

UNIVERSITY OF COPENHAGEN



Class Size, Teacher Hours and Educational Attainment

Browning, Martin; Heinesen, E.

Publication date:
2003

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Browning, M., & Heinesen, E. (2003). *Class Size, Teacher Hours and Educational Attainment*. Kbh.: Department of Economics, University of Copenhagen.



CAM

**Centre for Applied
Microeconometrics**

**Institute of Economics
University of Copenhagen**

<http://www.econ.ku.dk/CAM/>

Class size, teacher hours and educational attainment

Martin Browning
Eskil Heinesen

2003-15

Class size, teacher hours and educational attainment

Martin Browning

Centre for Applied Microeconometrics, University of Copenhagen
Studiestræde 6, DK-1455 Copenhagen K, Denmark; email: martin.browning@econ.ku.dk

Eskil Heinesen

AKF, Institute of Local Government Studies B Denmark
Nyropsgade 37, DK-1602 Copenhagen V, Denmark; email: esh@akf.dk

May 2003

Abstract

We employ a regression-discontinuity design to identify effects on educational attainment (years of education) of class size and the number of pupils per weekly teacher hour using administrative rules as instruments. We use a Danish administrative panel data set based on a 10% random sample of eight cohorts of pupils (and their parents). Restricting the sample to observations close to enrollment discontinuity points, instrumental variables estimates are consistently negative. Although the estimates are not very precisely determined, the point estimates indicate rather large positive effects of reducing class size, and especially of reducing the number of pupils per teacher hour.

Key words: School quality; Class size; Educational attainment; Instrumental variables; Regression-discontinuity design

JEL classification numbers: I2, I21

The Danish National Research Foundation is gratefully acknowledged for its support of this project, both directly and through its support of Centre for Applied Microeconometrics.

Class size, teacher hours and educational attainment

May 2003

Abstract

We employ a regression-discontinuity design to identify effects on educational attainment (years of education) of class size and the number of pupils per weekly teacher hour using administrative rules as instruments. We use a Danish administrative panel data set based on a 10% random sample of eight cohorts of pupils (and their parents). Restricting the sample to observations close to enrollment discontinuity points, instrumental variables estimates are consistently negative. Although the estimates are not very precisely determined, the point estimates indicate rather large positive effects of reducing class size, and especially of reducing the number of pupils per teacher hour.

Key words: School quality; Class size; Educational attainment; Instrumental variables; Regression-discontinuity design

JEL classification numbers: I2, I21

The Danish National Research Foundation is gratefully acknowledged for its support of this project.

1. Introduction

At some level one must expect that school inputs such as expenditure per pupil, class size, pupil-teacher ratio, etc., which are directly affected by government policy, have significant effects on pupils' educational and labour market outcomes. But this is nevertheless controversial in the literature. Based on his influential surveys of the literature, Hanushek (1986, 1996) concludes that measured school resources are not systematically related to pupil outcomes. However, this conclusion has been criticized, see for instance Card and Krueger (1996a, 1996b), Hedges and Greenwald (1996), and Krueger (2003).

A major difficulty in estimating effects of school inputs on pupil outcomes is the potential endogeneity of school resources. A large fraction of the variation in school inputs is the result of choices made by parents, school administrators, teachers, and politicians at both local and national levels. For instance, parents may choose to locate close to schools with more resources if they care about education and also invest much time in their children's education (creating an upward bias in estimated effects of school inputs), or parents located close to such schools may choose to invest less time in their children's education (creating a downward bias). Another important source of bias is local authorities' choices which are partly determined by external conditions. Local authorities may for instance allocate more resources to schools with a large fraction of pupils from socially disadvantaged backgrounds due to extended needs for special tuition, etc. This will tend to bias estimated effects of school inputs downwards, unless the analysis controls sufficiently for socioeconomic conditions. Another cause of such a negative bias is that the local socioeconomic environment may affect educational choices through role model and peer group effects. Downward bias will also be the result if school administrators tend to allocate more resources (in the form of small class sizes or extra teacher hours per class) to classes with a large fraction of pupils from disadvantaged backgrounds or with learning difficulties.

Several recent studies have tried to isolate exogenous variation in school resources using experiments (Krueger, 1999) or instruments based on administrative rules (Angrist and Lavy, 1999) and random population variation (Hoxby, 1999). The main focus of these studies has been to estimate effects of class size on pupil test scores.

In this paper we analyse effects of class size and the number of pupils per teacher hour on pupils' educational career after compulsory school using instruments for class size and pupils per

teacher hour based on a maximum-class-size rule, like Maimonides' rule used in Angrist and Lavy (1999), and a similar administrative rule for the allocation of teacher hours to schools and grades. Including the number of pupils per teacher hour as an input measure as an alternative to (or simultaneously with) class size may be important since larger classes are often assigned extra teacher hours. This compensation may take the form of more lessons per week or an extra teacher in some of the lessons, or both.¹ If it is not class size per se which is important for pupil outcomes, but rather the amount of teacher resources per pupil, compensatory teacher hours will have the effect of lowering the estimated class-size effect. Thus, if class size is used as the measure of school resources in an educational production function, its estimated effect on pupil outcomes may be a downward biased estimate of the effect of school resources.

The relative demand for unskilled labour is falling and there are strong efforts in most countries to increase the fraction of younger cohorts passing a qualifying education. One means to obtain this goal may be to improve the quality of primary schools. Therefore, it is important to estimate the effects of school resources on pupils' educational attainment after compulsory school. This is the focus of the present paper. The literature on this issue is much more limited than the literature on the effects of school resources on test scores.² Furthermore, according to our knowledge, there are no previous studies on the effects of class size (or pupil-teacher ratio) on educational attainment later in life which has addressed the endogeneity of school inputs using administrative rules (such as the maximum-class-size rule) as instruments.

The Danish data set which we use in this study is particularly well-suited for this kind of analysis since we have information on each individual's pathway through the educational system, and also high quality information on each individual's family background which is particularly important when studying educational attainment, given the overwhelming evidence of highly significant family background effects. The data set is based on administrative data merged from

¹ Actually, Maimonides= rule includes a similar compensation: If the number of pupils is higher than 25 (but less than 40) the teacher should have an assistant, see Angrist and Lavy (1999, p. 534).

² Two recent studies have used the British National Child Development Survey to study effects of the pupil-teacher ratio on educational attainment (and labour market outcomes) after compulsory schooling. Dustmann et al. (2003) find a significant positive effect of reducing the pupil-teacher ratio on the probability of staying on in full time schooling at the age of 16, whereas Dearden et al. (2002) find no effect on ultimate educational attainment. Both studies use the overall pupil-teacher ratio in the school, thus taking account of endogeneity due to intra-school allocation decisions. Endogeneity problems related to inter-school allocation and selection are taken account of by including controls for family background and neighbourhood characteristics.

several administrative registers by use of personal civil registration numbers. It is a panel data set for the period 1981-2000 covering a 10% random sample (almost 45,000 persons) of the cohorts who started in 8th grade in 1985-1992 (corresponding approximately to the cohorts born 1971-78) and their parents. These micro data are combined with data on school resources for each grade at each school (for each year we have data for about 1,100 schools) and socioeconomic conditions in the municipalities. The basic advantages using administrative data are that it is feasible to establish data sets covering a large number of individuals and variables, that data are generally highly reliable (e.g., there is no recall bias), and that it is possible to get data for each individual over a long period of time without the severe attrition problems common in longitudinal data sets based on repeated surveys.

Although we use basically the same regression-discontinuity approach as Angrist and Lavy (1999), our analysis is different in several important respects. First, we estimate effects of both class size and the number of pupils per teacher hour.³ Second, our outcome measure is not test scores but educational attainment (years of education). Third, we use data on outcomes and family background for each individual (whereas Angrist and Lavy use class averages). Fourth, we have a very large set of controls for family background and socioeconomic conditions at the local authority level (whereas Angrist and Lavy only use one control variable which is the percentage of pupils from disadvantaged backgrounds at school level). Finally, in the Israeli data used by Angrist and Lavy the maximum class size is 40 which is much higher than in most European countries and the US, whereas in Denmark it is effectively 24. This means that whereas Angrist and Lavy estimate effects of reducing class sizes from an average level of 32, we estimate effects given an average level of 20.

The rest of the paper is organized as follows. In section 2 we discuss econometric methods. Section 3 gives a short description of the Danish institutional setting. In section 4 we describe the data set and discuss the instruments for class size and the number of pupils per teacher hour. We present estimation results in section 5. Section 6 contains conclusions.

³ Lavy (1999) also includes a measure teacher hours as a school resource variable in addition to class size using basically the same data set as Angrist and Lavy (1999), but the administrative rules governing the allocation of teacher hours are very different from the ones used in this paper, and there is no compensatory allocation of extra teacher hours to larger classes.

2. Econometric methods

Our general model is given by

$$y_{iskt} = X_{iskt}\beta_1 + X_{skt}\beta_2 + X_{kt}\beta_3 + \beta_4 D_t + \alpha_1 n_{iskt} + \alpha_2 h_{iskt} + \varepsilon_{iskt} \quad (1)$$

$$n_{iskt} = X_{iskt}\gamma_1 + X_{skt}\gamma_2 + X_{kt}\gamma_3 + \gamma_4 D_t + \gamma_5 f_{skt} + \gamma_6 g_{skt} + \zeta_{iskt} \quad (2)$$

$$h_{iskt} = X_{iskt}\delta_1 + X_{skt}\delta_2 + X_{kt}\delta_3 + \delta_4 D_t + \delta_5 f_{skt} + \delta_6 g_{skt} + \eta_{iskt} \quad (3)$$

where: y_{iskt} is educational outcome for individual i of cohort t in school s of municipality k ; the cohort index t signifies the year in which individual i started in 8th grade; X_{iskt} is a vector of characteristics of pupil i (family background and gender); X_{skt} is a vector of characteristics of cohort t in school s in municipality k (e.g., functions of enrollment); X_{kt} is a vector of characteristics of municipality k for cohort t (socioeconomic variables at the municipal level in years relevant for cohort t); D_t is a dummy for cohort t ; n_{iskt} is average class size for cohort t in school s of municipality k ; h_{iskt} is the number of pupils per teacher hour (per week) for cohort t in school s of municipality k ; f_{skt} and g_{skt} are the mandated class size and number of pupils per teacher hour, respectively, determined by administrative rules; and ε_{iskt} , ζ_{iskt} and η_{iskt} are error terms.

