The effects of class size on educational attainment: Danish quasi-experimental evidence and evidence that controls for family, school and neighbourhood effects*

Paul Bingley

Department of Economics, Aarhus School of Business
& Department of Economics, University of Warwick

Vibeke Myrup Jensen

Department of Economics, Aarhus School of Business
& Danish National Institute for Social Research

and

Ian Walker

Department of Economics, University of Warwick
& Institute for Fiscal Studies

Keywords: class size, regression discontinuity, sibling differences
JEL Codes: I22, C23

Abstract:
This paper is concerned with the relationship between class size and the student outcome – length of time in post-compulsory schooling. Research on this topic has been problematic partly because omitted unobservables, like parents' incomes and education levels, are likely to be correlated with class size. Two potential ways to resolve this problem are to exploit either experimental or instrumental variation. In both cases, the methods require that the variation in both class size and the outcome should not be contaminated by other unobservable factors that affect the outcome – like family background. An alternative approach, which we pursue here, is to take advantage of variation in class size between siblings which allows unobservable family effects to be differenced out. Our aim is to provide estimates of the effect of class size and use these to conduct an evaluation of the costs and benefits of a reduction in class sizes.

* Her Majesty's Treasury Evidence Based Policy Fund in conjunction with the Inland Revenue, Department for Education and Skills, the Department of Work and Pensions, and the Department of Media, Culture and Sport funded the construction of the dataset used in this research. Our collaboration was supported by the Aarhus University Research Fund (AU-2004-533-033), the Danish Social Science Research Council (24-04-0240), and the UK Economic and Social Research Council (R.ECAA.0071). We are grateful to Eric Plug and other participants at ESPE 2006 in Verona for their comments. The usual disclaimer applies.

Corresponding author: Paul Bingley, Tel. +45 8948 6400, email pbi@asb.dk

1. Introduction

Academic and policy interest in improving schools comes from recognising the importance of human capital formation for individuals and society. This is based on theoretical models, and empirical evidence, that relates income, productivity and economic growth to the quantity of schooling - the most common proxy for the stock of human capital.

Class size is often a focus for both policy action and research interest because it is easy to measure and, apart from the opportunity cost of students' time, it is the most important cost of education. In Denmark, 80% of compulsory schooling expenditure goes to pay teachers' wages, and this factor alone explains 60% of the variance in expenditure between schools. Similar expenditure shares are accounted for by teachers pay in the US and the UK (Hanushek 2002).

There are many models of the effects of class size on learning outcomes, from economics and other disciplines. For example, Lazear (2001) postulates that children in smaller classes can learn more, because of the lower probability of interruption to teaching, if the probability of a student interrupting teaching is independent across students. Since an interruption in class requires that teaching be temporarily suspended this imposes a negative externality on everyone else in the class which is larger the larger is the class size. Of course there are other benefits to teaching in small classes too, but this model captures an important feature of class size and gives rise to a specific functional form for the educational production function.

One important implication of the Lazear (2001) model is that optimal class size is larger if students are well behaved, and/or if schools can assign weaker and/or more disruptive children to smaller classes, then local public education authorities should facilitate smaller class sizes in schools with a higher proportion of disruptive and/or weaker children. If such resource allocation occurs, but it is not able to entirely offset existing achievement differentials, then empirically this should give rise to a spurious association between smaller classes and lower student achievement. It is exactly this raw correlation which has been found in datasets from around the world (Hanushek, 2003). This motivates the need for sources of exogenous variation in class size in order to uncover the size of whatever causal mechanism is at work.

This paper is about the effect of school resources on length of completed education: in particular the effects of the Danish rules that determine class size and students per teacher hour. Like Browning and Heinesen (2004) (henceforth BH) for Denmark, and earlier work by Angrist and Lavy (1999) for Israel, our analysis applies a regression discontinuity design based on administrative rules for compulsory schooling. The variation in actual class size is driven by the interaction between random variation in cohort size and administrative rules that place a cap on class size. However, our analysis also pays attention to the importance of the home environment using a sibling difference approach. Indeed, we argue that the administrative rule results in class sizes that are systematically predictable by parents – except if attention is confined to samples close to discontinuities. However, close to these discontinuities actual class-size will be subject to considerable uncertainty, which is why it provides a valid IV. Thus it seems likely that the administrative rule will only provide a local average treatment effect for the children of low risk aversion parents.

The presumption in the approach based on regression discontinuities is that parents do not (or cannot) exploit the administrative rules because they do not know how large the cohort is. Whether this is true is arguable. Parents may be able to form a reasonable forecast, from pre-enrolment school meetings, of the likely number of classes in the cohort several months before enrolment. Parents who place a high value on education quality may be more likely to avail themselves of the private schooling option (which is relatively inexpensive in Denmark), or even the option of delaying entry for a year, when faced with a cohort is of a size likely to generate large class sizes. While it is true that there is some risk associated with such an action it seems unlikely that this would entirely dominate the systematic relationship between class size for your child and observable pre-entry data.

Thus, here we give results that are based on combining the administrative rules with an elimination unobservable family preferences by sibling differencing and restricts attention to siblings that attend the same school. We control for family effects by exploiting our ability to take sibling differences for the population of students attending 8th grade during the 1980's. Since schools are strictly associated with catchment areas (although this has been relaxed since 1993) this effectively controls for neighbourhood effects.

On the basis of a wide variety of sibling difference specifications, where we typically find statistically well determined effects, we conclude that it would be reasonable to presume that a 5% (one unit) reduction in class size in 8^{th} grade gives rise to approximately 0.04 more years of education (about 1% of the typical level of post-compulsory schooling) and a 5% reduction in the students per teacher hour ratio in 8^{th} grade gives around 0.05 more years on average. In contrast our results based simply on levels typically show, like earlier Danish research, that class size has an insignificantly positive effect on education length. We use our sibling difference results to conduct an elementary cost-benefit analysis of a policy of decreasing class sizes and find that, even in the most favourable circumstances, the costs outweigh the benefits. Thus, our conclusions are somewhat more pessimistic than the only other previous study - by Krueger (2003).

The remainder of the paper is organised as follows. The next section reviews the literature, and places our contribution within that. A data description is followed by estimation results, interpretation and discussion. We then compute the likely history of class size for individuals given what we know about class size in 8^{th} grade in recent data and the correlation between class sizes in consecutive grades. This allows us to compute the expected average class size throughout a child's education given what we observe at 8^{th} grade. We then investigate the effects of length of completed education on earnings, using results from a large sample of twins, so we can compute the present value of the financial returns to extending education. We then estimate the costs of lowering average class size both in terms of making each year of schooling more expensive, and making completed education length longer. This then leads to a present value of the costs of such a policy. Finally we conclude with an agenda for more research.

2. Literature

As befits the importance of the issue, the relevant literature is extensive. However, only one paper, Krueger (2003), draws out the implications of the findings for policy costs and benefits. Recent reviews of the literature can be found in Hanushek (2003) and Krueger (2003). Much of the literature consists of correlations between outcomes such as test scores, education length, or educational attainment and class size and related inputs using observational data and is therefore vulnerable to the criticism that the correlations are contaminated by unobservable heterogeneity. Here we highlight those contributions that are particularly relevant to our own research which focuses on the estimation of the causal effect of class size.

Hanushek (2003), based on a meta-analysis of many studies where each chosen *estimate* gets equal weight and the estimated standard error of each estimate is ignored, argues that input-based schooling policies have failed. Krueger (2003) on the other hand, conducts a meta-analysis based on the same set of studies, but gives each *paper* equal weight, and finds that reducing class size does improve educational outcomes.

The one and only truly experimental study is Krueger (1999) which analyses the Tennessee Student/Teacher Achievement Ratio (STAR) experiment which was conducted in the 1980's. This involved random assignment of approximately eleven thousand students during grades 1-4 into classes of either about 15 students or about 22. Students attending smaller classes obtained significantly higher test scores immediately after the experiment but that soon thereafter the effect approximately halved although it remained significant even 10 years after having left 4th grade. Criticisms of such experimental work include Hawthorne effects, cream skimming administrative placement, and charges of parental influence in student allocations. Moreover, Carneiro and Heckman (2003) argue that the test score effect is likely to be temporary. Thus, here, we focus of a permanent outcome – time spent in post-compulsory education.

Nonetheless, a substantive contribution of Krueger (2003) is to make cost-benefit calculations of class size reductions based on his earlier Tennessee STAR estimates. Krueger uses estimates of the effects of test scores on subsequent earnings together with his own estimates of class size on test scores to show that the internal rate of return that equates discounted costs and benefits, assuming a growth rate of 1%, is a relatively modest 6.2%.

Angrist and Lavy (1999) use the Maimonides' rule that limits the maximum class size in Israeli schools to be 40. The implied discontinuity in the relationship between grade enrolment and class size is used to provide exogenous identifying variation. In this regression discontinuity design, the administrative rule-based class size is used as an instrument for observed class size. Reductions in class size are found to increase end of grade test scores for 4th and 5th graders but not for 3rd graders.

Hoxby (2000) looks at Connecticut elementary schools and exploits cross county variation in the birth rate and the cross county variation in rules that determine the minimum and maximum class size to investigate student achievement (test scores). No class size effects are found. Case and Deaton (1999) analyse class size during the apartheid era in South Africa. Black parents were unable to choose their children's school and school resource allocation was (arguably) exogenous. On the basis of aggregated data at the district level, reductions in class size in the range 50-80 students were found to have positive effects on district level enrolment, literacy and numeracy tests, and years of completed schooling. Woessmann and West (2002) use the Third International Maths and Science Study (TIMSS) to examine the relation between class size and test scores for two classes in two consecutive grades in schools. They address within-school, between-class and between-school sorting: instrumenting actual class size with the school average class size within the grade as an instrument, and using school fixed effects to deal with sorting between schools. Sizeable beneficial effects of smaller class sizes are found only for Greece and Iceland, where teacher salaries are relatively low.

Recent work by BH follows Angrist and Lavy (1999) in using the Danish version of Maimonides' rule for maximum class size applied to Danish 8th grade students. A similar administrative rule that determines the ratios of students per teacher hour are also used. They find large, but imprecise, effects of reducing these resource measures on increasing length of completed education. Their results imply, for example, that a 5% reduction in 8^{th} grade class size and students per teacher hour ratio during 8th grade causes an insignificant 0.066 and 0.14 increase in the length of completed education[1]. Our results are approximately one-quarter of these magnitudes.

Our analysis is based on sibling pairs of students with the same mother, same father *and attending the same school*. This paper extends BH since we can then control for family, school and neighbourhood fixed effects using sibling differences. We know of no earlier research that attempts to identify class size effects from sibling differences. Indeed, if we could rely only on the difference in class sizes between the siblings this would typically be quite small. However, we can exploit the variation *between siblings* in the class sizes implied by the rules provided parents do not choose to send their children to *different* schools because of the *variation* in class size. Thus, we are assuming that parents make *long term* location choices and do not move from area to area to exploit variations in class size over time.

We estimate a variety of specifications and we typically find a small but very precisely determined positive effects of school resources on length of education: a reasonable view of our estimates would be that a 5% reduction in 8th grade class size causes a 0.015 increase in length of completed education (in years), and about the same effect for a 5% decrease in students per teacher hour years[2].

