

Effects of schooling and age on performance in mathematics and science: A between-grade regression discontinuity design with instrumental variables applied to Swedish TIMSS 1995 data

Christina Cliffordson^{a,b} & Jan-Eric Gustafsson^b

^aUniversity West, Trollhättan and ^bUniversity of Gothenburg, Göteborg, Sweden,

christina.cliffordson@hv.se & jan-eric.gustafsson@ped.gu.se

Effects of schooling and age on performance in mathematics and science: A between-grade regression discontinuity design with instrumental variables applied to Swedish TIMSS 1995 data

Christina Cliffordson, University West & University of Gothenburg, christina.cliffordson@hv.se

Jan-Eric Gustafsson, University of Gothenburg, jan-eric.gustafsson@ped.gu.se

Abstract

The main purpose of the study is to examine the relative effects of schooling and age on performance in mathematics and science by the use of a regression discontinuity approach augmented with an instrumental variable approach. The regression discontinuity design relies on the assumption that there is a sharp age-based decision rule for grade assignment, and a main purpose of the paper is to investigate approaches to relaxing this assumption. In a previous study it was shown that it is possible to bring together individuals born a particular year in the analysis, and to estimate the amount of bias caused by non-strict grade assignment on the within-year regression coefficients of achievement on age. Using the Swedish TIMSS 1995 data the present study demonstrates that IV-regression is a promising approach to obtain unbiased estimates of the grade and age effects when the assumption of a sharp age-based decision rule is violated.

Keywords: *Age effect, Schooling effect, Regression Discontinuity, Instrumental variables, TIMSS*

Introduction

The issue of the relative amount of influence of schooling and age on achievement is of considerable theoretical, methodological and practical interest. Among other things, this is of great importance for the design and interpretation of results of international comparative studies of school achievement, in which it is necessary to focus on comparability with respect to either age or to the number of years in school (Gustafsson, 2010). However, the effects of these two factors are not easily separated because experimental methods cannot be used for ethical and practical reasons. Instead, the problem has been studied using correlational and quasi-experimental methods. Of these methods, the between-grade regression discontinuity

design is held to be the most effective in disentangling the schooling effect from the effect of age on the devolvement of intellectual performance (Ceci, 1991).

The regression discontinuity design is a powerful non-experimental approach for estimating causal effects when assignment to treatment is done in such a way that persons with scores below a cut-point on a continuous variable are assigned to one treatment and persons with scores above the cut-point are assigned to another treatment. When decisions about school start are based strictly on age the regression discontinuity design is therefore a useful method to separate the causal influences on achievement of another year of schooling and another chronological year. The slope of the regression within grades on age estimates the effect of age, and the effect of schooling is represented by the discontinuity between the regression lines for the two adjacent grades.

However, the regression discontinuity design is based on strong assumptions. It is thus assumed that school start is based on age only, but in reality the admission of some individuals is delayed or made earlier in time, which is not random with respect to age and intellectual development. Early admission is more likely to occur for bright individuals born just after the cut-off, whereas delay is more likely for intellectually less developed individuals born just before the cut-off. These selection effects cause the regression on age to be underestimated. One way to deal with this problem is to exclude those students who are not normal-aged, and there is a recommendation that this can be done if the percentage does not exceed 5% (e.g., Shadish, Cook & Campbell, 2002; see Luyten, 2006). Another way to handle the problem is to exclude those students who are not normal-aged, as well as those whose birthday fall within two months around the cut-off date, that is, the birth dates with the highest proportion of missing students. The regression obtained from these data is then extrapolated to cover the whole range of one year (Cahan & Cohen, 1989), which assumes that the regression of performance on age is linear. A previous study by Cliffordson (2010) took advantage of the fact that the Swedish TIMSS 1995 design included three adjacent grades (6, 7 and 8), and showed that a linear function provides an adequate description of the relationship between age and performance in both mathematics and science, at least over two years. The results also showed that the exclusion approach applied, amongst others, by Luyten (2006) worked well in this example. However, these findings do not imply that this procedure would continue to yield reasonable results if the proportion of students who are not normal-aged within their grade cohort is excessive.