The school resource parameters α_1 and α_2 are the parameters of primary interest. They are interpreted as reflecting the average causal effects of changes in class size and the number of pupils per teacher hour, respectively, controlling for the X variables. To identify these parameters using instrumental variables we have to assume that, controlling for the X variables, educational outcomes and instruments are only correlated because of the correlation between instruments and school resources (n and h).

The instruments f and g are ‘discontinuous’ functions of enrollment, e , with discontinuity points at multiples of the maximum class size of 24, i.e. the discontinuity points are at enrollment counts 24, 48, 72, 96, etc. (see section 4). Enrollment, e , may be related to pupil educational outcomes for reasons other than effects of changing class size or changing the number of pupils per teacher hour. Therefore, the key identifying assumption is that any other effects of e on educational outcomes are adequately controlled for by the X variables. We have a very rich data

set allowing detailed control for individual family background and socioeconomic conditions in the municipalities (see section 4.2), so a priori this assumption may not seem unreasonable.

In addition, we may include smooth functions of e in X_{skt} to help ensure that the parameters α_1 and α_2 are primarily identified by the discontinuities in the instruments (considered as functions of e). This corresponds to identifying the parameters of interest using a regression-discontinuity design as in Angrist and Lavy (1999). This is a fuzzy regression-discontinuity design in the sense that the rule assigning treatment status (a small or a large class size), depending on the value of a covariate (enrollment), is not deterministic.⁴ Thus, to the right of the discontinuity points the probability of being treated (i.e. assigned to a small class) is higher than to the left of the discontinuity points, but treatment is not guaranteed.

Ideally, we would like to estimate the model (1)-(3) including both n and h simultaneously. However, it turns out that these variables are too highly correlated in our data set. Therefore, in this paper we only present results of estimating *either* the model consisting of (1) and (2) with n as the only school resource measure ($\alpha_2=0$) using f as an instrument ($\gamma_6=0$), *or* the model consisting of (1) and (3) with h as the only school resource measure ($\alpha_1=0$) using g as an instrument ($\delta_5=0$). Also, it is not obvious how to include both school resource variables simultaneously in our preferred econometric specification where we are identifying the parameters of interest using the regression-discontinuity design with a restriction of the sample to observations close to the discontinuity points (see the discussion below). Here, we would need different discontinuity points for the instruments for n and h (f and g , respectively) in order to include these variables simultaneously in the model. But in our data set the discontinuity points of f and g are identical, and even if f and g had different discontinuity points, we would have to estimate the effects of n and h using different sub-samples.

If the key assumption is valid (and given mild regularity conditions), α_1 and α_2 may be interpreted as weighted average causal effects of school resources on educational outcomes, depending on a heterogeneity of individual specific causal effects (see Angrist and Lavy, 1999, and Angrist and Imbens, 1995). Thus, we need not assume a constant effect for all pupils of class size or the number of pupils per teacher hour. In later work, we shall exploit the fact that we can follow individual students to address distributional issues. For example, does class size affect

⁴ Campbell (1969) used the regression-discontinuity design when the rule relating treatment status to a covariate, which is also related to the outcome analysed, is deterministic.

more and less able students differently?

There is an additional identifying assumption: parents do not selectively exploit the administrative rules so as to place their children in schools with small classes or few pupils per weekly teacher hour. This is a conditional independence assumption: Treatment (small class, few pupils per teacher hour) of pupil i is independent of α_{1i} and α_{2i} conditional on enrollment (and other covariates) near discontinuity points, see Hahn, Todd and van der Klaauw, 2001. This assumption should hold. There is no way to predict whether enrollment will be just above or just below a threshold, and even if there were, in Denmark in the eighties and early nineties, parents were not free to transfer children from one elementary school to another except by moving, and only a rather limited number of parents would consider the option of placing their children in private schools.

We estimate the model by two-stage least squares correcting the standard errors for intra-school correlation in educational outcomes.⁵ This correction is potentially very important as shown in Moulton (1990). In this paper it increases standard errors of the school resource coefficients by 4-14%.

The key identifying assumption (that conditional on the X variables, which may include smooth functions of e , y is related to e only through the instruments) is untestable. Therefore, we prefer alternative specifications of the model which focus explicitly on the variation at the discontinuity points. Thus, in addition to analyses on the full sample, we also present analyses using, as in Angrist and Lavy (1999), a ‘discontinuity sample’ of pupils at schools with enrollment within the ± 3 pupils intervals around the discontinuity points of f and g (i.e. around enrollments of 24, 48, 72, and 96). For this ± 3 discontinuity sample we also show estimation results where enrollment is replaced by dummy variables for enrollment segment: $\mathbf{1}(22 \leq e < 27)$, $\mathbf{1}(46 \leq e < 51)$, etc (leaving out one for reference category). Here the instruments f and g are replaced by dummy variables $\mathbf{1}(25 \leq e < 27)$, $\mathbf{1}(49 \leq e < 51)$, etc. This specification highlights the quasi-experimental identification strategy of the regression-discontinuity design: When e is to the right of the discontinuity points we have a high probability of treatment (the expected class size is

⁵ Since school resource variables vary by both school and year, clusters used in this correction are defined in terms of both schools and years, i.e. we assume that educational outcomes are uncorrelated between different cohorts at the same school. If we instead define clusters by schools only (allowing for correlation of outcomes between different cohorts at the same school) the standard errors of estimated school resource effects become smaller.

small), otherwise a low probability of treatment (the expected class size is large). It is the most appropriate specification in the sense that only the variation in the instrument at the discontinuity points is used. It resembles the Wald estimator discussed in Hahn, Todd and van der Klaauw (2001) except that in our analysis there are more than one discontinuity point and also additional controls.

3. The Danish institutional setting

This paper focuses on the effects on later educational attainment of school resources in public primary and lower secondary schools in Denmark. These schools are run by the 275 local authorities (municipalities) with on average about 18,500 inhabitants (ranging from less than 5000 up to almost half a million in the municipality of Copenhagen). Primary and lower secondary schools cover compulsory education which runs from 1st to 9th grade (corresponding approximately to age 7-15), and in addition one year of preschool class (age 6) attended by 98% of a cohort and one optional extra year of lower secondary school (10th grade, age 16) attended by about 60% of a cohort. Public schools are attended by almost 90% of a cohort, the remaining 10% attending private schools. However, a few percent of the pupils who start in public schools change to private schools during their school career. For the cohorts analysed in this paper, 84% attended public schools in 8th grade. Expenditure on primary and lower secondary schools in Denmark is (like most other municipal expenditure) financed by municipal taxes, primarily income taxes, but extensive grants and equalization schemes eliminate the greater part of financial inequalities between municipalities.

The options after having completed lower secondary school are: (1) to leave the (full-time) educational system in order to get a full-time job in the labour market, to take a sabbatical year or for other reasons; (2) to start a vocational education; or (3) to enter upper secondary school. Both vocational education and upper secondary school will typically take three years, although there are exceptions. There are numerous different vocational educations ranging from clerical education to training in trade skills such as carpentry or plumbing. They consist partly of schooling and partly of an apprenticeship with an employer. Difficulties finding appropriate apprenticeships are a main reason for the considerable dropout from vocational education. Admission to universities and other further or higher education is normally restricted to those who have passed the examination at the end of upper secondary school. However, passing some

forms of vocational education gives admission to specific forms of further education.

The Danish educational system is characterized by great flexibility in the sense that it is possible, and very common, to take one or more years out during the educational career, for instance immediately after completion of lower secondary school, during vocational education or after upper secondary school. It is also common to shift from one part of the educational system to another, for example to drop out of a vocational education and enter upper secondary school or another vocational education instead. One effect of this flexibility is that for many students it takes considerably more than three years from completion of lower secondary school (typically at the age of 15 or 16) to passing upper secondary school or vocational education. Of those who eventually pass, only about 60% have passed at the age of 19, about 93% at the age of 22, and about 97% at the age of 25.

4. Data and instruments

As discussed in the introduction, we have micro data based on administrative registers for a 10% random sample (44,974 persons) of the cohorts who started in 8th grade in public schools in 1985-92 and their parents. It is a panel data set for the period 1981-2000. A person who is in the data set in one year will also be in the data set the following year unless he or she died or emigrated. This data set is merged with data on enrollment and school resources (class size and the number of pupils per teacher hour) for each grade at each school, and data on socioeconomic conditions in the municipalities. For each year we have data on about 1,095 public schools. First we discuss data on class size, teacher hours and enrollment.

4.1 Class size, teacher hours and enrollment

For each year enrollment, class size and the number of teacher hours per week are recorded at the beginning of September (aprox. 1 month after the start of the school year). For each individual in the data set we have information on the school attended at 8th, 9th and 10th grade. We do not use information on school resources for 10th grade since this is an optional extra year at lower secondary school (as discussed in section 3). For simplicity we use data on class size and teacher hours in 8th grade only. For most pupils school resources (class size and the number of pupils per teacher hour) in 8th grade are correlated with school resources in earlier grades: most pupils attend the same school from preschool class to 8th grade. But it is rather common to change

school between 8th and 9th grade. Thus, in our data set, 6% of the pupils attending a public school in 8th grade changed to private boarding schools in 9th grade and 1% to other private schools. This choice is in principle endogenous with respect to school resources in 8th grade and earlier, so this is the reason why we do not use a (weighted) average of school resources in 8th and 9th grade.

Thus, for individual i of cohort t (where t is the year the individual started in 8th grade) in school s , school resources are measured as class size and the number of pupils per teacher hour (per week) at 8th grade in school s in year t . Since we only have data on school resources for public schools, persons attending a private school in 8th grade are excluded from the sample.

Table 1 shows data for enrollment, classes and teacher hours at 8th grade in public schools for the school years 1985/86 - 1992/93. These data are at school level (for 8th grade); we do not have data for individual classes. For most schools we have data for all 8 years, but a few schools do not have 8th grade pupils in all years. Average enrollment at 8th grade is 52, varying from 6 to 168. Average class size is 19.5 varying from 6 to 31.5. The average number of pupils per weekly teacher hour is 0.54 with a minimum and maximum of 0.19 and 0.87, respectively.

4.1.1 Instrument for class size

As in Angrist and Lavy (1999) we use a maximum-class-size rule as an instrument for class size:

$$f_{skt} = e_{skt} / [\text{int}((e_{skt} - 1)/q) + 1] \quad (4)$$

where f_{skt} is expected class size according to this rule (at 8th grade) in school s in municipality k for cohort t , e_{skt} is enrolment (at 8th grade), q is the maximum class size, and the function $\text{int}(x)$ is the largest integer smaller than or equal to x .