Although our results are statistically significant they are less than one-quarter of the size of the effects imprecisely estimated by BH. The greater precision of our estimates is due to our larger sample size, the whole population compared to BH's 10%, and our ability to control for more variation in the data which might otherwise compromise the experimental nature of the institutional setup that we are both exploiting. Controlling for all that is fixed (both observable and unobservable) about the school, family and neighbourhood distinguishes the effect of different (locally random) realisations of the rules from the confounding effects of allocations of resources and students between schools, families and neighbourhoods. For BH, a stochastic implementation of the rule, or a fuzzy design, reduces the explanatory power of their instrument (the class size predicted by the rule) but should not bias their estimated class size coefficient of interest. For us, applying the rule directly induces measurement error, which should bias the estimated coefficients of interest towards zero - at least if it were classical measurement error. For our differenced or within-family model, measurement error is much greater than in the levels and we attempt to address this problem in our analysis.

Most studies of class size examine the effect on test scores taken at then end of a grade. While immediate cognitive achievement changes are useful short run outcome measures, their persistence has been called into question. Educational attainment, or length of completed education, is the outcome we consider here. It is a long run outcome, which is strongly correlated with later earnings, and other adult outcomes.

[1] BH Table 6 for class size, and in Table 7 a coefficient of -4.923 for teacher hours times 0.028 which is a 5% reduction in the number of pupils per teacher

[2] These figures correspond to coefficients on log class size and on students per teacher hour in our completed years of education equations of about -0.3 evaluated at an average class size of 20.

3. Danish Education System

3.1 Financing public school expenditure.

Attendance at primary and lower secondary school (grades 1-9, corresponding roughly to ages 7-15) is compulsory in Denmark. Education is a requirement from 1 August in the year that the child turns seven years old until 31 July in the year which regular instruction has been received for 9 years. During the period 1981-1990 analysed in this paper, 89% of children attended public (i.e. state funded) schools. These 1826 (in 1990) schools are run by 275 municipalities, and are attended by an average of 309 students. Municipalities have a mean population of 36,094 residents, but this ranges from 2,512 to 466,723 (Copenhagen), and the number of schools per municipality ranges from 1 to 76 accordingly. Public school expenditure is financed through municipal income tax, together with a complex between-municipality redistribution scheme, which subsidises expenditures in low income municipalities. Average total expenditure per student per year was DKK 31,360 in 1990 (corresponding to €4,248 in 2005 prices), having risen steadily from DKK 18,447 in 1981 (€3,713 in 2005 prices). The total number of students fell consistently throughout the period, from 728,900 in 1981 to 559,600 in 1990 due to smaller birth cohorts. The net effect was a reduction in expenditure on public schools between 1981 and 1990 from €2.629 billion to €2.365 billion (2005 prices).[3]

There is a large variance in public school expenditure between municipalities (coefficient of variation of 0.13). Changes in expenditure can largely be attributed to reductions in agreed teacher working hours and increased seniority. Between-municipality variation in teacher salary weighting, proportions of school children of different ages, and students whose mother tongue is not Danish, explains some of the variation, but much of the variance cannot be explained by observable municipality characteristics (see Graversen and Heinessen (1999)).

3.2 Allocating students to schools and subsequent schooling choices.

During the analysis period, the allocation of public school places was on the basis of catchment area of place of residence at the beginning of the calendar year of first grade start. Parents are required to sign their children up to a school latest the start of

[3] See Danish Ministry of Education and Research (1993) for further details.

the year in which the child turns seven years old. Should a child move home to a different catchment area, that public school is obliged to offer a place from the beginning of the month following the move. 11% of children attended a private school and these are heavily subsidised (on average 85% of expenditures are provided by the municipality)[4]. Private schools are mostly found in urban areas and are disproportionately attended by the children of highly educated parents. While average educational attainment is higher for students having attended private school, this is no longer the case after allowing for selection into private schooling on the basis of observable characteristics (Rangvid, 2002). If it is the case that children attending private schools respond differently to class size then this may lead to bias in a class size coefficient estimated only on public school children. Private schools have a lower mean class size than public schools, and if parents are behaving rationally they ought to place children who respond better to class size in smaller private school classes. This ought to bias, towards zero, class size coefficients estimated on a sample where such students are selected out.

Students can leave lower secondary school after grades 7 (or 8) in order to attend a "continuation school", usually a private boarding school, and 1% (8%) take up this opportunity. In addition to the nine compulsory grades there is a voluntary 10th grade attended by 50% of those leaving 9th grade. On completing 9th or 10th grade respectively 95 and 90% of students take the public school final examinations.

Having completed lower secondary education, 7% never return to the educational system, 33% go to upper secondary school and 59% do vocational training. These transitions are most often immediately after a summer recess and the courses last two or three years, with completion rates of 88% for upper secondary school and 86% for vocational training. Upon completion, subsequent transitions to higher education occur on average after 18 and 13 months respectively. This study gap is explained by short term employment, travel, and admission criteria limiting places. Destinations from upper secondary are 26% vocational education, 62% higher education, 11% no further education. Times to completion average 2.4 years for vocational training and 3.6 years for higher education, with completion rates of 73% and 60% respectively. Education Ministry estimates of the average expected total time to completion of

[4] Danish Ministry of Education (2002).

education for those commencing first grade in 1981, 1990 and 2000 was 13.1, 14.0 and 15.1 years respectively.

In summary, post-compulsory education is in two broad phases with a study gap of more than a year on average between the two. Completion rates are lower for higher and longer courses. There is a large variance in times to completion, explained by different routes, gaps and course lengths. Of those entering 8^{th} grade in 1990 15% were enrolled at an educational institution in 2001, although less than 5% of the cohort were enrolled in 2003.

3.2 Class size and students per teacher hour rules

The student per teacher ratio averaged 11.9 in 1981 and fell gradually to 10.1 in 1990. Mean class size remained at 18.2 throughout. Primary and lower secondary public schools are comprehensive, whereby students are allocated to a class on entry, and most lessons will be taught to the same class group throughout all grades. A national curriculum stipulates the number of hours required teaching in each of 15 subjects at each grade level. Two hours of optional subjects are introduced first at 8th and 9th grades. Danish education law stipulates a maximum class size of 28 students for primary and lower secondary schools. Municipalities are free to implement their own class size rules subject to this restriction. In practice, BH and Heinesen and Rangvid (2003) show that an additional class is typically added at multiples of 24 students, making the effective class size maximum 24 students. This is to avoid the situation where new student enrolment at later grades would force a class to be divided in accordance with national law. The result is the discontinuous relationship between class size and school year group enrolment shown in Figure 1. Formally, the number of classes that a given school-grade-year needs to be split into, NCLASS = (INT (ENROL-1) / 24) + 1, where ENROL is the number of students enrolled in the given school-grade-year. Average class size for the school-grade-year is then CSIZE = ENROL / NCLASS. For example, enrolments 1-24, 25-48, 49-72 correspond to 1, 2, 3 classes respectively. Enrolments 24, 25, 48, and 49 correspond to average class sizes of 24, 12.5, 24, and 16.3 respectively. This is similar to Angrist and Lavy (1999) use of Maimonides' rule of class sizes of 40 in Israeli public schools.

Over and above the requirements for specific teaching hours dedicated to different curriculum subjects, the Danish Ministry of Education recommends a number of teacher hours per student per week, but does not impose these recommendations. Municipalities are free to interpret the guidelines for teacher hours, and their implementation varies between municipalities and within municipality over time. However, it is municipalities rather than schools that finance the incurred teacher and class expenditures associated with the class size rule and teacher hour recommendation. While municipalities themselves may trade-off, for example, books for teachers, this is not a substitution that is being made at the school level. Following BH and Heinesen and Rangvid (2003) we use the administrative rule for teacher hours which was in force in Copenhagen, the largest municipality, during the school year beginning 1991. We also consider the sensitivity of the results to using this rule (based on a maximum of 24) and to applying the same rule to all municipalities. The rule, based on 24, is a step function of enrolment per school-grade-year, ENROL, and the number of classes, NCLASS as indicated in Figure 2.

There is obviously a discontinuous relationship between number of students per teacher hour and enrolment - given in Figure 3. The maximum class size rule is behind the larger of the discontinuities, but compensatory allocation means that variation either side of the discontinuity is not as large as for the class size rule. Also, peaks in the students per teacher hour rule trend up with enrolment, whereas peaks in the class size rule do not. However, it is clear that, in expectation, schools with large enrolments have larger class sizes and more students per teacher hour.

