricted estimates, although this is a matter of luck in view of the large standard errors. In the college attendance case we obtain a large and implausibly negative value of ρ equal to -0.52 and an implausibly large but very imprecise estimate of α equal to 1.18. As Grogger and Neal (2000) note, a finding of

4.2 The Relative Amount of Selection on Unobservables Required to Eliminate the Catholic School Effect

In this section we provide a different, more informal way to use information about selection on the observables as a guide to selection on the unobservables that permits us to use the Catholic high school indicator directly. Consider the alternative restriction,

Condition 4

$$\frac{E(\varepsilon \mid CH=1) - E(\varepsilon \mid CH=0)}{var(\varepsilon)} = \frac{E(X'\gamma \mid CH=1) - E(X'\gamma \mid CH=0)}{var(X'\gamma)}.$$

This condition implies that the relationship between Catholic high school and the location of the distribution of the index of the observables that determine outcomes and the index of unobservables is the same, after adjusting for differences in the dispersion of these distributions. We justify this condition in Appendix A.7. The discussion in Section 3.2 pertains to Condition 4 as well.

For reasons discussed earlier, the standardized difference in the mean of the unobservables that determine is Y is likely to be smaller than the standardized difference in the index of observables. One way to gauge the strength of the evidence for a Catholic school effect is to see how much of it would remain if Condition 4 were true, and to ask how large the ratio on the left would have to be relative to the ratio on the right to eliminate the entire Catholic school effect. An advantage of this approach is that we do not have to simultaneously estimate the parameters of the CH and Y equations subject to (4.1). Consequently, we are able to use the full control set used in columns 4 and 8 of Tables 3 and 4. In Altonji, Elder, and Taber (2001) we expand on this approach by showing how it can be used to evaluate an instrumental variable.

To gauge the role of selection bias in a simple way we ignore the fact that Y is estimated by a probit but rather treat α as if it were estimated by a regression of the latent variable $Y^* = X'\gamma + \alpha CH + \varepsilon$ on X and CH. Let $X'\beta$ and $\widetilde{CH}$ represent the predicted value and residuals of a regression of CH on X so that $CH = X'\beta + \widetilde{CH}$. Then,

$$Y^* = X'[\gamma + \alpha\beta] + \alpha\widetilde{CH} + \varepsilon.$$

negative selection on unobservables based on bivariate probit models is not uncommon in the Catholic schools literature and is sometimes attributed to pre-existing differences in student motivation or discipline that are poorly captured in existing data sets. We are very skeptical of this interpretation because the rich set of 8th grade student behavior measures in NELS:88 point to positive selection more or less across the board. Our view is that without exclusion restrictions or a restriction such as Condition 1, identification of α and ρ is very tenuous. We place little weight on the unrestricted estimates.

Assuming that the bias in a probit is close to the bias in OLS applied to the above model and using the fact that $\widetilde{CH}$ is orthogonal to X leads to the familiar formula

$$\begin{aligned}\text{plim } \widehat{\alpha} &\simeq \alpha + \frac{cov(\widetilde{CH}, \varepsilon)}{var\left(\widetilde{CH}\right)} \\ &= \alpha + \frac{var\left(CH\right)}{var\left(\widetilde{CH}\right)}\left[E(\varepsilon \mid CH = 1) - E(\varepsilon \mid CH = 0)\right].\end{aligned}$$

Thus, subject to Condition 4 one can estimate $E(X'\gamma \mid CH = 1) - E(X'\gamma \mid CH = 0)$ and estimate the magnitude of this bias.

We use the single equation estimates of α obtained under the assumption that Catholic schooling is exogenous in the outcome equation. A problem with using Condition 4 is that bias in α will lead to bias in the estimates of γ, which are required to evaluate the left hand side of the equation. We believe that in many applications this problem will be minor. However, as a robustness check we try three alternative ways to obtain γ. The first method is use the γ from the public eighth grade sample to form the index $X'\gamma$ for each Catholic 8th grade student. The results are reported in the first row of Table 9. In the case of high school graduation, the estimate of $(E(X'\gamma \mid CH = 1) - E(X'\gamma \mid CH = 0))\ /Var(X'\gamma)$ is 0.30. That is, the mean/variance of the probit index of X variables that determine high school graduation is 0.30 higher for those who attend Catholic high school than for those who do not. Since the variance of ε_i is 1.00, the implied estimate of $E(\varepsilon \mid CH = 1) - E(\varepsilon \mid CH = 0)$ if Condition 4 holds is 0.30 (row 1, column 3). Multiplying by $var\left(CH_i\right)/var\left(\widetilde{CH}_i\right)$ yields a bias of 0.37, while the estimate of α is 1.03. The last column of the table reports that the ratio $\widehat{\alpha}/[\frac{var(CH)}{var(\widetilde{CH})}(E(\varepsilon \mid CH = 1) - E(\varepsilon \mid CH = 0))] = (1.03/.37) = 2.78$. That is, the normalized shift in the distribution of the unobservables would have to be 2.78 times as large as the shift in the observables to explain away the entire Catholic school effect. This seems highly unlikely.

The second row of Table 9 reports the results when the left hand side of Condition 4 is evaluating using the estimate of γ obtained from the single equation probit estimate of the high school graduation equation on the Catholic school sample. The third row uses the estimate of γ when α is constrained to be 0. For these methods, the implied ratios are 4.29 and 3.55 respectively. The results in Table 9 suggest that a substantial part of the effect of CH on high school graduation is real.

For college attendance the ratios range between 1.30 and 2.03 depending on how we estimate γ (rows 4, 5, and 6). Since the ratio of selection on unobservables relative to selection on observables is likely to be less than 1, part of the Catholic school effect on college graduation is probably real.

Table 10 presents 10th and 12th grade test score results using the same methodology described above. The coefficient on CH has a positive and statistically significant coefficient only in the case of 12th grade math scores. However, this effect is small (1.14) and would be almost completely eliminated assuming the upper bound Condition 4 holds. Even if selection on unobservables is only one half as strong as that on observables, the effect of Catholic schooling would be negligible. Given the weak evidence from the univariate models and the likelihood of some positive bias, we conclude that Catholic high school probably has little effect on test scores.

5 Results by minority status and urbanicity

A number of studies, including Evans and Schwab (1995), Neal (1997), and Grogger and Neal (2000) using NELS:88 have found much stronger effects of Catholic schooling for minority students in urban areas than for other students. Table 2 reports differences in the means of outcomes and control variables, by high school type, for all urban minority students and for urban minority students who attended Catholic eighth grades. Note that 54 of the 56 minority students who attended Catholic high school came from Catholic eighth grades. Only 15 of the 700 urban minority students in public 10th grades came from Catholic 8th grades, which is too few observations to support an analysis on the Catholic eighth grade subsample. In the full urban minority sample the control variables provide evidence of strong positive selection into Catholic high schools. The gaps in mother's education and father's education are 0.66 years and 1.69 years, respectively. The gap in the log of family income is 0.83. There are also very large discrepancies in the base year measures of parental expectations for schooling and student expectations for schooling and white-collar work, large gaps in the eighth-grade behavioral measures, and gaps of 6.49 and 3.28 in the eighth grade reading and math tests, respectively. Since there is more selection on observable variables for this subsample, one might expect more selection on unobservables as well. This could explain the large measured Catholic schooling effects.

In Table 5 we report models of the high school graduation probability estimated using the urban sample of white students as well as the urban sample of minorities. All of the regression models include our full set of controls. For the minority sample, the probit estimate implies that the average marginal effect of CH on high school graduation is 0.191.

The linear probability estimate is 0.133 (0.056).[32] One important caveat in interpreting these results is that of the 110 urban minority students who attend Catholic high school, only one subsequently drops out. There clearly appears to be a strong Catholic high school effect on graduation, but one should be wary of small sample bias in calculating the asymptotic standard errors. Turning to the bottom panel of Table 5, we find a substantial effect of Catholic high school on college attendance, with estimates for the urban minority sample varying from 0.144 to 0.182 depending on the estimation methods. Consistent with previous work, the effects are generally larger for minorities than for the samples of whites. However, since there is more selection on observable variables for this subsample it seems quite plausible that there could be more selection on unobservables as well and that this could explain the large measured Catholic schooling effects.

Table 6 presents test score results for the urban minority sample. As shown in the second column of the table, we obtain negative but small and statistically insignificant estimates of the effect of Catholic schooling on both the math and reading 10th grade tests, which agrees with the analysis based on both the full NELS:88 sample and the Catholic eighth grade subsample. We obtain a coefficient of -0.19 (1.39) for the 12th grade reading score as well, and a coefficient of 1.25 (1.09) for the 12th grade math score. Evidently, most or all of the substantial Catholic high school advantage for urban minorities in test scores disappears once we control for family background and 8th grade outcomes. This result reflects the large gap in the means of the controls in favor of minorities attending Catholic high school. As one can see in the table, we obtain similar results when we add suburbanites and extend our analysis to a pooled urban/suburban minority subsample.

We also perform a sensitivity analysis based on (2.1)-(2.3) for the urban minority sample. Turning again to Table 7, note that the raw differential in the high school graduation probability is 0.22 and the estimate of the Catholic school effect under the assumption $\rho = 0$ is 0.176. The estimate is 0.132 when $\rho = 0.2$, and 0.013 when $\rho = 0.5$. Thus, the correlation between the unobservables would have to be in the neighborhood of 0.5, a very large correlation, for one to conclude that the true effect of Catholic schools on the graduation rates of urban minorities is 0. This value seems unreasonable.

We also estimated the restricted bivariate probit model as in Table 8 for urban minorities. We experienced computational difficulties in estimating the model for high school

[32] The estimate including eighth grade school fixed effects is essentially zero, which leaves open the possibility that cross-school variation in the opportunities available to urban minority students may be responsible for the positive estimated Catholic high school effects. However, the standard error of the fixed effects estimate is quite large (.107), so one should not make too much of this result.

graduation that we suspect are related to the fact that only 1 Catholic school attendee failed to graduate. For college attendance, we obtained an estimate of ρ of 0.5 and a negative but insignificant estimate of α. Due to the computational problems, we focus on an analysis involving the differences in indices of observable variables based on Condition 4. In Table 11 rows 2 and 4 we form the measure of selection on observables using the estimates of γ from the urban minority public 8th grade sample. For this sample under Condition 4 the implied shift in $(E(\varepsilon \mid CH = 1) - E(\varepsilon \mid CH = 0))$ is 0.56 in the case of high school graduation and 0.72 in the case of college attendance, which reflects strong selection on the observables that influence these outcomes. Still, selection on the unobservables would have to be 2.37 times as strong as selection on the observables to explain away the entire high school graduation effect. This seems very unlikely to us; the evidence suggests that for urban minorities a substantial part of the estimated effect of Catholic schooling on graduation is real. On the other hand, we cannot rule out the possibility that much of the effect of CH on college attendance is due to selection bias.

In Table 12 we report the results of an analysis of test scores. As we have already noted, there is little evidence that Catholic high school improves the reading scores of minorities. The table shows that in the case of 12th grade reading scores $(E(X'\gamma \mid CH = 1) - E(X'\gamma \mid CH = 0))\ /Var(X'\gamma)$ is 0.090. Under Condition 4 this amount of favorable selection on the observables implies an estimate of $(E(\varepsilon \mid CH = 1) - E(\varepsilon \mid CH_i = 0))$ equal to 2.76. Since the point estimate of α is already negative, there is certainly no evidence that Catholic schools boost 12th grade reading scores.

In the case of 12th grade math, the point estimate of α is 1.82 and the implied estimate of $(E(\varepsilon \mid CH = 1) - E(\varepsilon \mid CH = 0))$ under Condition 4 is 1.17, and the implied ratio of selection on unobservables to selection on observables required to explain away the entire estimate of α is 0.89, which seems large given that a substantial part of the unexplained variance is due to unreliability in the tests. (See note 16.) Consequently, we would not rule out a small positive effect on math but there is little evidence that Catholic high schools substantially boost the test scores of urban minorities.[33]

[33] These test score findings are robust to the imputation procedures for dropouts described in Section 2.3. In contrast, Grogger and Neal (2000) find some evidence for a Catholic school effect on minority test scores using median regression, particularly when they restore high school dropouts with missing test score data to the sample by simply assigning them 0. We have not fully investigated the source of the discrepancy, but suspect that our use of a more extensive set of control variables, our imputation process, differences in the samples used, and differences between mean and median regression all play a role.