In Denmark there is a maximum class size of 28 given by national law. However, in practice class size is determined at the municipal level, and effectively municipalities have much smaller maximum class sizes than 28. Each municipality may have its own administrative rules determining class sizes. Some municipalities have a formal maximum-class-size rule of e.g. 23, 24 or 25 applying to all grades. In other municipalities such rules apply only for lower grades, but due to the fact that classes are seldom merged (unless class size gets very small) this will imply similar maximum class sizes at higher grades.⁶ Also, decisions about the number of classes at

⁶ For instance, the municipality of Gentofte had a rule saying that the class size at preschool and 1st grade levels should be 23 or less, but at higher grades larger class sizes would be acceptable (due to net Aimport@ of pupils into a

each grade in each school may be taken (typically about 6 months) before the start of the school year based on predicted enrollment. Thus there might be a precautionary motive for not being too close to the maximum. And, more importantly, municipalities do not want to split up classes in one year and merge them again the following year (or vice versa) just because enrollment in the district is increased by one or two pupils as a consequence of households moving into or out of a school district. Therefore, in practice the maximum class size of 28 given by the law is very seldom used, not even in municipalities where there is no formal rule determining a lower maximum class size than 28. In practice, municipalities typically do not want to have class sizes of more than 23-26. In the estimations we use a maximum-class-size rule of 24 (i.e. $q=24$ in (4)). Figure 1 shows observed average class size and predicted class size given by (4) with $q=24$ for each value of school enrollment. It will be seen that there is a high correlation between observed and predicted class size, especially for values of enrollment below 100. Also, at the four lower discontinuity points (corresponding to enrollments of 24, 48, 72 and 96) average observed class size is higher in a small interval to the left of these points than in a small interval to the right. This is a necessary condition for the regression-discontinuity design to work in this case.

4.1.2 *The relationship between the number of pupils per teacher hour and class size*

Let H be the total number of teacher hours per week at a given grade in a given school, and let e and c denote enrollment and number of classes, respectively, at that grade. If all teaching takes place within given classes and all lessons are with only one teacher, the number of teacher hours per week per pupil (H/e) is equal to the number of weekly lessons per pupil or class (P) divided by average class size (n) at this grade:

$$H/e = \ell c/e = \ell /n \tag{5}$$

However, if in m of the weekly lessons two teachers are teaching the class, H will be higher:

$$H/e = (\ell + m)/n \tag{6}$$

Similarly, H will be lower if in some of the lessons one teacher is teaching two classes (but this is

school district or net Aexport and merging of very small classes).

rather unusual in the Danish system). The inverse relationship between the number of pupils per teacher hour per week (h) and n is of course

$$h \equiv e/H = n/(\ell + m) \quad (7)$$

Like the relationship between average class size and enrollment shown in Figure 1, the average number of pupils per teacher hour for each value of enrollment is also influenced by the maximum-class-size rule, but there is a ‘trend’ in the peaks of the h -curve which is not present for the n -curve, and the variation in h around the discontinuity points is smaller than the variation in n . This is shown in Figure 2 (which also shows the fitted values of a regression of h on the instrument g , see the discussion in section 4.1.3 below).

There are several rules in the Danish system which affect the observed relationship between teacher hours, class size and enrollment: (A) there is a minimum (and a maximum) number of lessons per week for pupils at each grade given by law; (B) there are recommended numbers of teacher hours per week per pupil at each grade determined by the Ministry of Education; and (C) at grades 8-10 there are some minimum rules about the amount of optional subjects which should be supplied by the school (and which may be taught for pupils from several parallel classes together).

When enrollment is very low (about 12 pupils or less) the number of teacher hours per student will be above the recommendation of the Ministry of Education even when $m=0$ and P is equal to the minimum number of lessons given by law. When, from a very low level, e is increased by 1, n rises by 1 whereas m and P are typically constant, implying a constant increase in pupils per teacher hour of $1/P$, see (7). For levels of enrollment (and class size) above 12 or 15 an increase in e will typically lead to an increase in P and subsequently an increase in m implying that the increase in h in response to a unit increase in e will be smaller and smaller until e reaches the level where the pupils are divided into two classes. This will cause a fall in h , but the percentage fall will be less than the percentage fall in class size due to the fact that m (and possibly P) will typically be reduced discontinuously.

Rule (C) implies that small schools with only one class at grades 8-10 will have to teach very small groups of pupils in optional subjects. Similarly, it is easier for schools with three or four parallel classes at these grades to avoid teaching very small groups of pupils than it is for schools

with only two parallel classes. Therefore, rule (C) implies a tendency of a $A_{tend@}$ in h as a function of e . This trend in the peaks of the h curve is not present for class size as a function of e .

4.1.3 Instrument for the number of pupils per teacher hour

Besides the general rules described above, different municipalities have different specific rules for allocating teacher hours (and other resources) to schools, grades and classes, and these rules have been changed during the period 1985-93 for which we have data. Furthermore, the degree of autonomy of schools in determining the intra-school allocation on classes varies between municipalities. For simplicity, to form an instrument for h we use the basic administrative rule of the largest municipality, i.e. the municipality of Copenhagen, for the school years 1991/92 and 1992/93. This rule determines the number of teacher hours (H) as a function of the number of pupils (e) and classes (c) as follows

$$\begin{aligned}
 H(c,e) = & 16 c + 20 \mathbf{1}(e \leq 18) + 22 \mathbf{1}(18 < e \leq 22) + 24 \mathbf{1}(22 < e \leq 25) + 26 \mathbf{1}(25 < e \leq 28) \\
 & + 28 \mathbf{1}(28 < e \leq 35) + 30 \mathbf{1}(35 < e \leq 40) + 32 \mathbf{1}(40 < e \leq 45) + 36 \mathbf{1}(45 < e \leq 50) \\
 & + 42 \mathbf{1}(50 < e \leq 55) + 44 \mathbf{1}(55 < e \leq 60) + 48 \mathbf{1}(60 < e \leq 70) + 52 \mathbf{1}(70 < e \leq 80) \\
 & + 56 \mathbf{1}(80 < e \leq 90) + 60 \mathbf{1}(90 < e \leq 100) + 64 \mathbf{1}(100 < e \leq 110) + 68 \mathbf{1}(110 < e \leq 120) \\
 & + 72 \mathbf{1}(120 < e \leq 130) + 76 \mathbf{1}(130 < e \leq 140) + 80 \mathbf{1}(140 < e \leq 150) \\
 & + 84 \mathbf{1}(150 < e \leq 160) + 88 \mathbf{1}(160 < e)
 \end{aligned} \tag{8}$$

where the function $\mathbf{1}(\cdot)$ is equal to 1 if the expression inside the parentheses is true, and zero otherwise. According to this rule, an extra class induces a rise in teacher hours of 16. The step function determining H as a function of e (given c) implies that H increases (discontinuously) with e , but at a more or less decreasing rate (which also reflects that when e has a value for which one would expect big classes, a small increase in e induces a relatively large increase in H). When enrollment is above 60, an increase in enrollment of 10 induces an increase in teacher hours of 4.

Replacing the number of classes, c , with the expected number, e/f , where f is the expected class size given by (4) with $q=24$, equation (8) gives the administrative rule determining the expected number of teacher hours. Thus, the expected number of pupils per teacher hour given this rule is

$$g = e / H(e/f, e) \tag{9}$$

Figure 3 shows for each level of enrollment the average number pupils per teacher hour and the expected number given by (9) for schools in the municipality of Copenhagen in the school years 1991/92 and 1992/93 where the rule applied. It will be seen that the average number of pupils per teacher hour is very close to the predicted number for most values of enrollment. There are exceptions, but this is due to the fact that (8) is only the basic rule determining the number of teacher hours; there are some additional rules taking account of social conditions in the individual school districts and other criteria. For years before 1991 most observations for the municipality of Copenhagen lie below the curve for the expected number of pupils per teacher hour in Figure 3, but this mainly reflects a shift in the level. The general form of the curves for the average number of pupils per teacher hour for earlier years are similar to the one shown in Figure 3.

This also applies when data for other municipalities are compared to the instrument based on the Copenhagen administrative rule. As explained above, Figure 2 shows the average number of pupils per teacher hour for each level of enrollment and the fitted values from a regression of this variable on the instrument (9).

To explore further the correlation between instruments and observed class size and number of pupils per teacher hour for data at school level, Table 2 shows the results of OLS regressions of n and h on the two instruments f and g , given by (4) and (9), respectively. In the first two regressions n is regressed on f , and h on g . In both regressions R^2 is 0.22. The next two regressions are identical, except that year dummies are included. These are significant in both regressions (and especially in the h regression) reflecting an increase in class size and (especially) in the number of pupils per teacher hour during the sample period 1985/86 - 1992/1993. In the other regressions in Table 2 both f and g are included as regressors in both equations. Regressions (3) do not include additional regressors except a constant term (like regressions (1)). Regressions (4) include year dummies (like (2)). Although in (3) and (4) f is significant in the h regression and g in the n regression, R^2 is only marginally higher than in the corresponding regressions (1) and (2) because the instruments f and g are highly correlated (the correlation coefficient is 0.87). Regressions (5) also include enrollment and enrollment squared as explanatory variables. The last two sets of regressions includes these variables, and in addition municipality and school fixed effects, respectively. It will be seen that both the f and g instruments are significant in the class-size equation in all specifications, and when we control for a second order polynomial in

enrollment (and municipality or school fixed effects), the f instrument is more significant than the g instrument (as one would expect). In the h equation g is much more significant than f; the f instrument has a negative coefficient in specifications (3) and (4) and it is insignificant in the last three regressions. In the last four specifications R^2 is 0.24 in the class-size regression and 0.27 in the pupils per teacher hour regression.

4.2 Educational attainment and background variables

We use years of completed education in the last year in our data set (i.e. 2000) as the outcome variable measuring educational attainment.⁷ We can follow the eight cohorts of pupils included in the analyses 7-14 years after the year they completed 8th grade, corresponding for the vast majority of pupils to age 22-29. Since older cohorts will have more years to complete an education they have more years of education in 2000: The cohort completing 8th grade in 1986 has on average 12.3 years of completed education, whereas it is 11.3 years for the 1993 cohort. Therefore, in all estimations we include cohort dummies, and we report some estimations where the sample has been restricted to the oldest cohorts. Table 3 shows descriptive statistics for enrollment, class size, the number of pupils per teacher hour, and years of completed education for the full sample (with 44,974 observations) and for the +/-3 discontinuity sample (with 10,452 observations). The means of class size and pupils per teacher hour in the full sample are almost the same as in Table 1 with data at school level, whereas average enrollment is of course higher in Table 3 since larger schools are given a higher weight when observations are by pupil rather than by school. Figure 4 shows the estimated kernel density function for enrollment (for the full sample). There are only rather few observations on pupils with enrollment below 20 or above 100, whereas there are many observations around the discontinuity points 48 and 72.