Figure 1 Class size and 8th grade enrolment

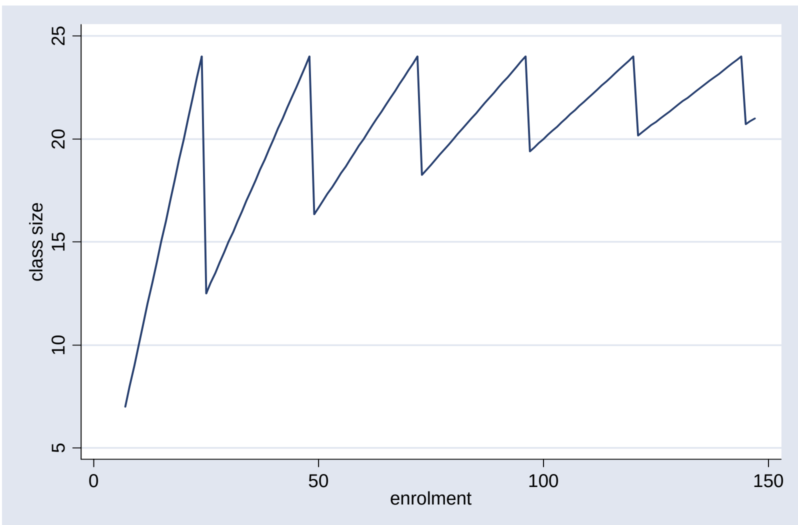


Figure 2   Teacher hours per student rule: by number of classes 1-4

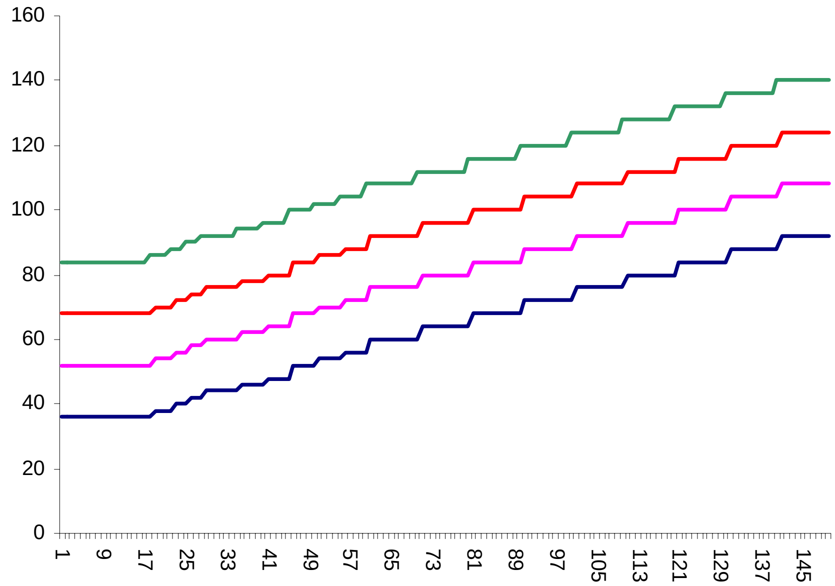
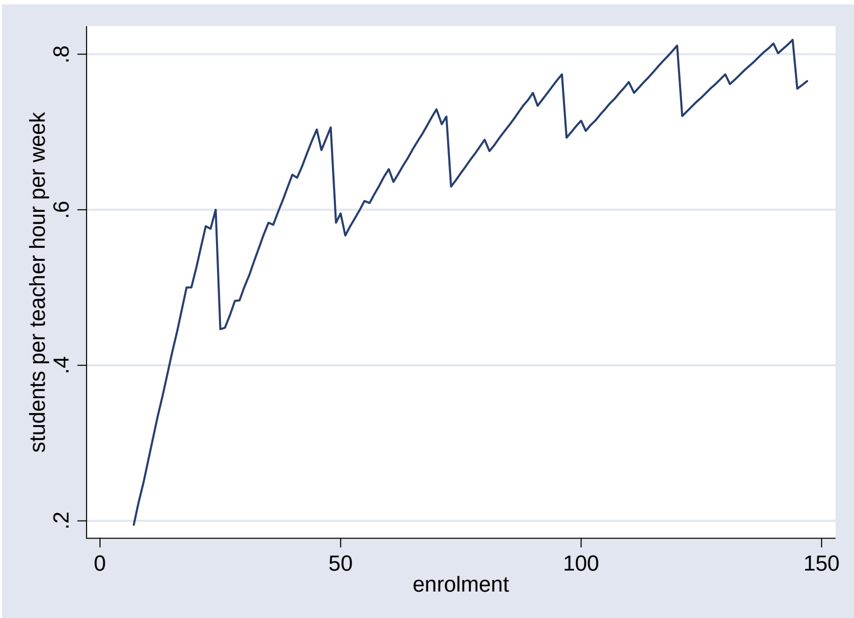


Figure 3   Students per teacher hour and 8th grade enrolment



4. Data Description

The dataset we use is based upon a very small number of variables from two administrative databases containing individual information for *all* residents of Denmark: the Central Person Register and the Integrated Student Register. The Central Person Register is a national administrative database that contains social security numbers that enable links between all children and their legal mother and legal father. Moreover, this enables us identify siblings. Our objective is to choose a sampling frame that controls for as many unobservable fixed school and family effects as possible, and allows us to estimate a relatively clean class size and students per teacher hour effect. Our motivation is to control for school (and, hence, neighbourhood), mother and father fixed effects at the same time by estimating class size difference effects within the group: same mother, same father and same school. Non-informative observations are dropped: (1) singletons (2) sibling groups where each goes to a different school at 8th grade (3) half-siblings (4) multiple-births. Finally dropping the 0.1% of remaining households with more than 6 siblings leaves an estimation sample which is described in Tables 1 and 2.

The student register links unique student social security numbers to school identifiers for 8th grade (children aged around 14) and above on 1 October each year, 2 months after the start of the school year. We are able to use this match and so calculate school enrolments in each grade-year consistently from 1981 until 1990.

It is important to note that, unlike BH, the data available to us, although much larger, does not contain actual class size and teacher hours. However, we do observe enrolment and can apply the administrative rules to compute the class size and teacher hours that should affect each child. BH use this information to create instrumental variables for actual class size and teacher hours. We use this information directly as explanatory variables, following Van der Klaauw (2003).

The crucial assumption in any analysis based on exploiting administrative rules for identification is that parents do NOT exploit them. In particular, it is assumed that the variation in, in this case, class size is uncorrelated with any other factor that affects the outcome of interest, in this case length of completed education. There are two pieces of evidence that could cast doubt on the validity of this identifying assumption: evidence that class size was predictable from observable information; and evidence that introducing covariates changed the effect of class size.

Inspection of Figure 1 in Angrist and Lavy (2002) or Figure 1 in BH shows that a knowledge of enrolment size allows one to predict the number of classes and hence average class size for your child's cohort. Since cohort size, within a catchment area, is likely to be relatively stable the number of classes is also likely to relatively stable and so too will class size. Figure 4 here, takes the Danish data on all 8th graders observed in 2002-2004 and shows the coefficients of regressing class size next year against dummy variables for the level of enrolment this year. There is clearly a strong positive effect of enrolment on future class size (R-squared is 0.25[5]), at least at low levels of class size and enrolment.

Thus, it seems unlikely that Maimonides' rule is entirely immune from the problem that parents may be able to exploit the rule to reduce the class size faced by their child. If the rule did genuinely produce experimental variation in class size then class size should be uncorrelated with observable and unobservable characteristics of parents. If this were true then estimates of class size effects should be robust to the inclusion of control variables. All that such control variables should do is to improve the precision of the estimated class size effect. In fact, BH do find that including an extensive list of observable family variables makes a large difference to the estimated effect of class size (in both their discontinuity sample and their whole sample), even though they presume that the class size variation is exogenous[6].

The outcome of interest, and dependent variable throughout, is number of years of schooling completed after beginning 8th grade. This is a long-run outcome measure which is not subject to the criticisms faced by immediate test score measures that they are not persistent. It is also simple to compute, and not subject to value judgements on the part of the researcher regarding the number of years a particular education is "worth" in comparison to other educations. However, although more years in education is positively correlated with obtaining higher qualifications, it is not unambiguously a good outcome. Late completers in typically short educations are

[5] Similarly, when one regresses class size next year against class size this year we get a R^2 of 0.17

[6] Angrist and Lavy (2000) use class level data but this is, nonetheless, also susceptible to this criticism because, on average, larger classes will be selected by parents with lower preferences for the outcome.

counted equally as those with average completion times in educations that typically take longer.[7]

Table 1 describes the dataset used in our analysis. There are 141,186 households containing 299,283 children (note that one child households have been dropped) – 77% of them are in 2-sibling, 20% in 3-sibling, 2.5% in 4-sibling, 0.3% in 5-sibling, and 0.08% in 6-sibling households. The distribution of our outcome (education length), and of the explanatory variables that are of primary interest (class size and students/teacher hour/week) is tabulated according to values of other explanatory variables used in the analysis. It is clear that neither class size not students per teacher hour are constant across groups which reflects the strong concentration of large households in rural areas where class sizes tend to be smaller. There are correspondingly large differences in education length across sibling sizes. Table 2 describes the sibling differenced data. There is a marked tendency for the differences between siblings to get larger in larger households, and inter-sibling differences in both class size and students/hour tend to be larger in larger households.

Figure 4 shows the distribution of 8th grade school enrolment - that is, the number of schools which have that number in the 8th grade cohort. Figures 5 and 6 show the distributions of class sizes and pupil/teacher hours. Comparing the enrolment distribution with Figures 1 and 3 shows that there are large discontinuities where the distribution of school year group sizes is quite dense. Thus, as can be seen, in Figures 5 and 6, there are many small schools but few schools which exceed an enrolment of 100 in 8th grade. Figure 7 shows the operation of the administrative rule applied to our data. There are clear class size discontinuities. However, there is a clear variance in class sizes that is systematic. Where we get close to the discontinuity the variance rises.

[7] In future work we intend to link qualifications obtained to “normal” completion times (Education Ministry 2002) in order to measure effects on normalized education length and timely completion.

Figure 1 Grade 1 class size predictions from lagged class size

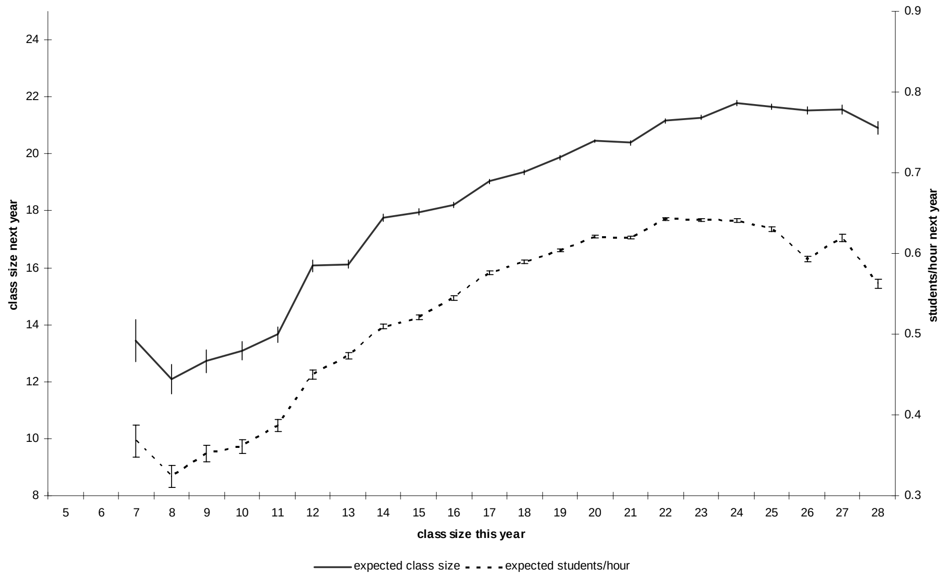


Figure 2 Grade 1 class size predictions from lagged enrolment size

Table 1 Summary Statistics for Levels

	frequency	education length %	education length Mean	education length *st.dev.*	students/hour mean	students/hour *st.dev.*	class size mean	class size *st.dev.*
# Siblings								
2	251050	83.9	7.18	*2.43*	0.653	*0.074*	20.17	*2.49*
3	43578	14.6	6.86	*2.52*	0.652	*0.076*	20.14	*2.53*
4	4136	1.4	6.34	*2.66*	0.645	*0.079*	19.98	*2.56*
5	435	0.15	5.51	*2.69*	0.642	*0.084*	19.89	*2.62*
6	84	0.03	5.82	*2.63*	0.605	*0.105*	18.88	*2.94*
Female	147839		7.21	*2.37*	0.653	*0.074*	20.16	*2.49*
Male	151444		7.03	*2.53*	0.653	*0.075*	20.16	*2.50*
Subsequent children	158097		7.08	*2.39*	0.650	*0.075*	20.09	*2.53*
First child	141186		7.17	*2.52*	0.656	*0.074*	20.24	*2.46*

Table 2 Summary Statistics for Sibling Differences: Differences from Family Mean

	frequency	education length mean	education length *st.dev.*	students/hour mean	students/hour *st.dev.*	class size mean	class size *st.dev.*
# Siblings							
2	251050	1.084	*0.866*	0.025	*0.021*	1.167	*0.937*
3	43578	1.282	*0.988*	0.030	*0.024*	1.383	*1.085*
4	4136	1.387	*1.074*	0.033	*0.026*	1.492	*1.154*
5	435	1.497	*1.183*	0.033	*0.028*	1.526	*1.212*
6	84	1.635	*1.286*	0.033	*0.030*	1.515	*1.054*
Female	147839	1.102	*0.876*	0.026	*0.021*	1.201	*0.963*
Male	151444	1.133	*0.906*	0.026	*0.021*	1.206	*0.971*
Subsequent children	158097	1.126	*0.897*	0.026	*0.021*	1.215	*0.976*
First child	141186	1.108	*0.885*	0.026	*0.021*	1.191	*0.958*

