Catholic School Effectiveness in Australia: A Reassessment Using Selection on Observed and Unobserved Variables*

Authors: Buly Cardak[+]
Joe Vecci[#]

La Trobe University: School of Economics

Working Paper 5 of 2013
Date : June 2013

ISSN: 1837-2198
ISBN:

Catholic School Effectiveness in Australia: A Reassessment Using Selection on Observed and Unobserved Variables*

Buly A. Cardak+ and Joe Vecci#

June 2013

Abstract

This paper provides new estimates of the effect of Catholic school attendance on high school completion and university commencement and completion for Australian students. First, an instrumental variables approach is adopted where the probability of Catholic affiliation is used as an instrument. Consistent with the recent US literature, results based on this instrument are mixed. Instead, bounds are placed on the Catholic school effect using the assumption of equality between selection on observables and unobservables. The effect of Catholic school attendance is found to be smaller than previous results and negative treatment effects cannot be ruled out. Recent improvements in public school outcomes may have contributed to the smaller Catholic school effects.

Keywords: Catholic Schools; High School Completion; University Attendance; Selection Bias;

JEL Codes: I21

---

* We thank Jeff Borland, David Johnston and Sandy Suardi for helpful comments and suggestions. We are very grateful to Chris Ryan for helpful discussions, comments and suggestions and thank Todd Elder for helpful advice with the estimation. Cardak acknowledges support from research grants provided by the Australian Research Council (DP0662909) and the Faculty of Business Economics and Law at La Trobe University.

+ Corresponding author: School of Economics, La Trobe University, 3086, Victoria, Australia. Ph:+61 3 9479 3419, Fax: +61 3 9479 1654, Email: b.cardak@latrobe.edu.au.

School of Economics, Monash University, 3800, Victoria, Australia. Email: joseph.vecci@monash.edu.au.

Catholic School Effectiveness in Australia: A Reassessment Using Selection on Observed and Unobserved Variables

June 2013

Abstract

This paper provides new estimates of the effect of Catholic school attendance on high school completion and university commencement and completion for Australian students. First, an instrumental variables approach is adopted where the probability of Catholic affiliation is used as an instrument. Consistent with the recent US literature, results based on this instrument are mixed. Instead, bounds are placed on the Catholic school effect using the assumption of equality between selection on observables and unobservables. The effect of Catholic school attendance is found to be smaller than previous results and negative treatment effects cannot be ruled out. Recent improvements in public school outcomes may have contributed to the smaller Catholic school effects.

Keywords: Catholic Schools; High School Completion; University Attendance; Selection Bias;

JEL Codes: I21

Introduction

This paper provides new estimates of the effect of Catholic school attendance on high school completion and university commencement and completion for Australian students. A probability of Catholic affiliation variable is constructed and used as an instrument to correct for selection bias, however, consistent with the recent US literature, results based on this instrument are mixed. The Catholic school effect is instead bounded using the assumption of equality between selection on observables and unobservables, providing a range for the Catholic school effect under different assumptions about the degree of selection into Catholic schools.

Catholic schools are a large provider of education services in many countries around the world. In the US, around 2.16 million students or 4% of elementary and secondary students attended Catholic schools in 2009; see Snyder and Dillow (2012). In Australia, around 750,000 students or 20.4% of elementary and secondary students attended Catholic schools in 2011; Australian Bureau of Statistics (2012). Similar to the UK, Ireland, Netherlands and many other European countries, Australian private schools, including Catholic schools, are the beneficiaries of government support based on student enrolments. Understanding the effectiveness of Catholic schools in Australia can contribute to our understanding of (i) effects at the margin when a larger proportion of the student cohort is enrolled in such schools and (ii) the effects of large scale public subsidizes directed to privately provided education. This knowledge can add valuable information to the debate on the subsidization of non-government schools and vouchers around the world.

The effect of Catholic school attendance on students in the US and Australia has been the subject of much research since the important work of Coleman *et al*. (1982) and Coleman and Hoffer (1987). While large benefits of Catholic school attendance have been found for both countries, one of the major concerns has been the possibility that students do not randomly select into Catholic schools, often referred to as selection bias. In order to correct for selection bias, an instrument or set of instrumental variables is typically required to identify the true Catholic school effect. However, finding valid instruments for the Catholic school setting has proved challenging. Catholic religious affiliation has been used, Evans and Schwab (1995), as have measures of the local availability of Catholic schools, Neal (1997). Serious concerns about the validity of such instruments are raised by Altonji *et al*. (2005b) (hereafter AET (2005b)) along with Cohen-Zada and Elder (2009), Figlio and Stone (1999) and Neal (1997). In related work, Altonji *et al*. (2005a) (hereafter AET (2005a)) propose an alternative approach to estimating Catholic school effects. The AET (2005a) approach can, under some conditions, be used to place bounds on the Catholic school effect through sensitivity analysis and by placing restrictions on the magnitude of the effect of unobservables relative to the effect of observables. A brief summary is provided in Section 3 below.

Following previous studies, Koutroumanes Hofrenning and Chiswick (1999) and Le and Miller (2003), we construct a probability of Catholic affiliation variable based on census data and ethnic ancestry. This variable is used as an instrument in an attempt to identify the Catholic school effect but similar to Cohen-Zada and Elder (2009), the Catholic school effect is found to be implausibly large, and much larger than the corresponding single equation Probit estimates. We are also concerned the instrument is correlated with unobservables suggesting the instrument is not excludable. These results and the absence of other instruments lead us to adopt the approach proposed in AET (2005a). Based on the estimated bounds, the marginal effects of Catholic school attendance, relative to public school attendance, range between -4.76% and 5.42% for high school completion, -3.47% and 6.23% for university commencement, and -4.79% and 7.04% for university completion.

While households choose to attend Catholic schools for many different reasons, we interpret this set of results as evidence that for Australia, the Catholic school effect on the outcome variables studied appears to be much lower than previously believed. Estimates of the marginal effects of Catholic school attendance on high school completion of 18% and 13% are provided by Vella (1999) and Le and Miller (2003) respectively. One explanation for the lower marginal effects is that our analysis includes education aspirations and expectations data not available in previous cohorts and studies. These additional variables pick up effects that seem to have been attributed to the Catholic school effect in past studies. As these results are based on earlier cohorts, our lower estimates may also be related to the changes in the school sector and education participation. Since these earlier studies, school completion rates have increased markedly, tripling (to 71%) in public schools and doubling (to 85%) for Catholic schools. In addition, Catholic school enrolments have grown by 13% over the relevant period while public school enrolments have grown by only 1.5%.