6 Conclusion

Our analysis of the Catholic school effect is guided by three premises. The first is that the exclusion restrictions used in previous studies do not provide a reliable means of identifying the Catholic school effect. The second premise is that in the absence of a bulletproof instrument, it is important to start with a rich set of control variables and with treatment and control groups who look similar in eighth grade. This leads us to focus on students from Catholic eighth grades. Conditioning on Catholic eighth graders allows us to avoid concerns about lack of comparability between the tiny fraction of students from public primary schools who attend Catholic high school and other students. It also allows us to isolate the effect of Catholic high school from the effect of Catholic primary school.

The third premise is that the degree of selection on the observables is informative about selection on unobserved characteristics. As we noted in the introduction, it is standard procedure to consider the relationship between an explanatory variable or an instrumental variable and the observed variables in the model in discussions of exogeneity. The methodological contribution of this paper is to formalize the use of such information and to provide a way to assess the degree of selection bias. We make the theoretical point that knowledge of how the observable variables are chosen from the full set of variables can be sufficient to identify the effect of an endogenous variable. We illustrate this by establishing identification in the case in which selection on observables and unobservables is the same in the sense that unit shifts in the indices of observables and unobservables that determine the outcome have the same effect on school choice. In the Catholic school case, selection on the observables is likely to be stronger than selection on the unobservables. Consequently, the estimates of our model subject to the restriction imposed by equal selection provides a lower bound estimate of the effect of Catholic schools while the single equation estimates provide an upper bound. We also propose an informal way to assess selectivity bias based on a measure of the ratio of selection on unobservables relative to selection on observables that would be required if one is to attribute the *entire* Catholic school effect to selection bias.

We have three main substantive findings regarding Catholic schools. First, attending Catholic high school substantially raises high school graduation rates. In the Catholic eighth grade sample, only 0.02 of the 0.105 Catholic high school advantage in graduation rates is explained by eighth grade outcomes or family background. We obtain a lower bound estimate of 0.07 when we impose equality of selection of observables and unobservables an upper bound estimate of 0.08 when we assume that there is no selection on unobservables.

While estimates that treat Catholic school attendance as exogenous almost certainly overstate the effect of Catholic high schools, the degree of selection on the unobservables would have to be much stronger than the degree of selection on the observables to explain away the entire effect. We also find that the upper bound effect of Catholic school on the probability of college attendance is very large (0.15) when Catholic school attendance is treated as exogenous, but the lower bound estimates ranges between 0.07 and 0.02 depending on estimation details. We conclude that part of the effect of CH on college attendance is probably real, but the evidence is less clear cut than in the high school graduation case.

Second, we find little evidence that Catholic high schools raise reading scores. In fact, some of our point estimates are negative. The single equation estimates point to a positive effect of about 0.1 standard deviations on the 12th grade math score. However, given sampling error and evidence of positive selection bias, we do not have much evidence that Catholic high schools boost test scores.

Third, our results for urban minorities suggest that Catholic high school attendance substantially raises the probability of high school graduation for this group. Single equation estimates of the impact on college attendance are also very large, but the degree of positive selection on the observables that determine college attendance is sufficiently large that one could not rule out selection bias as the full explanation for the Catholic school effect on college attendance. In the full urban minority sample, differences by high school sector in family background characteristics and eighth grade performance are very large. The assumption that the selection on the unobservables mirrors selection on the observables results in a larger selectivity bias correction for this group. Although in common with other recent studies we obtain larger single equation estimates of the Catholic school effect for urban minorities than other groups, these differences may be due to differences in the degree of selection bias.

A natural followup to our study would be an examination of the mechanism through which Catholic schools affect high school graduation. Such a study would draw on the literature on Catholic schools and the NELS:88 data on school characteristics and student behavior during the high school years. Multivariate analysis of the effect of differences in background and eighth grade social behavior suggests that such differences are more important for graduation than for test scores (not reported). Many of the traits of Catholic schools stressed by Bryk et al (1993) and Coleman and Hoffer (1987) may work to reduce the dropout probability among low achieving students or students with behavioral problems. The more structured and communitarian environment normally found in Catholic high schools may be effective in reducing dropout rates and increasing college attendance.

There is a long agenda for future research on the econometric methods that we propose. With regard to the theoretical foundations, high priorities include additional analysis of identification in both single equation and instrumental variables settings and a full analysis of heterogeneous effects case introduced in Appendix A.8. In our application, the measure of the relative degree of selection on observables and unobservables is not very sensitive to how we compute γ, the parameters of the outcome equation, and we were able to use the public 8th grade sample as a benchmark for γ in any case. However, a theoretical analysis of conditions under which bias in the estimates of γ is important would be helpful.

With regard to the art of assessing when and how to use the methods that we describe, a monte carlo analysis of how the methods perform in the context of real world examples would be informative, particularly in those cases in which concern about identification is a first order issue. Examples of hard to research questions that strike us as ripe for application of our methods include the effect of drugs and alcohol on future socioeconomic outcomes, the effect of criminal activity on future labor market success and the effects of peer characteristics on school outcomes. One could also carry out a monte carlo analysis in which one samples at random from the hundreds of 8th grade family background and student characteristics available in NELS:88, although this would be taking too literally the idea of random inclusion of variables. This idea could be formalized and used to bootstrap the estimates. More generally, we have used the model in Section 3.1 to justify the use of Condition 1 as a piece of identifying information. An alternative approach that should be developed is to use additional implications of the data generation model to perform inference. An important component of this would be to allow for more general patterns of dependence between the observables and unobservables. If the covariates are chosen at random and one can observe the covariance pattern among the observable covariates, one should be able to use this information to infer the dependence between the observable and unobservable covariates.

In closing, we caution against the potential for misuse of the idea of using observables to draw inferences about selection bias. The conditions required for Theorem 1 imply that it is dangerous to infer too much about selection on the unobservables from selection on the observables if the observables are small in number and explanatory power or if they are unlikely to be representative of the full range of factors that determine an outcome.

Appendix A

A.1 Proof of Theorem 1

Proof. We simplify the notation by defining

$$E^K(\cdot) \equiv E(\cdot \mid S_1, ..., S_K, \Gamma_1, ...\Gamma_K)$$

Define

$$
\begin{aligned}
\text{(A-1)} \qquad \phi &\equiv \frac{\operatorname*{plim}_{K\to\infty} E^K\left(CH_K^*\left(Y_K - \alpha CH_K\right)\right)}{\operatorname*{plim}_{K\to\infty} E^K\left(\left(Y_K - \alpha CH_K\right)^2\right)} \\
&= \frac{\operatorname*{plim}_{K\to\infty} \frac{1}{K}\sum_{j=1}^K \Gamma_j \sqrt{K} E^K\left(CH_K^* W_j\right)}{\operatorname*{plim}_{K\to\infty} \frac{1}{K}\sum_{j_1=1}^K \sum_{j_2=1}^K \Gamma_{j_1}\Gamma_{j_2} E^K\left(W_{j_1} W_{j_2}\right)} \\
&= \frac{\operatorname*{plim}_{K\to\infty} \frac{1}{K}\sum_{j=1}^K \Gamma_j V_j + \operatorname*{plim}_{K\to\infty} \frac{1}{K}\sum_{j=1}^K \Gamma_j \left(\sqrt{K} E^K\left(CH_K^* W_j\right) - V_j\right)}{\sum_{\ell=-\infty}^{\infty} E(\Gamma_j \Gamma_{j-\ell} W_j W_{j-\ell})} \\
&= \frac{\operatorname*{plim}_{K\to\infty} \left\{\frac{1}{K}\sum_{j=1}^K E\left(\Gamma_j V_j\right)\right\}}{\sum_{\ell=-\infty}^{\infty} E(\Gamma_j \Gamma_{j-\ell} W_j W_{j-\ell})}.
\end{aligned}
$$

The term $\frac{1}{K}\sum_{j=1}^K \Gamma_j \left(\sqrt{K} E^K\left(CH_K^* W_j\right) - V_j\right)$ goes to zero as a result of our assumption about $\operatorname*{plim}_{K\to\infty} \sup_j \left|\Gamma_j \left(V_j - \sqrt{K} E\left(CH_K^* W_j \mid \Gamma_1, ..., \Gamma_K\right)\right)\right|$. We apply the central limit theorem to $W_j \Gamma_j$ in deriving the denominator and apply the law of large numbers for the numerator. Under the assumptions of the theorem both the numerator and denominator are finite.

To simplify the exposition define

$$\Psi_K = \begin{bmatrix} \frac{1}{\sqrt{K}}\sum_{j=1}^K S_j W_j \Gamma_j \\ \frac{1}{\sqrt{K}}\sum_{j=1}^K \left(1 - S_j\right) W_j \Gamma_j \end{bmatrix}$$

By definition of the projection of interest

$$
\begin{aligned}
\begin{bmatrix} \phi_{X'\Gamma_X,K} \\ \phi_{\xi,K} \end{bmatrix} &= \left[E^K\left(\Psi_K \Psi_K'\right)\right]^{-1} E^K\left(\Psi_K CH_K^*\right) \\
&= \begin{bmatrix} \phi \\ \phi \end{bmatrix} + \left[E^K\left(\Psi_K \Psi_K'\right)\right]^{-1} E^K\left(\Psi_K\left(CH_K^* - \Psi_K' \begin{bmatrix} \phi \\ \phi \end{bmatrix}\right)\right).
\end{aligned}
$$

From the conditions in the theorem $E^K\left(\Psi_K \Psi_K'\right)$ is finite and positive definite.

To see that $E^K\left(\Psi_K\left(CH_K^* - \Psi_K'\begin{bmatrix}\phi\\ \phi\end{bmatrix}\right)\right)$converges to zero note that

$$
\begin{aligned}
&\operatorname*{plim}_{K\to\infty} E^K\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} S_j W_j \Gamma_j\left(CH_K^* - \Psi_K'\begin{bmatrix}\phi\\ \phi\end{bmatrix}\right)\right)\\
&=\operatorname*{plim}_{K\to\infty}\frac{1}{\sqrt{K}}\sum_{j=1}^{K} S_j\Gamma_j E^K\left(W_j CH_K^*\right)\\
&\quad-\operatorname*{plim}_{K\to\infty} E^K\left(\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} S_j W_j\Gamma_j\right)\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} W_j\Gamma_j\phi\right)\right)\\
&=\operatorname*{plim}_{K\to\infty}\frac{1}{K}\sum_{j=1}^{K} E\left(S_j\Gamma_j V_j\right)\\
&\quad-\operatorname*{plim}_{K\to\infty}\frac{1}{K}\sum_{j_1=1}^{K}\sum_{j_2=1}^{K} S_{j_1}\Gamma_{j_1}\Gamma_{j_2}E^K\left(W_{j_1}W_{j_2}\right)\phi\\
&=E\left(S_j\right)\operatorname*{plim}_{K\to\infty}\left\{\frac{1}{K}\sum_{j=1}^{K}E\left(\Gamma_j V_j\right)\right\} - E(S_j)\left(\sum_{\ell=-\infty}^{\infty}E(\Gamma_j\Gamma_{j-\ell}W_jW_{j-\ell})\right)\phi\\
&=0
\end{aligned}
$$

where the final equality follows from$(A-1)$. By virtually the same argument

$$
\operatorname*{plim}_{K\to\infty} E^K\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left(1-S_j\right)W_j\Gamma_j\left(CH_K^* - \Psi_K'\begin{bmatrix}\phi\\ \phi\end{bmatrix}\right)\right) = 0.
$$

Thus

$$
\begin{aligned}
\operatorname*{plim}_{K\to\infty}\left\{\phi_{X'\Gamma_X,K}\right\} &= \operatorname*{plim}_{K\to\infty}\left\{\phi_{\xi,K}\right\}\\
&= \phi
\end{aligned}
$$

■

A.2 Deriving Condition 1

As above treat the model as

$$
\begin{aligned}
CH_K^* &= \frac{1}{\sqrt{K}}\sum_{j=1}^{K} W_j\beta_j\\
Y_K^* &= \frac{1}{\sqrt{K}}\sum_{j=1}^{K} W_j\Gamma_j
\end{aligned}
$$

where we have incorporated αCH_K into Y_K^* to simplify the notation. In this section we use notation that differs somewhat from the text. Throughout this section we use "hats" to define the predicted value from a linear projection of a variable onto the observable covariates in W and "tildes" to denote the residual from that regression. For example $Y_K^* = \hat{Y}_K^* + \tilde{Y}_K^*$ where $\hat{Y}_K^*$ is the linear prediction from a regression of Y_K^* on the observables. In the notation of the text

$$\begin{aligned}\hat{Y}_K^* &= X'\gamma \\ \tilde{Y}_K^* &= \varepsilon\end{aligned}$$

Furthermore we simplify the notation by dropping the K subscript when we mean the probability limit of the variable so $Y^* \equiv \text{plim}\{Y_K^*\}$.