Figure 4 Distribution of 8th grade enrolment size

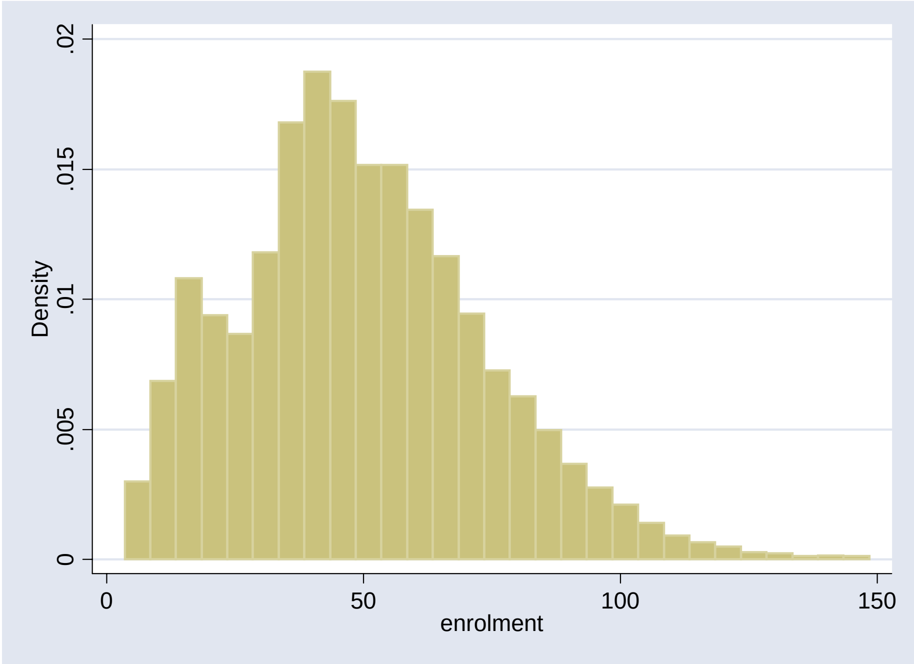


Figure 5 Distribution of rule given students per teacher hour at 8th grade

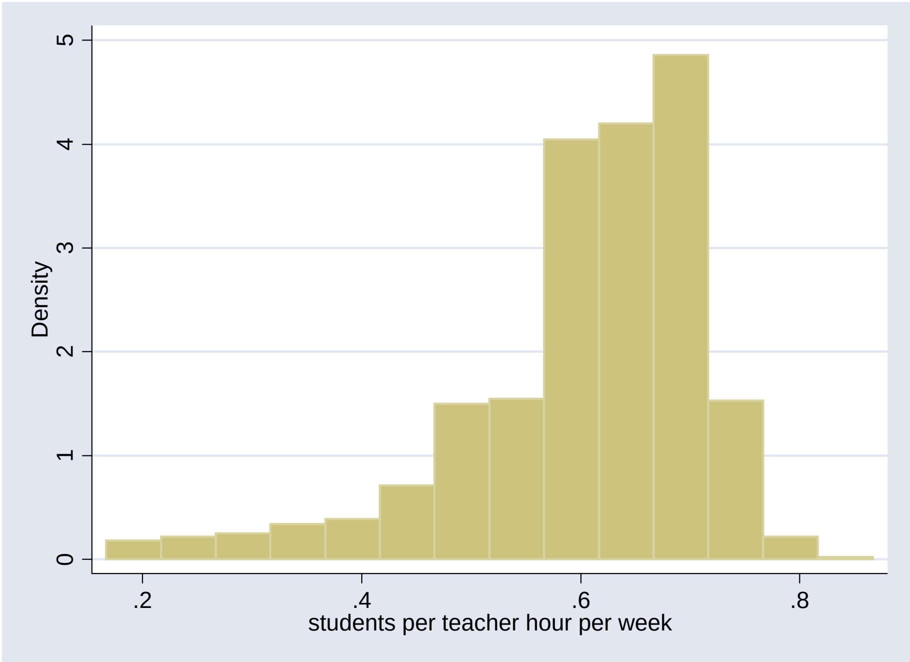


Figure 6 Distribution of (24 maximum) rule given class size at 8th grade

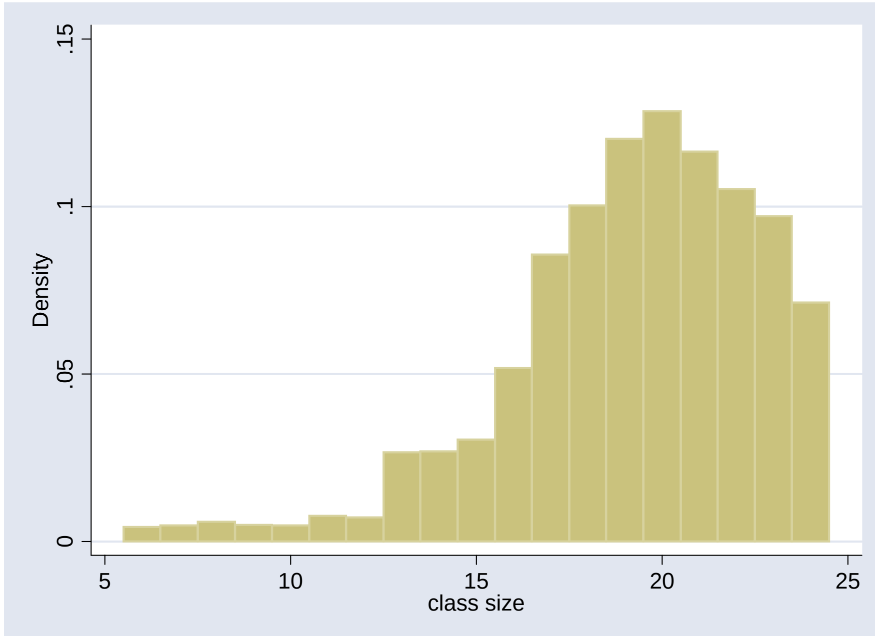
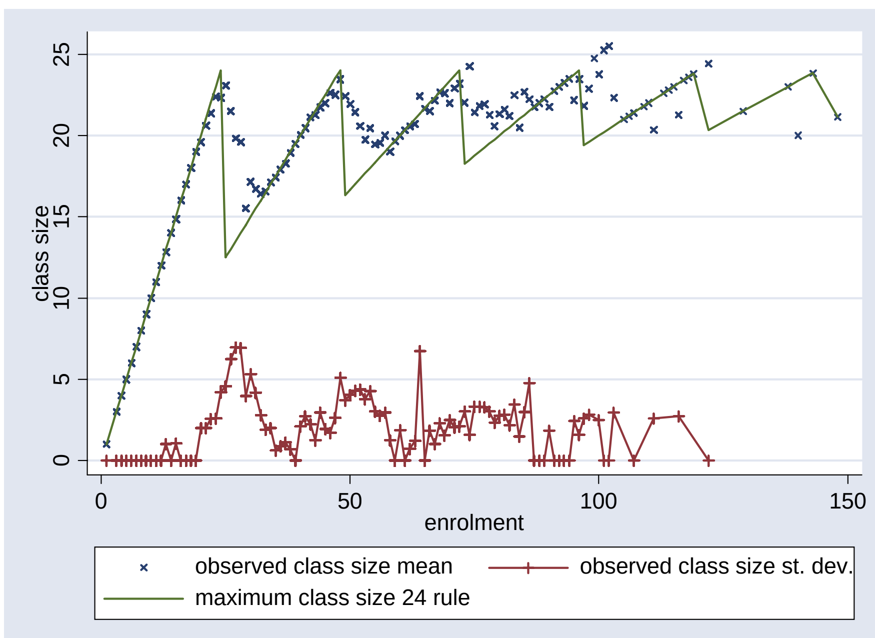


Figure 7. Class size rule and distribution of observed 8th grade class size 2000-2003



5. Estimation

Van der Klaauw (2002) uses a "fuzzy" regression-discontinuity design, albeit in a different context. With fuzzy rather than sharp designs, the treatment rule is non-deterministic. Van der Klaauw substitutes a non-parametric estimate of the conditional expectation of treatment for the endogenous regressor. In the present context, where true class size is not observed but enrolment is, expected class size, calculated from the above rule, is used. Similarly, in the case of students per teacher hour, rule-based expected students per teacher hour, is used. There are two important identifying assumptions:

1. Parents do not exploit administrative rules in order to place their children in schools with smaller classes or fewer students per teacher hour. This is the conditional independence assumption of Hahn, Todd and van der Klaauw (2001). It is usually argued that this seems plausible, as parents could not know which side of a discontinuity their school-grade-year would fall until after having signed up and enrolment was calculated. However, we showed earlier that, at least across the range of fairly low enrolment schools, lagged class size is a good predictor of actual class size. Choosing a class size amounts to choosing a public school catchment area. It seems likely that parents with higher preferences for the outcome are more likely to choose an area where the class size is likely to be low. However, the fixed costs of changing one's catchment area makes switching school, for reasons of a bad draw from the distribution of class sizes, unlikely. Thus, in levels it seems eminently possible that the conditional independence assumption is violated while across sibling differences it is not.

2. In levels, treatment effects are only locally identified at the point where the treatment probability changes discontinuously. This motivates BH to use data close to discontinuities in their analysis. In differences, or within groups, treatment effects are only locally identified where treatment probability differs within group.

In such fuzzy discontinuities there is the potential that our imputed class size will be measured with error and that the sibling difference estimates are then contaminated by sizeable measurement error. Thus, our sibling difference estimates should be regarded as a lower bound because of the attenuation bias they may exhibit. However, the fuzziness that infects our data is considerably smaller in urban areas and we try to push this bound by investigating the class size and teacher hour effects on the outcome in the Copenhagen, Aarhus, and Alborg areas which are a single administrations and our rule should then be exact and measurement error should disappear. We also consider estimates for subsets of the data broken down by the age difference between the siblings to allow for the possibility that municipalities my change the rule that they use.

6. Class Size Results

Here education length is measured up until 11 years after 8th grade. Our enrolment observation window spans the years 1981-1990 and the latest year for which we have educational institution registration is 2001, which dictates that 11 years is the longest time period for which we can be sure to observe all siblings after 8th grade. Thus, our data consists of observations for which we can observe a sibling whose age difference is no more than 9 years. In the first instance, we treat this data as a sample of family averages and the results are reported in Table 3[8]. We report specifications that include log class size, log teacher hours per student, both, and both plus their interaction. In each case we include controls for month of birth, number of classes, and start year. We also report the same specifications but also including a number of additional controls – child gender, whether the child was born after August 1st in the year, and whether the child is the first born child.

There is some multicollinearity between class size (i.e. the number of students in the 8th grade divided by the number of classrooms used by that grade) and teacher hours per student (which is essentially driven by the number of classes in the grade year). This is especially the case in the range immediately around a class size discontinuity since the teacher hours recommendation would then generate a discrete change in teacher hours. Indeed, the teacher hour recommendation is designed to partially compensate for the abrupt changes driven by the class size rule. This prevents BH from estimating the effect of class size, controlling for the number of teachers, so their estimates of class size should be interpreted as the effect of class size net of the effect of the compensation provided by the teacher hour recommendation.