We use in the estimations a large set of controls: cohort dummies, a gender dummy, immigrant/second generation immigrant dummies, family background variables (56 variables for

⁷ Specifically, completion of lower secondary school counts as 9 years, whether or not the pupil chose the optional 10th grade. Having completed upper secondary school or a vocational education corresponds to 12 years of education, and having completed a short further, a long further or a higher education corresponds to 14, 15, and 17 years of education, respectively; PhD is 20 years. A few persons do not complete compulsory school (9th grade); they will have 7 or 8 years of education depending on whether they completed 8th grade or not. Finally, this measure of years of completed education is modified: If a person was in the last year of the data set (year 2000) attending an education (beyond 9th grade) at a higher level than the one completed at that time, one year of education is added.

siblings, family structure, and parents' education, age, labour market status, income, wealth, unemployment and housing conditions), and variables for socioeconomic conditions in the municipality of the school attended in 8th grade (the fraction of bilingual pupils, the unemployment rate, the fraction of the population with a vocational or further education, respectively, the proportion of pupils from single-parent families, income per capita, and urbanization). Family background variables are measured in the year the child was 15, and variables at the municipality level in the year the child attended 8th grade. Descriptive statistics for all control variables are shown in Table A1 of Appendix A.

4.3 Instruments and years of education

Figure 5 shows, for enrollment intervals of 4, how the average length of education and the average expected class size (given by the maximum-class-size rule f) varies with enrollment. Figure 6 is similar, but with the expected number of pupils per teacher hour (given by the administrative rule g) instead of class size. It will be seen from Figure 5 that there is a positive 'trend association' between expected class size and years of education: Both variables increase with enrollment. This trend makes it hard to identify from the figure whether the marked ups and downs of the expected-class-size function are positively or negatively correlated with average length of education. A similar problem exists with respect to Figure 6. Also, the graph for average years of education is dominated by the very large variations for the smallest and largest schools where we have only few observations.

The positive correlation between average years of education and enrollment is due to the fact that large schools tend to be located in big cities where pupils have on average more favourable family backgrounds in terms of parental education, income, etc. Similarly, small schools are more likely to be located in small cities or rural areas where average parental background is less favourable.

Dropping observations with enrollment smaller than 13 or larger than 108, Figure 7 shows average residuals, in enrollment intervals of 4, from regressions of years of education and expected class size, respectively, on the full set of control variables (for family background, socioeconomic conditions in the municipalities, and cohorts) described above. After controlling for these variables and thereby removing the 'trend', it is still not clear whether years of education and expected class size are consistently correlated. They seem to be negatively correlated at low and high values of enrollment, but weakly positively correlated for enrollments

between about 55 and 75, an interval containing a large fraction of the observations, see the estimated density of enrollment in Figure 4. Average residuals from the years of education regression vary between -0.06 and 0.11 years, whereas the variation in class-size residuals is between -4.3 and 5.4 pupils. Figure 8 shows the same graph for residuals from the years of education regression, but here compared to residuals from regressing the instrument for pupils per teacher hour on the controls. Focussing on the discontinuity points in Figures 7 and 8, i.e. the points with a sharp fall in class size and pupils per teacher hour (at enrollments 24, 48, 72, and 96), it will be seen that three of these points (the first, second and fourth) are associated with a rise in average years of education, whereas the third is associated with a fall. Thus, at the discontinuity points there is an indication of a weak negative correlation between class size (or pupils per teacher hour) and years of education.

This graphical analysis indicates that there might be a positive effect on years of education from reducing class size or the number of pupils per teacher hour (at least when the sample is restricted to intervals of enrollment around the discontinuity points), but that this effect would not be determined with great precision. This is explored further in the next section where we present estimation results.

5. Estimation results

Since class size and the number of pupils per teacher hour are highly correlated we do not present estimation results where they are included simultaneously in the model for the outcome variable ‘years of education’. Instead, we present results where class size (instrumented by the maximum-class-size function f) or the number of pupils per teacher hour (instrumented by the administrative rule g) is included. In the presentation of results we focus on the estimates of class size, pupils per teacher hour and enrollment. We do not report results for the other variables included in the regressions. However, the last two columns of Table A1 in Appendix A show estimates for the full set of controls corresponding to the two-stage least squares regression reported in column (3) of Table 7 below. In general, the estimates of controls are not much affected by whether class size or pupils per teacher hour are included in the model for years of schooling, or whether it is estimated by OLS or two-stage least squares, or whether we use the full sample or the discontinuity sample.

5.1 OLS estimates

Table 4 shows the results from OLS regressions for the full sample and for the +/-3 discontinuity sample (i.e. for observations with enrollment in the intervals 22-27, 46-51, 70-75, and 94-99, see sections 2 and 4). All regressions include a constant term, dummies for the year the person attended 8th grade, and either class size or the number of pupils per teacher hour. Regressions in columns (1) and (4) do not include other regressors. The other regressions include, in addition to year dummies, the full set of controls (for family background and socioeconomic conditions in the school municipality, see section 4.2 and the appendix), and the regressions in columns (3) and (6) also include enrollment and enrollment squared. It will be seen that the coefficients of class size and pupils per teacher hour are consistently positive, although they are reduced when background controls are included, and they become insignificant when, in addition, enrollment and enrollment squared are included. Enrollment and enrollment squared are individually significant when estimating on the full sample, but not for the discontinuity sample. If we exclude observations with enrollment smaller than 13 and larger than 108 (as in Figures 7 and 8) the estimation results are very similar to those for the full sample.

Thus, even when we control for a very large set of family and municipal background variables, the OLS estimates of class size and pupils per teacher hour are positive indicating a negative effect on pupils' educational attainment of increasing per pupil school resources. This may however be due to selection on unobservables. For instance, extra resources may be allocated to schools with a large fraction of disadvantaged pupils (in terms of unobserved characteristics). Also, allocation of resources between grades within schools may be affected by the fraction of pupils at each grade with learning or disciplinary difficulties. Our instrumental variables approach can eliminate bias due to selection both between and within schools.

5.2 Reduced form estimates

Table 5 shows reduced form estimates for the class size and years of education regressions when the instrument for class size (the maximum-class-size rule f) is used as an explanatory variable. The upper half of the table shows estimation results for the full sample and the lower half for the +/- 3 discontinuity sample with enrollment in the segments 22-27, 46-51, 70-75, and 94-99. As expected, the instrument has a highly significant positive effect on class size for all specifications (i.e. for both the full sample and the discontinuity sample, and whether or not we control for background variables and enrollment). Regression (1) in Table 5 corresponds to the school level

class-size regression (2) in Table 2. The reason why the estimated coefficient of the instrument f is smaller in Table 5 is that in this regression larger schools weigh more (since there are more pupils at a large school compared to a small one), and the instrument is less powerful in predicting class size at larger schools (see figure 1).

Turning to the years of education regressions, it will be seen that for the full sample expected class size has a significantly positive coefficient if we do not control for enrollment (and enrollment squared); when we do control for enrollment it becomes clearly insignificant. In the discontinuity sample the coefficient is negative and close to being significant at the 5% level (it has a t -value of -1.88) when we control for background variables and enrollment.

These large differences between the results for the full sample and the discontinuity sample indicate that the results for the full sample are to a large extent determined by the variation for observations not close to the discontinuity points, i.e. at enrollments where the instrument is a continuous linear function of enrollment (although with different slopes in different segments, see section 4.1). To use this variation when estimating the effect of class size may be problematic since unobserved variables determining educational attainment may be correlated with enrollment. In a regression-discontinuity framework the maximum-class-size rule is only a valid instrument when it triggers a change in the number of classes and therefore a discontinuous fall in expected class size. However, since the design is Afuzzy@ in this case (see sections 2 and 4.1) and since the number of observations exactly at the discontinuity points is limited, it is not reasonable to restrict the analysis to exactly these points. This is the reason why we have chosen to focus on the +/-3 discontinuity sample (instead of the +/-1 sample). The Afuzzy@ character of the design implies that if we restricted the analysis to exactly the discontinuity points (the +/-1 discontinuity sample) there would be hardly any fall in observed average class size even at the 24 and 48 points (see figure 1). But averaging over a few values of enrollment at each side of these points, we observe a marked fall in class size.

The big differences between the results for the full sample and the +/-3 discontinuity sample are not changed when the polynomial in enrollment is replaced by a piecewise linear trend (a spline function) with slope equal to the slope of the instrument f at all points (except the discontinuity points).

Table 6 shows the results for the reduced form estimates for regressions involving the instrument for the number of pupils per teacher hour. The results are very similar to the results of Table 5: In the regressions for pupils per teacher hour the coefficient of the instrument is consis-

tently positive and highly significant; in the regressions for years of education the coefficient is either positive or insignificant, except for the regression on the discontinuity sample controlling for background variables and enrollment where it is negative and just significant at the 5% level (the t-value is -1.96).

5.3 Two-stage least squares estimations

Tables 7 and 8 show estimation results for two-stage least squares regressions of years of education on class size and the number of pupils per weekly teacher hour, respectively. Estimations (1)-(6) in these tables correspond to the reduced form estimations reported in Tables 5 and 6. For estimations without control for enrollment on the full sample, the coefficients of class size and pupils per teacher hour are significantly positive (which is not surprising given the reduced form results). Controlling for enrollment (and enrollment squared), the coefficients become clearly insignificant (also in accordance with the reduced form results).

For the discontinuity sample the coefficients of class size and pupils per teacher hour become negative when we control for background variables and enrollment, see estimation (6) in Tables 7 and 8, but they are not significant at the 5% level B the estimated t-values are -1.75 and -1.74, respectively. On the face of it one would expect numerical t-values much larger than 2 for any effects to be “truly significant” since even the discontinuity sample contains about 10,500 observations. But school resources do not vary by pupil, and not even by class, but only by school and cohort, and we take account of this school-by-cohort clustering when calculating standard errors (see section 2). In the discontinuity sample the number of independent observations of school resources is about 2,000. Furthermore, the fact that we only have outcome observations for a small fraction of the pupils of each cluster implies uncertainty in the estimations,⁸ and so does the use of instrumental variables methods. Other reasons for the low precision of the estimates are discussed in the concluding section.