Since $cov(\hat{Y}_K^*, \tilde{Y}_K^*) = 0$, in this notation Condition 1 can be written as

$$\text{(A-2)} \qquad \frac{cov(CH^*, \hat{Y}^*)}{var(\hat{Y}^*)} = \frac{cov(CH^*, \tilde{Y}^*)}{var(\tilde{Y}^*)}.$$

It is straightforward to verify that (A-2) is equivalent to

$$\text{(A-3)} \qquad \frac{cov(CH^*, Y^*)}{var(Y^*)} = \frac{cov(\widetilde{CH}^*, \tilde{Y}^*)}{var(\tilde{Y}^*)}.$$

When does our data generation process yield (A-2)? Since CH_K^* and Y_K^* are linear we can also write

$$\text{(A-4)} \qquad \begin{aligned}\widetilde{CH}_K^* &= \frac{1}{\sqrt{K}} \sum_{j=1}^{K} \widetilde{W}_j \beta_j \\ \tilde{Y}_K^* &= \frac{1}{\sqrt{K}} \sum_{j=1}^{K} \widetilde{W}_j \Gamma_j.\end{aligned}$$

Under the assumptions in Theorem 1 as the number of regressors gets large for any j,

$$\text{(A-5)} \qquad \begin{aligned}\frac{cov(CH^*, Y^*)}{var(Y^*)} &\approx \frac{\sum_{\ell=-\infty}^{\infty} E\left(W_j \beta_j W_{j-\ell} \Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(W_j \Gamma_j W_{j-\ell} \Gamma_{j-\ell}\right)} \\ &= \frac{\sum_{\ell=-\infty}^{\infty} E\left(W_j W_{j-\ell}\right) E\left(\beta_j \Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(W_j W_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right)},\end{aligned}$$

where the expectation is over both $(W_j W_{j-\ell})$ and $\left(\beta_j \Gamma_{j-\ell}\right)$.

Similarly,

$$\text{(A-6)} \qquad \frac{cov(\widetilde{CH}^*, \tilde{Y}^*)}{var(\tilde{Y}^*)} \approx \frac{\sum_{\ell=-\infty}^{\infty} E(\tilde{W}_j \tilde{W}_{j-\ell}) E\left(\beta_{jj-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\tilde{W}_j \tilde{W}_{j-\ell}\right) E\left(\Gamma_j \Gamma_{j-\ell}\right)}.$$

Comparison of (A-5) and (A-6) establishes the claim in the text that the conditions in Theorem 1 combined with (3.13) are sufficient for (A-2).

In general the autocovariance structure of W_j will be different from the autocovariance structure of $\tilde{W}_j$, and the restriction (A-2) will not be valid. We now provide an example (other than the standard case in which X is uncorrelated with the excluded variables) for which (3.13) and (A-2) hold.

Assume that for some constant τ

$$E\left(\beta_j \Gamma_{j-\ell}\right) = \tau E\left(\Gamma_j \Gamma_{j-\ell}\right).$$

In this case

$$\frac{cov(\widetilde{CH}^*, \tilde{Y}^*)}{var(\tilde{Y}^*)} = \frac{\sum_{\ell=-\infty}^{\infty} E\left(\tilde{W}_j \tilde{W}_{j-\ell}\right) \tau E\left(\Gamma_j \Gamma_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E(\tilde{W}_j \tilde{W}_{j-\ell}) E\left(\Gamma_j \Gamma_{j-\ell}\right)}$$
$$= \tau.$$

Such a case can occur when Γ_j and β_j have the same stationary ARMA process. To see this consider the MA(∞) process

$$\beta_j = \omega_j^1 + \theta_1 \omega_{j-1}^1 + \theta_2 \omega_{j-2}^1 + ...$$
$$\Gamma_j = \omega_j^2 + \theta_1 \omega_{j-1}^2 + \theta_2 \omega_{j-2}^2 + ...$$

where the joint distribution of (ω_j^1, ω_j^2) is serially uncorrelated with constant variance and $cov(\omega_j^1, \omega_k^2) = 0$ when $k \neq j$. Then (defining $\theta_0 = 1$)

$$E\left(\beta_j \Gamma_{j-\ell}\right) = cov(\omega_j^1, \omega_j^2) \sum_{r=0}^{\infty} \theta_r \theta_{r+\ell}$$
$$E\left(\Gamma_j \Gamma_{j-\ell}\right) = var(\omega_j^2) \sum_{r=0}^{\infty} \theta_r \theta_{r+\ell}$$

so

$$E\left(\beta_j \Gamma_{j-\ell}\right) = \frac{cov(\omega_j^1, \omega_j^2)}{var(\omega_j^2)} E\left(\Gamma_j \Gamma_{j-\ell}\right).$$

A.3 Proof of Theorem 2

Proof. Consider any value $\alpha^* \neq \alpha$ that is consistent with the model (3.16)-(3.18) and Condition 1-NP. Define $(g^*(X), \varepsilon^*)$ as the analogues of $(g(X), \varepsilon)$ that accompany α^*. Then

$$E(Y \mid CH = 0, X) = g^*(X) + E(\varepsilon^* \mid CH^* \leq 0, X)$$
$$= g(X) + E(\varepsilon \mid CH^* \leq 0, X),$$

$$E(Y \mid CH = 1, X) = \alpha^* + g^*(X) + E(\varepsilon^* \mid CH^* > 0, X)$$
$$= \alpha + g(X) + E(\varepsilon \mid CH^* > 0, X),$$

and

$$E\left(\varepsilon^* \mid X\right) = 0.$$

Solving these equations for g^* yields

$$g^*(X) = g(X) + p(X)\left(\alpha - \alpha^*\right),$$

where $p(X)$ is the propensity score (i.e. $p(X) = \Pr(CH = 1 \mid X)$ and thus

$$\varepsilon^* = \left(\alpha - \alpha^*\right)\left(CH - p(X)\right) + \varepsilon.$$

If the alternative model satisfies (3.19) then

$$\frac{cov(CH^*, g^*(X))}{var(g^*(X))} = \frac{cov(CH^*, \varepsilon^*)}{var(\varepsilon^*)},$$

or

$$
\begin{aligned}
&\frac{cov(CH^*, g(X)) + (\alpha - \alpha^*)\, cov(CH^*, p(X))}{var(g(X)) + 2\,(\alpha - \alpha^*)\, cov(g(X), p(X)) + (\alpha - \alpha^*)^2\, var(p(X))} \\
= \;& \frac{cov(CH^*, \varepsilon) + (\alpha - \alpha^*)\, cov\,(CH^*, CH - p(X))}{var(\varepsilon) + 2\,(\alpha - \alpha^*)\, cov(\varepsilon, CH - p(X)) + (\alpha - \alpha^*)^2\, var(CH - p(X))}.
\end{aligned}
$$

Defining

$$\phi \equiv \frac{cov(CH^*, g(X))}{var(g(X))} = \frac{cov(CH^*, \varepsilon)}{var(\varepsilon)},$$

and dividing top and bottom by $var(g(X))$ and $var(\varepsilon)$, we get

$$
\begin{aligned}
&\frac{\phi + (\alpha - \alpha^*)\, \frac{cov(CH^*, p(X))}{var(g(X))}}{1 + 2\,(\alpha - \alpha^*)\, \frac{cov(g(X), p(X))}{var(g(X))} + (\alpha - \alpha^*)^2\, \frac{var(p(X))}{var(g(X))}} \\
= \;& \frac{\phi + (\alpha - \alpha^*)\, \frac{cov(CH^*, CH - p(X))}{var(\varepsilon)}}{1 + 2\,(\alpha - \alpha^*)\, \frac{cov(\varepsilon, CH - p(X))}{var(\varepsilon)} + (\alpha - \alpha^*)^2\, \frac{var(CH - p(X))}{var(\varepsilon)}}.
\end{aligned}
$$

Algebraic manipulation yields

$$
\begin{aligned}
0 = (\alpha - \alpha^*)^3 & \left[\frac{var(CH - p(X))}{var(\varepsilon)} \frac{cov(CH^*, p(X))}{var(g(X))} - \frac{var(p(X))}{var(g(X))} \frac{cov\,(CH^*, CH - p(X))}{var(\varepsilon)} \right] \\
+ (\alpha - \alpha^*)^2 & \left[\phi \frac{var(CH - p(X))}{var(\varepsilon)} + 2 \frac{cov(CH^*, CH - p(X))}{var(\varepsilon)} \frac{cov(CH^*, p(X))}{var(g(X))} \right. \\
& \left. - \phi \frac{var(p(X))}{var(g(X))} - 2 \frac{cov(g(X), p(X))}{var(g(X))} \frac{cov\,(CH^*, CH - p(X))}{var(\varepsilon)} \right] \\
+ (\alpha - \alpha^*) & \left[\frac{cov(CH^*, p(X))}{var(g(X))} + 2\phi \frac{cov(\varepsilon, CH - p(X))}{var(\varepsilon)} \right. \\
& \left. - \frac{cov\,(CH^*, CH - p(X))}{var(\varepsilon)} - 2\phi \frac{cov(g(X), p(X))}{var(g(X))} \right].
\end{aligned}
$$

Thus the only values of α^* that are consistent with the observed data and Condition 1-NP of the model are members of the set $\mathcal{A}$, so the set is identified. Furthermore note that $\alpha^* = \alpha$ is a member of the set.

■

To understand why Condition 1-NP does not always yield point identification, note that in (3.19) the denominators, $var(g(X))$ and $var(\varepsilon)$, are not identified without knowledge of α. In particular, define $(\alpha^*, g^*, \varepsilon^*)$ to be an alternative possibility for (α, g, ε) that satisfies (3.16)-(3.19). For this possibility (3.19) can be rewritten as

$$cov(CH^*, g^*(X)) var(\varepsilon^*) = cov(u, \varepsilon^*) var(g^*(X)). \tag{A-7}$$

In that case one can show that

$$
\begin{aligned}
var(g^*(X)) &= var\,(g(X) + (\alpha - \alpha^*)\, p(X)) \\
&= var(g(X)) + 2\,(\alpha - \alpha^*)\, cov(g(X), p(X)) + (\alpha - \alpha^*)^2\, var(p(X)).
\end{aligned} \tag{A-8}
$$

The right hand side of (3.19) is the product of $var(g^*(X))$, which is quadratic in $(\alpha - \alpha^*)$, and $cov(u, \varepsilon^*)$, which is linear in $(\alpha - \alpha^*)$. This yields a cubic in $(\alpha - \alpha^*)$.

It is not clear how much we should worry about this potential problem even in the case that one is using the condition for point identification rather than a bound. Consider equation (A-8). We suspect that in typical applications, the contribution of $(\alpha^* - \alpha)p(X)$ to the variance of $g^*(X)$ will be small relative to $var(g(X))$ when α^* remains within a reasonable range. In this case the other two roots are not worrisome since they involve changes in $var(g^*(X))$ outside the range of plausibility. In our empirical work we have found that $var(g^*(X))$ is insensitive to reasonable values of α^*, but the question of whether this is true in most applications can only be answered through empirical implementation.

A.4 Proof of Theorem 3

Proof. Follow similar logic to Theorem 2. Consider any value $\alpha^* \neq \alpha$ that is consistent with the model and Condition 1-NP. Then define $(g^*(X), \varepsilon^*)$ as the analogues of $(g(X), \varepsilon)$ that accompany it. Then

$$Y = \alpha^* CH + g^*(X) + \varepsilon^*.$$

Since ε^* must be mean zero conditional on X,

$$g^*(X) = g(X) + (\alpha - \alpha^*)\, b(X),$$

and

$$\varepsilon^* = \varepsilon + (\alpha - \alpha^*)\, u\ ,$$

where $b(X)$ and u are defined in the text.