Table 4 shows the same specifications, except that the estimates are for sibling differences.

[8] See Appendix for results that include families with just a singleton child as well as the siblings data used in Table 3. The coefficients here on class size and students/hour of 0.51 and 0.67 become 0.81 and 1.04.

Table 3 Post-compulsory education length: Family Averages

Log Class size	0.6958	-	-0.1644	-0.1500	0.5082	-	0.2052	0.04541
	0.0489		*0.1146*	*0.1305*	*0.0539*		*0.2818*	*0.3067*
Log students/hours	-	0.7485	0.8873	0.8009	-	0.6712	0.4061	0.8598
		0.0456	*0.1069*	*0.3899*		*0.0708*	*0.3708*	*0.4316*
Log Size * Log Students/hours	-	-	-	0.0315	-	-	-	-0.3788
				0.1371				*0.1843*
Male	-	-	-	-	-0.0337	-0.0336	-0.0337	-0.0366
					0.0149	*0.0149*	*0.0149*	*0.0149*
First child	-	-	-	-	1.9115	1.9117	1.9116	1.9117
					0.945	*0.0945*	*0.0945*	*0.0945*
Age 1 August	-	-	-	-	-0.1437	-0.1437	-0.1437	-0.1437
					0.0020	*0.0020*	*0.0020*	*0,.0020*
Intercept	5.0631	7.4717	8.0245	7.9844	28.3563	30.3210	30.3210	28.3677
	0.1466	*0.0204*	*0.3857*	*0.4321*	*0.3815*	*0.3547*	*0.3547*	*1.2726*
R-squared	0.0014	0.0019	0.0019	0.0019	0.0541	0.0541	0.0541	0.0541
# observations	299283	299283	299283	299283	299283	299283	299283	299283
# families	141186	141186	141186	141186	141186	141186	141186	141186

Note: Standard errors in italics.

Table 4 Post-compulsory education length: Sibling differences

Log Class size	-0.0454	-	0.1384	0.1193	-0.0808	-	0.1218	0.1196
	0.0292		*0.0634*	*0.0725*	*0.0314*		*0.0954*	*0.0972*
Log students/hours	-	-0.1241	-0.3047	-0.1651	-	-0.1392	-0.2998	-0.2716
		0.0430	*0.0932*	*0.2733*		*0.0439*	*0.1333*	*0.2764*
Log Size * Log Students/hours	-	-	-	-0.0528	-	-	-	-0.0117
				0.0971				*0.1012*
Male	-	-	-	-	-0.1946	-0.1946	-0.1946	-0.1946
					0.0066	*0.0066*	*0.0066*	*0.0066*
First child	-	-	-	-	0.3319	0.3319	0.3319	0.3319
					0.0099	*0.0099*	*0.0099*	*0.0099*
Age 1 August	-	-	-	-	-0.0518	-0.0519	-0.0518	-0.0518
					0.0010	*0.0010*	*0.0010*	*0.0010*
Intercept	7.2629	7.0730	6.5800	6.6302	15.9141	15.5871	15.1164	15.1187
	0.0876	*0.0187*	*0.2266*	*0.2448*	*0.2099*	*0.1839*	*0.4121*	*0.4126*
R-squared	0.6612	0.6612	0.6612	0.6612	0.6661	0.6661	0.6661	0.6661
# observations	299283	299283	299283	299283	299283	299283	299283	299283
# families	141186	141186	141186	141186	141186	141186	141186	141186

Note: Standard errors in italics.

Our preferred specifications control for observable heterogeneity and, like BH we find substantially larger effects of the resources variables when we do this. However, like them, we find that it is difficult to find robust estimates when we include both class size and students/hour together. The case for including their interaction seems weak since it was never significant. Thus, hereafter, like BH we concentrate on specifications that include only class size and only students/hour and not both together[9].

An advantage of our sibling differences method relative to BH's instrumental variable method is that we do not have to rely on the identifying assumption that class size is not influenced by parents. It seems possible that this is a real problem for their estimates because they exhibit large changes when control variables are added suggesting that the administrative rules are not doing a good job of randomising resources. The suspicion is that there will remain, despite the large number of controls that they include, important unobservable effects that may still cause their IV estimates to be biased. This is a feature of their estimates that apply the rules to their complete 10% of the population sample, but it also applies to their much smaller sub-sample of pupils who are located close to the discontinuities. It is not clear what direction the remaining unobserved heterogeneity would bias the results. If the effect of class size on low ability children is higher than for high ability children, so that high ability children are more robust to large class size, then the bias will be negative. On the other hand, more able parents may have stronger preferences for low size and have more able children, in which case the bias would be positive.

However, a disadvantage of our method is that class size and teacher hours, as generated by the administrative rules, are a "fuzzy" measure of actual resources faced by a particular child. In particular, the actual practice of certain municipalities may differ from the federal rule. Densely populated municipalities will face lower variance in cohort sizes and so be able to adopt a practice that is closer to the national rule than a sparsely populated authority. To explore the sensitivity of the results we re-estimated our models using a variety of assumed maxima and we find, in Tables 5a and 5b, that a critical maximum of 24 actually does produce the most precise estimates.

[9] It should be noted that interactions of resources with gender, first child, and age at school entry were also insignificant, indicating that there are no differential resource effects along these dimensions. Thus, we restrict ourselves to this simple specification in subsequent analysis.

Table 5a *Post-compulsory education length: Sibling differences by Assumed Class Size Maxima*

	21	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Log Class size	-0.0640	-0.0726	-0.1162	-0.0863	-0.1102	-0.0751	-0.0189	-0.0050
	0.0345	*0.0038*	*0.0330*	*0.0325*	*0.0323*	*0.0318*	*0.0309*	*0.0303*
Male	-0.1948	-0.1947	-0.1947	-0.1947	-0.1945	-0.1946	-0.1947	-0.1947
	0.0066	*0.0067*	*0.0067*	*0.0067*	*0.0068*	*0.0068*	*0.0068*	*0.0068*
First child	0.3321	0.3322	0.3322	0.3319	0.3320	0.3321	0.3321	0.3321
	0.0099	*0.0099*	*0.0099*	*0.0099*	*0.0101*	*0.0101*	*0.0101*	*0.0102*
Age 1 August	-0.0519	-0.0519	-0.0519	-0.0519	-0.0519	-0.0519	-0.0519	-0.0519
	0.0011	*0.0011*	*0.0011*	*0.0011*	*0.0011*	*0.0011*	*0.0011*	*0.0011*
Intercept	15.852	15.814	15.998	15.932	16.022	15.907	15.726	15.676
	0.2133	*0.2132*	*0.2119*	*0.2116*	*0.2136*	*0.2132*	*0.2125*	*0.2129*
R-squared	0.5230	0.5303	0.5304	0.5302	0.5266	0.5263	0.5256	0.5236
# obs	299283	299283	299283	299283	299283	299283	299283	299283
# families	141186	141186	141186	141186	141186	141186	141186	141186

Note: Standard errors in italics.

Table 5b *Post-compulsory education length: Sibling differences by Assumed Students/hour Maxima*

	21	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Log (students/ hour)	-0.1279	-0.1359	-0.1786	-0.1392	-0.1526	-0.1279	-0.0574	-0.0458
	0.0480	*0.0461*	*0.0442*	*0.0440*	*0.0432*	*0.0440*	*0.0439*	*0.0442*
Male	-0.1948	-0.1947	-0.1947	-0.1947	-0.1945	-0.1946	-0.1947	-0.1947
	0.0066	*0.0067*	*0.0067*	*0.0067*	*0.0068*	*0.0068*	*0.0068*	*0.0068*
First child	0.3321	0.3322	0.3322	0.3319	0.3320	0.3321	0.3321	0.3321
	0.0099	*0.0099*	*0.0099*	*0.0099*	*0.0101*	*0.0101*	*0.0101*	*0.0102*
Age 1 August	-0.0519	-0.0518	-0.0519	-0.0519	-0.0519	-0.0519	-0.0519	-0.0519
	0.0011	*0.0011*	*0.0011*	*0.0011*	*0.0011*	*0.0011*	*0.0011*	*0.0011*
Intercept	15.583	15.515	15.538	15.587	15.594	15.604	15.640	15.641
	0.1846	*0.1851*	*0.1845*	*0.1840*	*0.1860*	*0.1859*	*0.1861*	*0.1872*
R-squared	0.5303	0.5304	0.5302	0.5266	0.5263	0.5263	0.5256	0.5236
# obs.	299283	299283	299283	299283	299283	299283	299283	299283
# families	141186	141186	141186	141186	141186	141186	141186	141186

Note: Standard errors in italics

Measurement error is, nonetheless, a problem for any sibling difference analysis. The primary source of measurement error in the levels of resources is due to pooling across individual municipalities that choose rules that differ from each other and from the national requirements. Thus, in an attempt to explore how far this lower bound could be pushed, we re-estimated our models for separate large municipalities. This leaves just time variation in the rules as our only remaining source of measurement error in the differences would be due to time variation in the local rules. We minimise this by including time effects. We also investigate the stability of the estimates to the length of the sibling difference - the age gap between siblings. Siblings that are close in age are likely to have experienced less instability in the rules.

Since actual resources are measured with error, sibling differences in resources may be measured with considerable error and this may lead to attenuation bias. Thus, in Tables 6a and 6b, we also estimate our sibling difference model for ten of the largest municipalities and different assumptions about the rule. We can see that our estimated class size and teacher hour inputs do indeed have a larger effect in the single municipality datasets, where there is little or no measurement error, than in the complete datasets where we have, undoubtedly incorrectly, assumed that the same maximum class size applies to all municipalities. It also seems to be the case that a 25 class maximum rule may be more appropriate, at least for larger municipalities. Our estimates class size effects are now much larger than the -0.09 figure from Tables 4 and 5a. Similarly, the students/hour effects are also much larger compared to -0.14 in Tables 4 and 5b.

One might argue that, even within a municipality, the practice may have changed over time and leave our sibling differences remain contaminated by some measurement error associated with changes in rules within municipalities. Thus, in Tables 7a and 7b, we estimate the models again by cutting the data into siblings whose age difference is 1, 2, 3….9 years. The group with the larger age difference faces a larger probability of the maximum class size practice having changed between siblings and so be more subject to measurement error in the class size change. Thus, we would expect the input effects to be subject to less attenuation bias, as so be larger, in the short difference case. This turns out to be the case: the Copenhagen estimates here should be compared to -0.11 for class size and -0.17 for students/hour; while the Aarhus figures should be compared to -0.26 and -0.47.