In addition, the approach to exclude students who are not normal-aged, along with students who are normal-aged and born early and late in the year, is an ad hoc solution which is not entirely

satisfactory. One problem is that this causes loss of cases and therefore loss of power. Another, and more serious problem, is that reliance on rules of thumb which lack a strong theoretical basis may lead to misapplication of the regression discontinuity design, either because it is used in cases where it should not be used, or because it is not used in cases when it could have been used. There is, therefore, a need for an alternative approach. In the present study we explore instrumental variable (IV) regression as an alternative to ordinary least squares (OLS) regression.

IV regression is used in situations when an independent variable (in this case age) is not truly independent but is affected by the dependent variable (in this case achievement). Such variables are called endogenous variables, and using them as independent variables typically causes considerable bias. Thus, over-aged students typically are low-achieving, while under-aged students typically are high-achieving, and the reason why they are not normal-aged is that indications of their expected level of achievement has caused school-start to be delayed or advanced. If these students are included in an OLS regression a negative regression of achievement on age is typically observed.

We can, however, obtain correct estimates of the effect of age on achievement if we can identify an *instrument* and employ that in an IV regression model, which can be estimated with two-step least squares (2SLS). An instrument for an endogenous variable must satisfy two conditions: it must be correlated with the endogenous variable and it must be uncorrelated with the residual of the regression model. In this case formal school start age is a suitable instrument which should satisfy these two conditions.

Thus, the main purpose of the study is to examine the relative effects of schooling and age on performance in mathematics and science by the use of a regression discontinuity approach with instrumental variables applied to the Swedish TIMSS 1995 samples. Given that we can compare the results from the IV analysis with the results obtained by Cliffordson (2010) we can evaluate the usefulness of this alternative analytical approach.

Important previous findings

Most of the previous research in the field has focused solely on effects of schooling on intellectual performance. Whilst reviews of research by, for example, Ceci (1991), Herrnstein and Murray (1994), and Winship and Korenman (1997) indicate that schooling can increase intellectual performance, conclusions about the strength of the schooling effect differ dramatically. In an influential review of 200 studies Ceci (1991) identified eight different types of designs in this area. Even though the estimates of the schooling effect varied widely over the

studies, between a low estimate of 0.25 IQ points per year of schooling to a high estimate of 6 IQ points, Ceci concluded nonetheless that the results from all the different approaches provide a clear indication that schooling exerts an effect on intellectual development. Further, even though Ceci (1991) observed that all of the studies suffered from different kinds of methodological limitations, his conclusion was that the regression discontinuity design was the strongest of all the designs investigated. In their seminal study using this design, Cahan and Cohen (1989) found that one year of schooling exerted an effect that was about twice as strong as the effect of one chronological year. This general pattern of findings has been confirmed in several subsequent studies (e. g., Cliffordson, 2010; Crone & Whitehurst, 1999; Gustafsson, 2009; Luyten, 2006; Stelzl, Merz, Ehlers & Remer, 1995).

The between-grade regression discontinuity design

Even though the between-grade regression discontinuity design can be regarded as the most powerful of the models available in this area of research, this does not imply that it is without its problems. Typically, the model is estimated using OLS regression techniques, where the slope of the regression within grades on age estimates the effect of age and where the effect of schooling is represented by the discontinuity between the regression lines for two adjacent grades, as represented by a dummy variable. Thus, conceptually, the differences between the oldest and the youngest students in each grade estimate the net effect of a one-year difference in age. The difference between the youngest student in any given grade level, and the oldest in the grade level immediately below, provides the estimate of the effect of one year of schooling.

However, admission to school is not based entirely on age. In Sweden, as in many other countries, for some students admission is delayed, whilst for others it may be brought forward. Thus, in any given grade, there are both students whose age would normally place them in either a higher or lower grade, as well as 'missing' students who are enrolled in either the grade above or the grade below. Furthermore, neither delayed nor accelerated admission is random with respect to age and intellectual development. Accelerated admission is more likely to occur for more gifted students born just after the cut-off date (in Sweden on January 1st), whereas delay is more likely for intellectually less developed students born just before the cut-off date (e.g., Svensson, 1993). These selection effects cause the regression on age to be underestimated, meaning in turn that the effects of schooling will be overestimated. In recognition of this problem, Cahan and Cohen (1989) excluded those students whose ages differed from the norm within their grade level, as well as those whose birthday fell within two months around the cut-off date, that is to say, the range of birth dates within which the highest proportion of missing students who are learning in either a higher or lower grade are to be found. The regression

obtained from these data was then extrapolated to cover the whole range of one year. That is, if the regression of performance on age is non-linear, the effect of age will be underestimated. With reference to e.g. Shadish, Cook and Campbell (2002), Luyten (2006) claimed that if the proportion of students who are not normal-aged within their grade cohort is not excessive (not exceeding 5%) it is still possible to obtain reliable effect estimates with the regression discontinuity design.