One of the earlier analyses of Catholic school effects on comparable outcome variables for the US was provided by Evans and Schwab (1995) who found a 13% marginal effect of Catholic school attendance on high school graduation and college commencement. Strong effects are also found by Neal (1997) and Figlio and Stone (1999) but these effects are primarily among urban students, particularly urban minorities. AET (2005a) find the marginal effect on high school graduation to range between 3% and 5% while the marginal effect on college attendance was found to range between 3% and 13%. Using the historical concentration of Catholics in counties as an instrument, Cohen-Zada and Elder (2009) find a 3% marginal effect on high school graduation and an 8% marginal effect on college attendance. Our upper bound estimates of the marginal effect on high school graduation and university commencement in Australia are similar to these more recent US results. However, our lower bound estimates are negative and quite a bit lower than the US results. While there are many differences between education systems in the US and Australia, the difference in the proportion of students attending Catholic schools may be part of the explanation, suggesting that as the Catholic school sector expands, marginal effects may decline.

The paper proceeds with a description of the data in Section 2. The empirical methods to be used are described in Section 3. Estimation results are presented in Section 4. This includes (i) estimates of the Catholic school effect from a single equation Probit, (ii) our analysis of the Catholic ancestry instrument and Catholic school effects from a bivariate Probit, and (iii) our estimated bounds on the Catholic school effect. Section 5 provides conclusions.

Data

This paper uses data from the 1998 cohort of the Longitudinal Surveys of Australian Youth (LSAY98). The 1998 cohort is based on a stratified sample of students that is nationally representative of school students in year 9 (9th grade) in 1998. The sample is stratified in each state by the three broad school sectors; Catholic, Government and Independent. The sample was chosen in two stages: first, within strata, schools were selected proportional to size; and second, within these schools, two Year 9 classes were randomly selected, again proportional to size, see Long and Fleming (2002) for more details. The result is a commencement sample of 14,117 students. Our analysis is conducted using weighted data to ensure the distribution of strata matches that of the population due to over sampling of certain demographics. The weights used are designed to also account for survey attrition; see LSAY (2012) for details on the weights and Rothman (2009) and Ryan (2011) for a discussion of attrition bias in this data.

Data from waves one (1998) to eleven (2008) are studied in this paper. As part of the first wave, student literacy and numeracy was tested, and students completed a questionnaire providing a range of background information about themselves and their families. Subsequent waves were conducted through telephone interviews (participants answered a mail questionnaire for the second wave), collecting information on post secondary school studies, labor market activities, social and community activities and further background characteristics.

Table 1: Means of variables in Catholic and public school sector, LSAY98

Demographics	Catholic	Public	Difference
Male	0.5526	0.4877	0.0649*
Aboriginal	0.0231	0.0511	-0.0279
Family Background			
No. of Siblings	2.0390	2.0920	-0.0528**
Mother born in Australia	0.7030	0.7071	-0.0040
Father born in Australia	0.6622	0.6797	-0.0175*
Student born in Australia	0.9209	0.8864	0.0345*
English Speaking Father	0.7574	0.8038	-0.0464*
English Speaking Mother	0.7873	0.8181	-0.0307*
Father completed HS	0.5923	0.4766	0.1156*
Mother completed HS	0.5998	0.4976	0.1021*
Father completed p. secondary	0.2540	0.1569	0.0971*
Mother completed p. secondary	0.2301	0.1532	0.0769*
SES Mother	38.6917	33.6367	5.0550*
SES Father	41.1523	34.1168	7.0354*
Geography			
Metropolitan area	0.6918	0.5108	0.1810*
Expectations			
Parent plans ft study	0.3325	0.2489	0.0835*
Student plans ft work	0.1913	0.2625	-0.0711*
Motivation	10.3129	9.8306	0.4823*
Effort	5.2508	5.1572	0.0936*
Ability	10.7703	10.4943	0.2759*
Repeated Primary School	0.0491	0.0567	-0.0076**
Outcomes			
Bachelor degree or higher	0.3222	0.2138	0.1083*
Commence University	0.5682	0.3810	0.1871*
Completed High School	0.8581	0.7141	0.1439*
Reading	10.8080	9.5010	1.3069*
Maths	10.4409	9.6031	0.8378*

Notes: (1) * and ** denote the difference in means is statistically significant at the 1% and 5% level respectively.
(2) The number of observations in the Catholic (*N*=2162-2989) and public (*N*=5591-8699) school sub-samples are based on the full commencing sample for all variables including maths and reading which were tested in the first wave. Outcome variables other than maths and reading are collected in subsequent waves and are therefore subject to attrition. Outcome variables for Catholic/public schools are based on the following number of observations: Bachelor degree (1,862/4,821), commence university (1,862/4,821) and completing high school (1,875/4,860). These summary statistics are computed without weights.

Our analysis focuses on whether respondents complete high school, commence university or complete university. The first two outcomes were surveyed in the fifth and subsequent waves while information on the latter outcome is taken from wave eight onwards. Students are awarded a High School Certificate (HSC) by completing secondary school and this is our measure of high school completion. Those students observed to drop out of school and who do not obtain a HSC by wave 9 are treated as not having completed high school. In addition students that did not complete high school and consequently never attempted university are treated as having failed to commence or complete university studies. These outcome variables are deliberately defined in the same way as the outcome variables studied in Evans and Schwab (1995) and AET (2005a) to ensure comparability.

Summary statistics of the LSAY98 data are provided in Table 1 with an explanation of variable definitions provided in Appendix A. On all outcome measures presented, average outcomes of Catholic school students are superior to their public school counterparts with the differences statistically significant at the 1% level. Catholic school students are on average around 50% more likely to commence and complete a university qualification. However, consistent with much of the literature on Catholic school effects, many student and family characteristics are found to favor Catholic school students outperforming their public school counterparts. The average educational attainment and socio-economic status of Catholic school parents is higher for mothers and fathers, while educational aspirations and motivation are also higher among Catholic school students. Similar results were found in Neal (1997), Evans and Schwab (1995) and AET (2005a). These summary statistics suggest some positive selection into Catholic schools based on observables. Controlling for this selection on observables in order to present a truer reflection of Catholic school effects is straightforward. However, what is more problematic is dealing with any possible selection on the basis of unobservables. We now turn to a discussion of the ways in which we can control for observables and unobservables when trying to identify Catholic school effects on educational outcomes.

Method

In this section we outline the widely used empirical approach to assessing the effectiveness of Catholic schools. In our context this approach ranges from single equation Probit models to bivariate Probit models with an exclusion restriction to control for selection bias. We then explain the approaches proposed by AET (2005a) to deal with the absence of reliable instruments by placing bounds on the effects of Catholic schools.

Our focus is on the effects of Catholic school attendance on high school completion, university participation and completion. Given the binary nature of these outcomes, we estimate the effects of Catholic school attendance using a Probit model. The naïve specification takes the form

$$outcome = 1\left(X^{'}\beta_0 + \alpha_0 C + u_0 > 0 \right), \qquad (1)$$

where $outcome$ is the probability of an educational outcome (e.g. high school completion), X is a vector of individual and household characteristics expected to influence $outcome$, C is an indicator variable of whether the student attended a Catholic (C=1) or Public (C=0) High School and u_0 is an independent disturbance term.