Table 6a Education length sibling differences estimates by largest 10 municipalities and different assumed class size maxima: Log class size

	24		25		26		Families	Observations
Copenhagen	-0.1056	*0.1679*	-0.2081	*0.1686*	0.0044	*0.1671*	13056	6336
Aarhus	-0.2635	*0.1634*	-0.6813	*0.1669*	-0.2128	*0.1616*	8973	4733
Odense	0.0315	*0.2031*	-0.4620	*0.2007*	-0.2043	*0.1965*	8530	4175
Aalborg	-0.2020	*0.2039*	-0.3493	*0.1924*	-0.5495	*0.1859*	8165	3984
Esbjerg	-0.9794	*0.2525*	-1.3146	*0.2545*	-0.9403	*0.2315*	4742	2303
Herning	0.4475	*0.3335*	0.2011	*0.3156*	0.2608	*0.2723*	4160	1977
Kolding	-0.3172	*0.2936*	-0.5703	*0.2927*	-0.4000	*0.2947*	3496	1723
Horsens	-0.0422	*0.3697*	0.1432	*0.3886*	-0.0341	*0.3883*	3267	1602
Silkeborg	-1.0884	*0.3590*	-1.0891	*0.3189*	-0.8916	*0.3114*	3136	1522
Randers	-0.1040	*0.3180*	-0.1773	*0.3168*	-0.1976	*0.3121*	3117	1524

Note: Standard errors in italics.

Table 6b Education length sibling differences estimates by largest 10 municipalities and different assumed class size maxima; Log students/hour

	24		25		26		Families	Observations
Copenhagen	-0.1733	*0.2281*	-0.2546	*0.2278*	-0.1233	*0.2308*	13056	6336
Aarhus	-0.4697	*0.2148*	-0.8724	*0.2179*	-0.4109	*0.2196*	8973	4733
Odense	-0.2093	*0.2832*	-0.6955	*0.2735*	-0.4649	*0.2749*	8530	4175
Aalborg	-0.2975	*0.2823*	-0.4726	*0.2634*	-0.8126	*0.2687*	8165	3984
Esbjerg	-1.3910	*0.3419*	-1.7004	*0.3420*	-1.2751	*0.3251*	4742	2303
Herning	0.6124	*0.4395*	0.3371	*0.4145*	0.3229	*0.3740*	4160	1977
Kolding	-0.6008	*0.3945*	-0.8749	*0.3933*	-0.6963	*0.4022*	3496	1723
Horsens	-0.1870	*0.5184*	0.0795	*0.5225*	-0.0141	*0.5519*	3267	1602
Silkeborg	-1.4274	*0.4829*	-1.5056	*0.4345*	-1.2418	*0.4262*	3136	1522
Randers	-0.3316	*0.4458*	-0.3414	*0.4271*	-0.5235	*0.4458*	3117	1524

Note: Standard errors in italics

Table 7a *Post-compulsory education length: Sibling differences: Copenhagen*

Max age difference	Log class size		Log (students/hour)		# observations	# families
1	-1.5960	*0.4891*	-2.1839	*0.6818*	1567	782
2	-0.5173	*0.2614*	-0.6392	*0.3549*	5212	2560
3	-0.3583	*0.2037*	-0.4280	*0.2759*	8904	4348
4	-0.2563	*0.1817*	-0.3213	*0.2475*	11126	5402
5	-0.0756	*0.1747*	-0.1037	*0.2373*	12163	5896
6	-0.1271	*0.1705*	-0.1770	*0.2315*	12679	6145
7	-0.1075	*0.1687*	-0.1671	*0.2292*	12950	6281
8	-0.1106	*0.1680*	-0.1762	*0.2283*	13024	6325
9	-0.1056	*0.1679*	-0.1733	*0.2281*	13056	6336

Note: Standard errors in italics.

Table 7b *Post-compulsory education length: Sibling differences: Aarhus*

Max age difference	Log class size		Log (students/hour)		# observations	# families
1	-0.5383	*0.3957*	-0.8052	*0.5188*	1707	879
2	-0.1773	*0.2349*	-0.3850	*0.3080*	4278	2214
3	-0.2450	*0.1916*	-0.4748	*0.2529*	6530	3402
4	-0.3216	*0.1757*	-0.6173	*0.2323*	7878	4125
5	-0.1952	*0.1680*	-0.3826	*0.2209*	8490	4459
6	-0.1861	*0.1652*	-0.3753	*0.2170*	8776	4618
7	-0.2580	*0.1641*	-0.4643	*0.2157*	8908	4693
8	-0.2446	*0.1635*	-0.4448	*0.2150*	8950	4720
9	-0.2635	*0.1634*	-0.4697	*0.2148*	8973	4733

Note: Standard errors in italics.

A further issue for us (and BH) is that of censoring in education length. We are taking data that is at least 11 years post grade 8 and no older than 20 years post grade 8, i.e. between the ages of 25 and 35. Many (15%) observations remain in education beyond even the age of 25 and so there is some censoring in the data. Table 8 presents the headline coefficients, using just the specifications that contains class size and teacher hours, for education length measured up to different numbers of years after the beginning of 8th grade. That is, this table acknowledges that there is censoring in our education length data – we only observe completed education for those whose education is less than the 2001 minus the year that they were in 8th grade. This could be as large as 20 years for those in 8th grade in 1981 and as little as 10 for those who took 8th grade in 1990. So, since many students do not complete their education until even older than 25, there is certainly some censoring in this data and the table shows the effects of resources on completed education using subsets of the data with different degrees of censoring. The distribution of the dependent variable for Table 8 is shown in Figure 8.

The first row corresponds to observations where individuals are followed until just one year out of 8th grade and subsequent years are ignored. Row 11 follows individuals up until 11 years out of 8th grade. This is the last year for which we can, with certainty, observe all members of the family for the same number of years. Row 12 may contain families with a mixture of some individuals for 12 years (8th graders 1981-1989) and perhaps one for 11 years who was an 8th grader in 1990. Therefore, row 11 is the last row without differential censoring within family. The last row follows individuals up until at most 20 years out of 8th grade. Here only those in 8th grade in 1981 are observed 20 years later in 2001, those in 8th grade in 1981 are observed 19 years later in 2001, etc.

A further concern in research based on sibling differences is family size. Table 9 shows estimations performed separately by number of siblings. Perhaps unsurprisingly in the light of the larger class size differences that we saw in larger households, it can be seen that our estimated resource effects are essentially being driven by 3 and 4 sibling households.

Finally, Table 10 investigates the importance of our chosen dependent variable. The presumption in our earlier results (and in BH) is that the outcome of interest is the number of years of post-compulsory schooling reflected in age at which individuals leave the education system. In fact, many young Danes take a break in their education, usually, between upper secondary and higher education. Moreover, there is a significant variance in the duration of secondary education even controlling for 3 or 5 year degree. In Table 10 we redefine the dependent variable to be a dummy variable which takes the value 1 if the individual had at least the indicated number of years of post-compulsory education. The mean shows that almost all Danes have some post-8th grade education while 25% end at or before 12th grade. Class size does seem to have a significant effect of getting to at least to 12th grade and there also seems to be an effect much later corresponding to the distinction between 3 year and 5 year degree. Students/hour does seem to have a beneficial effect throughout.

Figure 8. Distribution of education length for different observations windows after 8th grade

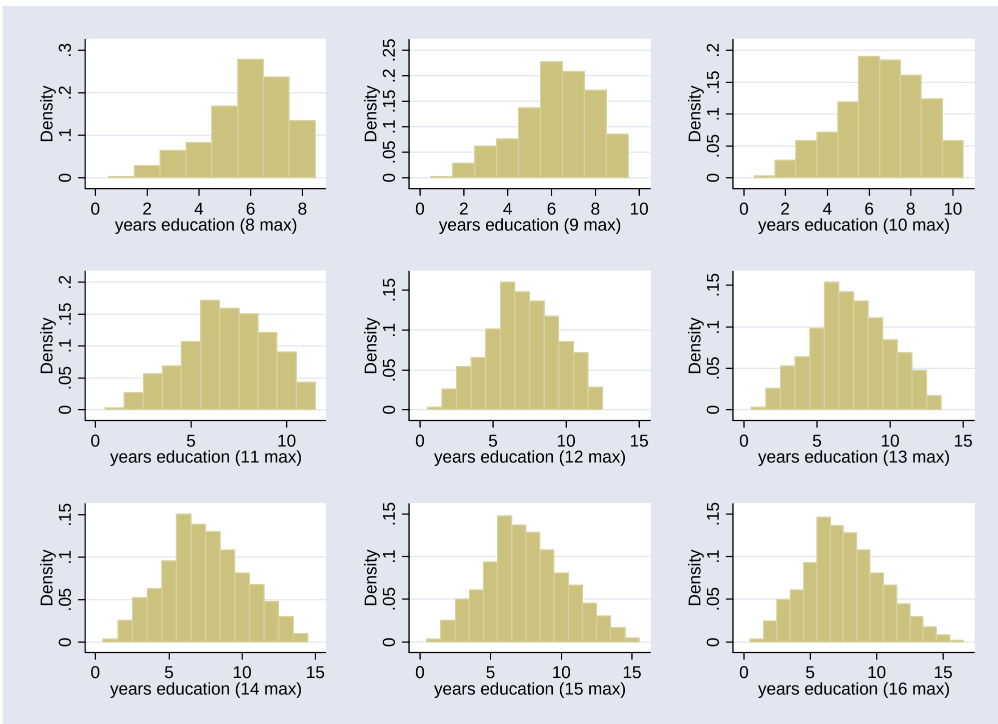


Table 8 *Coefficients on class size and students/hour by number of years after which the education length is censored*

	No additional controls				With additional controls			
# years later	class size		students/hour		class size		students/hour	
1	0.0011	*0.0015*	0.0021	*0.0021*	0.0018	*0.0014*	0.0036	*0.0021*
2	0.0074	*0.0043*	0.0128	*0.0061*	0.0092	*0.0040*	0.0126	*0.0059*
3	0.0003	*0.0078*	0.0065	*0.0109*	0.0024	*0.0072*	0.0060	*0.0106*
4	-0.0197	*0.0113*	-0.0222	*0.0158*	-0.0138	*0.0105*	-0.0292	*0.0154*
5	-0.0271	*0.0150*	-0.0270	*0.0209*	-0.0358	*0.0139*	-0.0802	*0.0205*
6	-0.0376	*0.0181*	-0.0497	*0.0253*	-0.0452	*0.0169*	-0.1033	*0.0248*
7	-0.0515	*0.0212*	-0.0779	*0.0296*	-0.0517	*0.0197*	-0.1321	*0.0289*
8	-0.0617	*0.0240*	-0.0998	*0.0335*	-0.0538	*0.0222*	-0.1470	*0.0327*
9	-0.0661	*0.0266*	-0.1131	*0.0371*	-0.0531	*0.0247*	-0.1548	*0.0362*
10	-0.0673	*0.0292*	-0.1180	*0.0407*	-0.0575	*0.0271*	-0.1689	*0.0398*
11	-0.0809	*0.0315*	-0.1392	*0.0440*	-0.0455	*0.0293*	-0.1241	*0.0430*
12	-0.1019	*0.0335*	-0.1729	*0.0467*	-0.0413	*0.0311*	-0.0998	*0.0458*
13	-0.1250	*0.0350*	-0.2078	*0.0489*	-0.0353	*0.0326*	-0.0664	*0.0480*
14	-0.1424	*0.0361*	-0.2325	*0.0504*	-0.0254	*0.0337*	-0.0289	*0.0496*
15	-0.1540	*0.0368*	-0.2500	*0.0514*	-0.0250	*0.0344*	-0.0196	*0.0506*
16	-0.1544	*0.0372*	-0.2511	*0.0520*	-0.0152	*0.0349*	0.0024	*0.0513*
17	-0.1562	*0.0374*	-0.2534	*0.0523*	-0.0094	*0.0351*	0.0211	*0.0516*
18	-0.1579	*0.0376*	-0.2550	*0.0524*	-0.0040	*0.0352*	0.0428	*0.0517*
19	-0.1595	*0.0376*	-0.2568	*0.0525*	0.0017	*0.0352*	0.0624	*0.0518*
20	-0.1595	*0.0376*	-0.2568	*0.0525*	0.0017	*0.0352*	0.0624	*0.0518*

Note: Standard errors in italics.