Using the Swedish sample of TIMSS 1995 which comprised three grades (6th, 7th, and 8th), Cliffordson (2010) investigated the robustness of the regression discontinuity approach in terms of deviations from underlying assumptions. The results indicate that the selection effect caused by the presence of under- and over-aged students on the estimated within-grade regression coefficient is generally relatively small. This result supports the practice of excluding students who are not normal-aged as applied, amongst others, by Luyten (2006). However, in the study, the total percentage of non normal-aged students was relatively small, about 3.5 %, that is, the finding does not imply that this procedure would continue to yield reasonable results if the proportion of students who are not normal-aged within their grade cohort is excessive.

Therefore, it is of great interest to investigate the robustness of previous findings using an instrumental variable approach to address the selection effects of a non-strict application of the age-based school admission principle on the estimated within-grade regression coefficients. IV regression is, in other disciplinary fields, such as economy, the most commonly used procedure for addressing selection bias (Heckman & Robb, 1985).

Methodology

Participants

The study is based on the Swedish TIMSS 1995 data, which comprises sample data for 8 816 students distributed amongst three successive grade cohorts, 6th, 7th and 8th (2 815, 4058 and 1943 students, respectively) and 270 schools (Beaton et al., 1996). Age means were 13.0, 14.0 and 15.0, respectively. The share of accelerated students was 0.4, 0.7 and 0.7% and of delayed students 2.9, 3.2 and 3.0%, respectively. Case weights were applied in the analysis. See Table 1 for detailed information based on weights.

Variables

Scores from performance tests in mathematics and science constituted the dependent variables, while age and schooling were the independent variables. Age was measured in terms of month of birth - from the youngest born in December 1983 (=1) to the oldest born in January 1979 (=60), and formal school start age (7 years) in Sweden was used as the instrument. That is, over- and under-aged students' ages were corrected to an age as if they were normal-aged for the appropriate grade. Schooling was measured in terms of grade membership represented using dummy variables.

Methods of analysis

The idea of IV regression is to find another variable (an instrument) which is correlated with the independent variable suspected to be endogenous but which does not exert any influence on the dependent variable on its own. Estimation is typically done with 2SLS, where, conceptually at least, the first step involves regression of the independent variable on the instrument, and the second step involves regression of the dependent variable on the residuals of the independent variable from the first step.

The analyses were conducted with STATA 10. In order to take account of clustering effects and the hierarchical structure of the sample data, the "cluster" option offered by the STATA program was used. With this approach the standard errors are corrected for the loss of information caused by the clustering, causing these to become larger and the t-values to become smaller. The extent of the information loss due to clustering effects is a function of the intra-class correlation and the cluster size (Muthén & Muthén, 2006).

Finding and Discussion

Descriptive statistics

In general the patterns of descriptive statistics, as shown in Table 1, were similar for mathematics and science. As may be expected, compared to the normal-aged students, over-aged students obtained significantly lower scores (Cohen's d : mathematics= 0.72, 0.92 and 0.84; science=0.69, 0.67 and 0.94, for grades 6, 7 and 8, respectively), and under-aged students obtained higher scores (Cohen's d : mathematics= 0.42, 0.44 and 1.15; science=0.09, 0.43 and 0.40), with the exception of science with regard to under-aged students in the 6th grade. Hence, if the problem with under-, and over-aged students is not taken care of in the

analyses, the results will be biased, so that the regression on age will become negative instead of positive, as expected.

Regression analysis with and without an instrumental variable

The analyses were initially based on all the available data, including over-aged, normal-aged and under-aged students in all three grades ($N = 8816$). Grade was identified with two dummy variables.

