WP-00-17

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

Joseph G. Altonji
Department of Economics,
Institute for Policy Research
Northwestern University
and NBER

Todd E. Elder
Department of Economics,
Institute for Policy Research
Northwestern University

Christopher R. Taber
Department of Economics,
Institute for Policy Research
Northwestern University
and NBER

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

Abstract

We develop estimation methods that use the amount of selection on the observables in a model as a guide to the amount of selection on the unobservables. We show that if the observed variables are a random subset of a large number of factors that influence the endogenous variable and the outcome of interest, then the relationship between the index of observables that determines the endogenous variable and the index that determines the outcome will be the same as the relationship between the indices of unobservables that determine the two variables. In some circumstances this fact may be used to identify the effect of the endogenous variable. 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 the endogenous variable 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 do not find much evidence for 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 do 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, restrictions on the functional form of the model, restrictions on the distribution of the unobserved variables, or restrictions on 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. But in most cases, doubt remains about the validity of the identifying assumptions and the inferences that are based on them.

Research on whether Catholic schools provide better education than public schools provides a good illustration of the difficulties. 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 school policies, are likely to

[1] The most influential examples are Coleman, Hoffer, and Kilgore (1982) and Coleman and Hoffer (1987). Other early example 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 (1999), 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.

be related to the effectiveness of the schools.

Several recent studies, including Evans and Schwab (1995), Neal (1997), Grogger and Neal (1999), Figlio and Stone (1999) 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 et al (1985) and Neal (1997) 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), 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 (1999) 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 recent 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 lead 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 and the National Longitudinal Survey of the High School Class of 1972. Neal (1997) uses the National Longitudinal Survey of Youth, 1979.

In this paper we make two contributions. The first is to develop estimation strategies that may be used 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 the vast majority of applications in economics and the other social sciences. The second is to use our strategy to assess the effectiveness of Catholic schools.

Our approach to estimation is based on the idea of using the degree of selection on observables as a guide to how much selection there is on the unobservables. It is routine for researchers to informally argue for the exogeneity of membership in a "treatment group" or of an instrumental variable by examining the relationship between group membership or 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 settings. More importantly, we provide a way to quantitatively assess the degree of estimation bias that is likely when a particular source of variation is used to identify a model.[5]

Not surprisingly, the empirical relevance of using the link between the endogenous variable and the observables as a guide to the link with the unobservables hinges on how the observables and unobservables are related. We make the theoretical point that prior information about how the observables are chosen from the full set of factors that determine the outcome can be used to identify an econometric model. Specifically, we consider the case in which the observed variables are a random subset of the factors that influence the endogenous variable and the outcome. Using our Catholic schools application, let the outcome Y be determined by

$$Y = \alpha CH + W'\Gamma \ , \tag{1.1}$$

where CH is an indicator for whether the student attends a Catholic high school, W is a vector containing the full set of other variables that influence Y, and α and Γ are coefficients. Let X represent the observed components of W and γ be the associated components of Γ. Then one may rewrite the above equation as

[4] See for example, Currie and Duncan (1995), Engen et al (1996), Poterba et al (1994), Angrist and Evans (1998), Jacobsen et al. (1999), Bronars and Grogger (1994), and Udry (1998).

[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.

$$Y = \alpha CH + X'\gamma + \varepsilon, \tag{1.2}$$

where the vector X is observed by the econometrician and the index ε is unobserved. Let the linear prediction equation for CH be

$$CH = X\beta + u, \tag{1.3}$$

where u is orthogonal to X by definition of β. We show that if (i) the elements of X are chosen at random from W and (ii) the number of elements in both X and W are large and none of the factors dominates the distribution of school choice CH or the outcome Y, then the relationship between the indices of observables in the school choice equation and the outcome equations will be the same as the relationship between the indices of unobservables in the two equations. This result implies that the coefficients of the linear projection of school choice on the indices of observables and unobservables that determine the outcome are equal. That is,

$$\text{Proj}(CH|X'\gamma, \varepsilon) = \phi_c X'\gamma + \phi_c \varepsilon. \tag{1.4}$$

We call this property equality of selection on the observables and unobservables. We show that equality of selection, prior knowledge about the sign of the bias, and an additional condition on the relationship among the included and excluded variables are sufficient to identify α. An important special case of that condition is when X is uncorrelated with the other elements W. This special case is the standard assumption in empirical analyses, which usually treat only one or two of the explanatory variables as potentially endogenous, but it not a natural assumption if one thinks of the X variables as a random subset of the regressors. When (1.1) and (1.3) are the equations for a latent variable CH^* and $CH = 1(CH^* > 0)$, then in some cases, including our application, our restrictions will identify α, while in others they are only sufficient to restrict α to be the root of a cubic equation.

Operationally, we estimate the joint model of Catholic school attendance and the outcome subject to the restriction

$$\frac{Cov(X'\beta, X'\gamma)}{Var(X'\gamma)} = \frac{Cov(u, \varepsilon)}{Var(\varepsilon)}, \tag{1.5}$$

which we show is equivalent to (1.4). This equation says that the coefficient of the regression relating the error terms in the equations for Catholic school attendance and the outcome is equal to coefficient of the regression relating the index of observables in the Catholic school attendance equation to the index of observables in the outcome equation. Similar ideas can be applied to "heterogeneous effects" models in which the benefits of Catholic school attendance and public school attendance vary across students.

We provide a similar estimation method that may be used 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 strictly exogenous. We show that if the observables are a random subset of the variables that determine the outcome, then the instrumental variable will have the same relationship with the regression index of the observables and the regression index of the unobservables. This condition can sometimes be used to identify α even though the instrumental variable is correlated with the error term in the outcome equation.

We also propose a 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. We compute $\frac{E(X'\gamma|CH=1)-E(X'\gamma|CH=0)}{Var(X'\gamma)}$, which is the normalized shift in the index of observables in the outcome equation that is associated with CH, and then ask how many times larger 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 CH 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

$$\text{(1.6)} \qquad \frac{E(\varepsilon|CH=1)-E(\varepsilon|CH=0)}{Var(\varepsilon)} = \frac{E(X'\gamma|CH=1)-E(X'\gamma|CH=0)}{Var(X'\gamma)}.$$

If selection on unobservables must be several times stronger than selection on the observables to explain for the entire estimate of α, then the case for a causal effect of Catholic school is strengthened.

In section 2 we set the stage for the development of our econometric methods in section 3 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. To reduce sample selection bias and to avoid confounding the effect of attending Catholic high school with the effect of Catholic elementary school, we use the Catholic eighth grade sample for much of our analysis, unlike most previous studies.

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 section 3 and 4 we develop and then 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 weaker than selection on observables. Consequently, our estimates of a joint model of Catholic high school attendance and educational attainment subject to (1.5) are likely to overstate selection and understate the Catholic school effect. 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 (1.6) 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 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 which serve as a benchmark for our subsequent analysis.

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, and the evolution of achievement throughout high school. 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.[6] The "College attendance" indicator is one if the respondent was enrolled in a four-year university at the date of the 1994 survey and zero otherwise.[7]

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.[8]

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, the

[6] 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.

[7] 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.

[8] 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 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.

results reported in the paper are weighted.[9] Details regarding construction of variables 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.[10] Looking at the full sample, the graduation and college attendance rates differ enormously between the two sectors. Catholic high school students are one fifth as likely to drop out of high school as their public school counterparts (0.03 versus 0.15), and are twice as likely to be enrolled in a four year college in 1994 (0.59 versus 0.29). 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 0.10), as is the gap in the college attendance rate (0.62 versus 0.39). The gap in the 12th grade math score is 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

[9] 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 weight declines with 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.

[10] 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.