If the alternative model satisfies the conditions in the theorem then

$$\frac{cov(CH^*, g^*(X))}{var(g^*(X))} = \frac{cov(CH^*, \varepsilon^*)}{var(\varepsilon^*)}.$$

Substituting in for g^* and ε^* leads to

$$\text{(A-9)} \qquad \frac{cov(b(x), g(X)) + (\alpha - \alpha^*)\, var(b(X))}{var(g(X)) + 2\,(\alpha - \alpha^*)\, cov(g(X), b(X)) + (\alpha - \alpha^*)^2\, var(b(X))}$$
$$= \frac{cov(u, \varepsilon) + (\alpha - \alpha^*)\, var\,(u)}{var(\varepsilon) + 2\,(\alpha - \alpha^*)\, cov(\varepsilon, u) + (\alpha - \alpha^*)^2\, var(u)}.$$

As above, defining

$$\phi \equiv \frac{cov(b(X), g(X))}{var(g(X))} = \frac{cov(u, \varepsilon)}{var(\varepsilon)}$$

and dividing top and bottom of the left hand side of (A-9) by $var(g(X))$ and the right hand side by $var(\varepsilon)$ (respectively), one finds that

$$\frac{\phi + (\alpha - \alpha^*)\, \frac{var(b(X))}{var(g(X))}}{1 + 2\,(\alpha - \alpha^*)\, \phi + (\alpha - \alpha^*)^2\, \frac{var(b(X))}{var(g(X))}} = \frac{\phi + (\alpha - \alpha^*)\, \frac{var(u)}{var(\varepsilon)}}{1 + 2\,(\alpha - \alpha^*)\, \phi + (\alpha - \alpha^*)^2\, \frac{var(u)}{var(\varepsilon)}}.$$

Algebraic manipulation leads to

$$0 = (\alpha - \alpha^*) \left[\frac{var(b(X))}{var(g(X))} - \frac{var(u)}{var(\varepsilon)}\right] + (\alpha - \alpha^*)^2\, \phi \left[\frac{var(b(X))}{var(g(X))} - \frac{var(u)}{var(\varepsilon)}\right].$$

This gives two roots

$$\begin{aligned} \alpha^* &= \alpha \\ \alpha^* &= \alpha + \frac{1}{\phi} = \alpha + \frac{var(\varepsilon)}{cov(u, \varepsilon)} \end{aligned}$$

Thus the only values of α^* that are consistent with the observed data and Condition 1-NP are members of the set $\mathcal{A}$, so it must be identified. Furthermore note that $\alpha^* = \alpha$ is a member of the set.

■

A.5 Cubic Solution from Instrumental Variables Approach

Following the text above, the question is whether the assumptions allow us to pin down the bias. Suppose it cannot. Then there would exist alternative values α^*, γ^*, and ε^* with $\alpha^* \neq \alpha$ so that for the same $\widehat{\alpha}$ as in the text

$$\widehat{\alpha} = \alpha^* + \frac{cov\left(v, \varepsilon^*\right)}{\lambda var(v)}.$$

Under these conditions note that

$$\begin{aligned} Y - \alpha^* CH &= \left(\alpha - \alpha^*\right) CH + X'\gamma + \varepsilon \\ &= \left(\alpha - \alpha^*\right) \left[X'\beta + u + \lambda\left(X'\pi + v\right)\right] + X'\gamma + \varepsilon, \end{aligned}$$

and thus

$$\begin{aligned} \gamma^* &= \gamma + \left(\alpha - \alpha^*\right)\left(\beta + \lambda\pi\right) \\ \varepsilon^* &= \varepsilon + \left(\alpha - \alpha^*\right)\left(u + \lambda v\right). \end{aligned}$$

But if this model satisfies the assumptions we know that

$$\frac{cov(X'\pi, X'\gamma^*)}{var(X'\gamma^*)} = \frac{cov(v, \varepsilon^*)}{var(\varepsilon^*)},$$

which is equivalent to

$$\begin{aligned} &\frac{cov\left(X'\pi, X'\gamma\right) + \left(\alpha - \alpha^*\right) cov\left(X'\pi, \left(X'\beta + \lambda X'\pi\right)\right)}{var\left(X'\gamma\right) + 2\left(\alpha - \alpha^*\right) cov(X'\gamma, \left(X'\beta + \lambda X'\pi\right)) + \left(\alpha - \alpha^*\right)^2 var(X'\beta + \lambda X'\pi)} \\ &= \frac{cov\left(v, \varepsilon\right) + \left(\alpha - \alpha^*\right) cov\left(v, \left(u + \lambda v\right)\right)}{var(\varepsilon) + 2\left(\alpha - \alpha^*\right) cov(\varepsilon, \left(u + \lambda v\right)) + \left(\alpha - \alpha^*\right)^2 var(u + \lambda v)}. \end{aligned}$$

Imposing the restriction from the true model

$$\phi \equiv \frac{cov(X'\pi, X'\gamma)}{var(X'\gamma)} = \frac{cov(v, \varepsilon)}{var(\varepsilon)},$$

yields

$$\begin{aligned} &\frac{\phi + \left(\alpha - \alpha^*\right) \frac{cov(X'\pi, (X'\beta + \lambda X'\pi))}{var(X'\gamma)}}{1 + 2\left(\alpha - \alpha^*\right) \frac{cov(X'\gamma, (X'\beta + \lambda X'\pi))}{var(X'\gamma)} + \left(\alpha - \alpha^*\right)^2 \frac{var(X'\beta + \lambda X'\pi)}{var(X'\gamma)}} \\ &= \frac{\phi + \left(\alpha - \alpha^*\right) \frac{cov(v, (u + \lambda v))}{var(\varepsilon)}}{1 + 2\left(\alpha - \alpha^*\right) \frac{cov(\varepsilon, (u + \lambda v))}{var(\varepsilon)} + \left(\alpha - \alpha^*\right)^2 \frac{var(u + \lambda v)}{var(\varepsilon)}}. \end{aligned}$$

Solving out yields

$$
\begin{aligned}
0 = (\alpha-\alpha^*)^3 & \left[\frac{cov\,(v,(u+\lambda v))}{var(\varepsilon)}\frac{var(X'\beta+\lambda X'\pi)}{var(X'\gamma)} - \frac{cov\,(X'\pi,(X'\beta+\lambda X'\pi))}{var(X'\gamma)}\frac{var(u+\lambda v)}{var(\varepsilon)}\right] \\
+(\alpha-\alpha^*)^2 & \left[\phi\frac{var(X'\beta+\lambda X'\pi)}{var(X'\gamma)} + 2\frac{cov\,(v,(u+\lambda v))}{var(\varepsilon)}\frac{cov(X'\gamma,(X'\beta+\lambda X'\pi))}{var(X'\gamma)}\right. \\
& \left. -\phi\frac{var(u+\lambda v)}{var(\varepsilon)} - 2\frac{cov\,(X'\pi,(X'\beta+\lambda X'\pi))}{var(X'\gamma)}\frac{cov(\varepsilon,(u+\lambda v))}{var(\varepsilon)}\right] \\
+(\alpha-\alpha^*) & \left[\frac{cov\,(v,(u+\lambda v))}{var(\varepsilon)} + 2\phi\frac{cov(X'\gamma,(X'\beta+\lambda X'\pi))}{var(X'\gamma)}\right. \\
& \left. -\frac{cov\,(X'\pi,(X'\beta+\lambda X'\pi))}{var(X'\gamma)} - 2\phi\frac{cov(\varepsilon,(u+\lambda v))}{var(\varepsilon)}\right].
\end{aligned}
$$

One solution to this cubic is the true α (i.e. $\alpha=\alpha^*$). Depending on whether the solution to the remaining quadratic is real or not, this value may be the only solution or there may be two others.

A.6 Proof of Theorem 4

Proof.

For any value $\alpha^* \neq \alpha$, consider an alternative model

$$Y^* = \alpha^* CH + g_{\alpha^*}(X) + \varepsilon_{\alpha^*}$$

with $E(\varepsilon_{\alpha^*} \mid X) = 0$ in which $g_{\alpha^*}$and ε_{α^*} satisfy the conditions of (3.16)-(3.18) .

Then it must be the case that

$$
\begin{aligned}
g_{\alpha^*}(X) &= E(Y-\alpha^* CH \mid X) \\
&= E(\alpha CH + g(X) - \alpha^* CH \mid X) \\
&= (\alpha-\alpha^*)\, b(X) + g(X).
\end{aligned}
$$

Thus g_{α^*} is identified and

$$\varepsilon_{\alpha^*} = Y - \alpha^* CH - g_{\alpha^*}(X)$$

is also identified.

Since $g_{\alpha^*}$and ε_{α^*} are identified, the set

$$\mathcal{A} = \left\{\alpha^* \in \Re : 0 \leq \frac{cov(CH^*,\varepsilon_{\alpha^*})}{var(\varepsilon_{\alpha^*}))} \leq \frac{cov(CH^*, g_{\alpha^*}(X))}{var(g_{\alpha^*}(X))}\right\}$$

must be identified.

At the true value of α, $g_{\alpha^*}(X) = g(X)$ and $\varepsilon_{\alpha^*} = \varepsilon$. Since for this model Condition 3 implies that

$$0 \leq \frac{cov(CH^*,\varepsilon)}{var(\varepsilon))} \leq \frac{cov(CH^*, g(X))}{var(g(X))},$$

the true value of α is a member of this set.

■

A.7 Justifying Condition 4

Consider applying Theorem 1 and (3.13) in the case in which $CH^* = CH$. Following the notation in the text this would imply that

$$\text{Proj}(CH|X'\gamma, \varepsilon) = \phi_0 + \phi_{X'\gamma}X'\gamma + \phi_\varepsilon\varepsilon; \phi_{X'\gamma} = \phi_\varepsilon \ .$$

Since ε is uncorrelated with X by definition of γ, the above equation implies

$$\frac{cov(CH, X'\gamma)}{var(X'\gamma)} = \frac{cov(CH, \varepsilon)}{var(\varepsilon)}. \tag{A-10}$$

Let $P = \Pr(CH = 1)$. Since CH is binary, for any random variable Q,

$$cov(CH, Q) = P(1 - P)\left[E(Q \mid CH = 1) - E(Q \mid CH = 0)\right].$$

Applying this to (A-10) for $Q = \varepsilon$ and $Q = X'\gamma$ yields

$$\frac{E(X'\gamma \mid CH = 1) - E(X'\gamma \mid CH = 0)}{var(X'\gamma)} = \frac{E(\varepsilon \mid CH = 1) - E(\varepsilon \mid CH = 0)}{var(\varepsilon)}$$

which is Condition 4.

A.8 Heterogeneity in the Effects of Catholic Schools

Our analysis extends in a natural way to the case of heterogeneity in the effect of attending Catholic school. Let Y_{ch}^* and Y_p^* be the outcomes conditional on choice of Catholic high school and public high school, respectively, for a given student. As above let W be the set of covariates that fully determine Y_{ch}^* and Y_p^* and let X be the observed components of W. The heterogeneous effects model may be written as

$$\begin{aligned} CH^* &= b(X) + u \\ Y_{ch}^* &= g_{ch}(X) + \varepsilon_{ch} \\ Y_p^* &= g_p(X) + \varepsilon_p, \end{aligned}$$

where Y_{ch}^* is observed if $CH^* \geq 0$, in which case $CH = 1$, and Y_p is observed otherwise. Our previous specification is a special case of this model in which $g_{ch}(X) - g_p(X)$ is constant and $\varepsilon_{ch} = \varepsilon_p$. Treating the data generation processes for Y_{ch}^* and Y_p^* as equivalent to the data generation process for $Y^* - \alpha CH$ above and applying Theorem 1, we obtain the restrictions

$$\begin{aligned} \text{Proj}(CH^* | g_{ch}(X), \varepsilon_{ch}) &= \phi_{ch} g_{ch}(X) + \phi_{ch} \varepsilon_{ch} \\ \text{Proj}(CH^* | g_p(X), \varepsilon_p) &= \phi_p g_p(X) + \phi_p \varepsilon_p. \end{aligned}$$

These restrictions can be used to help identify the model in a way that is directly analogous to our use of Condition 1 to identify the model in the homogeneous effects case. We conjecture that if the components of X are a random subset of the components of W and if the number of elements of W and X are large, then the joint distribution of $(b(X), g_{ch}(X), g_p(X))$ is the same as the joint distribution of $(u, \varepsilon_{ch}, \varepsilon_p)$ up to a scale parameter that depends on the fraction of elements of W that are observed. If this is the case, then a nonparametric or semiparametric analysis may be possible, at least in theory. We leave a full analysis of this case to future work.

Appendix B: Sample Creation and Variables Used

B.1 Description of all variables used

The variables used in the empirical analysis can be classified into several categories: demographics, family background, geography, eighth grade test scores, eighth grade performance in school, and outcomes. We describe each of these in turn, with NELS:88 variables used in the creation of our measures shown in *italics*.

Demographic Variables:
These include indicators for female, asian, hispanic, black, and whether catholic, which is created from parental responses regarding religion (*byp29*).

School Sector:
Eighth Grade Sector (*g8ctrl1*)
High School Sector (CH) (*g10ctrl1*)

Family Background Measures:
Household composition: Separate 0-1 indicators for whether the student lives with his/her mother and father, mother and male guardian, father and female guardian, mother only, or father only. Excluded category is "other relative or non-relative". Created from *byfcomp*.