According to the point estimate of the class-size effect of -0.07, reducing class size by one pupil (about 5% of the average class size) will increase predicted length of education by 0.07 years. The estimated effect of reducing the number of pupils per weekly teacher hour by 0.028

⁸ For comparison, Angrist and Lavy (1999) have a +/-3 discontinuity sample of about 300 classes with an average class size of 30 and class-average pupil outcome measures based on all pupils in these classes (corresponding to observations of about 9,000 pupils).

(corresponding to 5% of the average level) is an increase in length of education by 0.14 years. The fact that the estimated effect of a given percentage reduction in the number of pupils per teacher hour is much larger than the estimated effect of the same percentage reduction in class size is not surprising since the relative reduction in pupils per teacher hour at the discontinuity points is much smaller than the relative reduction in class size (compare Figures 1 and 2 in section 4 with respect to the data points around the discontinuity points). This indicates that using class size as the measure of school resources may result in downward biased estimates of the effect of school resources on pupil outcomes for school systems characterized by compensatory allocation of extra teacher hours to larger classes (as we hypothesized in the introduction). The concluding section contains a further discussion of the point estimates of the effects of school resources which are rather large.

One concern about the outcome measure used is that there may be rather little variation in years of education for the younger cohorts for which we only have data on educational outcomes rather few years after compulsory school (see section 4.3). Therefore, we have estimated the model for a sub-sample consisting only of the four oldest cohorts in the data set (i.e. the cohorts who attended 8th grade in 1985/86 B 88/89). The results are basically the same in the sense that restricting the estimation to the discontinuity sample and controlling for background variables and enrollment the estimates of class size and pupils per teacher hour become negative (otherwise they are positive or insignificant). Column (7) in Tables 7 and 8 show the result corresponding to column (6) but only for the four oldest cohorts. Numerically the estimates of class size and pupils per teacher hour are about twice as high in the reduced sample of older cohorts, but the standard errors are also much higher. The class-size coefficient is just significant at the 5% level, whereas the coefficient of pupils per teacher hour is just significant at the 10% level.

5.4 Estimates with dummy instrumental variables

In the +/-3 discontinuity sample the variation in the instruments is dominated by the discontinuous falls in expected class size. However, there is also some (positive) variation just before and just after the discontinuity points. In this section we present results from an alternative dummy instrumental variables specification like the one used in Angrist and Lavy (1999) which only use the information from the instruments that expected class size (or pupils per teacher hour) is smaller to right of the discontinuity points than to the left of these points. This captures the treatment idea of the identification strategy: The probability of being treated (i.e. of being placed

in a small class) is higher for enrollments just to the right of the discontinuity points (see section 2).

Instead of using a polynomial in enrollment as controls we use dummy variables for the enrollment segments of the ± 3 discontinuity sample. Letting the first segment be the reference category, we include three dummies for enrollment intervals: $\mathbf{1}(46\#e\#51)$, $\mathbf{1}(70\#e\#75)$, and $\mathbf{1}(94\#e\#99)$. The instruments are dummy variables for enrollment being larger than the discontinuity point of each segment: $\mathbf{1}(25\#e\#27)$, $\mathbf{1}(49\#e\#51)$, $\mathbf{1}(73\#e\#75)$, and $\mathbf{1}(97\#e\#99)$. We include different dummy instruments for each segment because the expected class-size reduction at the discontinuity points is larger for lower segments than for higher segments according to the expected-class-size function f (and similarly for pupils per teacher hour).

Table 9 shows reduced form results of using this set of instrumental variables in regressions of class size, pupils per teacher hour and years of education. The first four dummies are the instruments, while the last three are the controls for enrollment interval. The table presents results both with and without controls for background variables. For the class-size regressions all dummy instruments have negative coefficients as expected, and three of them (the first, second and fourth) are significant, while the third is insignificant. Looking at Figure 1, this is not surprising since the reduction in observed class size at the enrollment discontinuity point 72 is very small. The controls for enrollment segment are not significant reflecting the very similar levels of the peaks of observed average class size left of the discontinuity points, see Figure 1.

In the regressions for pupils per teacher hour the two first dummy instruments are significant (or almost significant) with negative coefficients as expected, while the third and fourth are clearly insignificant (and the coefficient of the third is even positive). Thus, the dummy instruments are more powerful predictors of class size than of the number of pupils per teacher hour. The dummies for enrollment segment are highly significant and their coefficients reflect the trend in the peaks of the observed number of pupils per teacher hour (see Figure 2).

In the reduced form regressions for years of education in Table 9 the coefficients of the dummy instruments are positive as expected (treatment in the form of more school resources per pupil imply more years of education), but they are individually insignificant; especially the third dummy instrument is clearly insignificant in both regressions, but as discussed above the observed class-size discontinuity is very weak at this point. The controls for enrollment intervals are highly significant without controls for background variables (for family and municipalities), but they are smaller and less significant when background controls are included. Including

background controls also has an interesting effect on the structure of the estimated treatment effects (although these are as noted individually insignificant) which is consistent with the hypothesis that the effect of school resources is larger for pupils from disadvantaged backgrounds: The treatment effect at the lowest level of enrollment (small schools primarily in rural areas where a smaller fractions of the pupils have parents with high education and earnings) becomes larger, whereas the effect at the highest level of enrollment becomes smaller.

Table 10 shows two-stage least squares estimates for regressions with the dummy instrumental variables. Regressions (1) and (2) do not include background controls, while the other regressions do include these controls. Regressions (1) and (2) of Table 10 correspond to regression (5) of Table 9 with the dummy variables used to instrument class size and pupils per teacher hour, respectively. Similarly, regressions (3) and (4) of Table 10 correspond to regression (6) of Table 9. Regressions (5) and (6) of Table 10 correspond to regressions (3) and (4) except that only the four oldest cohorts are included in the estimations; similarly regressions (7) and (8) are further restricted to include only the two oldest cohorts. It will be seen that all estimates of the effects of class size and pupils per teacher hour are negative as expected. When background controls are included, the class-size effect is just significant at the 5% level for the sample including all eight cohorts (column (3)) and the sample restricted to the four oldest cohorts (column (5)). Restricting the sample to include only the two oldest cohorts, the standard error of the class-size effect becomes larger and the effect is insignificant (see column (7)), but the size of the estimate is very similar to that in column (5). The size of the estimated effects of class size and pupils per teacher hour in columns (3)-(6) are very similar to the corresponding estimates in columns (6)-(7) of Tables 7 and 8 where we use the original instruments and the discontinuity sample.

6. Conclusion

The focus of this paper is the effect of class size and the number of pupils per weekly teacher hour on educational attainment measured by years of education. We use a panel data set based on administrative registers for a 10% random sample of eight cohorts of pupils (and their parents) in public schools in Denmark. Educational attainment is measured at age 22-29 for the different cohorts. As an instrument for class size we use a maximum-class-size rule like the one used in Angrist and Lavy (1999). The instrument for the number of pupils per teacher hour is based on an administrative rule for the largest municipality. Both instruments are discontinuous functions of

enrollment with discontinuities at enrollments where the maximum-class-size rule triggers an extra class.

For the full sample of pupils, OLS estimates of class size and pupils per teacher hour are positive; they are significant if we do not control for background variables or enrollment, but insignificant with controls for both background variables and enrollment. Instrumental variables estimates are significantly positive without these controls, but clearly insignificant when they are included.

Restricting the sample to ± 3 intervals of enrollment around class-size discontinuity points, instrumental variables estimates using the full set of controls are consistently negative, but the estimates are statistically weakly determined: only some of them are just significant at the 5% level even though we have a large data set. However, restricting the estimations to the “discontinuity sample” reduces the number of observations, and precision is also reduced due to the use of IV methods and the fact that school resources are measured at school-by-cohort level, see the discussion in section 5.3. Other possible reasons for the low precision of the estimates are discussed below, but first we will discuss the interpretation of the size of the point estimates.

For the discontinuity sample, the point estimate of class size is -0.07 indicating that the effect of reducing class size by 1 pupil (about 5% of average class size) will, on average, increase years of education by 0.07. The estimated effect of a 5% reduction in the number of pupils per teacher hour (a reduction by 0.028 pupils per weekly teacher hour) is an increase in years of education by 0.14. Restricting the sample to the four oldest cohorts for whom the variation in educational attainment is larger (because they have more years to complete an education in the sample period), the estimated effects are larger, but the relative size of the effects of class size and pupils per teacher hour is not changed.

The relative size of the estimated effects indicates that using class size as measure of school resources may bias estimates of the effect of school resources on pupil outcomes towards zero for school systems characterized by compensatory allocation of extra teacher hours to larger classes.

The estimates on the discontinuity sample indicate sizeable effects of school resources on educational attainment compared to findings in other studies. Thus, in a survey of US studies on the effect of school resources on educational attainment, Betts (1996) summary of the evidence on the effect of school expenditure indicates that a 5% increase in expenditure per pupil increases years of education by 0.05 on average in studies using state-level data, and by 0.007 in studies using district-level data. We use class size and pupils per teacher hour as resource measures

instead of expenditure, but it is reasonable to assume that a 5% decrease in pupils per teacher hour is equivalent to a 5% increase in expenditure per pupil (teacher salaries account for about 80% of school expenditure in Denmark), and that a 5% decrease in class size is equivalent to an increase in expenditure of less than 5% (since the need for compensatory teacher hours to larger classes would be reduced).

However, even though the point estimates seem to be rather large (and we use a rather large data set), they are not precisely determined statistically. We discussed some reasons for this above. Another reason may be that we measure school resources for 8th grade only, whereas school resources for all grades at primary school may be important for later educational attainment. However, most pupils are enrolled at the same school for all compulsory school years, and school resources at different grades are highly correlated. At least this applies for class size whereas there may be more variation in pupils per teacher hour between grades at the same school for a given cohort (which may explain why class-size estimates are more significant in spite of their smaller relative size).

Another reason why the estimated effects are not very significant may be that average class size in Denmark is rather small, about 20 pupils. Thus, one may expect that there are more significant effects of reducing class size if this is done from a higher level. Using the same methods, Angrist and Lavy (1999) find significant class-size effects for Israeli schools, where class sizes are much larger than the average being 32 pupils. However, the analyses are not directly comparable, since Angrist and Lavy estimate class-size effects on test scores which are recorded for the same grade as the one for which class size is measured. This more narrow connection between measurement of class size and outcome could be another reason for the significant class-size effects in their study. Furthermore, the maximum-class-size instrument is more powerful in predicting observed class size in Angrist and Lavy's data set (where the maximum class size is 40) than in our data set. Finally, the results of Angrist and Lavy are not unequivocal: when their analysis is restricted to a +/-3 discontinuity sample there are only significant class-size effects (at the 5% level) for reading test scores for 5th graders, whereas the effects are insignificant for math test scores for 5th graders and reading and math test scores for 4th graders; using the full sample or a +/- 5 discontinuity sample the effects on test scores are significant for 5th graders but insignificant for 4th graders.