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.[11] The high school sector differential in father's education is about one year in both samples, but for mother's education 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 very broad set of 8th grade observables is that selection bias is positive in both samples. These considerations, concern about selection bias arising from the fact that only a 0.3% of public school eighth graders in our effective sample go to Catholic high school, and the desire to avoid confounding the Catholic high school effect with the effect of Catholic elementary school lead us to focus on the sample of Catholic eighth graders in much of our analysis, in contrast to most previous studies.[12]

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

[12] 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

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, the Catholic high school attendance dummy, in univariate probit, OLS, and school fixed effects models for high school graduation.[13] The difference in means is 0.08 when no controls are included. In the probit model, the coefficient is 0.88 (0.25), with an associated average marginal effect on the graduation rate of 0.084, which is a huge effect given that the graduation rate is 0.947 among students from Catholic eighth grades. In this sample the family background and geographic controls explain *none* of the raw difference of 0.08 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 5, 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 6 are quite powerful, explaining 0.32 percent of the variance in the latent variable for high school graduation when CH is excluded from the model.

The second row in Table 3 is based on linear probability models of high school graduation. The coefficient on CH varies from 0.080 to 0.081 and closely agrees with the probit estimates. Row 3 of columns 4-6 adds eighth grade fixed effects to the specifications reported in row 2.[14] The fixed effects estimate is .115 for the basic specification and 0.102 when the full set of controls is included.[15]

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

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.

[13] 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 serial correlation among students from the same eighth grade.

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

[15] We report fixed effects results not because the use of fixed effects is necessarily a more appropriate estimator but rather 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 than the cross-school analysis.

specification (column 4) the probit estimate implies that CH raises the college enrollment probability by 0.154, which compares to a raw difference of 0.23. This estimate falls to 0.149 when we add detailed controls to the model. Linear probability models yield similar estimates.

In Table 4 we report estimates of the effect of CH on 10th and 12th grade reading and 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 (1999) 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.[16]

To facilitate a comparison to other studies, we also present estimates for the combined sample of students from Catholic and public eighth grades. For this sample the effect of Catholic high school attendance is reduced from 0.081 to 0.052 after we add the full set of controls (Table 3, columns 1-3). 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 derivative. The college attendance results largely mirror the high

[16] 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.

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 0.31. Note that the controls make a much larger difference in the full sample than in the Catholic eighth grade sample, which is consistent with the evidence that selection on the observables is more severe 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 Lessons from the Preliminary Analysis

Our preferred results, which are based on the Catholic eighth grade sample, suggest a strong positive effect of CH on high school graduation and college attendance. The estimates of the effect on 12th grade test scores are much smaller (Tables 5 and 6 present qualitatively similar results for urban minorities, which we consider in detail in section 5). The key question is how much of the estimated high school effect on educational attainment is real, and how much is due to selection bias. We have taken advantage of the fact that the NELS:88 data set contains an unusually rich set of family background variables and eighth grade outcomes that are likely to be relevant for educational attainment and achievement to provide some guidance regarding the extent of selection bias. The means of favorable variables are typically higher for Catholic high school students, suggesting positive selection bias. However, positive selection is more modest in the sample of Catholic eighth graders, and in this sample the estimates of the effect of CH on high school graduation and college attendance, are very large. Furthermore, the estimates are not very sensitive to the addition of a powerful set control variables, especially in the high school graduation case. We would conclude that part of the Catholic school effect on educational attainment is real, but could not go much beyond such a statement. This is where the typical analysis of bias due to selection on unobservables based on patterns in the observables would end.

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 estimation strategy. We then apply our methods to study the effect of CH.

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

In this section we consider ways to use the relationship among the observed determinants of Catholic school attendance and educational outcomes to provide a quantitative assessment of the importance of the bias resulting from a relationship among the unobserved determinants. In particular we show that modeling how the set of observed variables is arrived at can yield conditions that are useful for identification. In section 3.1 we provide a benchmark model of the data generation process in which the observed variables are a random subset of the determinants of the outcome. We establish that this implies (1.4), which we refer to as Condition 1. 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, although that is not the approach we take here. In sections 3.3 and 3.4 we consider the implications of Condition 1 for identification. In section 3.5 we extend the analysis to cases in which a possibly invalid instrumental variable is available and in section 3.6 we consider the case in which the effect of Catholic schools is heterogeneous. In section 3.7 we discuss the relevance of our analysis for studies of the Catholic school effect.

3.1 Random Choice of Observed Variables

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 some of the elements of W are observable to the econometrician and some are unobservable. Denote the observable portion of W as X and the corresponding elements of Γ as γ so that

$$Y = \alpha CH + X'\gamma + \varepsilon, \tag{3.2}$$

where the vector X is observed by the econometrician and ε is unobserved. That is for each potential covariate, W_j, let s_j be a dummy variable indicating whether W_j is observable. Then

$$X'\gamma = \sum_{j=1}^{K} s_j W_j \Gamma_j, \quad \varepsilon = \sum_{j=1}^{K} (1 - s_j) W_j \Gamma_j.$$

A key assumption for our approach is that the observables are a "random subset" of the 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.[17]

Random choice of the observables from W guarantees that the degree of selection on the observables and unobservables is the same. In particular we will show that it implies

Condition 1

$$\begin{aligned} \text{(3.3)} \qquad Proj(CH^*|X'\gamma, \varepsilon) &= Proj(CH^*|W'\Gamma) \\ &\equiv \phi_c W'\Gamma \\ &= \phi_c X'\gamma + \phi_c \varepsilon, \end{aligned}$$

where Proj(.|.) denotes a linear projection.The symmetry of unobservables and observables arises since the coefficients on $X'\gamma$ and ε are the same. This condition relates selection into Catholic school to the factors that influence the outcome Y. This formalization has direct relevance to the problem of omitted variables bias, as we shall see in section 3.4.

To see the intuition for Condition 1, define ϕ_{c1}, ϕ_{c2}, and ω such that

$$\text{(3.4)} \qquad CH^* = \phi_c W'\Gamma + \omega$$

$$\text{(3.5)} \qquad \text{Proj}\left(CH^* \mid X'\gamma, \varepsilon\right) = \phi_{c1} X'\gamma + \phi_{c2}\varepsilon$$

where ϕ_c is defined in equation (3.3). One can see that ϕ_{c1} and ϕ_{c2} would be equal to ϕ_c if $X'\gamma$ and ε were uncorrelated with ω. We can write this condition as

$$\text{(3.6)} \qquad E(X'\gamma\omega) = E(\varepsilon\omega) = 0.$$

[17] If this covariate is observable it will be observable for every person in the data set.

Notice that

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

Similar logic yields $E(\varepsilon\omega) = 0$. Consequently, Condition 1 holds on average over draws of the vector $\{s_1....s_K\}$.. This result in itself is not useful in practice because we only observe one draw of the sequence of s_j. To justify our condition we now show that as the number of covariates W gets large, a similar form of equality of selection on observables and unobservables holds.

In order to keep Y well behaved as K gets large, we model Y_K as

$$(3.8) \qquad Y_K = \alpha CH_K^* + \frac{1}{\sqrt{K}} \sum_{j=1}^{K} W_j \Gamma_j.$$

One can interpret this expression as specifying that either the scale of each covariate or the scale of γ is inversely proportional to $\sqrt{K}$, with $W_j^K = W_j/\sqrt{K}$ or $\Gamma_j^K = \Gamma_j/\sqrt{K}$.

We also need to guarantee that CH_K^* is well behaved as the number of covariates gets large. We do this by defining β_j to be the coefficients from a linear projection of CH_K^* on W, so that

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

where by definition u_K is orthogonal to W_j. Once again we scale by $\sqrt{K}$ to guarantee that CH_K^* is well behaved as K gets large. In the outcome equation, Γ is a "structural" or causal parameter defined in the model. In contrast, for the selection equation β is defined simply as the coefficient from a linear projection.