Parents' marital status: Separate 0-1 indicators for divorced, widowed, separated, never married, and not married but living in a marriage-like relationship. The excluded category is married. Created from *byparmar*.

Mother's and father's education: Continuous variables ranging from 8-18 years created from parental questionnaires (*byp30* and *byp31*). If these variables are missing, student responses from *bys34a* and *bys34b* are used.

Log family income: Continuous variable created using the midpoints of the ranges of the categorical variable *byfaminc*.

Missing value treatment: All family background variables are set equal to the sample mean when missing, and new 0-1 indicators for missing values are created for each of the original variables.

Geographic Variables:
Region and Urbanicity: These are 0-1 indicator variables taken from the urbanicity and region controls for the 8th grade school the student attended, variables *g8urban* and *g8region*. There are a total of 3 urbanicity and 8 region categories.

Distance Measures: 6 categories of distance from the student to the nearest catholic high school, ranging from 0-1 mile, 1-3, 3-6, 6-10, 10-20, and over 20. Student's residence was taken as the center of the 8th grade school's zip code. The zip code was determined by matching on zipcode population in NELS with the Census of Population and Housing zip code level data. High school locations were assigned the center of the zip code as reported in Ganley's *Catholic Schools in America*, 1988 edition. We obtain the minimum distance for each student by first computing distance to all of the Catholic schools using a program from the National Oceanic and Atmospheric Administration.

Eighth Grade Test Score Measures:
All test scores were taken from NELS standardized values from Item Response Theory scaled scores—*by2xrstd*, *by2xmstd*, *by2xsstd*, and *by2xhstd*.

Eighth Grade Performance-in-School Measures:
Delinquency Index: Created from student self-reports of whether sent to the office for misbehavior (*bys55a*) or parents contacted because of a behavior problem (*bys55e*). This variable ranges in value from 0-4.

Student got in a fight: Created from student self-reported variable *bys55f,* this variable ranges from 0 ("never") to 2 ("more than twice in the past semester").

Student performs below ability: 0-1 indicator variable taken from teacher surveys (*byt1_2* and *byt4_2*).

Student rarely completes homework: 0-1 indicator variable taken from teacher surveys (*byt1_3* and *byt4_3*).

Student frequently absent: 0-1 indicator variable taken from teacher surveys (*byt1_4* and *byt4_4*).

Student frequently tardy: 0-1 indicator variable taken from teacher surveys (*byt1_5* and *byt4_5*).

Student inattentive in class: 0-1 indicator variable taken from teacher surveys (*byt1_6* and *byt4_6*).

Student frequently disruptive in class: 0-1 indicator variable taken from teacher surveys (*byt1_8* and *byt4_8*).

Trouble-Maker: 0-1 indicator variable created from *bys56e*, and coded as 1 if the student report indicates that other students see the respondent as a "very big" trouble-maker.

Behavior problem: 0-1 indicator variable created from *byp50*, regarding whether the parent considers their child to have a behavior problem in school.

Parents Contacted About Behavior: Created from *byp57e*, this variable corresponds to how often parents report being contacted about behavior problems in the past school year, ranging from 0 ("never") to 3 ("more than four times").

Limited English Proficiency Composite: 0-1 indicator variable *bylep*, a NELS composite variable created from student and teacher reports.

Repeated Grade: 0-1 indicator of whether a student repeated any grade 4-8, taken as the maximum of the student (*bys74e*-*bys74i*) and parent (*byp46e*-*byp46i*) reports.

Grade trouble index: Created from student self-reports of whether sent to the office for grade problems (*bys55b*) or parents contacted because of a grade problem (*bys55d*). This variable ranges in value from 0-4.

Risk index: Taken from NELS composite variable *byrisk*, ranging from 0-6. This variable was constructed using NELS coding, from the following 6 questionnaire variables: *byfcomp*, *bypared*, *byp6*, *bys41*, *bylep*, and *byfaminc*.

Grade Index: Taken from NELS variable *bygrads*, ranging from 0-4.

Unpreparedness Index: Taken from student self-reports regarding how often the respondent comes to class without pencil or paper (*bys78a*), books (*bys78b*), and homework (*bys78c*), each of which range from 1 (usually) to 4 (never). These variables are summed so that the index ranges from 3 to 12.

Gifted: 0-1 indicator of whether parent reported student to be currently enrolled in a gifted/talented program (*byp51*).

Outcome Measures:
Test Scores: All 10th and 12th grade test scores were taken from NELS standardized values from Item Response Theory scaled scores—*f12xrstd*, *f12xmstd*, *f22xrstd*, and *f22xmstd.*

High School Graduation: 0-1 indicator for whether received high school diploma as of the third follow-up, coded equal to one if *hsstat*=1.

College Attendance: 0-1 indicator for whether enrolled in a 4-year college as of April 1994, coded equal to one if third follow-up variable *enrl0494*=15 or 16.

B.2 Sample Creation

The final sizes for the three samples used were 11,278 for the pooled (Catholic and public 8th grade) sample, 973 for Catholic 8th graders only, and 844 for the urban minorities, although sample sizes in the empirical work will differ slightly due to nonresponse in the outcome measures. Observations were excluded for one of several reasons: the 8th or 10th grade school sector could not be determined to be either public or Catholic, or one or more of the previously described demographic variables, location variables, eighth grade test scores, or eighth grade performance-in-school measures were missing. Attrition rates based on these grounds are presented below for each of the three samples. The sample sizes given in tables 3-12 reflect additional observations lost due to missing data on the particular dependent variable used.

Sample Attrition in NELS:88

Reason for Excluded Observations	Remaining Sample Size		
	Full NELS:88	Cath. 8th Grade	Urban Minority
No excluded cases	27,805	2,602	2,999
Student attended a non-Catholic private 8th Grade	25,233	2,602	2,895
8th grade school type missing	21,674	2,602	2,825
Student excluded from 2nd followup sample	16,460	1,416	1,648
Student in 2nd followup sample but not interviewed	16,168	1,398	1,574
Student attended a non-Catholic private 10th grade	16,130	1,393	1,571
10th grade school type missing	15,852	1,388	1,507
Missing location or demographic variables	14,367	1,174	1,327
Missing 8th grade test scores	13,648	1,146	1,195
Missing 8th grade performance-in-school variables	11,278	973	844

7 References

Alexander, Karl L., and Aaron M. Pallas, "School Sector and Cognitive Performance: When is a Little a Little" *Sociology of Education*, LVIII (1985), 115-128.

Altonji, Joseph G., "The Effects of Family Background and School Characteristics on Education and Labor Market Outcomes," unpublished manuscript, Northwestern University, 1988.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "The Effectiveness of Catholic School," under revision, Northwestern University, 1999.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "IV Strategies for Evaluating the Catholic School Effect", (2001, in progress).

Angrist and Evans, "Children and their Parent's Labor Supply: Evidence from Exogenous Variation in Family Size" *American Economic Review*, 88 (1988), 450-477.

Angrist, Joshua D., and Alan B. Krueger, "Empirical Strategies in Labor Economics," *Handbook of Labor Economics Vol. 3A,* Ashenfelter and Card (eds.), North Holland, 1999.

Bryk, Anthony S., Valerie E. Lee, and Peter B. Holland, *Catholic schools and the Common Good*, Cambridge, Mass. : Harvard University Press, 1993.

Bronars, Stephen G. , and Jeff Grogger, "The Economic Consequences of Unwed Motherhood: Using Twins as a Natural Experiment," *American Economic Review* 80(1994), 1141-56.

Cameron, Stephen V. and Heckman, James J., "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males," *Journal of Political Economy,* 106(1998), 262-333.

Cameron, Stephen V. and Christopher R. Taber, "Estimating Borrowing Constraints Using the Returns to Schooling," unpublished manuscript, Nortwestern University, 2001.

Chubb, John E., and Terry M. Moe, *Politics, Markets, and America's Schools* (Washington, D.C.: The Brookings Institution, 1990).

Coleman, James S., Thomas Hoffer, and Sally Kilgore, *High School Achievement: Public, Catholic, and Private Schools Compared* (New York, NY: Basic Books, Inc., 1982).

Coleman, James S., and Thomas Hoffer, *Public and Private Schools: The Impact of Communities* (New York, NY: Basic Books, Inc., 1987).

Cookson, Peter W., Jr., "Assessing Private School Effects: Implications for School Choice," in Edith Rasell and Richard Rothstein, eds., *School Choice: Examining the Evidence* (Washington, D.C.: Economic Policy Institute, 1993).

Currie, Janet, and Thomas Duncan, "Does Head Start Make a Difference?" *American Economic Review*, 85 (1990), 341-64.

Engen, Eric, William Gale, and John Karl Sholz, "The Illusory Effects of Saving Incentives on Saving," *Journal of Economic Perspectives,* 10 (1996), 113-138.

Evans, William N., and Robert M. Schwab, "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *Quarterly Journal of Economics,* 110 (1995), 947-974.

Figlio, David N., and Joe A. Stone, "Are Private Schools Really Better?," forthcoming *Research in Labour Economics*, (2000).

Goldberger, Arthur S., and Glen C. Cain, "The Causal Analysis of Cognitive Outcomes in the Coleman, Hoffer and Kilgore Report," *Sociology of Education*, LV (1982), 103-122.

Grogger, Jeff, and Derek Neal, "Further Evidence on the Benefits of Catholic Secondary Schooling," *Brookings-Wharton Papers on Urban Affairs* (2000), 151-193.

Heckman, James .J., "Varieties of Selection Bias," *American Economic Review,* 80(1990).

Heckman, J., and Robb, R., "Alternative Methods for Evaluating the Impact of Interventions," in J. Heckman and B. Singer eds.,*Longitudinal Analysis of Labor Market Data.* Cambridge, Cambridge University Press, 1985

Imbens, G., and Angrist, J., "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62 (1994), 467-75.

Jacobsen, Joyce P., James W. Pearce III, and Joshua L. Rosenbloom, "The Effect of Childbearing on Married Women's Labor Supply and Earnings," *Journal of Human Resources* 34(3), Summer 1999, pp. 449-474.

Jepsen, Christopher, "The Effectiveness of Catholic Primary Schools, " unpublished manuscript, Northwestern University, 2000.

Ludwig, Jens, "Educational Achievement in Public, Private, and Catholic Schools: New Evidence on What We Know (and Don't Know)," working paper, Georgetown University, 1997.

Manski, C., "Anatomy of the Selection Problem," *Journal of Human Resources,* 24 (1989), 343-360.

Manski, C., "The Selection Problem," in C. Sims (ed) *Advances in Econometrics: Sixth World Congress,* (Cambridge: Cambridge University Press, 1994).

McLeish, D. L., "A Maximal Inequality and Dependent Strong Laws," *The Annals of Probability*, 3 (1975), 829-839.

Murnane, Richard J., "A Review Essay—Comparisons of Public and Private Schools: Lessons from the Uproar," *Journal of Human Resources* 19 (1984), 263–77.

Murphy, Kevin M., and Robert H. Topel, "Efficiency Wages Reconsidered: Theory and Evidence," in Y. Weiss and R. Topel eds., *Advances in the Theory and Measurement of Unemployment.* New York, St. Martin's Press, 1990, 204-40.

Neal, Derek, "The Effects of Catholic Secondary Schooling on Educational Attainment," *Journal of Labor Economics* 15 (1997), 98-123.

Neal, Derek, "What Have We Learned About the Benefits of Private Schooling?" *Economic Policy Review,* Vol. 4, (1998), 79-86.

Noell, J., "Public and Catholic Schools: A Reanalysis of Public and Private Schools," *Sociology of Education* 55 (1982), 123-32.

Poterba, James, Steven Venti, and David Wise, "Targeted Retirement Saving and the Net Worth of Elderly Americans," *American Economic Review,* 84 (1994), 180-185.

Rosenbaum, Paul R., *Observational Studies*, Springer-Verlag, New York, (1995).

Sander, William, "Parochial Schools and Students Achievement: Findings for Older Adults," *Education Economics* 8 (2000), 259-68.

Staiger, Douglas, and James Stock, "Instrumental Variables Regression with Weak Instruments," *Econometrica,* (65) No. 3 (1997), 557-586.

Stock, James, "Unit Roots, Structural Breates and Trends," *Handbook of Econometrics, Volume 4,* Engle and Mcfaddend eds., Elsevier Science, (1994), 2740-2841.

Tyler, Brian S., "An Analysis of Public and Catholic Secondary Education and the Earnings of Men," University of Chicago Ph.D. Thesis (1994).

Udry, Christopher, "Gender, Agricultural Production, and the Theory of the Household", *Journal of Political Economy*, 104 (1996), 1010-1046.