This model is used in the first instance to establish a baseline effect of Catholic School attendance. However, as is well understood in the literature, estimating this single equation Probit model is likely to provide estimates that suffer from selection bias. The selection problem is one of Catholic school attendance not being random within the sample. Either parents or schools or both may be selecting students into Catholic schools systematically, based on unobservable characteristics. The implication is that the effect of such unobservable characteristics on educational outcomes will be incorrectly attributed to Catholic school attendance.

The typical approach to addressing the problem of selection bias due to a binary endogenous explanatory variable is to estimate a bivariate Probit model of the form

$$C = 1\left(X^{'}\beta_1 + u_1 > 0 \right), \qquad (2)$$

$$outcome = 1\left(X^{'}\beta_2 + \alpha_2 C + u_2 > 0 \right), \qquad (3)$$

where u_1 and u_2 are the usual standard Normal disturbance terms with the correlation between disturbances given by $\rho_{1,2}$. Due to the effects of unobservable characteristics already noted, these two disturbances are not necessarily independent, $\rho_{1,2} \neq 0$.

The parameters of this model are in theory identified through the nonlinearities of the bivariate Probit model. However, results from the estimation of this model without any exclusion restrictions are not generally viewed as reliable; see AET (2005a) and Dujardin and Goffette-Nagot (2010).[1] It is typical to include an instrument or set of instruments in the following manner

[1] AET (2005b) show that much of the identification in bivariate Probit estimation of Catholic school effects seems to be due to the nonlinearities of the model rather than the exclusion of instruments.

$$C = 1\left(X^{'}\delta_1 + \gamma Z + \varepsilon_1 > 0 \right), \tag{4}$$

$$outcome = 1\left(X^{'}\delta_2 + \phi C + \varepsilon_2 > 0 \right), \tag{5}$$

where Z is either one or a vector of instrument(s) that affect the decision to attend a Catholic school but are uncorrelated with $outcome$. Examples of instruments used in the identification of this system include Catholic religious affiliation of the family, probability of Catholic affiliation based on the family's ethnic background, measures of ease of access and proximity to Catholic schools and historical shares of Catholics; see for example Evans and Schwab (1995), Neal (1997), Vella (1999), and Cohen Zada & Elder (2009), Cohen-Zada (2009), Koutroumanes Hofrenning and Chiswick (1999), Le and Miller (2003). Since the LSAY98 does not elicit religious affiliation of the student's family, we construct an instrument based on the ethnicity of parents and national data on population proportions of Catholic affiliations of different ethnic ancestries; see Koutroumanes Hofrenning and Chiswick (1999) and Le and Miller (2003). The resulting instrument provides a probability of Catholic affiliation. More details about the instrument, its construction and assessment of its validity are provided in the next section.

Notwithstanding the numerous studies that use the range of instruments based on Catholic religious affiliation, finding reliable instruments to deal with the selection bias in the decision to attend a Catholic school is problematic as the instrument is often found to be correlated with unobservables. Concerns about the reliability of instruments based on Catholic religious affiliation, Neal (1997), Figlio and Stone (1999), AET (2005b), Cohen-Zada (2009) and Cohen-Zada and Elder (2009), and the lack of other potential instrumental variables leads us to employ the AET (2005a) "selection on observed and unobserved variables" technique to place bounds on the actual treatment effect of Catholic schools and to verify the reliability of the instrument employed.

The AET (2005a) approach can be used to estimate the effects of selection bias when valid instruments are unavailable. It involves three broad approaches to placing credible bounds on a treatment effect in the presence of selection bias. The first is to estimate the system in equations (2) and (3) under the assumption that the correlation between the disturbance terms, $\rho_{1,2}$, is not identified. As a consequence, a constrained model is estimated for different exogenously chosen values of $\rho_{1,2}$. The results are presented for a range of values from $\rho_{1,2} = 0$, the case of independence between the disturbance terms or no selection bias, to $\rho_{1,2} = 0.5$ as it is typically believed that there is a positive selection effect into Catholic schools implying a positive correlation. The estimates of the Catholic school effect, $\widehat{\alpha}_2$, provide a guide to the effects at different levels of endogeneity, informing us of the sensitivity of the estimated treatment effect to the perceived selection bias.

The second approach proposed by AET (2005a) involves the assumption that selection on the unobservables is equal in importance to selection on the observables. As outlined in AET (2005a), the assumption of equality between observables and unobservables requires that (i) the included covariates are randomly chosen from the full set of factors that determine *outcome*, (ii) the number of observed and unobserved factors are large and none dominate the distribution of *C* or *outcome*, and (iii) the effect of Catholic school attendance (*C*) has a similar relationship to the explained and unexplained components of *outcome*. The data to be used provides a relatively large number of control variables that are wider ranging than in previous data and research used to study Catholic school effects in Australia; see for example Vella (1999) which does not include information on aspiration variables. Given the set of included covariates, the assumption of equality between selection on observables and unobservables may be too strong and as pointed out in AET (2005a), estimates based on such an assumption are treated as a lower bound for the Catholic school effect.[2]

While the formal analysis and more details of the assumption of equal selection on observables and unobservables can be found in Section III of AET (2005a), the implementation of this assumption requires the following restriction be incorporated into the estimation of the bivariate Probit model in equations (2) and (3),

$$\rho_{12} = \frac{Cov\left(X^{'}\beta_1, X^{'}\beta_2\right)}{Var\left(X^{'}\beta_2\right)}. \qquad (6)$$

This restriction acts as an indentifying assumption and allows the model in (2) and (3) to be estimated without exclusion restrictions. As already noted, the resulting estimates of the treatment effect can be treated as a lower bound on the effects of Catholic school attendance on the outcome under study. In addition to this lower bound, an upper bound on the treatment effect is identified from the estimation of equations (2) and (3) under the restriction that $\rho_{1,2} = 0$, as outlined above.

The third and again related approach proposed by AET (2005a) considers the possibility that the whole treatment effect is driven by selection on the unobservables. This sensitivity

[2] One could take a more or less conservative approach to this lower bound by assuming unobservables are respectively greater or smaller than observables but we follow the assumption adopted by AET (2005a) and followed in the subsequent literature of equality between selection on observables and unobservables.

analysis is based on Condition 4 from AET (2005a), reproduced here as

$$\frac{E\left(u_2 \mid C=1\right)-E\left(u_2 \mid C=0\right)}{Var\left(u_2\right)}=\frac{E\left(X^{'}\beta_2 \mid C=1\right)-E\left(X^{'}\beta_2 \mid C=0\right)}{Var\left(X^{'}\beta_2\right)}, \tag{7}$$

which is a formal statement of the assumption that the selection on unobservables is the same as the selection on observables in the system in equations (2) and (3). Following the analysis and discussion in AET (2005a) and Dujardin and Goffette-Nagot (2010), this condition can be used to develop the following ratio

$$\lambda=\hat{\alpha}_2\left(\frac{Var\left(\tilde{C}\right)}{Var\left(C\right)}\frac{Var\left(X^{'}\beta_2\right)}{\left[E\left(X^{'}\beta_2 \mid C=1\right)-E\left(X^{'}\beta_2 \mid C=0\right)\right]}\right), \tag{8}$$

where $X^{'}\beta_2$ and $\tilde{C}$ denote respectively the predicted values and residuals of a regression of C on X. The ratio λ tells us the relative magnitude of the role of unobservables to observables required for all of the estimated effect of Catholic school attendance to be attributable to unobservable characteristics. Values of $\lambda > 1$ imply that the role of unobservables needs to be greater than that of observables.