We now show that under certain assumptions Condition 1 will hold as the number of elements of W gets large. Note that our asymptotics are nonstandard. First, we are allowing the number of regressors K to get large. Second, W_j is different in a sense than (β_j, Γ_j) and s_j. For each j we draw one observation on (β_j, Γ_j) and s_j which are the same for every person in the population; however, each person will draw their own W_j. Consider the projection of CH_K^* on the observable portion of Y_K, $\frac{1}{\sqrt{K}} \sum_{j=1}^{K} s_j W_j \Gamma_j$, and the unobservable portion,

$\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(1-s_j)\,W_j\Gamma_j$. This projection is meant to be the population projection (i.e., for a very large number of persons) with K fixed, conditional on a particular realization of (β_j,Γ_j) and s_j, $j=1...K$. The theorem states that as K gets large the projection coefficients on $\frac{1}{\sqrt{K}}\sum_{j=1}^{K}s_jW_j\Gamma_j$ and $\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(1-s_j)\,W_j\Gamma_j$ will approach each other with probability one.

Theorem 1 *Let*

$$CH_K^* = \frac{1}{\sqrt{K}}\sum_{j=1}^{K}W_j\beta_j + u_K$$

$$Y_K - \alpha CH_K = \frac{1}{\sqrt{K}}\sum_{j=1}^{K}W_j\Gamma_j,$$

where W_j and (β_j,Γ_j) are independent nondegenerate, stationary, ergodic processes that satisfy the conditions for White's (1984) Central Limit Theorem 5.15, $E(W_j)=0$[18] *for $j=1,...,K$, and u_K is uncorrelated with W.*

Define ϕ_{1K} and ϕ_{2K} such that conditional on $s_1,...,s_K,\Gamma_1,...,\Gamma_K,\beta_1,...,\beta_K$,

$$\text{Proj}\left(Y_K - \alpha CH_K \mid \frac{1}{\sqrt{K}}\sum_{j=1}^{K}s_jW_j\Gamma_j, \frac{1}{\sqrt{K}}\sum_{j=1}^{K}(1-s_j)\,W_j\Gamma_j\right)$$
$$= \phi_{1K}\frac{1}{\sqrt{K}}\sum_{j=1}^{K}s_jW_j\Gamma_j + \phi_{2K}\frac{1}{\sqrt{K}}\sum_{j=1}^{K}(1-s_j)\,W_j\Gamma_j,$$

where s_j is an iid binary random variable. Then as K gets large, $(\phi_{1K}-\phi_{2K})$ converges in probability to zero.

(Proof in Appendix A)

We will apply this notion of equality of observables and unobservables to the Catholic school problem and show how it can aid identification.

3.2 Structural Models of School Choice and Condition 1

A conventional path to identification of causal effects in the presence of endogneous variables is through the use of an economic model as a source of a priori restrictions. Here we digress briefly to show that this kind of approach can also deliver restrictions like (3.3). Suppose that Catholic school attendance depends on X and ε only through Y. In addition, CH^*

[18] Assuming that the variables have zero expected value just simplifies the exposition; we could relax this restriction by including an intercept in the model without changing the results.

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:

$$CH^* = a_1 (Y - \alpha CH) \ + \varsigma, \tag{3.9}$$

where ς is uncorrelated with X and ε. Combining these equations one obtains

$$CH^* = \phi_c X'\gamma + \phi_c \varepsilon + \varsigma, \tag{3.10}$$

where $\phi_c = a_1$, and Condition 1 is satisfied.

This might be a plausible approximation to the decision making process by 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 the secondary school outcomes.[19] Our point is simply to establish that structural models of school choice and outcomes may also lead to Condition 1, even though we arrive at the Condition in a very different way. It also highlights the fact that our model of the data generation process is only sufficient for Condition 1, it is not necessary.

3.3 Identification Based on Condition 1

In this section we show how Condition 1 can help solve the identification problem. Model (3.2) developed above is linear;[20] however, in studying identification we want to isolate the contribution of Condition 1 from the role of linearity or other large sample properties (e.g., normality of ε).[21] We only want to rely on an assumption about the relationship

[19] 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 (3.9) may not be sufficient and one would require a priori information about which variables to enter. The implicit restrictions on the unobservable embodied in (3.9) 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.

[20] Linearity of 3.2 is not essential for Theorem 1, but separability between observables is. We treat unobservables symmetrically with observables, so if we allowed for interactions between observables we would have to allow for interactions between observables and unobservables as well. This is beyond the scope of this paper.

[21] For example, as long as the probability of going to a Catholic school is nonlinear, linearity of g in (3.11) below is sufficient for identification of α and one does not need an exclusion restriction. The propensity score could be used as an instrument.

between unobservables and observables rather than all of the implications of the model.[22] Consequently, we study identification of α using the following "treatment effect" model without exclusion restrictions:

$$
\begin{aligned}
(3.11) \qquad CH^* &= b(X) + u \\
CH &= 1(CH^* \geq 0) \\
Y &= \alpha CH + g(X) + \varepsilon,
\end{aligned}
$$

where $1(\cdot)$ is the indicator function taking the value one if its argument is true and zero otherwise.[23] The econometrician observes (X, CH, Y), but not the unobservables (u, ε) or the latent variable CH^*. The errors (u, ε) are independent of X.

Model (3.11) drops linearity but imposes the standard assumption that the observables X are independent of ε, which we did not need for Condition 1. In the independence case Condition 1 is equivalent to

Condition 2

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

While independence is sufficient for this condition, it is not necessary. Independence or mean independence is maintained in almost all studies of selection problems, but it is hard to justify in most settings, including ours.[24] The spirit of our analysis is that if there are correlations among the observables, which is the case in most applications, including ours,

[22]The use of a subset of restrictions implied by a model for identification is common in applied work. For example, consider a standard linear model with one endogenous variable such as

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

where we are concerned about endogeneity of CH but not X. One could always use nonlinear functions of covariates in X as instruments, but this is not deemed appropriate. 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.

[23]We abstract from most of the recent literature on program evaluation by assuming that α does not vary across individual. 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 discuss how one might extend our model into this framework in section 3.6.

[24]One can always rewrite $g(X) + \varepsilon$ as $\widetilde{g}(X) + \widetilde{\varepsilon}$ where $\widetilde{\varepsilon}$ is mean independent of X and $\widetilde{g}(X)$ is the sum of the structural effect of X on Y and its indirect association with Y through ε. However, if CH is correlated with X, then in most circumstances CH will not be mean independent of $\widetilde{\varepsilon}$, even if CH is mean independent of ε. Suppose one has an instrumental variable Z that determines CH and is independent of ε. Unless Z is also mean independent of X, it is unlikely to be mean independent of $\widetilde{\varepsilon}$. In nonexperimental situations X is unlikely to be independent of CH and Z.

correlations between the observables and unobserved variables are likely. In Appendix A we show that if the processes defining CH^* and $Y-\alpha CH$ are similar in a manner we make precise, then Condition 2 will hold after $g(X)$ and ε are redefined so that $g(X)$ is the sum of the structural effect of X on Y and the indirect effect through ε. Furthermore, in monte carlo results not reported, we did not obtain large biases even when the unobservables were correlated with the observables in the original data generating process, which provides some additional reassurance.

It is well known that (3.11) is not identified without further restrictions. We are essentially only one parameter (or one equation) short of identification.[25] This result suggests that one more restriction on this set of equations may deliver identification of α. We now show that prior information about how the observables are chosen can suffice.

It turns out that Condition 2 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 above let α be the true value of the treatment effect. Under Condition 2, in the data we can identify the set $\mathcal{A}$ for which α is a member. Define $p(X)$ as the propensity score $p(X) \equiv \Pr(CH=1 \mid X)$. The set of $\alpha^* \in \mathcal{A}$ are roots of the cubic*

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

(Proof in Appendix)

Since there is no constant term $\alpha^*=\alpha$ is one root of the cubic. Except for pathological

[25] In particular, if α were known, the system of equations would be identified. To see this, first note that with knowledge of α, we could identify $g(X)$ since

$$E(Y-\alpha CH \mid X) = g(X). \tag{3.12}$$

One can identify $b(X)$ and the distribution of u under sufficient normalizations since this is a standard binary choice model. Given $b(X)$, $g(X)$, and α, the joint distribution of (u,ε) can be identified by varying $b(X)$. For a discussion see Heckman and Robb (1985).

cases, there will be either no other real roots, or two others.[26]

The fact that we sometimes obtain three solutions is not a consequence of assuming the "unobservables are like the observables" per se, but rather a consequence of the particular form of restriction that we use. To see this consider the following alternative condition which is very easy to derive using the same basic methodology as in the previous subsection:

Condition 3

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

With Condition 3 we need one additional restriction which will fail to hold only in special cases:

Assumption 1

$$\frac{cov(b(X),p(X))}{var(b(X))}\neq\frac{cov(u,CH-p(X))}{var(u))}.$$

In this case we arrive at an exact identification result.