In the first step an OLS regression model was estimated with mathematics as the dependent variable. According to this model the estimated age effect was -9.1 for one chronological year, the grade effect 50.2 for grade 7 and 44.6 for grade 8. Corresponding estimates for science were for age -7.1 and for grade 53.9 and 42.3, respectively. These estimates are absurd, and they clearly show that OLS estimates based on all data cannot be trusted.

In the next step the IV-approach was applied to the entire data set, and the main results are presented in Table 2. It may be noted that the age-effect is relatively strong for both mathematics ($b = 1.69$) and science ($b = 1.72$), and amounts to about 20 score points for one chronological year. The effect of grade 7 compared to grade 6 (i.e., b_7) is 21 for mathematics and 26.5 for science. The effect of grade 8 compared to grade 7 (i.e., b_8) is 15 for mathematics and 14 for science. Thus, in these analyses the age effect for mathematics is about equally strong, and for science somewhat weaker, compared to the schooling effect for grade 7, and it is somewhat stronger than the schooling effect for grade 8.

It is interesting to compare these results with those obtained by Cliffordson (2010). The correction technique applied in that study could best be applied in grade 7, where accelerated students from grade 8 and delayed students from grade 6 could be combined with normal-aged students to cover the age-group. For mathematics the corrected regression coefficient for one month of age based on this data set was 1.23, and for science the coefficient was 1.51, which could be compared with regression coefficients of 0.97 and 1.34, respectively, in the analysis based on normal-aged grade 7 students only. These results thus indicate that regression discontinuity analyses based on normal-aged students only may yield somewhat biased results. The IV-estimates for one month of age were 1.69 for mathematics and 1.72 for science, which are higher than the corrected estimates for grade 7 computed by Cliffordson (2010) by 0.46 and 0.21, respectively. This may be taken as an indication that correction methods used by Cliffordson (2010) do not quite manage to guard against threats to the validity of regression discontinuity design results caused by deviations from a strict decision rule. However, the IV estimate is based on data from all students in the three grades, and the analysis assumes that the

regression slope on age is homogenous over grades.

A separate IV-regression for mathematics and grades 6 and 7 yielded a regression coefficient on age of 1.24 (se 0.33), and for grades 7 and 8 a regression coefficient of 1.88 (se 0.43) was obtained. Even though the difference in regression slopes is not statistically significant these results nevertheless indicate that the steeper slope in the overall IV analysis is partially due to the steeper slope for the grade 8 students. In fact, the corrected estimates for the age regression presented by Cliffordson (2010) for grades 6 ($b = 1.34$) and grade 8 ($b = 1.84$) agree excellently with the results from the IV-analyses done separately for grades 6-7 and grades 7-8.

Conclusion and Implications

One main result was that the IV approach based on all students in each grade agreed reasonably well with the results from the OLS approach restricted to the normal aged students (Cliffordson, 2010), while the OLS estimates based on all students differed widely from both the estimates presented by Cliffordson (2010) and from the IV-estimates.

One main advantage of the IV-approach is that it can take advantage of all the data, which eliminates the need to make more or less arbitrary decisions about whether to apply the regression discontinuity design or not, or if to apply some kind of adjustment. However, it may be noted that the possibility to estimate a common regression slope on age in all three grades in this case caused some differences compared to the Cliffordson (2010) results, and these differences almost vanished when the assumption of a common within-grade regression slope was relaxed. Given the large standard errors of estimates from the regression discontinuity designs, and which are further increased by the IV-approach, it may be necessary to be careful in applying the IV-approach with such strict assumptions of homogeneity over the three years.

The results obtained in the current study show that IV-regression is a promising approach to deal with the problem that there are threats to correct inference from regression discontinuity analyses when decisions about school start are made on the basis of student performance, rather than a strict age-based decision rule. The study is limited however, with its focus on a single country, so further research should investigate its applicability to data from other countries.