Results

Preliminary Probit Results

Table 2 presents the results of naïve Probit estimates of the effect of Catholic school attendance, relative to public school attendance, on high school completion and university commencement and completion. The first column presents results of a model where *X* has been excluded from equation (1) implying the only covariate is *C*. The results show a strong, statistically significant unconditional effect of Catholic school attendance on all outcomes of interest. However, controlling for a range of demographic and family background covariates (second column) lowers the magnitude of these marginal effects, though they remain highly statistically significant with a positive marginal effect ranging from 7% to 10%.

The next two columns add year 9 literacy and numeracy test scores (third column) and student and parental educational expectations and student's own perceptions of ability (fourth column). With this full set of controls, students attending a Catholic school have a statistically significant advantage as they are between 5% and 7% more likely to achieve the respective outcomes than a student at a public school.

Table 2: Results of Probit estimates, based on equation (1), of the effect of Catholic high school attendance on high school completion and university commencement and completion.

	Included Covariates[1]			
	None	Family background, residence and demographics	Col. 2 plus 9th Grade Tests	Col. 3 plus expectations
A. High School Completion				
Coefficient	0.4038	0.3045	0.2898	0.2534
Marginal effect	0.1186	0.0736	0.0662	0.0542
Z stat	7.73*	4.18*	3.95*	3.20*
Pseudo R^2	0.0122	0.0742	0.1188	0.1797
Observations				3327
B. University Commencement				
Coefficient	0.4692	0.2516	0.2275	0.2214
Marginal effect	0.1791	0.0878	0.0703	0.0623
Z stat	11.72*	4.52*	3.94*	3.58*
Pseudo R^2	0.0180	0.1167	0.2119	0.2793
Observations				3327
C. University Completion				
Coefficient	0.4508	0.2981	0.2896	0.2423
Marginal effect	0.1585	0.1016	0.0905	0.0704
Z stat	9.13*	4.44*	4.16*	3.31*
Pseudo R^2	0.0172	0.1000	0.1694	0.2235
Observations				2336

Notes: (1) Each column includes the Catholic high school attendance dummy variable as a covariate.
(2) * and ** denote significance at the 1% and 5% levels respectively.

These marginal effects are much lower than those estimated in earlier studies of Australian students by Vella (1999) who find 16-18% marginal effect of Catholic school attendance on high school completion and university commencement and Le and Miller (2003) who find a 12% marginal effect of Catholic school attendance.[3] These previous studies did not have available data on educational aspirations and expectations, and as can be seen from columns 3 and 4 of Table 2, these variables have some impact on the marginal effects and seem to pick up effects that have been attributed to the Catholic school effect in past studies. Since these previous studies, Australian school completion rates have tripled (to 71%) in public schools and doubled (to 85%) for Catholic schools. In addition, Catholic school enrolments have grown by 13% over the relevant period while public school enrolments have grown by only 1.5%. These decreasing differences between public school and Catholic school completion rates, along with relative growth in the Catholic sector might also be contributing to smaller marginal effects.

While these results point to strong benefits arising from Catholic school attendance even after conditioning on a wide range of covariates, we cannot rule out selection based on unobservables which may lead to biased estimates of the Catholic school effect. The typical response to such concerns is to estimate a bivariate Probit model, outlined in equations (4) and (5), with instruments to identify an unbiased estimate of the Catholic school effect. We attempt this exercise in the next subsection.

Instrumental Variable and Bivariate Probit Results

The bivariate Probit model outlined in equations (4) and (5) can be estimated in order to correct for bias arising from selection into Catholic schools based on unobservables. This typically involves the use of an instrument or set of instruments, as discussed above. Previous studies have used religious affiliation as an instrument based on the idea that Catholic religious affiliation is likely to influence the decision to attend a Catholic school but may be unlikely to be correlated with the educational outcomes of interest; Vella (1999) for Australia and Evans and Schwab (1995) in the United States. More recent work by AET (2005b) has shown that Catholic religious affiliation is not a valid instrument for US National Educational Longitudinal Survey of 1988 (NELS:88) and National Longitudinal Study of the High School Class of 1972 (NLS-72).

[3] This previous literature did not have access to expectations data and as a consequence comparisons to the third column of Table 2 may be more appropriate. The results of our preferred specification in column four of Table 2 show the inclusion of expectations data does not markedly change the marginal effects.

The LSAY questionnaires do not include questions about religious affiliation. Instead we construct a variable based on parental nationality and census data on the proportions of each nationality that are Catholic; see Koutroumanes Hofrenning and Chiswick (1999) for a discussion and an algorithm to construct the variable and Le and Miller (2003) for analysis of a similar instrument for earlier cohorts of Australian students born in the 1960's and 1970's. The rationale for the construction of this variable is that a parent's ancestry may be an important predictor of Catholic religious affiliation and therefore, attendance at a Catholic School.

The religious ancestry instrument is based on data about the country of birth of each parent, reported by respondents (students) as part of the first wave of LSAY98. Each parent's country of birth is then matched with census data on ancestry and religion from Australian Bureau of Statistics (2001). This data provides the proportion of Australians from each ethnic background who identify as Catholic, and with other religions, and is used to create a new variable measuring probability of Catholic affiliation for each parent.[4] The instrument is then created by multiplying father's and mother's ancestry variables, providing a continuous variable ranging from 0-1 where a value of 0 implies no probability of Catholic religious affiliation for that student's family and higher values imply a higher probability of Catholic religious affiliation.

In order to verify the validity of our instrument, we initially estimate the correlation between the Catholic ancestry instrument and catholic school attendance, $Corr(C,Z)$. A relatively high correlation is required for a valid instrument. It is found that $Corr(C,Z)=0.073$, which is quite low when compared to similar instruments used in the literature (see Evans and Schwab (1995) and Cameron and Trivedi (2005)) and could indicate an inconsistent estimator and ultimately a weak instrument.