Theorem 3 *Assumptions 1 and Condition 3 are sufficient to identify α in the selection model defined above.*

(Proof in Appendix A)

The source of the difference in the results between using Condition 2 and Condition 3 is in the denominators. In Condition 3 the denominators on the two sides of the equation, $var(b(X))$ and $var(u)$, are identified directly from the binary choice model for CH. Changing α does not change the value of these. However, in Condition 2 the denominators, $var(g(X))$ and $var(\varepsilon)$, are not identified without knowledge of α. In particular, defining $(\alpha^*,g^*,\varepsilon^*)$ to be an alternative possibility for (α,g,ε), one may write $var(g^*(X))$ as

$$\begin{aligned}(3.13)\qquad var(g^*(X)) &= var\left(g(X)+(\alpha-\alpha^*)\,p(X)\right)\\ &= var(g(X))+2\,(\alpha-\alpha^*)\,cov(g(X),p(X))+(\alpha-\alpha^*)^2\,var(p(X)).\end{aligned}$$

Condition 2 may be rewritten as

$$cov(b(X),g^*(X))var(\varepsilon^*)=cov(u,\varepsilon^*)var(g^*(X)).$$

[26] 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.

The right hand side 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.

It is not clear how much we should worry about this potential problem. Consider equation (3.13). 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. While $var(g^*(X))$ is insensitive to reasonable values of α^* in our empirical work, ultimately the question of whether this is true in most applications can only be answered through empirical implementation.

3.4 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, the link between CH and CH^* in the theory section is not restricted to $CH = 1(CH^* > 0)$. Many potential applications of the idea involve continuous endogenous variables. We can incorporate this into the model by assuming that $CH = CH^*$. Consider the model

$$\begin{aligned} CH &= b(X) + u \\ Y &= \alpha CH + g(X) + \varepsilon, \end{aligned} \tag{3.14}$$

where (u, ε) have zero expected value conditional on X but may be correlated with each other, which implies that ε may be correlated with CH. In this case we obtain a stronger identification result using Condition 2 and one additional assumption:

Assumption 2

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

Theorem 4 *With the continuous endogenous variable model (3.14), under Condition 2 and Assumption 2 we can identify the set $\mathcal{A}$ which includes two values, the true α and $\alpha + \frac{var(\varepsilon)}{cov(u,\varepsilon)}$.*

(Proof in Appendix A)

Although there are two roots, this result is very useful. In most cases in which an applied researcher is worried about the bias in a regression type estimator, he or she 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.

When CH is binary, Model (3.14) is similar to a linear probability model. In the notation of the previous section we can write this as $b(X) = p(X)$. The problem in interpreting the linear probability model is that the error term u takes on two values, $1 - p(X)$ with probability $p(X)$ and $-p(X)$ with probability $(1 - p(x))$. Our underlying model treats the relationship between the unobservables as if it were similar to the relationship between the observables. In the linear probability case $p(X)$ is continuous while u is binary with a distribution that is determined by $p(X)$, which is difficult to justify using our condition. This problem does not arise for continuous endogenous variables.

3.5 Using an Invalid Instrumental Variable

In this section we extend the results above to the case in which the researcher works with an invalid instrumental variable Z that is 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. Once again we assume that X is uncorrelated with ε, but CH is potentially endogenous and thus correlated with ε. We assume our 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 for identification by showing that if the relationship between $X'\gamma$ and Z is similar to the relationship between ε and Z, then we can obtain identification.

Define β and π to come from least squares projection such that

$$\text{Proj}\,(Z \mid X) = X'\pi, \tag{3.15}$$

$$\text{Proj}\,(CH \mid X, Z) = X'\beta + \lambda Z, \tag{3.16}$$

and define v and u as the residuals of these regressions, so that

$$Z = X'\pi + v, \tag{3.17}$$

$$CH = X'\beta + \lambda Z + u. \tag{3.18}$$

Consider the regression of Y onto the predicted value Proj$(CH \mid X, Z)$ and X. The coefficient on the predicted value in this regression converges to

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

If Z is a valid instrument, v would be uncorrelated with ε and $\widehat{\alpha}$ would equal α.

Can our assumption about the relationship between unobservables and observables pin down this bias? In general it helps but may or may not be sufficient for identification. As above, the condition

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

restricts the solutions α^* to be the solutions of a cubic equation, one of which is a. 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 provided in section A.6 of Appendix A.

3.6 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

$$Y_{ch} = g_{ch}(X) + \varepsilon_{ch} \tag{3.19}$$

$$Y_p = g_p(X) + \varepsilon_p \tag{3.20}$$

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 process 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 X variables are a random subset of W and the number of elements of W and X is 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 (i.e., part of X). 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.

3.7 Relevance to the Study of the Catholic School Effect

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 observables may provide a useful guide to the unobservables.

However, there are strong reasons to expect that the relationship among the unobservables will be weaker than the relationship among the observables. The most important is that shocks that occur after eighth grade are excluded from X.[27] These will influence high school outcomes but not the probability of starting a Catholic high school.[28] To see this, augment the model by rewriting ε as $\varepsilon = \varepsilon_1 + \varepsilon_2$, where ε_1 is determined during eighth grade and ε_2 is the independent innovation in the error term that is determined during high school:

$$\begin{aligned} CH^* &= b(X) + u \\ Y &= \alpha CH + g(X) + \varepsilon_1 + \varepsilon_2. \end{aligned}$$

Since the observables X and the other unobservable u are determined during eighth grade,

[27] A second reason is that it is quite possible that among Catholic eighth graders the decision to attend Catholic high school is influenced by highly idiosyncratic preference variables, such as the religious beliefs of the parents, whether close friends of the student are going to Catholic high school, whether the parents attended Catholic high school, the influence of a particular eighth grade teacher or minister, the quality of the school band or sports teams, the logistics of getting to and from the school, or transitory variation in the finances of the family. We suspect that many of these factors have coefficients in the outcome and school choice equations that are quite different with some being positively correlated with outcomes and others negatively correlated. This would also lead us to overestimate the amount of correlation.

[28] In the case of the 10th and 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.

we can impose our data generation condition to the variable determined prior to high school so that

$$\begin{aligned} \frac{cov(b(X), g(X))}{var(g(X))} &= \frac{cov(u, \varepsilon_1)}{var(\varepsilon_1)} \\ &> \frac{cov(u, \varepsilon_1 + \varepsilon_2)}{var(\varepsilon_1 + \varepsilon_2)}. \end{aligned}$$

In the empirical work below, we use estimates of α that incorporate Condition 2 as an informal lower bound for α, and single equation estimates with CH treated as exogenous as an upper bound. 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 Adjusting for Selection Bias Using Selection on the Observables

We now examine the sensitivity of the estimates of the Catholic high school effect to assumptions about the correlation between the unobserved factors that determine CH and the various outcomes Y. We begin by simply displaying estimates of the Catholic school effect for a range of values of the correlation between the unobserved determinants of school choice in the outcome. We then use the extreme assumption that the relationship between the unobserved and observed determinants of school choice and the outcomes are the same to provide a lower bound on the Catholic school effect, given that selection on the unobservables is likely to be weaker than selection on the observables.