Table 9 Education Length Model Estimates and standard errors: by number of siblings

2 siblings				
Log class size	-0.0006	*0.0360*		
Log students/hour			-0.0192	*0.0488*
Male	-0.1630	*0.0073*	-0.1630	*0.0073*
First child	0.4289	*0.0123*	0.4289	*0.0123*
Age August 1	-0.0538	*0.0012*	-0.0538	*0.0012*
Intercept	15.8818	*0.2346*	15.8711	*0.2042*
R-squared	0.6800		0.6800	
# obs/families	251050 / 125525		251050 / 125525	
3 siblings				
Log class size	-0.3534	*0.0802*		
Log students/hour			-0.5172	*0.1078*
Male	-0.3142	*0.0169*	-0.3142	*0.0169*
First child	0.1721	*0.0267*	0.1721	*0.0267*
Age August 1	-0.0510	*0.0026*	-0.0510	*0.0026*
Intercept	16.5896	*0.5246*	15.2047	*0.4554*
R-squared	0.5960		0.5961	
# obs/families	43578 / 14526		43578 / 14526	
4 siblings				
Log class size	-0.7945	*0.2508*		
Log students/hour			-0.9837	*0.3315*
Male	-0.4175	*0.0544*	-0.4174	*0.0544*
First child	0.0641	*0.0980*	0.0624	*0.0980*
Age August 1	-0.0512	*0.0080*	-0.0514	*0.0080*
Intercept	17.6174	*1.6299*	14.5965	*1.4059*
R-squared	0.5802		0.5801	
# obs/families	4136 / 1034		4136 / 1034	
5 siblings				
Log class size	-0.7120	*0.7594*		
Log students/hour			-0.7349	*0.9478*
Male	-0.2217	*0.1727*	-0.2247	*0.1727*
First child	0.4031	*0.3655*	0.4104	*0.3658*
Age August 1	-0.0716	*0.0243*	-0.0722	*0.0243*
Intercept	21.1268	*4.8543*	18.5644	*4.2663*
R-squared	0.5685		0.5683	
# obs/families	435 / 87		435 / 87	
6 siblings				
Log class size	-2.2225	*1.9279*		
Log students/hour			-2.6249	*2.2470*
Male	0.5417	*0.4190*	0.5443	*0.4181*
First child	2.1489	*1.1142*	2.1459	*1.1140*
Age August 1	0.0314	*0.0063*	0.0297	*0.0661*
Intercept	2.8710	*12.2264*	-5.2294	*11.3748*
R-squared	0.6948		0.6949	
# obs/families	84 / 14		84 / 14	

Note: Standard errors in italics.

Table 10 Sibling difference linear probability model on "at least" years of post compulsory schooling

Dep var:	Mean of dep var	Log class size			Log students/hour		
		Coeff	Std error	R squared	Coeff	Std error	R squared
1 more years	0.9701	-0.0019	0.0025	0.5725	-0.0012	0.0034	0.5725
2 more years	0.9157	-0.0136	0.0040	0.5960	-0.0168	0.0055	0.5960
3 more years	0.8497	-0.0149	0.0052	0.6026	-0.0206	0.0070	0.6026
4 more years	0.7482	-0.0098	0.0063	0.5936	-0.0155	0.0086	0.5936
5 more years	0.5881	-0.0040	0.0071	0.6015	-0.0096	0.0096	0.6015
6 more years	0.4401	-0.0116	0.0071	0.6076	-0.0215	0.0096	0.6076
7 more years	0.3038	-0.0041	0.0067	0.5985	-0.0092	0.0090	0.5985
8 more years	0.1859	-0.0058	0.0058	0.5832	-0.0133	0.0078	0.5832
9 more years	0.1001	-0.0123	0.0046	0.5602	-0.0199	0.0062	0.5602
10 more years	0.0283	-0.0070	0.0026	0.5195	-0.0091	0.0036	0.5195

7. Costs and Benefits of Reducing Class Sizes

Above we have identified the effect of class size at 8th grade on the length of completed education. An overview of the results would suggest that the coefficient on log class size would be about -0.3 and on log students per teacher hour about -0.5. There are two further difficulties in turning this result into a cost-benefit analysis of a class size reduction policy.

First, our estimate is interpreted as the effect of a rule-induced unit increase in class size at 8th grade given *the correlation that exists in class size across grades*. Other things being equal, we would expect a higher class size at grade 8 to be associated with a higher class size at grade 7. Thus, we need to investigate the correlation in class sizes across grades to be able to say what our estimate is an estimate of. If there is no correlation across grades in class size then our estimate is the effect of raising class size in grade 8 alone. If there is a perfect correlation across years then our estimate is an estimate of the effect of an increased class size every grade. These two extremes will have very different cost implications. In the sub-section below we investigate the cross grade correlations in class size (and teacher hours).

Secondly, to compute the benefits we need to know how variations in length of completed schooling affects subsequent earnings. This is the subject of the literature on the returns to education. Card (1999) reviews the literature with special attention on the issues of ability bias and bias due to measurement errors in education. Ability bias may arise in least squares estimates because the effect of education on wages is contaminated by the correlations that are thought to exist between omitted ability and both education and wages. Since ability is thought to be positively correlated with wages and positively correlated with education this implies that least squares estimates of the returns to education are biased upwards. On the other hand, measurement error in education causes attenuation in the least squares estimates – that is, it biases the estimated return downwards. Card declared that evidence from twins data represents the "gold standard", although other researchers have expressed some reservations about the appropriateness of twins data for this purpose[10].

[10] See Neumark (1999) and Bound and Solon (1999).

7.1 The correlation between class size across grades

The data used above contained length of completed education but told us only about the rules used to generate class size. However, more recent data available for 8th graders in 2002 to 2004 tells us about their actual class size in 8th grade and in their previous two grades[11]. Table 11 shows the correlations between each pair of two consecutive years for this data. There is a close correlation (typically between 0.7 and 0.8) for each adjacent grade up to grade 8. Thereafter, there is a much weaker correlation because this is the point where students begin to switch to higher secondary schooling which typically involves a change of school. Table 12 shows similar figures for pupils per teacher hour and enrolment.

Table 11 Correlation between class sizes across adjacent grades

Grade:	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10
Class size	0.79	0.84	0.85	0.84	0.85	0.83	0.76	0.59	0.57	0.21
Teacher hours	0.90	0.93	0.93	0.93	0.92	0.91	0.88	0.76	0.78	0.39
Enrolment	0.97	0.98	0.98	0.98	0.98	0.97	0.89	0.80	0.94	0.37

Thus, we find that the correlation between class size at grade 8 and at grade 7 is 0.57, and between grades 7 and 6 is 0.59, etc. The implication is that a unit increase in class size at grade 8 also implies a 0.57 larger class size at grade 7, a 0.34 larger class size at grade 6, and so on. Thus a unit increase at grade 8 is associated with a cumulative increase of 3.996 units across all 10 grades, and so this would be equivalent to a 0.4 class size increase in every grade in lower secondary schooling. The corresponding cumulative figure for pupil per teacher hour is 9.968 across all grades or, approximately, a unit increase per grade.

7.2 The effect of completed education on wages

Card's assertion that evidence from data on twins represents the "gold standard" on the ability bias issue was motivated by the view that within twin differencing removes that bias. On the other hand differencing exacerbates measurement error and

[11] Unfortunately we have little information on length of completed education in this sample because few have yet completed. Thus, we assume that the correlation between grades in this recent data can be applied to our older data.

the innovation in Ashenfelter and Krueger (1994) was to use one twin's cross-reported education as an instrument for the other twin's education to eliminate this problem. Neumark (1999) notes that if differencing does not remove all of the omitted ability then the within-twin estimator may still be biased, and may even be more biased than least squares applied to the individuals. Moreover, Neumark is concerned about the non-classical nature of the measurement error when education is constructed from qualifications information which undermines the value of IV. Finally, Bound and Solon (1999) is concerned that differences in schooling are themselves endogenous.

The alternative to twins is instrumental variables. Card (1999) lists several studies that use instruments for education to deal with both measurement error and endogeneity induced by ability bias. However, Angrist and Imbens (1994) note that, in the context of a model where the returns to education is a random parameter, IV provides an unbiased estimate only of a *local* average treatment effect – that is, the effect of education on those individuals whose education has been affected by the instrument. For some policy purposes and some instruments this may an appropriate parameter but, in general, it will not be informative. In contrast, the twins method provides an estimate of the average return across the population (of twins, at least) which is what we require for our analysis here. Thus, here we adopt the twins method.

Our data is a sample of twins is constructed from matching children to their mothers, identifying which children have the same mothers, and which of those have the same date of birth. One advantage of register data over survey data is that education is not self-reported but, rather, is the official record of the individual's activity. In the Danish case this is recorded as the month of completing education so, in addition to their being no recall problem, rounding errors are likely to be small and we therefore feel able to ignore the measurement error issue.

One shortcoming of our data here is that we do not have the zygosity indicator and so cannot tell which twins are identical (MZ) and which are fraternal (DZ). However, we do know that different gender twins are necessarily DZ and the proportion of same sex twins which are MZ is about 50%[12]. Thus, if all of ability bias is genetically

[12] In other Danish data we find that 22% of all twins (alive at 1970 or later) are MZ, 31% are DZ of different sex and 38% are DZ of the same sex, with the remainder being missing, triplets or quads.

determined, then we would expect our same sex twins to exhibit half of the bias that we would get from estimates where we treat the twins as individuals. Thus, we apply OLS to the individual data and then to the twin differences and can infer the MZ estimate by adding half the difference between the individual and within-twin estimates to the individual estimates.

Table 12 reports the results[13] where the dependent variable is the log of annual labour earnings at age 26 for those that report earnings. Since we assume that age-earnings profiles are parallel this estimate at age 26 is sufficient to compute the present value across the lifecycle. The female estimates are both 5% and so suggest no ability bias and so we infer that the MZ estimate, which we think of as the average causal effect, would also be 5%. In contrast, the male equation suggests that the ability bias in OLS is approximately 3% (double 0.031-0.016) which, when added to the OLS estimate, we infer the MZ male estimate would also be approximately 5%.

An important assumption in this specification is that age - log wage profiles are parallel across the lifecycle. Thus, to compute the present value of the gain from additional education we need to know what the shape of age–log wage profiles are. In Table 13 we provide estimates of a regression of age-specific average log annual earnings against a quadratic in age.

Table 12 Estimated Education Returns

	Male	Female
OLS	0.016 (0.006)	0.050 (0.006)
Within-twin	0.031 (0.008)	0.051 (0.009)
Sample sizes	2542/1271	2150/1075

Note: Figures in parentheses are standard errors.