As suggested by Angrist and Pischke (2009) and following common practice in the literature, we initially estimate the model in equations (4) and (5) ignoring the dichotomous nature of the dependant variables using two stage least squares (2SLS). The results of the 2SLS estimation for all three outcome variables are used to further test the validity of the Catholic ancestry instrument. Table 3 provides a summary of key diagnostics obtained from the first stage of the 2SLS estimation, equation (4), that are useful in identifying weak instruments.

[4] For example, a mother with New Zealand ancestry has a 16.3% probability of being Catholic in contrast to a mother with Maltese ancestry who has an 87% probability of being Catholic. For respondents who indicated a parent was born in Australia, the parent was allocated the Australian ancestry probability of Catholic religious affiliation of 0.246, also found in Australian Bureau of Statistics (2001).

Table 3: Diagnostic tests of validity of Catholic ancestry instrument based on 2SLS.

	High School Completion	University Commencement	University Completion
First stage R^2	0.0907	0.0907	0.0977
Partial R^2	0.0083	0.0083	0.0064
First Stage *F*-Statistic	20.74	20.74	10.0513
Sample Size	*N*=3327	*N*=3327	*N*=2336

Notes: (1) As the set of covariates are identical for all three models, the high school completion and university commencement equations are identical but the university completion equation differs because of sample attrition.

The first statistic is the R^2 from the estimation of this equation for each outcome variable. These are all around 0.09, indicating that the instrument contributes very little to the fit of the model and that there maybe be a substantial loss of identification due to the instrumental variables estimation (Cameron and Trivedi 2010, page 198). The partial R^2 provides the goodness of fit between Catholic school attendance, *C*, and the instrument, *Z*, after controlling for the explanatory variables which helps to isolate the explanatory power of the instrument in explaining why students attend a Catholic school. At between 0.006-0.008 this is very low and could mean the ancestry instrument performs poorly in explaining students' attendance at a Catholic school.

The first stage *F*-Statistic tests the significance of the instrument, *Z*, in the first stage equation. As we are using robust standard errors and the model is just identified, we can compare this statistic with the critical values of Stock and Yogo (2005). If we assume only a 5% relative bias tolerance so that the true size can be at most 10% we can reject the null hypothesis that the instrument is weak if the *F*-statistic exceeds 16.38. By this criterion, the instrument is considered a weak instrument only for university completion. We can reject the null of a weak instrument for high school completion and university commencement. However, these results are marginal and suggest the need for caution when using the Catholic ancestry instrument as it seems unlikely to have strong explanatory power of Catholic school attendance. While we have some reservations about the quality of our instrument, we nonetheless estimate our models using it, keeping in mind that we intend to conduct analysis along the lines of AET (2005a) to deal with the lack of good quality instruments.

Results of estimating the bivariate Probit model in equations (4) and (5) where the outcomes include high school completion and university commencement and completion, with Catholic ancestry used as an instrument are presented in columns A, B and C of Table 4 respectively. The marginal effects of the included covariates are broadly consistent with existing findings in the literature though we are surprised by the negative effect of an English speaking mother on high school and university completion. Many of the expectations variables are significant and the signs of the marginal effects are consistent with intuition.

The Catholic school attendance variable is no longer significant for university attendance and completion, implying no effect, while the marginal effect in the high school completion equation is significant at the 1% level and is much higher (0.1669) than the marginal effect in the last column of panel A of Table 2 (0.0542).[5]

More important however, is the fact that the estimated correlation between the two disturbance terms in each bivariate model is not statistically significant. It is surprising that these correlations, while not statistically significant, point to a negative correlation as it is commonly believed that there is a positive selection into Catholic schools. These results are similar in nature to those of Vella (1999) where it is found that Catholic school attendance is exogenous and single equation results can be used. Assuming we have a valid instrument, the results imply the null that each of the three outcome variables and Catholic high school attendance are independently determined cannot be rejected. The implication is that there does not appear to be a problem with selection and that we can rely on the naïve Probit estimates presented in Table 2 as a genuine measure of the effects of Catholic school attendance on these outcomes.

The results in Table 4 show the Catholic ancestry instrument is significantly correlated with Catholic school attendance and therefore satisfies one of the requirements for a valid instrument. In order to provide some evidence about the correlation between the instrument and the disturbance term, ε_2, in equation (5) we follow AET (2005b) and compute additional summary statistics. Table 9 in Appendix B reports the relationship between the full set of control variables used in this paper and the Catholic ancestry instrument. Column 2 (column 3) provides the mean of these variables when the probability of Catholic ancestry is greater

[5]As discussed in AET (2005b) and noted above, the identification observed in the model estimating high school completion is due to the nonlinearity inherent in the functional form of the bivariate Probit rather than the efficacy of the included instrument. They also point out that this is more likely when there is only one included instrument.

Table 4 Results of bivariate Probit regression of Catholic school attendance on three outcome variables using Catholic ancestry instrument, estimated model given by equations (4) and (5).

	A. High School Completion		**B. University Commencement**		**C. University Completion**	
	Marginal Effects	*Z* stat	Marginal Effects	*Z* stat	Marginal Effects	*Z* stat
Demographics						
Male	-0.0681	-4.52*	-0.1113	-6.84*	-0.1471	-7.02 *
Aboriginal or Torres Strait Islander	-0.0600	-0.94	0.0507	0.63	0.1626	1.36
Family Background						
No. of Siblings	-0.0109	-1.81***	-0.0195	-3.15*	-0.0287	-3.61*
Mother born in Australia	0.0274	1.13	0.0667	2.09**	0.0632	1.97 **
Father born in Australia	0.0078	0.32	-0.0243	-0.90	0.0347	0.72
Student born in Australia	-0.0375	-0.98	-0.0445	-1.44	-0.0945	-2.04**
English Speaking Father	-0.0203	-0.55	-0.0095	-0.26	-0.0672	-1.59
English Speaking Mother	-0.0885	-2.18**	-0.0960	-2.18**	-0.0522	-1.30
Father completed HS	0.0270	1.65***	0.0250	1.37	0.0274	0.91
Mother completed HS	-0.0067	-0.43	0.0185	0.99	-0.0007	-0.01
Fatherpost secondary	0.0247	1.08	0.0603	2.34*	0.0051	0.00
Motherpost secondary	-0.0042	0.20	0.0263	1.18	0.0048	-0.05
SES Mother	0.0013	2.87*	0.0014	2.39*	0.0019	3.27 *
SES Father	0.0002	0.62	0.0011	1.58***	0.0016	2.66 *
Geography						
Residential Location	-0.0071	-0.32	-0.0310	-0.75	-0.0098	0.08
Expectations						
Parent plans ft study	0.0469	2.65*	0.1165	5.46*	0.0891	4.29 *
Student plans ft work	-0.0643	-3.29*	-0.0929	-2.75*	-0.0848	- 1.87 **
Motivation	-0.0010	-0.19	0.0083	0.85	0.0161	1.87 **
Effort	0.0213	2.60*	0.0011	0.12	0.0173	1.80 **
Ability	0.0265	6.60*	0.0378	7.59*	0.0273	4.47*
Repeated Primary School	-0.0594	-1.75***	-0.0393	-0.87	-0.0864	-1.29
Outcomes						
Test scores	0.0331	4.23*	0.0832	5.64*	0.0798	7.32*
Catholic School Attendance	0.1669	2.13*	0.2409	1.07	0.1208	1.08
Catholic Ancestry Instrument[(2)]	0.3816	4.47*	0.4084	4.52*	0.3442	3.13*
$\rho_{12}\left(\chi^2\right)$	-0.3079	(1.2787)	-0.3935	(0.4999)	-0.1283	(0.2708)

Notes: (1) *, ** and *** denote significance at the 1%, 5% and 10% levels respectively.
(2) The Catholic ancestry instrument results come from the estimation of the school choice equation, given by equation (4).