About the representativeness of the estimates, it should be noted that the regression-discontinuity estimates, using the maximum-class-size rule and the administrative rule for teacher

hours, are driven primarily by smaller schools since the difference in average class size (and pupils per teacher hour) just left of a discontinuity point compared to just right of this point is much larger at the first discontinuity point than at the second, and larger at the second than at the third, etc.

Appendix A. Descriptive statistics and estimates for the full set of control variables

The first four columns of Table A1 show descriptive statistics for all control variables for the full sample of 44,974 pupils (the statistics for the discontinuity sample are very similar). The two last columns show, also for the full sample, two-stage least squares estimates and t-values (calculated from robust standard errors taking into account intra-school-by-cohort correlation in outcomes) corresponding to the estimation reported in column (3) of Table 7. All family background variables are recorded at the time the child was 15 years old. It should be noted that the degree of unemployment (i.e. the fraction of the year a person is unemployed) is only defined for people in the work force. For persons not in the work force the degree of unemployment is set to zero. Due to lack of reliable income data for self-employed persons, income from employment is set to zero for self-employed, and a dummy for being self-employed is included in the estimations. Income, wealth (defined as taxable wealth) and school expenditure are measured at 1996 prices using a wage index.

The estimated cohort dummies have the expected pattern: Older cohorts have completed more years of education; the dummy for the oldest cohort (with age 29 in 2000 which is the last year of the sample period) has an estimated effect of 0.9 years compared to the youngest cohort (of age 22 in 2000). Many family background variables are highly significant and the vast majority of the estimated coefficients have the expected signs. The variables representing socioeconomic conditions in the municipalities are less significant; only the variables for the fraction of pupils living with a single parent and the fraction of the population living in rural areas are significant at the 5% level (and they have, as expected, negative effects on educational attainment).

References

Angrist, J.D. and G.W. Imbens (1995): Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90: 431-442.

Angrist, J.D. and V. Lavy (1999): Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*: 533-575.

Betts, J.R. (1996): Is there a link between school inputs and earnings? Fresh scrutiny of an old literature. In Burtless (1996).

Burtless, G. (ed.) (1996): *Does money matter? The effect of school resources on student achievement and adult success*. Brookings Institution Press. Washington D.C.

Campbell, D.T. (1969): Reforms as experiments. *American Psychologist* 24: 409-429.

Card, D. and A.B. Krueger (1996a): *Labor market effects of school quality: Theory and evidence*. In Burtless (1996).

Card, D. and A.B. Krueger (1996b): School resources and student outcomes: An overview of the literature and new evidence from North and South Carolina. *Journal of Economic Perspectives* 10 (4): 31-50.

Dearden, L., J. Ferri and C. Meghir (2002): The effect of school quality on educational attainment and wages. *The Review of Economics and Statistics* 84 (1): 1-20.

Dustmann, C., N. Rajah and A. van Soest (2003): Class size, education, and wages. *The Economic Journal* 113: F99-F120.

Hahn, J., P. Todd and W. van der Klaauw (2001): Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69: 201-209.

Hanushek, E.A. (1986): The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature* 24: 1141-77.

Hanushek, E.A. (1996): *School resources and student performance*. In Burtless (1996).

Hedges, L.V. and R. Greenwald (1996): Have times changed? The relation between school resources and student performance. In Burtless (1996).

Hoxby, C. (1999): The effects of class size on student achievement: New evidence from population variation. Working paper. Harvard University and National Bureau of Economic Research.

Krueger, A.B. (1999): Experimental estimates of educational production functions. *Quarterly Journal of Economics*: 497-532.

Krueger, A.B. (2003): Economic considerations and class size. *The Economic Journal* 113: F34-F63.

Lavy, V. (1999): Using dual natural quasi-experimental designs to evaluate the effect of school hours and class size on student achievement. Working paper, Department of Economics, The Hebrew University of Jerusalem, Israel.

Moulton, B.R. (1990): An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics* 72 (2): 334-338.

Table 1. Descriptive statistics for schools 1985/86 - 1992/93: Enrollment, classes and teachers at 8th grade. (8,764 observations)

Variable	Mean	Std.dev.	Minimum	Maximum
Enrollment	51.99	20.4	6	168
# classes	2.68	0.98	1	8
# teacher hours per week	95.04	34.49	16	275
class size	19.45	2.98	6	31.5
pupils per teacher hour per week	0.54	0.07	0.19	0.87

Table 2. Regressions of class size and pupils per teacher hour on the instruments f and g, and other regressors. All public schools with 8th grade pupils, 1985/86 - 1992/93, school-level data (8,764 observations). (t-values in parentheses)

	f	g	Other regressors	R ² (adj.)
(1) Class size Pupils/hour	0.5089 (49.1)	0.4163 (49.8)	constant	0.22 0.22
(2) Class size Pupils/hour	0.5147 (49.9)	0.4335 (53.0)	constant, year dummies	0.23 0.26
(3) Class size Pupils/hour	0.2909 (13.9) -0.0024 (5.3)	9.0925 (12.0) 0.4938 (29.2)	constant	0.23 0.22
(4) Class size Pupils/hour	0.2783 (13.4) -0.0030 (6.6)	9.8962 (13.1) 0.5286 (31.9)	constant, year dummies	0.24 0.27
(5) Class size Pupils/hour	0.3586 (12.6) 0.0004 (0.7)	4.6079 (3.1) 0.3015 (9.4)	const., year dum., e, e ² (e is enrollment)	0.24 0.27
(6) Class size Pupils/hour	0.3450 (12.0) 0.0002 (0.4)	4.7766 (3.2) 0.3973 (10.1)	const., year dum., e, e ² , mun. FE	0.24 0.27
(7) Class size Pupils/hour	0.2925 (8.9) -0.0003 (0.4)	6.2204 (3.6) 0.3153 (9.3)	const., year dum., e, e ² , school FE	0.24 0.27

Table 3. Descriptive statistics: Enrollment, school resources and years of education

Variable	Mean	Std.dev.	Minimum	Maximum
Full sample: 44,974 pupils				
Enrollment	59.75	21.51	7	168
Class size	19.80	2.72	7	31.5
Pupils per teacher hour	0.55	0.06	0.19	0.87
Years of education	11.91	1.87	7	20
Discontinuity sample: 10,452 pupils				
Enrollment	59.95	20.03	22	99
Class size	20.37	3.66	11	27
Pupils per teacher hour	0.56	0.07	0.30	0.87
Years of education	11.93	1.85	7	20

Table 4. OLS estimation results. Dependent variable: years of education

Regressors	(1)	(2)	(3)	(4)	(5)	(6)
	Full sample: 44,974 observations			Discontinuity sample: 10,452 observations		
Class size	0.0273 (0.0036)	0.0095 (0.0031)	0.0054 (0.0032)	0.0225 (0.0054)	0.0085 (0.0048)	0.0082 (0.0048)
e/100			0.453 (0.163)			0.635 (0.431)
e ² /10,000			-0.204 (0.115)			-0.417 (0.340)
Background	No	Yes	Yes	No	Yes	Yes
Root MSE	1.84	1.69	1.69	1.82	1.67	1.67
R ²	0.033	0.190	0.190	0.035	0.190	0.190
Pupils/hour	1.142 (0.157)	0.442 (0.138)	0.216 (0.147)	1.401 (0.290)	0.647 (0.255)	0.569 (0.261)
e/100			0.450 (0.164)			0.492 (0.433)
e ² /10,000			-0.203 (0.116)			-0.333 (0.340)
Background	No	Yes	Yes	No	Yes	Yes
Root MSE	1.84	1.69	1.69	1.82	1.67	1.67
R ²	0.033	0.190	0.190	0.036	0.190	0.190

Notes: The variable e is enrollment. Robust standard errors corrected for within-school correlation between pupils of the same cohort are reported in parentheses. All regressions include a constant and dummies for year of 8th grade school attendance. AOther regressors@ are the full set of individual and municipal background controls described in section 4.2 and the appendix.

Table 5. Reduced form estimations involving the instrument for class size

Regressors	Class size (first stage regression)			Years of education (reduced form)		
	(1)	(2)	(3)	(4)	(5)	(6)
Full sample: 44,974 observations						
f (instr.)	0.458 (0.016)	0.447 (0.016)	0.383 (0.018)	0.025 (0.004)	0.008 (0.003)	-0.0002 (0.0038)
e/100			5.459 (0.583)			0.512 (0.166)
e ² /10,000			-2.989 (0.400)			-0.233 (0.116)
Background	No	Yes	Yes	No	Yes	Yes
Root MSE	2.47	2.44	2.42	1.84	1.69	1.69
R ²	0.177	0.192	0.206	0.032	0.190	0.190
+/- 3 discontinuity sample: 10,452 observations						
f (instr.)	0.156 (0.027)	0.147 (0.027)	0.146 (0.027)	0.002 (0.005)	-0.007 (0.005)	-0.0096 (0.0051)
e/100			0.392 (2.278)			0.752 (0.437)
e ² /10,000			-0.272 (1.745)			-0.467 (0.341)
Background	No	Yes	Yes	No	Yes	Yes
Root MSE	3.56	3.47	3.47	1.82	1.67	1.67
R ²	0.053	0.107	0.107	0.033	0.189	0.190

Notes: The function f is the maximum class size rule given by equation (4) with q=24; e is enrollment. Robust standard errors corrected for within-school correlation between pupils of the same cohort are reported in parentheses. All regressions include a constant and dummies for year of 8th grade school attendance. ABackground@ regressors are the full set of individual and municipal background controls described in section 4.2 and the appendix.