Table 13 Estimated Age Earnings Profiles

	Male	Female
Age	0.1552 (0.0011)	0.1356 (0.0013)
Age squared	-0.00188 (0.00002)	-0.00161 (0.00002)
Sample size	250054	214209

Note: Figures in parentheses are standard errors.

[13] We omit the mixed gender twins because we cannot identify sex specific rates of return from such data.

7.3 Costs and benefits

The only previous study to have conducted a cost-benefit analysis of class size reduction is Krueger (2004) who exploits findings from the STAR experiment. In our analysis we take the estimates from above and infer the effective average change in class size that would be associated with a 1 unit change at grade 8 and, under certain assumptions, we can estimate the cost of providing a unit change in class size across all grades.

We assume that the new teachers required would, in steady state, cost as much per unit as the stock of existing teachers. We assume that there is an infinitely elastic supply of teachers at existing wage rates to facilitate this expansion and we assume that there would be no additional costs besides the teachers. The average annual cost of a pupil year of lower-secondary education in 2002 is Dkkr 51,300 and approximately 80% of this is accounted for by teaching staff, according to Ministry of Education (2000). Reflating this, and the corresponding figure for higher education, by the rise in the cost of living to January 2005 we get Dkkr 53,900 for a secondary school child year, and Dkkr 57,680 for a higher education student year.

Table 11 allows us to infer that the estimates of grade 8 class size imply a cumulative class size effect that is equivalent to a change in class size across all grades of 3.996. Similarly the inferred teacher hour effect is equivalent to a change of 1.00 across all grades.

Since the average class size is close to 20 a convenient and realistic policy to consider would be a 5% reduction in class size but our analysis is linear and our estimated effects can simply be scaled for any other change[14].

Our estimate of the grade 8 log class size effect is about -0.3 which implies that a class size effect of approximately -0.015 since a typical class size is about 20. Moreover a unit increase in grade 8 is, in the data, associated with a cumulative difference across all grades that is equivalent to about a 0.4 increase in class size at all grades. So, effectively, our estimates are estimates of the effect of increasing class size in all grades of 0.4. That is, we multiply our coefficient of -0.015 on class size by 2.5 we find that we imply that a one unit change in average class size across all grades

[14] Krueger (2004) considers the effects or reducing class size from 22 to 15 since this is what the STAR experiment did for the treatment group.

would raise average length of education by 0.0375 years. This additional class size raises the costs of providing the compulsory schooling through the larger teaching inputs, and it also raises costs through additional post-compulsory schooling costs because it extends the level of average post compulsory schooling. This latter effect is simply 0.0375 multiplied by Dkkr 57,680 – that is, Dkkr 2163 per student which needs to be discounted back to the start of grade 0 at age 6 - that is, by 16 years from the average age of leaving education of 22 . The former effect is 5% of 80% of Dkkr 53,900 – an annual flow of Dkkr 2156 per child per grade of compulsory schooling and so this also needs to be discounted to grade 0.

An additional cost of this additional education length is the opportunity costs of a fall of 0.0375 years worth of earnings which we cost at the average earnings for a 22 year old education leaver of Dkkr 177,000 for men and Dkkr 138,000 for women, to give Dkkr 6638 for men and Dkkr 5175 for women, which again need to be discounted back to grade 0.

Table 15 reproduces the analysis in Table 5 of Krueger using the same range of discount rates and annual rates of productivity growth. We assume that 1 unit increase in class size at all grades 1-9 raise the average length of completed schooling from by 0.0375 years and decreases the date at which earnings start by the same amount. We assume that there is a 5% effect of one year of schooling on annual earnings and that retirement occurs at 63. Real wages grow at some assumed rate of productivity and follow the age earnings profile given by Table 13.

If the return to education were twice as high at 0.10 (or the effects of class size were double at 0.075), then with the productivity growth of 2%, the ratio of benefits to costs is 2.17 for men and 0.66 for women with no discounting, and 0.68 for men and 0.25 for women if the discount rate is 0.04.

Table 15 Ratio of Discounted Present Value of Benefits to Costs of Reducing Class size by 5% (2005 Dkkr) per child

Discount rate	Increase in income assuming annual productivity growth of:							
	0		1%		2%		3%	
	M	F	M	F	M	F	M	F
0.00	1.07	0.32	1.08	0.33	1.09	0.33	1.10	0.33
0.02	0.58	0.19	0.58	0.19	0.59	0.20	0.81	0.26
0.04	0.34	0.12	0.34	0.12	0.34	0.12	0.35	0.13
0.06	0.20	0.08	0.20	0.08	0.21	0.08	0.21	0.08

Note: Rate of return to education assumed to be 0.05.

8. Conclusions and Further Research

Our sibling differences analysis suggests that class size and students per teacher hour rules do have statistically and economically significant effects on education length in Denmark. Smaller classes and more teacher hours have been shown to increase length of education. An overview of our results suggest that it is reasonable to assume that reducing class size during compulsory schooling by 5% (about a unit reduction from the current mean class size) would increase mean length of education by about 0.0375 years (about 8 days) - which is about a one per cent change in the length of post-compulsory schooling . According to our estimated returns to education this translates to approximately a 0.2% increase in lifetime earnings which, undiscounted, amounts to approximately Dkkr 30,000 for an average man and around half of that for an average women. The undiscounted costs, including the opportunity costs, of such a policy seem likely to be approximately Dkkr 30,000 per person. When discounted, these figures seem somewhat more pessimistic that even the relatively modest net benefits in Krueger (2003).

There are several avenues for development of this work. Firstly, individual 9th grade test scores and teacher assessments for the years 2002-4 has recently been made available. This will enable us to place our measures in the wider literature on immediate test score outcomes. Moreover, this data contains information on actual class size so we would be able to exploit the rules as instrumental variables to contrast IV results with sibling differences. However, this data is too recent to enable us to look also at completed education length.

Secondly, while our sibling differences controls for unobservable family effects, and limiting ourselves to siblings that attended the same school allows us to control for school *fixed* effects, we have not exploited the information that we have about *peer* parental background which is not removed by differencing even holding the school constant. In particular, we would like to know the effect of being young or old, or more or less able, relative to the average class member, since teachers may focus their attention on the average or, alternatively, teachers might focus on the youngest, or most able. Moreover, we would like to identify the effects of the parental backgrounds of other children in the class - for example, the proportion with working mothers, or the distribution of parental education levels.

References

Angrist, Joshua and Guido Imbens (1994) “Identification and Estimation of Local Average Treatment Effects”, Econometrica, 62, 467-75.

Angrist, Joshua and David Lavy (1999) "Using Miamonides' rule to estimate the effect of class size on scholastic achievement", Quarterly Journal of Economics, 114, 533-575.

Bound, John and Gary Solon (1999), “Double trouble: on the value of twin-based estimation of the return to schooling”, Economics of Education Review, 18, 169-182.

Browning, Martin and Eskil Heinesen (2004) "Class size, teacher hours and educational attainment", CAM Discussion Paper.

Card, David (1999), “The causal effect of education on wages”, in O. Ashenfelter and D. Card (eds), Handbook of Labor Economics Vol 3, Elsevier, Amsterdam.

Carneiro, Pedro and James Heckman (2003), “Human Capital Policy” in James Heckman and Alan Krueger (eds.), Inequality in America, MIT Press.

Case, Anne and Angus Deaton (1999) "School inputs and educational outcomes in South Africa", Quarterly Journal of Economics, 114, 1047-1084.

Danish Ministry of Education and Research (1993) "Facts and Figures (in Danish)", Danish Ministry of Education and Research: Copenhagen.

Danish Ministry of Education (2002) "Education key facts and figures for 2002", Danish Ministry of Education: Copenhagen.

Graversen, Brian and Eskil Heinesen (1999) "Resource use in public schools: Differences between municipalities" (in Danish), memo from Danish Institute for Local Government Studies.

Graversen, Brian and Eskil Heinesen (2005) "The effect of school resources on educational attainment: Evidence from Denmark" forthcoming in Bulletin of Economic Research.

Hahn, Jinjong, Petra Todd and Wilbert van der Klaauw (2001) "Identification and estimation of treatment effects with a regression-discontinuity design" Econometrica, 69, 201-209.

Hanushek, Eric (2003) "The failure of input-based schooling policies", Economic Journal, 113, F64-F98.

Hanushek, Eric (2002) "Publicly provided education” in Alan Auerbach and Martin Feldstein (eds.) Handbook of Public Economics, Volume 4, Elsevier, Amsterdam.

Hoxby, Caroline (2000) "The effects of class size on student achievement: New evidence from population variation", Quarterly Journal of Economics, 115, 1239-1285.

Isacsson, Gunnar (2004), “Estimating the economic returns to educational levels using data on twins”, Journal of Applied Econometrics, 19, 99-119.

Krueger, Alan (1999) ""Experimental estimates of education production functions" Quarterly Journal of Economics, 114, 497-532.

Krueger, Alan (2003) "Economic considerations and class size", Economic Journal, 113, F34-F63.

Lazear, Edward (2001) "Educational production" Quarterly Journal of Economics, 116, 777-803.

Neumark, David (1999), “Biases in twin estimates of the return to schooling”, Economics of Education Review, 18, 143-148.

Rangvid, Beatrice (2002) "Evaluating private school quality in Denmark", mimeo Danish Institute of Local Government Studies.

Heinesen, Eskil and Beatrice Rangvid (2003) "Class size effects on early career employment", mimeo Danish Institute of Local Government Studies.

van der Klaauw, Wilbert (2002) "Estimating the effects of financial aid offers on college enrolment: A regression-discontinuity approach" International Economic Review, 43, 1249-1287.

Woessmann, Ludger and Martin West (2002) "Class-size effects in school systems around the world: Evidence from between-grade variation in TIMSS", IZA Discussion Paper 485.

Appendix

Table 3A *Post-compulsory education length: Family averages only singletons*

Log Class size	1.4698	-	-1.4388	-1.3647	0.8142	-	-0.1954	-0.2052
	0.0177		*0.0460*	*0.0494*	*0.0194*		*0.0884*	*0.0971*
Log students/hour	-	1.7044	2.9476	2.5551	-	1.0368	1.2709	1.2605
		0.0166	*0.0431*	*0.1047*		*0.0238*	*0.1085*	*0.1167*
Log Size*Student/hr	-	-	-	0.1588	-	-	-	0.0125
				0.0386				*0.0514*
Male	-	-	-	-	-0.0337	-0.0811	-0.0807	-0.0807
					0.0149	*0.0062*	*0.0062*	*0.0062*
Age 1 August	-	-	-	-	-0.1437	-0.1206	-0.1204	-0.1204
					0.0020	*0.0008*	*0.0008*	*0.0008*
Intercept	2.4024	7.5635	12.417	12.229	28.356	31.456	34.416	34.455
	0.0528	*0.0084*	*0.1566*	*0.1621*	*0.3815*	*1.1045*	*1.1994*	*1.1010*
R-squared	0.0109	0.0167	0.0182	0.0183	0.0681	0.0691	0.0692	0.0692
# observations	392010	392010	392010	392010	392010	392010	392010	392010

Note: Standard errors in italics.