(less) than the median value of *Z*. Among expectations, with the exception of parental expectations which are higher among those less likely to be Catholic, most means are very similar. There is also a negative relationship between higher Catholic ancestry and living in a metropolitian area. While important family background characteristics such as parent high school completion and SES point to a positive effect on those less likely to have Catholic ancestry. Similar to AET (2005b), systematic differences in background characteristics suggest the possibility that there may also be unmeasured differences which could affect outcomes and lead to bias in models using the Catholic ancestry instrument. Coupled with doubts from instrument checks in Table 3 and further doubts from the existing literature (AET (2005b), Figlio and Stone (1999) and Neal (1997)) about the efficacy of Catholic religion variables as instruments in such settings, we turn to the alternative approach proposed by AET (2005a) and outlined above.

Selection on Observables and Unoberservables: Bounding the Catholic School Effect

As outlined above, faced with poor information about the exogeneity of Catholic religious affiliation as an instrument, AET (2005a) hypothesize that it may be possible to use the relationship between an endogenous variable and observable characteristics to make inferences about the relationship between the endogenous variable, observed variables and the unobservables. Details of the three elements of their approach are provided above. Used together these three elements can provide useful inferences about the Catholic school effect.

The first element is a sensitivity analysis, the results of which are provided in Table 5. These results provide information on the strength of the Catholic school effect under the assumption of different degrees of selection on unobservables, the stronger the correlation between the residuals, the stronger the selection effect. The sensitivity analysis is carried out for all three outcomes being studied. The marginal effect of Catholic school attendance on high school completion when it is assumed $\rho_{1,2} = 0$ is 5.42%. The marginal effect decreases to 1.74% when we assume $\rho_{1,2} = 0.1$ and becomes negative for stronger positive correlations, decreasing to -14.78% with $\rho_{1,2} = 0.5$. This implies a small amount of correlation between the unobservables affecting Catholic attendance and high school completion can explain the Catholic attendance effect on high school completion. Results are very similar for university commencement and completion with the marginal effects becoming negative when we assume $\rho_{1,2} = 0.2$.

While difficult to interpret definitively, these results suggest a small amount of selection on unobservables could explain much of the Catholic school effects identified in Table 2. Drawing stronger inferences from these results is problematic because it is difficult to judge the true value of the (unobservable) correlation, $\rho_{1,2}$. However, as outlined above, it is possible to use the degree of selection on observables as a guide to the degree of selection on unobservables. This is the second element of the AET (2005a) approach and the results are presented in Table 6.

The first row of Table 6 presents estimates of the model in equations (2) and (3), where the constraint in equation (6) has been imposed on the estimation. This imposes the assumption that the selection on unobservables is of the same magnitude as the selection on observables.

Table 5 Estimates of the effect of Catholic high school attendance given different assumptions about the correlation of disturbances in estimated bivariate Probit model.

	Correlation of Disturbances, $\rho_{1,2}$					
	0.0	0.1	0.2	0.3	0.4	0.5
A. High School Completion	0.2534 (0.0792)* [0.0542]	0.0807 (0.0789) [0.0174]	-0.0959 (0.0782) [-0.0206]	-0.2767 (0.0769)* [-0.0605]	-0.4619 (0.0751)* [-0.1026]	-0.6528 (0.0727)* [-0.1478]
B. Commence University	0.2214 (0.0618)* [0.0623]	0.0517 (0.0616) [0.0143]	-0.1189 (0.0609)** [-0.0340]	-0.2905 (0.0599)* [-0.0830]	-0.4636 (0.0585)* [-0.1325]	-0.6362 (0.0566)* [-0.1827]
C. University Completion	0.2423 (0.0742) [0.0704]	0.0684 (0.0738) [0.0213]	-0.1014 (0.0730) [-0.0284]	-0.2716 (0.0718)* [-0.0786]	-0.4380 (0.07024)* [-0.1291]	-0.6083 (0.0678)* [-0.1797]

Notes: (1) Standard errors in parentheses, marginal effects in brackets.
(2) * and ** denote significance at the 1% and 5% levels respectively.

Given a relatively comprehensive set of observables, this provides a plausible lower bound estimate on the Catholic school effect. The results include the coefficients and implied correlations, along with standard errors in parentheses and marginal effects in brackets. The parameter estimates are negative and significant at the 5% level or better for all three outcome variables, while the 95% confidence intervals for each of the three outcomes exclude zero. The marginal effects are negative in all cases, -4.76% for high school completion, -3.47% for university commencement and -4.79% for university completion. As can be seen from the estimated correlations which range between 0.20 and 0.27, these results are consistent with those in Table 5 and suggest that correlations of this magnitude imply the role of unobservables matches that of observables. If this were indeed the case, the results suggest that Catholic school student outperformance is driven by unobservables and that the true effect of Catholic schools is indeed negative.

Table 6 Estimates of the Catholic school effect on outcome variables under the assumption (i) that selection of observables is equal to unobservables and (ii) of independent disturbances.

	High School Completion		University Commencement		University Completion	
Constraint	$\rho_{1,2}$	$\hat{\alpha}_2$	$\rho_{1,2}$	$\hat{\alpha}_2$	$\rho_{1,2}$	$\hat{\alpha}_2$
(i) $\rho_{1,2}$: Unobservables = Observables	0.2682	-0.2187* (0.0774) [-0.0476]	0.2012	-0.1224** (0.0609) [-0.0347]	0.2434	-0.1631* (0.0729) [-0.0479]
(ii) $\rho_{1,2} = 0$	0	0.2534* (0.0792) [0.0542]	0	0.2214* (0.0618) [0.0623]	0	0.2432* (0.0742) [0.0704]

Notes: (1) Standard errors in parentheses, marginal effects in brackets.
(2) * and ** denote significance at the 1% and 5% levels respectively.