Table 6. Reduced form estimations involving the instrument for the number of pupils per teacher hour

Regressors	Pupils per teacher hour (first stage regr.)			Years of education (reduced form)		
	(1)	(2)	(3)	(4)	(5)	(6)
Full sample: 44,974 observations						
g (instr.)	0.384 (0.010)	0.358 (0.010)	0.291 (0.017)	1.417 (0.141)	0.512 (0.126)	0.049 (0.195)
e/100			0.140 (0.017)			0.487 (0.182)
e ² /10,000			-0.088 (0.012)			-0.224 (0.118)
Background	No	Yes	Yes	No	Yes	Yes
Root MSE	0.055	0.053	0.053	1.84	1.69	1.69
R ²	0.200	0.239	0.249	0.033	0.190	0.190
+/- 3 discontinuity sample: 10,452 observations						
g (instr.)	0.224 (0.021)	0.210 (0.021)	0.121 (0.029)	0.558 (0.239)	-0.083 (0.215)	-0.596 (0.304)
e/100			0.210 (0.044)			0.989 (0.470)
e ² /10,000			-0.131 (0.033)			-0.571 (0.350)
Background	No	Yes	Yes	No	Yes	Yes
Root MSE	0.066	0.063	0.063	1.82	1.67	1.67
R ²	0.103	0.176	0.190	0.034	0.189	0.190

Notes: The function g is the administrative rule for the number of pupils per teacher hour given by equation (9); e is enrollment. Robust standard errors corrected for within-school correlation between pupils of the same cohort are reported in parentheses. All regressions include a constant and dummies for year of 8th grade school attendance. ABackground@ regressors are the full set of individual and municipal background controls described in section 4.2 and the appendix.

Table 7. Two-stage least squares regressions of years of education on class size and covariates

Regressors	Full sample			+/- 3 discontinuity sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)*
Class size	0.055 (0.008)	0.017 (0.008)	-0.0006 (0.0100)	0.010 (0.035)	-0.045 (0.034)	-0.066 (0.037)	-0.129 (0.066)
e/100			0.515 (0.190)			0.778 (0.480)	0.804 (0.856)
e ² /10,000			-0.235 (0.125)			-0.485 (0.372)	-0.473 (0.654)
Backgr.	No	Yes	Yes	No	Yes	Yes	Yes
Root MSE	1.84	1.69	1.69	1.82	1.68	1.69	1.91
# obs.	44,974	44,974	44,974	10,452	10,452	10,452	5,328

Notes: Robust standard errors corrected for within-school correlation between pupils of the same cohort are reported in parentheses. All regressions include a constant and dummies for year of 8th grade school attendance. The variable e is enrollment. ABackground@ regressors are the full set of individual and municipal background controls described in section 4.2 and the appendix.

* The sample of estimation (7) is restricted to the cohorts attending 8th grade in 1985/86 - 1988/89.

Table 8. Two-stage least squares regressions of years of education on the number of pupils per teacher hour and covariates

Regressors	Full sample			+/- 3 discontinuity sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)*
Pupils/hour	3.694 (0.373)	1.432 (0.352)	0.167 (0.671)	2.492 (1.066)	-0.394 (1.027)	-4.923 (2.833)	-12.634 (7.566)
e/100			0.464 (0.242)			2.020 (0.945)	4.267 (2.233)
e ² /10,000			-0.209 (0.145)			-1.217 (0.606)	-2.424 (1.320)
Backgr.	No	Yes	Yes	No	Yes	Yes	Yes
Root MSE	1.85	1.69	1.69	1.82	1.68	1.71	2.02
# obs.	44,974	44,974	44,974	10,452	10,452	10,452	5,328

Notes: Robust standard errors corrected for within-school correlation between pupils of the same cohort are reported in parentheses. All regressions include a constant and dummies for year of 8th grade school attendance. The variable e is enrollment. ABackground@ regressors are the full set of individual and municipal background controls described in section 4.2 and the appendix.

* The sample of estimation (7) is restricted to the cohorts attending 8th grade in 1985/86 - 1988/89.

Table 9. Reduced form estimations with dummy instrumental variables for the +/- 3 discontinuity sample (10,452 observations).

Regressors	Class size		Pupils per teacher hour		Years of education	
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbf{1}(25\#e\#27)$	-2.299 (0.616)	-2.160 (0.602)	-0.0210 (0.0104)	-0.0177 (0.0102)	0.089 (0.113)	0.140 (0.101)
$\mathbf{1}(49\#e\#51)$	-1.144 (0.263)	-1.093 (0.255)	-0.0160 (0.0049)	-0.0153 (0.0047)	0.061 (0.059)	0.069 (0.051)
$\mathbf{1}(73\#e\#75)$	-0.310 (0.288)	-0.252 (0.279)	0.0017 (0.0056)	0.0027 (0.0053)	-0.000 (0.072)	0.029 (0.062)
$\mathbf{1}(97\#e\#99)$	-0.915 (0.382)	-0.918 (0.387)	-0.0062 (0.0084)	-0.0073 (0.0080)	0.142 (0.110)	0.115 (0.090)
$\mathbf{1}(46\#e\#51)$	-0.248 (0.393)	-0.168 (0.390)	0.0424 (0.0081)	0.0469 (0.0080)	0.257 (0.098)	0.154 (0.086)
$\mathbf{1}(70\#e\#75)$	-0.458 (0.397)	-0.334 (0.401)	0.0468 (0.0080)	0.0506 (0.0080)	0.327 (0.098)	0.169 (0.088)
$\mathbf{1}(94\#e\#99)$	-0.032 (0.445)	-0.097 (0.455)	0.0638 (0.0090)	0.0637 (0.0091)	0.352 (0.109)	0.163 (0.097)
Background	No	Yes	No	Yes	No	Yes
Root MSE	3.56	3.46	0.065	0.062	1.82	1.67
R ²	0.058	0.111	0.121	0.196	0.036	0.190

Notes: The function $\mathbf{1}(x\#e\#z)$ is equal to 1 if $x\#e\#z$, and 0 otherwise, where e is enrollment. The first four dummy variables are the instruments, whereas the last three are controls for enrollment segment. Robust standard errors corrected for within-school correlation between pupils of the same cohort are reported in parentheses. All regressions include a constant and dummies for year of 8th grade school attendance. ABackground@ regressors are the full set of individual and municipal background controls described in section 4.2 and the appendix.

Table 10. Two-stage least squares estimations for years of education with dummy instrumental variables; +/- 3 discontinuity sample.

Cohorts	1985/86B92/93				1985/86B88/89		1985/86B86/87	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Class size	-0.054 (0.036)		-0.070 (0.034)		-0.137 (0.060)		-0.116 (0.071)	
Pupils/hour		-4.441 (3.325)		-5.494 (3.209)		-10.080 (6.053)		-5.908 (5.761)
$\mathbf{1}(46\#e\#51)$	0.262 (0.070)	0.443 (0.169)	0.145 (0.066)	0.381 (0.173)	0.084 (0.125)	0.587 (0.319)	0.324 (0.189)	0.590 (0.347)
$\mathbf{1}(70\#e\#75)$	0.314 (0.076)	0.540 (0.212)	0.157 (0.073)	0.443 (0.212)	0.170 (0.131)	0.793 (0.414)	0.391 (0.202)	0.737 (0.547)
$\mathbf{1}(94\#e\#99)$	0.404 (0.089)	0.681 (0.260)	0.182 (0.082)	0.519 (0.244)	0.144 (0.141)	0.867 (0.483)	0.413 (0.214)	0.793 (0.507)
Background	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Root MSE	1.84	1.86	1.70	1.72	1.92	1.96	1.98	1.97
# obs.	10,452	10,452	10,452	10,452	5,328	5,328	2,752	2,752

Notes: The function $\mathbf{1}(x\#e\#z)$ is equal to 1 if $x\#e\#z$, and 0 otherwise, where e is enrollment. Robust standard errors corrected for within-school correlation between pupils of the same cohort are reported in parentheses. All regressions include a constant and dummies for year of 8th grade school attendance. ABackground@ regressors are the full set of individual and municipal background controls described in section 4.2 and the appendix.

Table A1. Descriptive statistics and two-stage least squares estimates for the full sample (44,974 pupils). The estimation result corresponds to the one in Table 7, column (3)

Variable	Descriptive statistics				Estimation result	
	Mean	Std. Dev.	Min	Max	Coef.	t-value
Class size	19.801	2.718	7	31.5	-0.001	-0.06
Attended 8th grade 1985/86	0.133	0.340	0	1	0.927	19.33
Attended 8th grade 1986/87	0.136	0.342	0	1	0.920	20.43
Attended 8th grade 1987/88	0.128	0.334	0	1	0.797	19.43
Attended 8th grade 1988/89	0.125	0.331	0	1	0.678	18.59
Attended 8th grade 1989/90	0.129	0.335	0	1	0.508	14.7
Attended 8th grade 1990/91	0.120	0.325	0	1	0.412	13.55
Attended 8th grade 1991/92	0.116	0.320	0	1	0.305	10.6
Female	0.495	0.500	0	1	0.239	14.78
Immigrant	0.010	0.098	0	1	0.051	0.52
Second generation immigrant	0.008	0.091	0	1	0.669	6.24
Number of siblings aged 0-17	0.685	0.815	0	9	-0.013	-0.77
Has younger siblings aged 0-14	0.460	0.498	0	1	0.160	6.11
Lives with single mother	0.132	0.339	0	1	-0.316	-7.05
Lives with mother and stepfather	0.092	0.289	0	1	-0.393	-8.69
Lives with single father	0.024	0.154	0	1	-0.322	-3.86
Lives with father and stepmother	0.019	0.137	0	1	-0.380	-4.26
Does not live with father or mother (FM lives together)H(Child not with FM)	0.019	0.135	0	1	-0.949	-9.79
Mother not in the register	0.002	0.043	0	1	-0.141	-0.69
Father not in the register	0.014	0.118	0	1	0.460	4.58
Teenage mother (at time of birth)	0.055	0.228	0	1	0.509	8.96
Teenage father (at time of birth)	0.087	0.282	0	1	-0.265	-8.44
Mother upper secondary school	0.019	0.137	0	1	-0.002	-0.02
Father upper secondary school	0.014	0.119	0	1	0.507	6.85
Mother vocational education	0.017	0.131	0	1	0.634	9.47
Father vocational education	0.330	0.470	0	1	0.329	16.74
Mother short further education	0.405	0.491	0	1	0.257	12.15
Father short further education	0.074	0.262	0	1	0.567	16.58
Mother long further education	0.049	0.215	0	1	0.448	11.13
Father long further education	0.090	0.287	0	1	0.746	22.12
Mother higher education	0.078	0.268	0	1	0.737	20.24
Father higher education	0.019	0.136	0	1	0.805	11.67
Mother's education unknown	0.054	0.227	0	1	0.926	20.04
Father's education unknown	0.096	0.294	0	1	-0.040	-1.17
Mother self-employed	0.096	0.295	0	1	0.061	1.78
Father self-employed	0.085	0.279	0	1	0.000	0.01
Mother student	0.156	0.363	0	1	0.231	6.88
Father student	0.003	0.057	0	1	0.009	0.06
Mother receives social assistance	0.001	0.028	0	1	-0.022	-0.09
Father receives social assistance	0.013	0.114	0	1	-0.525	-6.67
Mother not in the labour market	0.007	0.084	0	1	-0.194	-1.81
Father not in the labour market	0.091	0.287	0	1	-0.285	-7.35
Mother's income from employment	0.052	0.222	0	1	-0.076	-1.66
Father's income from employment	14.605	10.612	0	132.60	0.002	1.15
(Mother's income)H(not with mother)	21.589	18.355	0	407.14	0.008	9.14
	0.612	3.711	0	74.66	-0.003	-0.79