White, Halbert, *Asymptotic Theory for Econometricians,* Academic Press, Inc. (1984).

Witte, John F., "Private School Versus Public School Achievement: Are There Findings That Should Affect the Educational Choice Debate," *Economics of Education Review*, XI (1992), 371-394.

Wooldridge, Jeffrey, *Introductory Econometrics*, South-Western College Publishing, 2000.

Table 1
Comparison of Means of Key Variables by Sector

Variable	Full Sample			Catholic 8th Grade		
	Public 10th (N=11,167)	Cath 10th (N=672)	Difference	Public 10th (N=366)	Cath 10th (N=640)	Difference
Demographics						
FEMALE	0.52	0.45	-0.07	0.61	0.50	-0.11
ASIAN	0.03	0.04	0.01	0.05	0.05	0.00
HISPANIC	0.09	0.09	0.00	0.08	0.09	0.01
BLACK	0.10	0.09	-0.01	0.07	0.11	0.04
WHITE	0.78	0.78	0.00	0.80	0.74	-0.06
Family Background						
MOTHER'S EDUCATION IN YEARS	13.21	13.96	0.75	13.34	13.88	0.54
FATHER'S EDUCATION IN YEARS	13.49	14.51	1.01	13.39	14.38	0.99
LOG OF FAMILY INCOME	10.23	10.72	0.49	10.47	10.66	0.19
MOTHER ONLY IN HOUSE	0.14	0.09	-0.05	0.07	0.09	0.02
PARENT MARRIED	0.79	0.89	0.10	0.90	0.88	-0.02
PARENTS CATHOLIC	0.28	0.82	0.54	0.84	0.84	0.00
Geography						
RURAL	0.36	0.03	-0.33	0.13	0.01	-0.12
SUBURBAN	0.45	0.51	0.06	0.40	0.48	0.08
URBAN	0.19	0.46	0.27	0.47	0.51	0.04
DISTANCE TO CLOSEST CATHOLIC HS, MILES	22.16	2.97	-19.19	6.91	2.37	-4.53
Expectations [1]						
SCHOOLING EXPECTATIONS IN YEARS	15.25	15.97	0.72	15.52	15.92	0.40
VERY SURE TO GRADUATE HS	0.84	0.89	0.05	0.84	0.90	0.06
PARENTS EXPECT AT LEAST SOME COLLEGE	0.89	0.98	0.09	0.94	0.98	0.04
PARENTS EXPECT AT LEAST COLLEGE GRAD	0.79	0.92	0.13	0.88	0.91	0.03
STUDENT EXPECTS WHITE-COLLAR JOB	0.47	0.61	0.14	0.55	0.59	0.04
8th Grade Variables						
DELINQUENCY INDEX, RANGE FROM 0 TO 4	0.64	0.53	-0.11	0.54	0.46	-0.08
STUDENT GOT INTO FIGHT	0.24	0.23	-0.02	0.20	0.19	-0.01
STUDENT RARELY COMPLETES HOMEWORK	0.19	0.08	-0.11	0.08	0.06	-0.01
STUDENT FREQUENTLY DISRUPTIVE	0.12	0.08	-0.05	0.08	0.08	0.00
STUDENT REPEATED GRADE 4-8	0.06	0.02	-0.05	0.03	0.02	-0.01
RISK INDEX, RANGE FROM 0 TO 4	0.69	0.35	-0.34	0.39	0.39	0.00
GRADES COMPOSITE	2.94	3.16	0.22	3.09	3.20	0.11
UNPREPAREDNESS INDEX, FROM 0 TO 25	10.77	11.08	0.31	10.84	11.02	0.17
8TH GRADE READING SCORE	51.19	55.05	3.86	54.12	55.59	1.47
8TH GRADE MATHEMATICS SCORE	51.13	54.57	3.44	52.89	53.98	1.09
Outcomes						
10TH GRADE READING STANDARDIZED SCORE	51.02	54.69	3.66	54.63	54.62	-0.01
10TH GRADE MATH STANDARDIZED SCORE	51.12	55.03	3.91	53.40	54.52	1.12
12TH GRADE READING STANDARDIZED SCORE	51.20	54.60	3.40	53.25	54.70	1.45
12TH GRADE MATH STANDARDIZED SCORE	51.20	55.54	4.34	53.13	55.63	2.49
ENROLLED IN 4 YEAR COLLEGE IN 1994	0.31	0.59	0.28	0.38	0.61	0.23
HS GRADUATE	0.85	0.98	0.13	0.88	0.98	0.10

Notes:
(1) The Expectations variables are not included in our empirical models

Table 2
Comparison of Means of Key Variables by Sector, NELS:88 Urban Minority Subsample

Variable	Full Sample			Catholic 8th Grade		
	Public 10th	Cath 10th	Difference	Public 10th	Cath 10th	Difference
Demographics	(N=700)	(N=56)		(N=15)	(N=54)	
FEMALE	0.57	0.57	0.00	0.60	0.61	0.01
ASIAN	0.00	0.00	0.00	0.00	0.00	0.00
HISPANIC	0.44	0.49	0.05	0.34	0.45	0.11
BLACK	0.56	0.51	-0.05	0.66	0.55	-0.11
WHITE	0.00	0.00	0.00	0.00	0.00	0.00
Family Background						
MOTHER'S EDUCATION, IN YEARS	12.61	13.27	0.66	13.58	13.21	-0.37
FATHER'S EDUCATION, IN YEARS	12.64	14.33	1.69	12.66	14.36	1.70
LOG OF FAMILY INCOME	9.62	10.45	0.83	10.16	10.38	0.22
MOTHER ONLY IN HOUSE	0.29	0.27	-0.02	0.29	0.23	-0.06
PARENT MARRIED	0.57	0.74	0.18	0.71	0.79	0.08
PARENTS CATHOLIC	0.39	0.58	0.19	0.39	0.55	0.16
Geography						
RURAL	0.00	0.00	0.00	0.00	0.00	0.00
SUBURBAN	0.00	0.00	0.00	0.00	0.00	0.00
URBAN	1.00	1.00	0.00	1.00	1.00	0.00
DISTANCE TO CLOSEST CATHOLIC HS, MILES	6.04	1.90	-4.14	1.90	2.01	0.11
Expectations [1]						
SCHOOLING EXPECTATIONS, IN YEARS	15.27	16.10	0.83	16.48	16.05	-0.43
VERY SURE TO GRADUATE HS	0.80	0.94	0.14	0.88	0.94	0.06
PARENTS EXPECT AT LEAST SOME COLLEGE	0.90	0.99	0.09	0.95	0.99	0.04
PARENTS EXPECT AT LEAST COLLEGE GRAD	0.78	0.86	0.08	0.84	0.85	0.01
STUDENT EXPECTS WHITE-COLLAR JOB	0.53	0.72	0.19	0.50	0.70	0.20
8th Grade Variables						
DELINQUENCY INDEX, RANGE FROM 0 TO 4	0.88	0.63	-0.25	1.22	0.65	-0.57
STUDENT GOT INTO FIGHT	0.34	0.19	-0.15	0.05	0.19	0.15
STUDENT RARELY COMPLETES HOMEWORK	0.25	0.13	-0.12	0.23	0.14	-0.09
STUDENT FREQUENTLY DISRUPTIVE	0.19	0.17	-0.02	0.14	0.17	0.03
STUDENT REPEATED GRADE 4-8	0.11	0.05	-0.06	0.10	0.05	-0.05
RISK INDEX, RANGE FROM 0 TO 4	1.30	0.90	-0.40	1.05	0.91	-0.14
GRADES COMPOSITE	2.78	2.88	0.09	3.01	2.88	-0.13
UNPREPAREDNESS INDEX, FROM 0 TO 25	10.99	11.28	0.29	11.10	11.27	0.17
8TH GRADE READING SCORE	46.76	53.25	6.49	49.99	52.88	2.89
8TH GRADE MATHEMATICS SCORE	45.43	48.71	3.28	48.88	48.61	-0.27
Outcomes						
10TH GRADE READING STANDARDIZED SCORE	47.14	51.46	4.32	48.62	50.75	2.13
10TH GRADE MATH STANDARDIZED SCORE	45.80	48.92	3.12	48.16	48.09	-0.07
12TH GRADE READING STANDARDIZED SCORE	47.29	50.78	3.49	52.74	50.17	-2.57
12TH GRADE MATH STANDARDIZED SCORE	46.40	51.71	5.31	51.46	50.92	-0.54
ENROLLED IN 4 YEAR COLLEGE IN 1994	0.23	0.52	0.28	0.28	0.56	0.28
HS GRADUATE	0.78	0.99	0.21	0.89	1.00	0.11

Notes:
(1) The Expectations variables are not included in our empirical models

Table 3

OLS, Fixed Effect, and Probit Estimates of Catholic High School Effects[5,6]
in Subsamples of NELS:88
Weighted, (Huber-White Standard Errors in Parentheses)
[Marginal Effects in Brackets[4]]

	Full Sample				*Catholic 8th Grade Attendees*			
	Controls							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	None	Fam. BG, city size, and region.[1]	(2) plus 8th grade tests	(3) plus other 8th grade measures[2]	None	Fam. BG, city size, and region.[1]	(1) plus 8th grade tests	(2) plus other 8th grade measures[2]
HS Graduation								
Probit	0.97	0.57	0.48	0.41	0.99	0.88	0.95	1.27
	(0.17)	(0.19)	(0.22)	(0.21)	(0.24)	(0.25)	(0.27)	(0.29)
	[0.123]	[0.081]	[0.068]	[0.052]	[0.105]	[0.084]	[0.081]	[0.088]
(Pseudo R^2)	0.01	0.16	0.21	0.34	0.11	0.35	0.44	0.58
OLS	0.123	0.041	0.033	0.023	0.105	0.081	0.080	0.080
	(0.011)	(0.015)	(0.015)	(0.015)	(0.035)	(0.025)	(0.024)	(0.021)
Fixed Effects[3]	0.097	0.063	0.061	0.066	0.143	0.115	0.113	0.102
	(0.032)	(0.031)	(0.030)	(0.026)	(0.045)	(0.035)	(0.033)	(0.027)
College in 1994								
Probit	0.73	0.37	0.33	0.32	0.60	0.48	0.56	0.60
	(0.08)	(0.09)	(0.09)	(0.09)	(0.13)	(0.15)	(0.15)	(0.15)
	[0.283]	[0.106]	[0.084]	[0.074]	[0.236]	[0.154]	[0.154]	[0.149]
(Pseudo R^2)	0.02	0.19	0.29	0.34	0.04	0.18	0.29	0.36
OLS	0.283	0.135	0.113	0.111	0.236	0.159	0.162	0.145
	(0.030)	(0.032)	(0.028)	(0.026)	(0.048)	(0.050)	(0.044)	(0.043)
Fixed Effects[3]	0.146	0.084	0.086	0.097	0.269	0.219	0.220	0.207
	(0.050)	(0.049)	(0.044)	(0.043)	(0.058)	(0.056)	(0.049)	(0.048)

Notes:
(1) Control sets (2)-(4) include race (white/nonwhite), hispanic origin, gender, urbanicity (3 categories), region (8 categories), and distance to the nearest Catholic high school (5 categories). Family background variables used as controls include log family income, mother's and father's education, 5 dummy variables for marital status of the parents, and 8 dummy variables for household composition.
(2) "Other 8th grade measures" include measures of attendance, attitudes toward school, academic track, achievement, and behavioral problems (from teacher, parent, and student surveys). See Appendix Table 1.
(3) Fixed effects models include 807 and 75 dummy variables, respectively, for each 8th grade school represented in the 2 samples.
(4) Marginal effects of probit models are computed as average derivatives of the probability of an outcome with respect to catholic high school attendance.
(5) NELS:88 3rd follow-up questionnaire weights used in the computations.
(6) Sample sizes for Full sample: N=8560 (HS Graduation), N=8315 (College Attendance). For Catholic 8th Grade sample, N=859 (HS Graduation), N=834 (College Attendance)

Table 4

OLS and Fixed Effect Estimates of Catholic High School Effects[4,5]
in Subsamples of NELS:88
Weighted, (Huber-White Standard Errors in Parentheses)