Consider the bivariate probit model

$$CH_i = 1(X_i'\beta + u > 0) \tag{4.1}$$

$$Y_i = 1(X_i'\gamma + \alpha CH_i + \varepsilon > 0) \tag{4.2}$$

$$(u, \varepsilon) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right). \tag{4.3}$$

While this model is formally identified without an exclusion restriction, semiparametric identification requires such an excluded variable. In the spirit of the previous section, 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. In section 4.1 we provide a sensitivity analysis in which we estimate the model restricting ρ to different values. In section 4.2 we use Condition 2 to pin down the value of ρ.

4.1 A Sensitivity Analysis.

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.[29] 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 derivatives of 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 probit case that we have already considered above) to 0.5 by estimating probit models constraining ρ to the specified value. For the full sample, the raw difference in the high school graduation probability is 0.12. 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 the estimates that take Catholic school attendance as exogenous suggest.

For our preferred sample of Catholic 8th graders, the results are less sensitive to ρ. 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 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

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

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. Since Table 1 suggests that there is only limited selection on observables, it appears that the amount of "selection on unobservables" would have to be a higher than the amount of "selection on observables" to account for the graduation effect. We explore this issue more formally in the next two sections.

4.2 Using the indices of Observables in the School Choice and Outcome Equations as a value for ρ.

Now we estimate the model given by (4.1), (4.2), and (4.3), subject to Condition 2, which in the bivariate probit case may be re-written[30] as

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

We take two approaches to estimating the model while imposing this restriction. The first is to use the Catholic eighth grade sample directly. To improve precision of the estimates of α and as a check on the robustness of the results, we employ an alternative method using 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 the same category. 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 then 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.[31]

In section 3 we discussed a set of conditions that imply Condition 2, but argued that $Cov(X_i'\beta, X_i'\gamma)/Var(X_i'\gamma)$ is likely to exceed the true value of ρ. Consequently, we view

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

[31] The restrictions 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, largely because the restriction fails for the coefficients on distance from Catholic school. Details are in Table 8 note 4.

$Cov(X_i'\beta, X_i'\gamma)/Var(X_i'\gamma)$ as an upper bound for ρ and treat it as such in assessing the evidence.

In Table 8, we present estimates using methods 1 and 2 to impose the restriction, focusing on the results for the Catholic eighth grade sample. 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 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 given our sample size.

When we use method 2, we obtain qualitatively similar results that point to an even larger effect of Catholic schooling on high school graduation—in this specification, ρ is only 0.09 and the estimate of the effect on the graduate probability is 0.09. The college effect is only 0.02. The restrictions on γ restrictions 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.[32]

4.3 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,

[32] For completeness, we also present 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 (1999) note, a finding of 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 2, identification of α and ρ is very tenuous. We place little weight on the unrestricted estimates.

Condition 4

$$\frac{E(\epsilon_i \mid CH_i = 1) - E(\epsilon_i \mid CH_i = 0)}{var(\epsilon_i)} = \frac{E(X_i'\gamma \mid CH_i = 1) - E(X_i'\gamma \mid CH_i = 0)}{var(X_i'\gamma)}.$$

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 informally in section A.7 of Appendix A. It will hold under the assumptions that lead to Conditions 1 and 2. However, it requires that X be uncorrelated with ε.

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, because many post-eighth grade factors influence the outcomes, and many hard-to-observe factors influence high school choice. 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 3 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 Condition 2. Consequently, we are able to use the full control set used in columns 3 and 6 of Tables 3 and 4. A similar approach may used to evaluate an instrumental variable.

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

$$Y_i^* = X_i'[\gamma + \alpha\beta] + \alpha\widetilde{CH}_i + \varepsilon_i.$$

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

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

Thus, subject to Condition 4 one can estimate $E(X_i'\gamma \mid CH_i = 1) - E(X_i'\gamma \mid CH_i = 0)$ 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 $X_i'\gamma$ 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 $\left(E(X_i'\gamma \mid CH_i = 1) - E(X_i'\gamma \mid CH_i = 0)\right) / Var(X_i'\gamma)$ is 0.30. That is, the mean/variance of the probit index of X variables that determine high school graduation is 0.28 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 $E(\epsilon_i \mid CH_i = 1) - E(\epsilon_i \mid CH_i = 0)$ 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 the α is 1.03. The last column of the table reports that the ratio $\widehat{\alpha}/[\frac{var(CH_i)}{var\left(\widetilde{CH}_i\right)}(E(\epsilon_i \mid CH_i = 1) - E(\epsilon_i \mid CH_i = 0))] = (1.03/.37) = 2.78$. 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. 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_i 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 (1999) 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 it is quite plausible that there could be more selection on unobservables as well and that 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 average derivative implied by the probit estimate of the Catholic high school effect on high school graduation is 0.191, while the linear probability model estimate is 0.133 (0.056).[33] 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

[33]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.

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 of the kind described above 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 ρ is 0.2, and 0.013 when ρ is 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.

In Table 11, we conduct an analysis involving the differences in indices of observable variables based on Condition 4. In rows 2 and 4 we form the index 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 $(E(\varepsilon_i \mid CH_i = 1) - E(\varepsilon_i \mid CH_i = 0))$ 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 a substantial part of the estimated effect of Catholic schooling on graduation would remain for this group, even if there was a high degree of sample selection bias. 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 $(E(X_i'\gamma \mid CH_i = 1) - E(X_i'\gamma \mid CH_i = 0)) / Var(X_i'\gamma)$ is 0.090. Under Condition 4 this amount of favorable selection on the observables implies an estimate of $(E(\varepsilon_i \mid CH_i = 1) - E(\varepsilon_i \mid CH_i = 0))$ 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 a is 1.82 and the implied estimate of $(E(\varepsilon_i \mid CH_i = 1) - E(\varepsilon_i \mid CH_i = 0))$ 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. Consequently, we would not rule out a small positive effect on math but overall conclude that there is not much evidence that Catholic high schools boost the test scores of urban minorities.[34]

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, including Altonji, Elder and Taber (1999), 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 a group of students who do not differ dramatically by whether or not they attended Catholic high school. This leads us to focus on students from Catholic eighth grades. Focusing 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 the 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 main

[34]These test score findings are robust to the imputation procedures for dropouts described in Section 2.3. In contrast, Grogger and Neal (1999) 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.

contribution of this paper is to formalize the use of such information and to provide a way to assess quantitatively 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. We estimate our model subject to the restrictions imposed by equal selection. We argue that in the Catholic school case, selection on the observables is likely to be stronger than selection on the unobservables. Consequently, we interpret the results as a lower bound estimate of the effect of Catholic schools and use the single equation estimates as 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. First, attending Catholic high school substantially raises high school graduation rates. In the Catholic eighth grade sample, none of the 0.08 Catholic high school advantage in graduation rates is explained by eighth grade outcomes or family background and we obtain a lower bound estimate of 0.07 when we impose equality of selection of observables and 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 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 range 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 strong than in high school graduation case.

Second, we find little evidence that Catholic high schools raise reading scores. In fact, most 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 as well as high school graduation rates.

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. One problem is that our sample of urban minorities who attended Catholic eighth grade is not big enough to permit us to perform the analysis on the Catholic eighth grade sample. Unfortunately, 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. While we believe that selection on the unobservables is less strong, the evidence for a Catholic school effect on college attendance is weaker for this group. In general, we find smaller differences between urban minorities and other groups in the Catholic school effect than other recent studies.

The next step on the empirical side of the project is to examine the mechanism through which Catholic schools affect high school graduation in light of the literature on Catholic schools and the data on school characteristics and student behavior during the high school years in NELS:88. 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 the heterogeneous effects case introduced in section 3.6. Our theoretical analysis suggests that the observables may not have much to say about bias from selection on unobservables in situations in which only a handful of variables dominate the distribution of the outcome (a situation in which structural economic model may be feasible to develop and estimate), or in which the set of observables is small. 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 prove informative, particularly in those cases in which concern about identification is a first order issue. One could also do a monte carlo analysis in which one samples at random from the hundreds of 8th grade family background and student characteristics available in the NELS 88.