The second row of Table 6 presents estimates of the model in equations (2) and (3) under the assumption $\rho_{1,2} = 0$. The estimates under this independence assumption are identical to the naïve Probit estimates of Table 2 and are considered the upper bound for the Catholic school effect on the outcomes studied. Results include the coefficients with standard errors in parentheses and marginal effects in brackets. These upper bounds are positive and statistically significant at the 1% level in all cases. The marginal effects are 5.42% for high school completion, 6.23% and 7.04% for university commencement and completion respectively.

These results tell us the Catholic school effect could be negative if the selection into Catholic schools based on unobservables is sufficiently strong. Conversely, if there is no selection into Catholic schools based on unobservables, the Catholic school effect is in the 5-7% range. One interesting aspect of these findings is that the estimated range for the Catholic school effect includes zero.

The third element of the AET (2005a) approach seeks to identify the relative role of unobservables required in order for the Catholic school effect to indeed be zero, thereby providing further guidance as to the confidence we should have in positive estimates of the Catholic school effect. Estimates of the ratio λ, given by equation (8), are provided for each of the three outcomes being studied in Table 7. The values of λ are 0.912 for high school completion, 1.167 for university commencement and 1.063 for university completion. These ratios imply that in order for the whole estimated Catholic school effect ($\hat{\alpha}_2$) to be attributed to unobservables, the role of unobservables would need to be slightly smaller (slightly larger) than the role of observables for high school completion (university commencement and completion).

Table 7 Estimates of amount of selection on unobservables relative to selection on observables required to attribute the entire Catholic school effect to selection bias

Outcome	$\hat{\alpha}_2$	Bias	λ
High School Completion	0.2534* (0.0781)	0.2777* (0.0675)	0.912
Commence University	0.2214* (0.0620)	0.1896* (0.0219)	1.167
University Completion	0.2423* (0.0704)	0.2237* (0.0414)	1.083

Notes: (1) Standard errors in parentheses.

(2)* denotes significance at the 1% level.

For comparison, the equivalent ratios found in AET (2005a) are 3.55 for high school graduation and 1.43 for college attendance. It is highly unlikely that the role of unobservables could be 3.55 times that of observables as the relevant Psuedo $R^2 = 0.58$. In our study, the Pseudo R^2 = 0.18, 0.28 and 0.22 for high school completion, university enrollment and completion respectively. While, these results suggest that it is possible that unobservables could be as important as observables in explaining the respective outcomes, the relatively strong explanatory power of the observables here leads us to believe it is unlikely that *all* of the Catholic school effect could be attributable to unobservables.

Summarizing these results, Catholic school marginal effects range between -4.76% and

5.42% for high school completion, -3.47% and 6.23% for university commencement and -4.79% and 7.04% for university completion. These results imply that the Catholic school effect is at best, positive but lower than previously believed for Australian students. It is possible that the Catholic school effect is zero or even negative but such findings depend on assumptions that imply large unobservable factors are behind the Catholic school effects presented in Table 2.

Conclusion

We provide new estimates of the effects of Catholic high school attendance on high school completion and university commencement and completion for Australian students. These estimates are based on the approach suggested by AET (2005a) to deal with a lack of instrumental variables. While an instrument based on Catholic ancestry was constructed to deal with the typical concerns about selection bias when estimating Catholic school effects, results based on this instrument are mixed. Instead, the AET (2005a) approach was used to provide plausible upper and lower bounds for estimates of the treatment effect of Catholic school attendance.

The evidence suggests that the marginal effect of Catholic school attendance on high school completion ranges from -4.76% to 5.42%. Similar marginal effects were identified for the effects of Catholic school attendance on university commencement (-3.47% to 6.23%) and completion (-4.79% to 7.04%). If the Catholic school effect is in fact negative, the results imply that on average, Catholic school students would achieve better outcomes in public schools. Notwithstanding the relatively large government subsidies directed to Catholic schools in Australia, there is a common perception of low levels of resources, consistent with the relatively low tuition fees. This evidence of possible negative Catholic school effects might lead to calls for increased public resources to be directed to Catholic schools in order to improve outcomes.

The importance of our results is that we identify a smaller Catholic school effect for Australian students than previous studies. Using data on a cohort that attended high school 13 years earlier than LSAY98, Vella (1999) found the marginal effect on high school completion to be 18%, around 3 times as large as our upper bound estimate. Our estimate is less than half of the 13% Catholic school marginal effect on high school completion found for an earlier (1961) cohort studied by Le and Miller (2003). Two factors allow us to reconcile these results. First, our analysis includes a set of education aspiration and expectation variables which are not available in the surveys used in previous studies. These variables appear to pick up some unobservable effects which seem to have been attributed to Catholic school effects in past studies. Second, there has been dramatic change in the schooling landscape in Australia between the 1985 cohort studied by Vella (1999) and the LSAY 1998 cohort. Between the two cohorts, Catholic school enrolments grew by 13% while public school enrolments grew by only 1.5%; ABS (1985, 2011). In addition, high school completion rates have risen from 26% (43%) in public (Catholic) schools for the 1985 Cohort to 71% (86%) in public (Catholic) schools for the 1998 LSAY Cohort. These changes suggest smaller differences in high school completion rates and as a result, the possibility of smaller Catholic school effects on high school completion.

These estimated Catholic school effects are at the lower end of results for the US but quite similar to the more recent estimates from AET (2005a) and Cohen-Zada and Elder (2009). An important difference is that the lower bounds on the Catholic school effect are negative. This implies we cannot rule out that all of the Catholic school effect may be due to selection on unobservables. This is in contrast to the positive lower bound estimates provided by AET (2005a). While many differences between Australia and the US exist, one possible reason for this difference may be the relative magnitude of the Catholic school sector in the two countries. In Australia, almost 20% of students attend Catholic schools while in the US only 4% of students attend Catholic schools. This large Catholic school sector in Australia is driven at least in part by per student government subsidies in the order of 50-60% of recurrent public school expenditures per student. While it is notable that Australian Catholic schools may exhibit positive marginal effects even when the sector is five times as large as in the US, negative marginal effects cannot be ruled out. Further analysis at the school and regional level may shed more light on relative performance of public and Catholic schools. This will require more detailed data about schools and their local competition which is not available in LSAY98. Such analysis is left for future work should the required data become available.

References

Altonji, J.G., Elder, T.E. and C.R. Taber, (2005a), “Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools”, *Journal of Political Economy*, 113(1), 151-184.

Altonji, J.G., Elder, T.E. and C.R. Taber, (2005b), “An evaluation of instrumental variable strategies for estimating the effects of Catholic schools”, *Journal of Human Resources* 40, 791–821.

Angrist, J. and J.S. Pischke, (2009), *Mostly harmless econometrics: an empiricist’s companion*, Princeton University Press.

Australian Bureau of Statistics, (1986), *National Schools Statistics Collection Australia1985*, Cat. no. 4221.0, ABS, Canberra.

Australian Bureau of Statistics, (2001), *Australian Census Analytic Program: Australians' Ancestries (Corrigendum)*, Cat. no. 2054.0, ABS, Canberra.