	Full Sample				*Catholic 8th Grade Attendees*			
	Controls							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	None	Fam. BG, city size, and region.[1]	(1) plus 8th grade tests	(2) plus other 8th grade measures[2]	None	Fam. BG, city size, and region.[1]	(1) plus 8th grade tests	(2) plus other 8th grade measures[2]
10th Grade Reading Score								
OLS	3.65	1.69	0.17	0.15	0.22	-1.07	-1.27	-1.32
	(0.51)	(0.62)	(0.38)	(0.37)	(0.86)	(0.97)	(0.61)	(0.59)
R^2	0.01	0.18	0.69	0.69	0.00	0.18	0.64	0.67
Fixed Effects[3]	0.20	-0.63	0.52	0.64	-0.11	-0.31	-0.60	-0.87
	(1.00)	(0.99)	(0.72)	(0.70)	(1.16)	(1.19)	(0.76)	(0.77)
10th Grade Math Score								
OLS	3.94	1.31	0.06	0.16	0.88	-0.32	-0.16	-0.11
	(0.56)	(0.54)	(0.47)	(0.45)	(0.96)	(1.01)	(0.51)	(0.49)
R^2	0.01	0.24	0.78	0.79	0.00	0.22	0.77	0.79
Fixed Effects[3]	1.43	-0.13	0.17	0.30	0.50	-0.32	-0.15	-0.09
	(0.97)	(0.93)	(0.52)	(0.50)	(1.20)	(1.21)	(0.60)	(0.55)
12th Grade Reading Score								
OLS	4.28	2.08	1.18	1.14	1.92	0.17	0.37	0.33
	(0.47)	(0.54)	(0.38)	(0.38)	(0.82)	(0.98)	(0.63)	(0.62)
R^2	0.01	0.19	0.60	0.60	0.01	0.19	0.59	0.62
Fixed Effects[3]	2.21	0.74	1.55	1.64	1.75	0.35	0.48	0.27
	(0.95)	(1.21)	(0.84)	(0.81)	(1.06)	(1.33)	(0.82)	(0.79)
12th Grade Math Score								
OLS	4.86	1.98	1.07	0.92	2.79	1.10	1.46	1.14
	(0.44)	(0.54)	(0.34)	(0.32)	(0.77)	(1.00)	(0.53)	(0.46)
R^2	0.01	0.26	0.72	0.74	0.02	0.26	0.73	0.77
Fixed Effects[3]	3.50	0.75	1.07	1.24	3.30	1.29	1.66	1.19
	(0.94)	(1.09)	(0.66)	(0.61)	(1.01)	(1.29)	(0.67)	(0.58)

Notes:
(1) Control sets (2)-(4) include race (white/nonwhite), hispanic origin, gender, urbanicity (3 categories), region (8 categories), and distance to the nearest Catholic high school (5 categories). Family background variables used as controls include log family income, mother's and father's education, 5 dummy variables for marital status of the parents, and 8 dummy variables for household composition.
(2) "Other 8th grade measures" include measures of attendance, attitudes toward school, academic track, achievement, and behavioral problems (from teacher, parent, and student surveys). See Table 2 and Appendix Table 1.
(3) Fixed effects models include 807 and 75 dummy variables, respectively, for each 8th grade school represented in the 2 samples.
(4) NELS 1st follow-up and 2nd follow-up panel weights used for the 10th and 12th grade models, respectively.
(5) Sample sizes for Full sample: N=10,180 (10th Reading), N=10,166 (10th Math), N=8116 (12th Reading), N=8119 (12th Math). For Catholic 8th Grade sample, N=878 (10th Reading), N=878 (10th Math), N=739 (12th Reading), N=739 (12th Math).

Table 5

OLS, Fixed Effect, and Probit Estimates of Catholic High School Effects by Race and Urban Residence. Full Set of Controls[1,3]
(Huber-White Standard Errors in Parentheses)
[Marginal Effects in Brackets[4]]

	Sample			
	(1)	(2)	(3)	(4)
	Urban and Suburban White Only	Urban and Suburban Minorities Only	Urban White Only	Urban Minorities Only
HS Graduate	(N=3799)	(N=1308)	(N=1002)	(N=697)
Sample Mean	0.88	0.80	0.88	0.80
Probit	0.443	0.524	1.176	1.592
	(0.279)	(0.338)	(0.417)	(0.673)
	[0.046]	[0.085]	[0.091]	[0.191]
OLS	0.022	0.058	0.069	0.133
	(0.016)	(0.040)	(0.023)	(0.056)
Fixed Effects[2]	0.062	0.023	0.136	-0.009
	(0.031)	(0.089)	(0.048)	(0.107)
College in 1994	(N=3695)	(N=1258)	(N=981)	(N=666)
Sample Mean	0.37	0.26	0.32	0.26
Probit	0.354	0.697	0.506	0.677
	(0.107)	(0.201)	(0.167)	(0.303)
	[0.087]	[0.158]	[0.110]	[0.144]
OLS	0.115	0.168	0.151	0.146
	(0.030)	(0.058)	(0.044)	(0.090)
Fixed Effects[2]	0.119	0.180	0.160	0.182
	(0.052)	(0.115)	(0.070)	(0.158)

Notes:
(1) All models include controls for hispanic origin, gender, region, citysize, distance to the nearest Catholic school (5 categories), family background, 8th grade tests, and other 8th grade measures. (from teacher, parent, and student surveys). See Table 3 notes 1 and 2.
(2) Fixed effects models include a dummy variable for each 8th grade school attended by members of the sample.
(3) NELS:88 third follow-up sampling weights used in the computations.
(4) Marginal effects of probit models are computed as average derivatives of the probability of an outcome with respect to Catholic school attendance.

Table 6

OLS, Fixed Effect, and Probit Estimates of Catholic High School Effects by Race and Urban Residence[3]. Full Set of Controls[1] Weighted, (Huber-White Standard Errors in Parentheses) [Marginal Effects in Brackets]

	Sample			
	(1)	(2)	(3)	(4)
	Urban/Suburban White Only	Urban/Suburban Minorities Only	Urban Whites Only	Urban Minorities Only
10th Grade Reading Score	(N=4637)	(N=1386)	(N=1194)	(N=734)
Sample Mean	52.82	47.78	52.73	47.52
OLS	0.40	-1.38	0.01	-0.92
	(0.45)	(0.74)	(0.60)	(1.21)
Fixed Effects	0.24	-3.06	0.49	-2.68
	(0.81)	(2.07)	(0.92)	(2.21)
10th Grade Math Score	(N=4633)	(N=1382)	(N=1195)	(N=733)
Sample Mean	53.11	46.69	52.67	46.08
OLS	0.56	-1.04	0.16	-0.65
	(0.32)	(0.83)	(0.49)	(1.21)
Fixed Effects	0.19	-0.08	0.03	-0.50
	(0.51)	(1.32)	(0.66)	(1.86)
12th Grade Reading Score	(N=3638)	(N=1051)	(N=978)	(N=561)
Sample Mean	52.94	47.72	53.33	47.61
OLS	1.30	-0.72	1.59	-0.19
	(0.44)	(0.98)	(0.67)	(1.39)
Fixed Effects	1.33	-3.05	2.24	-1.86
	(1.00)	(3.04)	(1.15)	(2.81)
12th Grade Math Score	(N=3638)	(N=1053)	(N=979)	(N=563)
Sample Mean	53.09	47.33	52.90	48.88
OLS	1.07	1.17	1.69	1.25
	(0.35)	(0.76)	(0.52)	(1.09)
Fixed Effects	0.81	1.84	1.33	0.37
	(0.61)	(1.57)	(0.86)	(2.18)

Notes:
(1) All models include controls for hispanic origin, gender, region, city size, distance to the nearest Catholic school (5 categories), family background, 8th grade tests, and other 8th grade measures. (from teacher, parent, and student surveys). See Table 3 notes 1 and 2.
(2) Fixed effects models include a dummy variable for each 8th grade school attended by members of the sample
(3) NELS 1st follow-up and 2nd follow-up panel weights used for the 10th and 12th grade models, respectively.

Table 7

Sensitivity Analysis: Estimates of Catholic High School Effects Given Different Assumptions on The Correlation of Disturbances in Bivariate Probit Models in Subsamples of NELS:88[2]. Modified Control Set[3]. (Huber-White Standard Errors in Parentheses) [Marginal Effects in Brackets]

	Correlation of Disturbances[1]					
	$\rho = 0$	$\rho = 0.1$	$\rho = 0.2$	$\rho = 0.3$	$\rho = 0.4$	$\rho = 0.5$
HS Graduation:						
Full Sample	0.459	0.271	0.074	-0.132	-0.349	-0.581
Raw difference=0.12)	(0.150)	(0.150)	(0.150)	(0.148)	(0.145)	(0.140)
	[0.058]	[0.037]	[0.011]	[-0.021]	[-0.060]	[-0.109]
Catholic 8th Graders	1.036	0.869	0.697	0.520	0.335	0.142
(Raw difference=0.08)	(0.314)	(0.313)	(0.310)	(0.306)	(0.299)	(0.290)
	[0.078]	[0.064]	[0.050]	[0.038]	[0.025]	[0.011]
Urban	1.095	0.905	0.706	0.499	0.282	0.053
Minorities	(0.526)	(0.538)	(0.549)	(0.560)	(0.570)	(0.578)
(Raw difference=0.22)	[0.176]	[0.157]	[0.132]	[0.101]	[0.062]	[0.013]
College Attendance:						
Full Sample	0.331	0.157	-0.019	-0.196	-0.376	-0.558
(Raw difference=0.31)	(0.070)	(0.070)	(0.070)	(0.068)	(0.067)	(0.064)
	[0.084]	[0.039]	[-0.005]	[-0.047]	[-0.087]	[-0.125]
Catholic 8th Graders	0.505	0.336	0.165	-0.008	-0.184	-0.362
(Raw difference=0.23)	(0.121)	(0.120)	(0.119)	(0.117)	(0.114)	(0.110)
	[0.140]	[0.093]	[0.045]	[-0.002]	[-0.050]	[-0.099]
Urban	0.447	0.269	0.090	-0.091	-0.272	-0.455
Minorities	(0.282)	(0.282)	(0.280)	(0.276)	(0.269)	(0.259)
(Raw difference=0.30)	[0.116]	[0.062]	[0.020]	[-0.020]	[-0.057]	[-0.091]

Notes:
(1) Models estimated as bivariate probits with the correlation ρ between u and ε set to the values in column headings.
(2) NELS:88 3rd follow-up sampling weights used in the computations.
(3) Due to computational difficulties, several variables were excluded from the control sets in the bivariate probit models: all dummy variables for household composition, urbanicity and region, indicators for "student rarely completes homework", "student performs below ability", "student inattentive in class", "parents contacted about behavior", and a limited-English proficiency index. Other than these exclusions, the controls are identical to those described in Table 3 notes 1 and 2.

Table 8

Sensitivity of Estimates of Catholic Schooling Effects on College Attendance and HS Graduation to Assumptions about Selection Bias in NELS:88, Catholic 8th Grade Subsample[2], Modified Control Set[3] (Huber-White Standard Errors in Parentheses) [Marginal Effects in Brackets]

Model (Estimated as a Bivariate Probit)

$$CH_i = 1(X_i'\beta + u > 0)$$

$$Y_i = 1(X_i'\gamma + \alpha CH_i + \epsilon > 0)$$

Estimation Method 1: β, γ, and α estimated simultaneously as a constrained bivariate probit model:

Model	Constraint on ρ	HS Graduation Coefficients		College Attendance Coefficients	
		$\widehat{\rho}$	$\widehat{\alpha}$	$\widehat{\rho}$	$\widehat{\alpha}$
(1)	$\rho = \frac{cov(X\beta, X\gamma)}{var(X\gamma)}$	0.24	0.59	0.24	0.11
			(0.33)		(0.16)
			[0.07]		[0.07]
(2)	$\rho = 0$	0	1.04	0	0.51
			(0.31)		(0.12)
			[0.08]		[0.14]

Estimation Method 2: 2-step, with β obtained from a univariate probit, γ from a univariate probit on the public 8th grade subsample. Next, α is computed from a bivariate probit with β fixed at this initial value and γ fixed up to 6 proportionality factors.[4]

Model	Constraint on ρ	HS Graduation Coefficients		College Attendance Coefficients	
		$\widehat{\rho}$	$\widehat{\alpha}$	$\widehat{\rho}$	$\widehat{\alpha}$
(3)	$\rho = \frac{cov(X\beta, X\gamma)}{var(X\gamma)}$	0.09	0.94	0.27	0.06
			(0.30)		(0.10)
			[0.09]		[0.02]

Estimation Method 3: β, α, γ, and ρ estimated from an unrestricted bivariate probit model.

Model	Constraint on ρ	HS Graduation Coefficients		College Attendance Coefficients	
		$\widehat{\rho}$	$\widehat{\alpha}$	$\widehat{\rho}$	$\widehat{\alpha}$
(4)	none	0.13	0.77	-0.52	1.18
			(1.12)		(0.50)
			[0.07]		[0.27]

Notes:

(1) Estimation performed on a sample of Catholic 8th grade attendees from NELS:88. N=859 for the HS graduation sample, and N=834 for the college attendance sample.