Appendix A

A.1 Proof of Theorem 1

Proof. Define ϕ and ω_K so that

$$\phi = \frac{cov(\operatorname*{plim}_{K\to\infty}\{Y_K - \alpha CH_K\}, \operatorname*{plim}_{K\to\infty}\{CH_K^*\})}{var(\operatorname*{plim}_{K\to\infty}\{Y_K - \alpha CH_K\})}$$

$$\begin{aligned}\omega_K &= CH_K^* - \phi(Y_K - \alpha CH_K) - u_K \\ &= \frac{1}{\sqrt{K}}\sum_{j=1}^{K} W_j\left(\beta_j - \phi\gamma_j\right)\end{aligned}$$

Under the assumptions stated in the theorem

$$\frac{1}{\sqrt{K}}\sum_{j=1}^{K}\left[\begin{array}{l} W_j\Gamma_j \\ W_j\left(\beta_j - \phi\Gamma_j\right)\end{array}\right] \xrightarrow{d} N\left(0, V\right).$$

By definition of ϕ, $Y_K - \alpha CH_K$ is uncorrelated with ω_K asymptotically, so it must be the case that the covariance of $\operatorname*{plim}_{K\to\infty}\{Y_K - \alpha CH_K\}$ and $\operatorname*{plim}_{K\to\infty}\{\omega_K\}$ is zero:

$$\begin{aligned}0 &= cov\left(\operatorname*{plim}_{K\to\infty}\{Y_K - \alpha CH_K\}, \operatorname*{plim}_{K\to\infty}\{\omega_K\}\right) \\ &= \sum_{\ell=-\infty}^{\infty} E\left[(W_j\Gamma_j)E(W_{j-\ell}\left(\beta_{j-\ell} - \phi\Gamma_{j-\ell}\right))\right] \\ &= \sum_{\ell=-\infty}^{\infty} E\left(W_jW_{j-\ell}\right)E(\Gamma_j\left(\beta_{j-\ell} - \phi\Gamma_{j-\ell}\right))\end{aligned} \tag{A-1}$$

Now consider the conditional covariance between ω_K and $\frac{1}{\sqrt{K}}\sum_{j=1}^{K} s_jW_j\Gamma_j$:

$$\begin{aligned}&E\left(\omega_K\frac{1}{\sqrt{K}}\sum_{j=1}^{K} s_jW_j\Gamma_j \mid s_1,...,s_K,\Gamma_1,...,\Gamma_K,\beta_1,...,\beta_K\right) \\ &= E\left(\frac{1}{K}\sum_{j=1}^{K} s_jW_j\Gamma_j\sum_{r=1}^{K} W_r(\beta_r - \phi\Gamma_r) \mid s_1,...,s_K,\Gamma_1,...,\Gamma_K,\beta_1,...,\beta_K\right) \\ &= \frac{1}{K}\sum_{j=1}^{K} s_j\sum_{r=1}^{K}\Gamma_j(\beta_r - \phi\Gamma_r)E(W_jW_r)\end{aligned}$$

Notice that

$$\begin{aligned}&\lim_{k\to\infty}\sum_{r=1}^{K}\Gamma_j(\beta_r - \phi\Gamma_r)E(W_jW_r) \\ &= \sum_{\ell=-\infty}^{\infty}\Gamma_j\left(\beta_{j-\ell} - \phi\Gamma_{j-\ell}\right)E(W_jW_{j-\ell}) - \sum_{\ell=-\infty}^{-j}\Gamma_j\left(\beta_{j-\ell} - \phi\Gamma_{j-\ell}\right)E(W_jW_{j-\ell}),\end{aligned}$$

but by the assumptions in White's Theorem 5.15, the second summation after the equality tends to zero as $j \to \infty$ so its sample mean goes to zero. By (A-1), the first summation after the equality has expected value zero. Thus, applying the law of large numbers and (A-1)

$$\begin{aligned} & \lim_{K\to\infty} E\left(\omega_K \frac{1}{\sqrt{K}} \sum_{j=1}^{K} s_j W_j \Gamma_j \mid s_1, ..., s_K, \Gamma_1, ..., \Gamma_K, \beta_1, ..., \beta_K\right) \\ = \ & E(s_j) \sum_{\ell=-\infty}^{\infty} E(\Gamma_j \left(\beta_{j-\ell} - \phi\Gamma_{j-\ell}\right)) E(W_j W_{j-\ell}) \\ = \ & 0, \end{aligned}$$

and by the same logic

$$\lim_{K\to\infty} E\left(\omega_K \frac{1}{\sqrt{K}} \sum_{j=1}^{K} (1-s_j) W_j \Gamma_j \mid s_1, ..., s_K, \Gamma_1, ..., \Gamma_K, \beta_1, ..., \beta_K\right) = 0.$$

Recalling that ϕ_{1k} and ϕ_{2k} are coefficients from a projection of $Y_K - \alpha CH_K$ on $\frac{1}{\sqrt{K}} \sum_{j=1}^{K} s_j W_j \Gamma_j$ and $\frac{1}{\sqrt{K}} \sum_{j=1}^{K} (1-s_j) W_j \Gamma_j$, respectively, it follows that since ϕ satisfies the limiting version (as $K \to \infty$) of the moment conditions associated with linear projection, then

$$\lim_{K\to\infty} \phi_{1K} = \lim_{K\to\infty} \phi_{2K} = \phi.$$

■

A.2 Justification for Condition 2

When implementing our model we have assumed that the error terms are uncorrelated with the regressors, but this is not a property of the data generation process that we defined in Theorem 1. We briefly discuss the conditions under which our assumption is consistent with the previous analysis.

As above treat the model as

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

where β_j are the parameters from a linear regression of CH_K^* onto W and we have incorporated $\alpha - CH_K$ into Y_K to simplify the notation. Throughout this section we use "hats" to define the predicted value from a least square regression of a variable onto the observable covariates in W and "tildes" to denote the residual from that regression. For example $Y_K = \widehat{Y}_K + \widetilde{Y}_K$ where $\widehat{Y}_K$ is the linear prediction from a regression of Y_K on the observables. Furthermore we simplify the notation so that when we drop the K subscript we mean the probability limit of the variable so $Y \equiv \text{plim}\{Y_K\}$.

In a regression context rather than imposing Condition 2 for identification we might actually impose that

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

The question is when our data generation process yields this condition. It is straightforward to verify that A-2 is equivalent to

$$\frac{cov(CH^*,Y)}{var(CH^*)} = \frac{cov(\widetilde{CH^*},\widetilde{Y})}{var(\widetilde{CH^*})}.$$

Since CH_K^* and Y_K are linear we can also write

$$\text{(A-3)} \qquad \begin{aligned} \widetilde{CH}_K^* &= \frac{1}{\sqrt{K}}\sum_{j=1}^{K}\widetilde{W}_j\beta_j + u_K \\ \widetilde{Y}_K &= \frac{1}{\sqrt{K}}\sum_{j=1}^{K}\widetilde{W}_j\gamma_j \end{aligned}$$

Under the assumptions in Theorem 1 as the number of regressors gets large

$$\begin{aligned} \frac{cov(CH^*,Y)}{var(CH^*)} &\approx \frac{\sum_{\ell=-\infty}^{\infty} E\left(W_j\gamma_j W_{j-\ell}\beta_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(W_j\beta_j W_{j-\ell}\beta_{j-\ell}\right)} \\ &= \frac{\sum_{\ell=-\infty}^{\infty} E\left(W_j W_{j-\ell}\right) E\left(\gamma_j\beta_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(W_j W_{j-\ell}\right) E\left(\beta_j\beta_{j-\ell}\right)}, \end{aligned}$$

and similarly

$$\frac{cov(\widetilde{CH^*},\widetilde{Y})}{var(\widetilde{CH^*})} \approx \frac{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{W}_j\widetilde{W_{j-\ell}}\right) E\left(\gamma_j\beta_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{W}_j\widetilde{W_{j-\ell}}\right) E\left(\beta_j\beta_{j-\ell}\right)}.$$

Since in general the autocovariance structure of W_j will be different from the autocovariance structure of $\widetilde{W}_j$, these will be different and the restriction (A-2) will not be valid. However, we can give two examples for which (A-2) holds.