References

- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA, USA: Boston College.
- Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development, 60*, 1239-1249.
- Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*(5), 703-722.
- Cliffordson, C. (2010). Methodological issues in investigations of the relative effects of schooling and age on school performance: The between-grade regression discontinuity design applied to Swedish TIMSS 1995 data. *Educational Research and Evaluation, 16*(1), 39-52.
- Crone, D. A., & Whitehurst, G. J. (1999). Age and schooling effects on emergent literacy and early reading skills. *Journal of Educational Psychology, 91*(4), 604-614.
- Gustafsson, J.-E. (2009). Strukturell ekvationsmodellering [Structural equation modeling]. I G. Djurfeldt & M. Barmark (red). *Statistisk verktyglåda II* [Statistical tool-box II]. Lund: Studentlitteratur.
- Heckman, J. J., & Robb, R. (1986). Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes. In H. Wainer (ed.), *Drawing Inferences from Self-Selected Samples* (pp. 63-107). New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.
- Herrnstein, R. J., & Murray, C. (1994). *The Bell Curve. Intelligence and Class Structure in American Life*. New York: The Free Press.
- Luyten, H. (2006). An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95. *Oxford Review of Education, 32*(3), 397-429.
- Luyten, H., Peschar, J., & Coe, R. 2008. Effects of Schooling on Reading Performance, Reading Engagement and Reading Activities of 15-year-olds in England. *American Educational Research Journal 45*(2), 319-342.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, NY: Houghton Mifflin Company.
- Stelzl, I., Merz, F., Remer, H., & Ehlers, T. (1995). The effect of schooling on the development of fluid and crystallized intelligence: A quasi-experimental study. *Intelligence, 21*, 279-296.
- Svensson, A. (1993). Skolanpassning och skolframgång bland elever födda i början respektive slutet av året [Adjustment to school and school performance among students born in the beginning and the end of the year, respectively]. (*Report 1993:04*). Göteborg: Göteborg University, Department of Education.
- Winship, C., & Korenman, S. (1997). Does staying in school make you smarter? The effect of education on IQ in The Bell Curve. In B. Devlin, S. E. Fienberg, D. P. Resnick, & K. Roeder (Eds.), *Intelligence, Genes, & Success. Scientists Respond to the Bell Curve*. New York: Springer-Verlag, Inc.

Table 1: Means and standard deviations per grade for mathematics and Science performance.

<i>Grade</i>		<i>Year of birth</i>	<i>N</i>	<i>%</i>	<i>Mathematics</i>		<i>Science</i>	
					<i>Mean</i>	<i>S.D.</i>	<i>Mean</i>	<i>S.D.</i>
6	Accelerated	83	10	0.3	507.52	59.04	496.83	59.24
	Normal	82	2783	96.7	479.12	75.82	490.32	83.39
	Delayed	81	84	2.9	422.38	80.97	429.44	92.27
7	Accelerated	82	20	0.7	553.87	66.50	572.68	75.14
	Normal	81	2816	96.2	520.84	84.72	537.14	89.59
	Delayed	80	92	3.1	446.79	75.94	476.17	93.22
8	Accelerated	81	20	0.7	647.24	69.39	610.21	94.66
	Normal	80	2898	96.4	555.86	89.25	572.83	92.64
	Delayed	79	89	3.0	475.10	103.44	485.85	93.09

Note. Cases weighted by HOUSE WEIGHT.

Table 2: Unstandardized regression coefficients, standard errors, and t-values from IV-regression models estimated with 2SLS, based on normal-aged, accelerated and delayed students from the grades they really completed, for mathematics and science.

Grade	Subject		Age				Schooling		
	N	N	b	S.E.	t-value	One year of age	b	S.E.	t-value
6	2877	Math	1.32	0.48	2.75	15.89			
		Science	0.56	0.52	1.07	6.71			
7	2928	Math	1.15	0.44	2.60	13.81			
		Science	1.51	0.44	3.45	18.16			
8	3007	Math	2.63	0.75	3.50	31.55			
		Science	3.13	0.76	4.12	37.55			
6 - 7	5805	Math	1.24	0.33	3.78	14.85	26.48	5.53	4.79
		Science	1.04	0.34	3.04	12.44	34.59	5.88	5.88
7 - 8	5935	Math	1.88	0.43	4.38	22.50	12.53	7.66	1.64
		Science	2.30	0.43	5.40	27.66	7.00	6.95	1.01
6 - 7 - 8	8816	Math	1.69	0.33	5.12	20.28	b_7 21.12	5.59	3.78
		Math					b_8 35.90	10.26	3.50
		Science	1.72	0.34	5.10	20.64	b_7 26.53	5.76	4.60
						b_8 40.68	10.03	4.06	

Note. Cases weighted by HOUSE WEIGHT; The lowest grade included in the model is the reference.