(2) NELS:88 3rd follow-up sampling weights used in the computations.

(3) Due to computational difficulties, several variables were excluded from the control sets in the bivariate probit models. See Table 7, note 3.

(4) The categories of proportionality factors are demographics/family background, test scores, behavioral problems, school attendance and attitudes toward school, grades and achievement, and distance measures. The coefficients and (standard errors) of the proportionality factors for these categories are 0.82 (0.19), 0.87 (0.22), 0.92 (0.03), 1.07 (0.04), 0.59 (0.08), and 0.90 (6.08) respectively, in the high school graduation case. For college attendance, the coefficients and (standard errors) are 0.80 (0.01) , 1.01 (0.04), 0.95 (0.15), 0.43 (0.17), 1.44 (0.03), and 1.04 (1.59), respectively.

Table 9

The Amount of Selection on Unobservables Relative to Selection on Observables
Required to Attribute the Entire Catholic School Effect to Selection Bias[5]
(Huber-White Standard Errors in Parentheses)

Model: $Y_i = 1(X_i'\gamma + \alpha CH_i + \epsilon_i)$, estimated as a probit

Estimates of $\widehat{\alpha}$ from univariate probit on the Catholic 8th Grade Subsample, with γ freely estimated, Full Set of Controls[3]

Outcome:		$\frac{\widehat{E}(X_i'\widehat{\gamma}\mid CH_i=1)-\widehat{E}(X_i'\widehat{\gamma}\mid CH_i=0)}{\widehat{Var}(X_i'\widehat{\gamma})}$ (1)	$\widehat{Var}\ (\widehat{\epsilon})$ (2)	$E(\epsilon_i \mid CH_i=1)$ $-E(\epsilon_i \mid CH_i=0)$ if Cond. 4 Holds (3)	$\frac{Cov(\epsilon_i,\widehat{CH}_i)}{Var(\widehat{CH}_i)}$ (4)	$\widehat{\alpha}$ (5)	Implied Ratio (6)
HS Graduation (N=859)							
$\widehat{\gamma}$ in col. 1 from public 8th grade sample[1]	(1)	0.30	1.00	0.30	0.37	1.03	2.78
						(0.31)	
$\widehat{\gamma}$ in col. 1 from Cath. 8th grade, freely estimated[1]	(2)	0.20	...	0.20	0.24	...	4.29
$\widehat{\gamma}$ in col. 1 from Cath. 8th grade, $\alpha = 0$[1]	(3)	0.24	...	0.24	0.29	...	3.55
College Attendance (N=834)							
$\widehat{\gamma}$ in col. 1 from public 8th grade sample[1]	(4)	0.42	...	0.42	0.51	0.67	1.30
						(0.16)	
$\widehat{\gamma}$ in col. 1 from Cath. 8th grade, freely estimated[1]	(5)	0.27	...	0.27	0.33	...	2.03
$\widehat{\gamma}$ in col. 1 from Cath. 8th grade, $\alpha = 0$[1]	(6)	0.39	...	0.39	0.47	...	1.43

Notes:

(1) In rows (1) and (4) the $\widehat{\gamma}$ used to evaluate $\frac{\widehat{E}(X_i'\widehat{\gamma}\mid CH_i=1)-\widehat{E}(X_i'\widehat{\gamma}\mid CH_i=0)}{\widehat{Var}(X_i'\widehat{\gamma})}$ in column (1) is estimated using the public school sample. In rows (2) and (5) $\widehat{\gamma}$ is estimated using the Catholic school sample, and and in rows (3) and (6) $\widehat{\gamma}$ is estimated from the catholic school sample under the restriction $\alpha = 0$.

(2) See Table 3 notes 1 and 2 for a description of the controls.

(3) Condition 4 states that the standardized selection on unobservables is equal to the standardized selection on observables.
i.e. $\frac{E(\epsilon_i\mid CH_i=1)-E(\epsilon_i\mid CH_i=0)}{Var(\epsilon_i)} = \frac{E(X_i'\gamma\mid CH_i=1)-E(X_i'\gamma\mid CH_i=0)}{Var(X_i'\gamma)}$.

(4) "Implied Ratio" in column 6 is the ratio of standardized selection on unobservables to observables under the hypothesis that there is no catholic school effect.

(5) NELS:88 3rd follow-up sampling weights used in the computations.

Table 10

Selection Bias Estimates Using Differences by CH in Means of the Index of Observables from the Outcome Equations[4] (Huber-White Standard Errors in Parentheses), Full Control Set

Model: $Y_i = X_i'\gamma + \alpha CH_i + \epsilon_i$, estimated by OLS

Estimates of $\widehat{\alpha}$ taken from the Catholic 8th Grade Subsample with γ freely estimated

Outcome:	$\frac{\widehat{E}(X_i'\widehat{\gamma}\|CH_i=1)-\widehat{E}(X_i'\widehat{\gamma}\|CH_i=0)}{\widehat{Var}(X_i'\widehat{\gamma})}$ $\widehat{\gamma}$ from public. 8th grade sample[1]	$\widehat{Var}\ (\widehat{\epsilon})$	$E(\epsilon_i \mid CH_i=1)$ $-E(\epsilon_i \mid CH_i=0)$ if (Cond 4) Holds	$\frac{Cov(\epsilon_i,\widetilde{CH}_i)}{Var(\widetilde{CH}_i)}$	$\widehat{\alpha}$	Implied Ratio
10th Grade Reading Score (N=888)	0.029	28.52	0.83	1.00	-1.32 (0.56)	-1.33
10th Grade Math Score (N=878)	0.023	19.71	0.45	0.54	-0.11 (0.45)	-0.20
12th Grade Reading Score (N=739)	0.091	36.00	3.28	3.94	0.33 (0.62)	0.08
12th Grade Math Score (N=739)	0.038	24.01	0.91	1.09	1.14 (0.46)	1.04

Notes:

(1) Estimates formed using the full control set, and $\widehat{\gamma}$ estimated from the public 8th grade sample.

(2) Condition 4, used in constructing column 3 is that the standardized selection on unobservables is equal to the standardized selection on observables, i.e. $\frac{E(\epsilon_i|CH_i=1)-E(\epsilon_i|CH_i=0)}{Var(\epsilon_i)} = \frac{E(X_i'\gamma|CH_i=1)-E(X_i'\gamma|CH_i=0)}{Var(X_i'\gamma)}$.

(3) "Implied Ratio" in column 6 is the ratio of standardized selection on unobservables to observables under the hypothesis that there is no catholic school effect.

(4) NELS:88 1st follow-up and 2nd follow-up panel weights used for the 10th and 12th grade models, respectively.

Table 11

Selection Bias Estimates Using Differences by CH in Means of the Index of Observables from the Outcome Equations, Urban Minority Subsample[2,5]
(Huber-White Standard Errors in Parentheses)

Model: $Y_i = 1(X_i'\gamma + \alpha CH_i + \epsilon_i)$, estimated as a probit

Estimates of $\widehat{\alpha}$ taken from the Urban Minority Subsample with γ freely estimated

Outcome:		$\frac{\widehat{E}(X_i'\widehat{\gamma}\|CH_i=1)-\widehat{E}(X_i'\widehat{\gamma}\|CH_i=0)}{\widehat{Var}(X_i'\widehat{\gamma})}$ (1)	$\widehat{Var}\ (\widehat{\epsilon})$ (2)	$E(\epsilon_i \mid CH_i=1)$ $-E(\epsilon_i \mid CH_i=0)$ if (Cond 4) Holds[3] (3)	$\frac{Cov(\epsilon_i,\widehat{CH}_i)}{Var(\widehat{CH}_i)}$ (4)	$\widehat{\alpha}$ (5)	Implied Ratio[4] (6)
HS Graduation (N=698)							
$\widehat{\gamma}$ in col. 1 from public 8th grade UM sample[1]	(1)	0.56	1.00	0.56	0.67	1.59 (0.67)	2.37
$\widehat{\gamma}$ in col. 1 from full UM. sample, freely estimated[1]	(2)	0.56	...	0.56	0.68	...	2.34
$\widehat{\gamma}$ in col. 1 from full UM sample, $\alpha = 0$[1]	(3)	0.73	...	0.73	0.88	...	1.81
College Attendance (N=698)							
$\widehat{\gamma}$ in col. 1 from public 8th grade UM sample[1]	(4)	0.72	1.00	0.72	0.87	0.68 (0.30)	0.78
$\widehat{\gamma}$ in col. 1 from full UM. sample, freely estimated[1]	(5)	0.54	...	0.54	0.65	...	1.05
$\widehat{\gamma}$ in col. 1 from full UM sample, , $\alpha = 0$[1]	(6)	0.58	...	0.58	0.69	...	0.99

Notes:

(1) In rows (1) and (4) the $\widehat{\gamma}$ used to evaluate $\frac{\widehat{E}(X_i'\widehat{\gamma}|CH_i=1)-\widehat{E}(X_i'\widehat{\gamma}|CH_i=0)}{\widehat{Var}(X_i'\widehat{\gamma})}$ in column (1) is estimated using the public school urban minority sample. In rows (2) and (5) $\widehat{\gamma}$ is estimated using the full urban minority sample, and in rows (3) and (6) $\widehat{\gamma}$ is estimated from the full urban minority sample under the restriction $\alpha = 0$.

(2) Full Set of Control Variables.with city size and Black excluded. See Table 3, notes 1 and 2.

(3) Condition 4 states that the standardized selection on unobservables is equal to the standardized selection on observables, i.e. that $\frac{E(\epsilon_i|CH_i=1)-E(\epsilon_i|CH_i=0)}{Var(\epsilon_i)} = \frac{E(X_i'\gamma|CH_i=1)-E(X_i'\gamma|CH_i=0)}{Var(X_i'\gamma)}$.

(4) "Implied Ratio" in column (6) is column (5)/column (4). It corresponds to the ratio of standardized selection on unobservables to observables that is consistent with the hypothesis that there is no Catholic school effect.

(5) NELS:88 3rd follow-up sampling weights used in the computations.

Table 12

Selection Bias Estimates Using Differences by CH in Means of the Index of Observables from the Outcome Equations, Urban Minority Subsample[2,5] (Huber-White Standard Errors in Parentheses) Full Control Set

Model: $Y_i = X_i'\gamma + \alpha CH_i + \epsilon_i$, estimated by OLS

Estimates of $\widehat{\alpha}$ Taken from the Urban Minority Subsample

Outcome:	$\frac{\widehat{E}(X_i'\widehat{\gamma}\|CH_i=1)-\widehat{E}(X_i'\widehat{\gamma}\|CH_i=0)}{\widehat{Var}(X_i'\widehat{\gamma})}$	$\widehat{Var}\ (\widehat{\epsilon})$	$E(\epsilon_i \mid CH_i = 1)$ $-E(\epsilon_i \mid CH_i = 0)$ if (Cond 4) Holds[3]	$\frac{Cov(\epsilon_i,\widetilde{CH}_i)}{Var(\widetilde{CH}_i)}$	$\widehat{\alpha}$	Implied Ratio[4]
10th Grade Reading Score (N=734)	0.097	28.62	2.78	3.34	-0.92 (1.21)	-0.28
10th Grade Math Score (N=733)	0.074	19.36	1.44	1.73	-0.65 (1.21)	-0.38
12th Grade Reading Score (N=733)	0.090	30.58	2.76	3.31	-0.19 (1.39)	-0.06
12th Grade Math Score (N=561)	0.058	20.25	1.17	1.40	1.25 (1.09)	0.89

Notes:

(1) Estimates formed using the full control set, and $\widehat{\gamma}$ estimated from the urban minority public 8th grade sample.

(2) Full Set of Control Variables.with city size and Black excluded. See Table 3, notes 1 and 2.

(3) Condition 4 states that the standardized selection on unobservables is equal to the standardized selection on observables, i.e. that $\frac{E(\epsilon_i|CH_i=1)-E(\epsilon_i|CH_i=0)}{Var(\epsilon_i)} = \frac{E(X_i'\gamma|CH_i=1)-E(X_i'\gamma|CH_i=0)}{Var(X_i'\gamma)}$.

(4) "Implied Ratio" in column (6) is column (5)/column (4). It is the ratio of standardized selection on unobservables to observables under the hypothesis that there is no catholic school effect.

(5) NELS:88 1st follow-up and 2nd follow-up panel weights used for the 10th and 12th grade models, respectively.