The first example is what motivated Condition 2. Suppose there is no serial correlation in the W's so that the unobservables are uncorrelated with the observables. If that is the case

$$\begin{aligned} \frac{cov(CH^*,Y)}{var(CH*)} &\approx \frac{E\left(W_jW_j\right)E\left(\gamma_j\beta_j\right)}{E\left(W_jW_j\right)E\left(\beta_j\beta_j\right)} \\ &= \frac{E\left(\gamma_j\beta_j\right)}{E\left(\beta_j\beta_j\right)} \\ &\approx \frac{cov(\widetilde{CH^*},\widetilde{Y})}{var(\widetilde{CH^*})}. \end{aligned}$$

The second example is if there exists some constant τ such that

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

In this case

$$\frac{cov(\widetilde{CH^*},\widetilde{Y})}{var(\widetilde{CH^*})} = \frac{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{W_j}\widetilde{W_{j-\ell}}\right)\tau E\left(\beta_j\beta_{j-\ell}\right)}{\sum_{\ell=-\infty}^{\infty} E\left(\widetilde{W_j}\widetilde{W_{j-\ell}}\right) E\left(\beta_j\beta_{j-\ell}\right)} = \tau.$$

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

$$\begin{aligned}\beta_j &= \omega_j^1 + \theta_1\omega_{j-1}^1 + \theta_2\omega_{j-2}^1 + ...\\ \gamma_j &= \omega_j^2 + \theta_1\omega_{j-1}^2 + \theta_2\omega_{j-2}^2 + ...\end{aligned}$$

If the joint distribution of (ω_j^1, ω_j^2) is stationary and $cov(\omega_j^1, \omega_k^2) = 0$ when $k \neq j$, then (defining $\theta_0 = 1$)

$$\begin{aligned}E\left(\gamma_j\beta_{j-\ell}\right) &= cov(\omega_j^1, \omega_j^2)\sum_{r=0}^{\infty}\theta_r\theta_{r+\ell}\\ E\left(\beta_j\beta_{j-\ell}\right) &= var(\omega_j^1)\sum_{r=0}^{\infty}\theta_r\theta_{r+\ell}\end{aligned}$$

so

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

A.3 Proof of Theorem 2

Proof. Suppose α is not identified. In particular suppose there exist $\alpha^* \neq \alpha$ and associated $(g^*(X), \varepsilon^*)$ for which

$$\begin{aligned}E(Y \mid CH = 0, X) &= g^*(X) + E(\varepsilon^* \mid b(X) + u \leq 0)\\ &= g(X) + E(\varepsilon \mid b(X) + u \leq 0),\end{aligned}$$

$$\begin{aligned}E(Y \mid CH = 1, X) &= \alpha^* + g^*(X) + E(\varepsilon^* \mid b(X) + u > 0)\\ &= \alpha + g(X) + E(\varepsilon \mid b(X) + u > 0),\end{aligned}$$

and

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

Solving these equations for g^* yields

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

where $p(X)$ is the propensity score and thus

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

If the alternative model satisfies Condition 2 then

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

or

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

Defining

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

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

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

Algebraic manipulation yields

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

■

A.4 Proof of Theorem 3

Proof. Following the same logic as in the proof of Theorem 2, if α^* yields the same observed data as α then

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

where

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

This means that

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

Under Condition 3 and Assumption 1, it is impossible for the items on the right hand side of these two equations to be equal to each other unless $\alpha = \alpha^*$. Thus we obtain a contradiction.

■

A.5 Proof of Theorem 4

Proof. Follow similar logic to Theorem 2. Suppose α is not identified. In particular suppose there exist $\alpha^* \neq \alpha$, g^*, and ε^* that satisfy our condition, so that

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

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

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

and

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

If the alternative model satisfies the condition then

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

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

$$\begin{aligned}
\text{(A-4)} \qquad & \frac{cov(b(X), g(X)) + (\alpha - \alpha^*)\, var(b(X))}{var(g(X)) + 2\,(\alpha - \alpha^*)\, cov(g(X), b(X)) + (\alpha - \alpha^*)^2\, var(b(X))} \\
= \ & \frac{cov(u, \varepsilon) + (\alpha - \alpha^*)\, var\,(u)}{var(\varepsilon) + 2\,(\alpha - \alpha^*)\, cov(\varepsilon, u) + (\alpha - \alpha^*)^2\, var(u)}.
\end{aligned}$$

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-4 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

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

■

A.6 Cubic Solution from Instrumental Variable

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

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

and thus

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

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

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

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

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

Solving out yields

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

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.7 Theorem 4: Justifying Condition 4

We do not formally justify Condition 4 with large sample theory as in the previous section, but just lay out the intuition. Let

$$
\begin{aligned}
X'\gamma &= \sum_{j=1}^{K} s_j W_j \Gamma_j, \\
\varepsilon &= \sum_{j=1}^{K} (1-s_j) W_j \Gamma_j,
\end{aligned}
$$

where $W_j\Gamma_j$ is independently distributed and s_j is *iid* binary and independent of $W_j\Gamma_j$. As before assume that $E(W_j)=0$. Then

$$
\begin{aligned}
\frac{E\left(X'\gamma \mid CH\right)}{Var(X'\gamma)} - \frac{E\left(\varepsilon \mid CH\right)}{Var(\varepsilon)} &= \frac{\sum_{j=1}^{K} E\left(s_j W_j \Gamma_j \mid CH\right)}{\sum_{j=1}^{K} E\left(s_j W_j \Gamma_j\right)^2} - \frac{\sum_{j=1}^{K} E\left((1-s_j) W_j \Gamma_j \mid CH\right)}{\sum_{j=1}^{K} E\left((1-s_j) W_j \Gamma_j\right)^2} \\
&= \frac{E\left(s_j\right) E\left(\sum_{j=1}^{K} W_j \Gamma_j \mid CH\right)}{E\left(s_j\right)\sum_{j=1}^{K} E\left(W_j\Gamma_j\right)^2} - \frac{E\left(1-s_j\right) E\left(\sum_{j=1}^{K} W_j \Gamma_j \mid CH\right)}{E\left(1-s_j\right)\sum_{j=1}^{K} E\left(W_j\Gamma_j\right)^2} \\
&= 0.
\end{aligned}
$$

Appendix B: Sample Creation and Variables Used

B.1 Description of all variables used

The variables used in the empirical analyses 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 3-11 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, "The Effects of Family Background and School Characteristics on Education and Labor Market Outcomes," unpublished manuscript, Northwestern University, 1988.

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

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

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.

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 Effects of Catholic Secondary Schooling," forthcoming *Brookings Wharton Papers on Urban Affairs*, (2000).

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).

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).

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.

Table 1
Comparison of Means of Key Variables by Sector

Variable	Full Sample			Catholic 8th Grade		
	Public 10th	Cath 10th	Difference	Public 10th	Cath 10th	Difference
Demographics	(N=11,167)	(N=672)		(N=366)	(N=640)	
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	0.29	0.59	0.31	0.39	0.62	0.23
HS GRADUATE	0.85	0.97	0.12	0.90	0.98	0.08

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

	Full Sample			Catholic 8th Grade		
Variable	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]

(Huber-White Standard Errors in Parentheses)
[Marginal Effects in Brackets[4]]