Australian Bureau of Statistics, (2012), *Schools, Australia*, Cat. no. 4221.0, ABS, Canberra.

Cameron , A.C. and P.K. Trivedi, (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press, New York, NY.

Cameron, A.C. and P.K. Trivedi, (2010), *Microeconometrics Using Stata*, Revised Edition, Stata Press: College Station, Texas.

Cohen-Zada, D., (2009), “An alternative instrument for private school competition”, *Economics of Education Review* 28(1), 29-37.

Cohen-Zada, D. and T. Elder, (2009), “Historical religious concentrations and the effects of catholic schooling”, *Journal of Urban Economics* 66(1), 65-74.

Cohen-Zada, D. and W. Sander, (2008), “Religion, religiosity and private school choice: Implications for estimating the effectiveness of private schools”, *Journal of Urban Economics* 64(1), 85-100.

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

Coleman, J S, T Hoffer, and S Kilgore, (1982), *High School Achievement: Public, Catholic, and Private Schools Compare,* Basic Books, New York, NY.

Dujardin, C. and F. Goffette-Nagot, (2010), “Neighborhood effects on unemployment? A test à la Altonji”, *Regional Science and Urban Economics* 40, 380-396.

Evans, R. and R. Schwab, (1995), “Finishing high school and starting college: do Catholic schools make a difference?”, *Quarterly Journal of Economics* 110, 947–974.

Figlio, D.N. and J.A. Stone, (1999), “Are private schools really better?”, *Research in Labor Economics* 18, 115-140.

Koutroumanes Hofrenning, S. and B.R. Chiswick, (1999), “A method for proxying a respondent’s religious background: An application to school choice decisions”, *Journal of Human Resources* 34, 193–207.

Le, A. and P. Miller, (2003), “Choice of school in Australia: Determinants and consequences”, *Australian Economic Review* 36(1), 55–78.

Long, M. and N. Fleming, (2002), “The Designed and Achieved Sample of the 1998 LSAY Sample, LSAY Technical Paper no.16, ACER Melbourne.

LSAY, (2012), LSAY Y98 Cohort – Technical notes,
http://www.lsay.edu.au/cohort/1998/103.html#LSAY_weightings

Neal, D., (1997), “The effects of Catholic secondary schooling on educational achievement”, *Journal of Labor Economics* 15(1), 98–123.

Rothman, S., (2009), “Estimating Attrition Bias in the Year 9 Cohorts of the Longitudinal Surveys of Australian Youth”, LSAY Technical Paper no.48, ACER Melbourne.

Ryan, C., (2011), “Year 12 Completion and Youth Transition”, LSAY Research Report no.56, ACER Melbourne.

Snyder, T.D., and Dillow, S.A. (2012). *Digest of Education Statistics 2011* (NCES 2012-001). National Center for Education Statistics, Institute of Education Sciences, U.S Department of Education. Washington, DC.

Stock, J.H. and M. Yogo, (2005), “Testing for weak instruments in linear IV regression”, In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. D.W.K. Andrews and J.H. Stock, 80-108, Cambridge University Press: Cambridge.

Vella, F., (1999), “Do Catholic schools make a difference? Evidence from Australia”, *Journal of Human Resources* 34, 208–224.

Appendix A: Variable Definitions

Table 8 Definition of variables used throughout analysis and summarised in Table 1

Variable	Definition
Demographics	
Male	Dummy Variable = 1 if student is male.
Aboriginal	Dummy variable = 1 if students is Aboriginal or Torres Strait Islander.
Family Background	
No. of Siblings	
Mother born in Australia	Dummy variable = 1 if mother was born in Australia.
Father born in Australia	Dummy variable = 1 if father was born in Australia.
Student born in Australia	Dummy variable = 1 if student was born in Australia.
English Speaking Father	Dummy variable = 1 if Father was born in a predominately English speaking country.
English Speaking Mother	Dummy variable = 1 if Mother was born in a predominately English speaking country.
Father completed HS	Dummy variable = 1 if Father completed high school.
Mother completed HS	Dummy variable = 1 if Mother completed high school.
Father completed p. secondary	Dummy variable = 1 if Father completed post secondary education.
Mother completed p. secondary	Dummy variable = 1 if Mother completed post secondary education.
SES Mother	Continuous variable for the Mother ranging from 0-100 where low (high) numbers indicate low (high) job status.
SES Father	Continuous variable for the Father ranging from 0-100 where low (high) numbers indicate low (high) job status.
Geography	
Metropolitan area	Dummy variable = 1 if Student lives in metropolitan area during year 9.
Expectations	
Parent plans ft study	Dummy variable = 1 if parent’s plans were for the respondent to participate in full time study.
Student plans ft work	Dummy variable = 1 if students plans were to participate in full time work.
Motivation	Ordinal data ranging from 1-16 where high numbers indicate high student motivation.
Effort	Ordinal data ranging from 1-8 where high numbers indicate high student effort.
Ability	Ordinal data ranging from 1-15 where high numbers indicate high student self-assessed level of ability.
Repeated Primary School	Dummy variable = 1 if student repeated one or more years of primary school.
Outcomes	
Bachelor degree or higher	Dummy variable = 1 if student completes a bachelor degree.
Commence University	Dummy variable = 1 if student commences a bachelor degree
Completed High School	Dummy variable=1 if student completed high school by 2006
Reading	Discrete quartile data(1-4), where 4 is highest achievement.
Maths	Discrete quartile data(1-4), where 4 is highest achievement.

Appendix B: Summary Statistics by Catholic Ancestry

Table 9: Means of variables by Catholic ancestry (*Z*).

Demographics	*Z*> median	*Z*< median
Male	0.4620	0.4553
Aboriginal	0.0140	0.0125
Family Background		
No. of Siblings	1.9760	1.8260
Mother born in Australia	0.8843	0.3069
Father born in Australia	0.8699	0.2515
Student born in Australia	0.9634	0.7748
English Speaking Father	0.8858	0.6754
English Speaking Mother	0.8976	0.6767
Father completed HS	0.5461	0.6238
Mother completed HS	0.5672	0.6490
Father completed p. secondary	0.2331	0.3106
Mother completed p. secondary	0.2313	0.2943
SES Mother	35.93	36.11
SES Father	38.28	40.28
Geography		
Metropolitan area	0.5048	0.6591
Expectations		
Parent plans ft study	0.3270	0.3672
Student plans ft work	0.1977	0.1949
Motivation	10.19	10.16
Effort	5.27	5.23
Ability	10.96	11.04
Repeated Primary School	0.0376	0.0264
Outcomes		
Bachelor degree or higher	0.3035	0.3084
Commence University	0.5005	0.5461
Completed High School	0.8059	0.8314
Reading	11.29	11.25
Maths	10.95	10.98
Catholic	0.3407	0.2805

Note: Sample size *N*= 3501, Median is *Z* = 0.0605.