	Full Sample			*Catholic 8th Grade Attendees*		
	Controls					
	(1)	(2)	(3)	(4)	(5)	(6)
	Fam. BG, city size, and region[1].	(1) plus 8th grade tests	(2) plus other 8th grade measures[2]	Fam. BG, city size, and region[1].	(1) plus 8th grade tests	(2) plus other 8th grade measures[2]
HS Graduation						
Sample Means	CH=0: 0.85, CH=1: 0.97			CH=0: 0.90, CH=1: 0.98		
Probit	0.57	0.48	0.41	0.88	0.95	1.27
	(0.19)	(0.22)	(0.21)	(0.25)	(0.27)	(0.29)
	[0.081]	[0.068]	[0.052]	[0.084]	[0.081]	[0.088]
OLS	0.041	0.033	0.023	0.081	0.080	0.080
	(0.015)	(0.015)	(0.015)	(0.025)	(0.024)	(0.021)
Fixed Effects[3]	0.063	0.061	0.066	0.115	0.113	0.102
	(0.031)	(0.030)	(0.026)	(0.035)	(0.033)	(0.027)
College in 1994						
Sample Means	CH=0: 0.29, CH=1: 0.59			CH=0: 0.39, CH=1: 0.62		
Probit	0.37	0.33	0.32	0.48	0.56	0.60
	(0.09)	(0.09)	(0.09)	(0.15)	(0.15)	(0.15)
	[0.106]	[0.084]	[0.074]	[0.154]	[0.154]	[0.149]
OLS	0.135	0.113	0.111	0.159	0.162	0.145
	(0.032)	(0.028)	(0.026)	(0.050)	(0.044)	(0.043)
Fixed Effects[3]	0.084	0.086	0.097	0.219	0.220	0.207
	(0.049)	(0.044)	(0.043)	(0.056)	(0.049)	(0.048)

Notes:

(1) All models control for race (white/nonwhite), hispanic origin, gender, urbanicity (3 categories), region (8 categories), and distance to the nearest Catholic high school (5 categories). The family background variables used as controls in the regressions 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 B, "Eighth Grade Performance-in-School Measures".

(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]

(Huber-White Standard Errors in Parentheses)

	Full Sample			*Catholic 8th Grade Attendees*		
	Controls					
	(1)	(2)	(3)	(4)	(5)	(6)
	Fam. BG, city size, and region.[1]	(1a) plus 8th grade tests	(2a) plus other 8th grade measures.[2]	Fam. BG, city size, and region.[1]	(1b) plus 8th grade tests	(2b) plus other 8th grade measures[2]
10th Grade Reading Score						
Sample Means	CH=0: 51.02, CH=1: 54.69			CH=0: 54.63, CH=1: 54.62		
OLS	1.69	0.17	0.15	-1.07	-1.27	-1.32
	(0.62)	(0.38)	(0.37)	(.97)	(0.61)	(0.59)
Fixed Effects[3]	-0.63	0.52	0.64	-0.31	-0.60	-0.87
	(0.99)	(0.72)	(0.70)	(1.19)	(0.76)	(0.77)
10th Grade Math Score						
Sample Means	CH=0: 51.12, CH=1: 55.03			CH=0: 53.40, CH=1: 54.52		
OLS	1.31	0.06	0.16	-0.32	-0.16	-0.11
	(0.54)	(0.47)	(0.45)	(1.01)	(0.51)	(0.49)
Fixed Effects[3]	-0.13	0.17	0.30	-0.32	-0.15	-0.09
	(0.93)	(0.52)	(0.50)	(1.21)	(0.60)	(0.55)
12th Grade Reading Score						
Sample Means	CH=0: 51.20, CH=1: 54.60			CH=0: 53.25, CH=1: 54.70		
OLS	2.08	1.18	1.14	0.17	0.37	0.33
	(0.54)	(0.38)	(0.38)	(0.98)	(0.63)	(0.62)
Fixed Effects[3]	0.74	1.55	1.64	0.35	0.48	0.27
	(1.21)	(0.84)	(0.81)	(1.33)	(0.82)	(0.79)
12th Grade Math Score						
Sample Means	CH=0: 51.20, CH=1: 55.54			CH=0: 53.13, CH=1: 55.63		
OLS	1.98	1.07	0.92	1.10	1.46	1.14
	(0.54)	(0.34)	(0.32)	(1.00)	(0.53)	(0.46)
Fixed Effects[3]	0.75	1.07	1.24	1.29	1.66	1.19
	(1.09)	(0.66)	(0.61)	(1.29)	(0.67)	(0.58)

Notes:

(1) All models control for race (white/nonwhite), hispanic origin, gender, urbanicity (3 categories), region (8 categories), and distance to the nearest Catholic high school (5 categories). The family background variables used as controls in the regressions 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 B, "Eighth Grade Performance-in-School Measures".

(3) Fixed effects models include a 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:		$\frac{\widehat{E}(X_i'\widehat{\gamma}\|CH_i=1)-\widehat{E}(X_i'\widehat{\gamma}\|CH_i=0)}{\widehat{Var}(X_i'\widehat{\gamma})}$	$\widehat{Var}\ (\widehat{\epsilon})$	$E(\epsilon_i \mid CH_i=1)$ $-E(\epsilon_i \mid CH_i=0)$ if Cond. 3 Holds	$\frac{Cov(\epsilon_i,\widetilde{CH}_i)}{Var(\widetilde{CH}_i)}$	$\widehat{\alpha}$	Implied Ratio
		(1)	(2)	(3)	(4)	(5)	(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 $\frac{\widehat{E}(X_i'\widehat{\gamma}|CH_i=1)-\widehat{E}(X_i'\widehat{\gamma}|CH_i=0)}{\widehat{Var}(X_i'\widehat{\gamma})}$ in column (1) is estimated using the public school 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 3 states that the standardized selection on unobservables is equal to the standardized selection on observables.
i.e. $\frac{E(\epsilon_i|CH_i=1)-E(\epsilon_i|CH_i=0)}{Var(\epsilon_i)} = \frac{E(X_i'\gamma|CH_i=1)-E(X_i'\gamma|CH_i=0)}{Var(X_i'\gamma)}$.

(4) "Implied Ratio" in column 5 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:	$\frac{\widehat{E}(X_i'\widehat{\gamma}\mid CH_i=1)-\widehat{E}(X_i'\widehat{\gamma}\mid CH_i=0)}{\widehat{Var}(X_i'\widehat{\gamma})}$ $\widehat{\gamma}$ 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 3) 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 (3), used in constructing column 3 is that the standardized selection on unobservables is equal to the standardized selection on observables, i.e. $\frac{E(\epsilon_i\mid CH_i=1)-E(\epsilon_i\mid CH_i=0)}{Var(\epsilon_i)} = \frac{E(X_i'\gamma\mid CH_i=1)-E(X_i'\gamma\mid CH_i=0)}{Var(X_i'\gamma)}$.

(3) "Implied Ratio" in column 5 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:		$\frac{\widehat{E}(X_i'\widehat{\gamma}\mid CH_i=1)-\widehat{E}(X_i'\widehat{\gamma}\mid CH_i=0)}{\widehat{Var}(X_i'\widehat{\gamma})}$ (1)	$\widehat{Var}\ (\widehat{\epsilon})$ (2)	$E(\epsilon_i \mid CH_i=1)$ $-E(\epsilon_i \mid CH_i=0)$ if (Cond 3) 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 $\frac{\widehat{E}(X_i'\widehat{\gamma}\mid CH_i=1)-\widehat{E}(X_i'\widehat{\gamma}\mid CH_i=0)}{\widehat{Var}(X_i'\widehat{\gamma})}$ in column (1) is estimated using the public school 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 3 states that the standardized selection on unobservables is equal to the standardized selection on observables, i.e. that $\frac{E(\epsilon_i\mid CH_i=1)-E(\epsilon_i\mid CH_i=0)}{Var(\epsilon_i)} = \frac{E(X_i'\gamma\mid CH_i=1)-E(X_i'\gamma\mid CH_i=0)}{Var(X_i'\gamma)}$.

(4) "Implied Ratio" in column 6 is 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:	$\frac{\widehat{E}(X_i'\widehat{\gamma}\mid CH_i=1)-\widehat{E}(X_i'\widehat{\gamma}\mid CH_i=0)}{\widehat{Var}(X_i'\widehat{\gamma})}$	$\widehat{Var}\ (\widehat{\epsilon})$	$E(\epsilon_i \mid CH_i = 1)$ $-E(\epsilon_i \mid CH_i = 0)$ if (Cond 3) 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 3 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.