(Father's income)H(not with father)	3.728	10.847	0	282.22	0.000	0.19
Mother's degree of unemployment	7.070	20.268	0	100.00	-0.004	-6.88
Father's degree of unemployment	4.934	16.952	0	100.00	-0.003	-4.92
(M's degree of unempl.)H(not with M)	0.617	6.461	0	100.00	0.002	1.48
(F's degree of unempl.)H(not with F)	1.986	11.422	0	100.00	0.003	2.54
Mother's taxable wealth	0.436	8.144	-87.68	1228.29	0.000	0.05
Father's taxable wealth	1.937	16.364	-277.15	1845.95	0.001	1.89
(Mother's wealth)H(not with Mother)	0.004	0.533	-36.08	38.61	-0.015	-0.94
(Father's wealth)H(not with father)	0.029	5.270	-140.09	696.22	0.000	0.19
Lives in rented dwelling	0.213	0.410	0.00	1.00	-0.364	-14.06
Lives in not-categorized dwelling	0.016	0.124	0.00	1.00	-0.063	-0.88
Type of dwelling unknown	0.007	0.085	0.00	1.00	-0.139	-1.2
No. of rooms per person	1.251	0.490	0.00	5.50	0.208	10.21
No. of rooms per person unknown	0.004	0.060	0.00	1.00	0.096	0.59
Lives in socially deprived area	0.037	0.189	0.00	1.00	-0.165	-3.31
Percent of bilingual children in municipality	2.389	2.875	0.00	18.00	-0.011	-1.8
Rate of unemployment in municipality	9.452	3.034	1.97	23.29	-0.005	-1.03
Percent with vocational education in mun.	33.907	4.022	19.71	46.05	-0.003	-1.04
Percent with further educat. in municipal.	15.046	5.505	5.04	40.60	0.002	0.54
Metropolitan region of Copenhagen	0.275	0.447	0.00	1.00	0.061	1.55
Per cent of pupils living with single parent	13.052	5.340	2.85	30.89	-0.012	-3.23
Average income per capita in municipal.	83.791	16.229	55.65	162.79	-0.001	-0.79
Percent of population living in rural areas	16.807	15.976	0.00	77.00	-0.004	-3.25
Percent in towns with 200-800 inhabitants	7.499	8.585	0.00	58.00	0.000	0.13
Percent in towns with 800-1,500 inhabitants	5.096	8.608	0.00	55.00	0.001	0.54
Percent in towns with 1,500-3,000 inhabit.	7.703	13.914	0.00	91.00	0.001	0.96
Percent in towns with 3,000-5,000 inhabit.	6.092	13.312	0.00	70.00	-0.001	-1.85
Percent in towns with 5,000-10,000 inhabit.	6.987	17.681	0.00	93.00	-0.000	-0.41
Enrollment	59.750	21.508	7.00	168.00	0.005	2.72
Enrollment squared/100	40.327	29.498	0.49	282.24	-0.002	-1.88
Constant term					10.676	35.88

Figure 1. Mean class size (n) for each value of enrollment and expected class size (f)

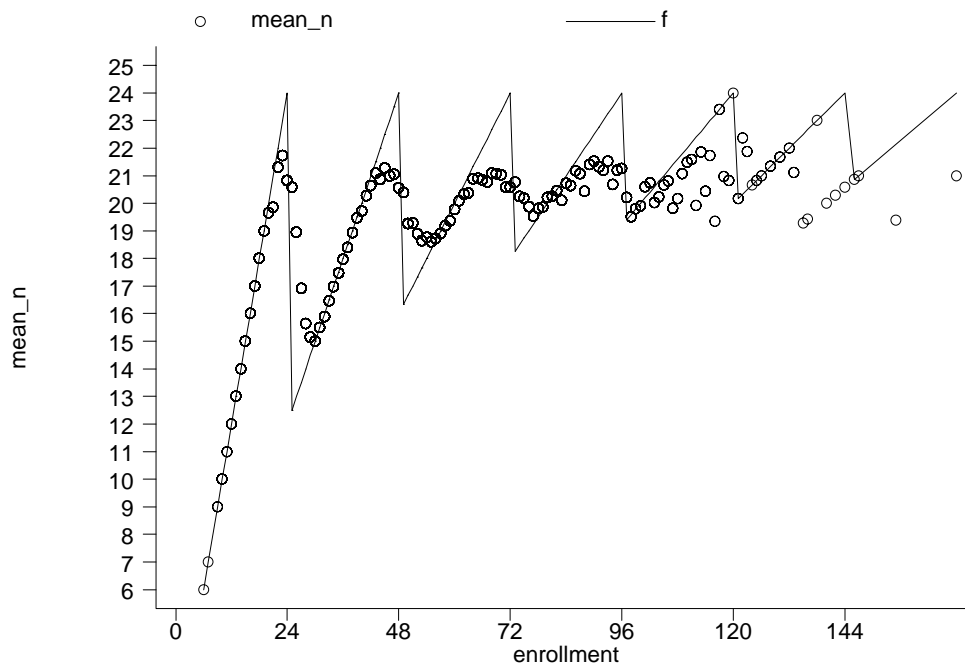


Figure 2. Average number of pupils per teacher hour (h) for each value of enrollment and fitted values from a regression of this variable on the instrument g.

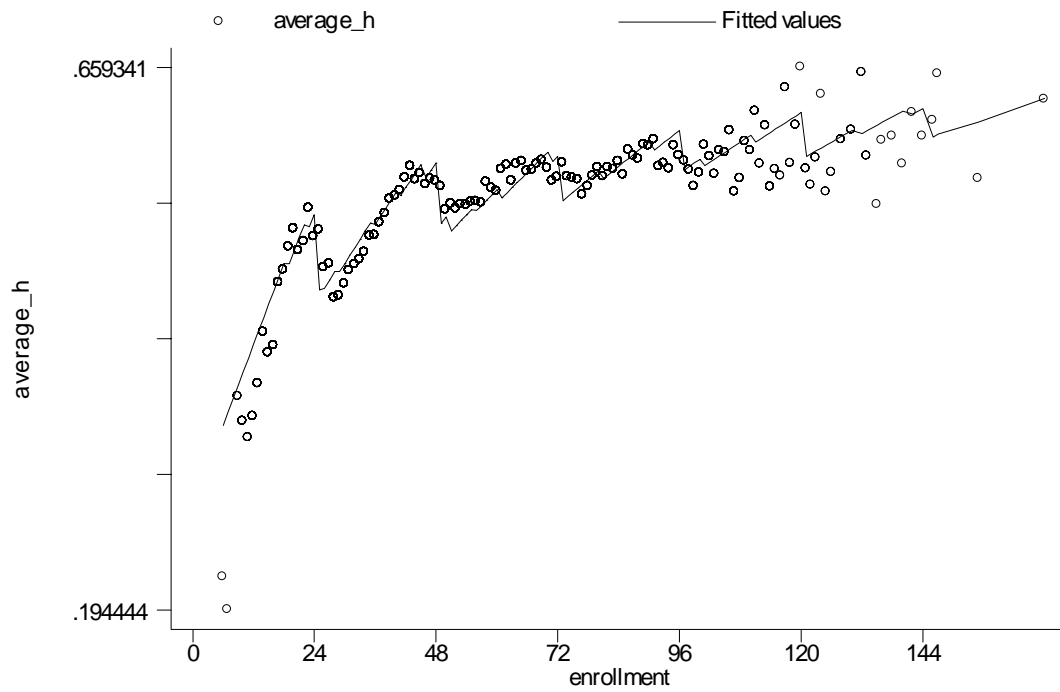


Figure 3. The average number of pupils per teacher hour for each level of enrollment (mean h) for schools in the municipality of Copenhagen in 1991/92 and 1992/1993, and the instrument g based on the Copenhagen administrative rule.

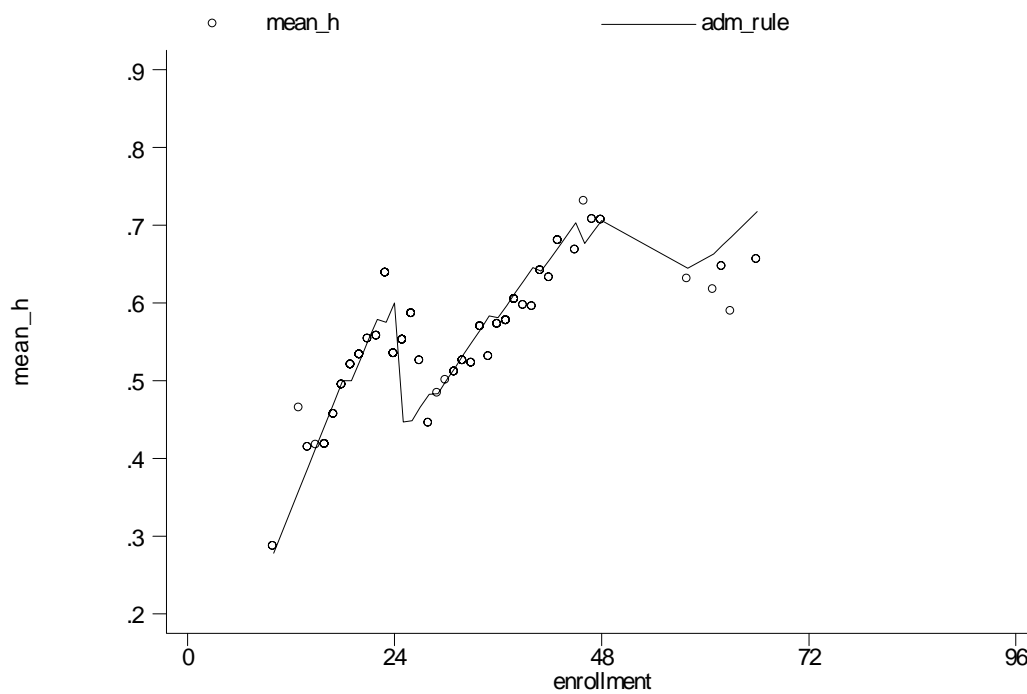


Figure 4 . Kernel density estimate for enrollment

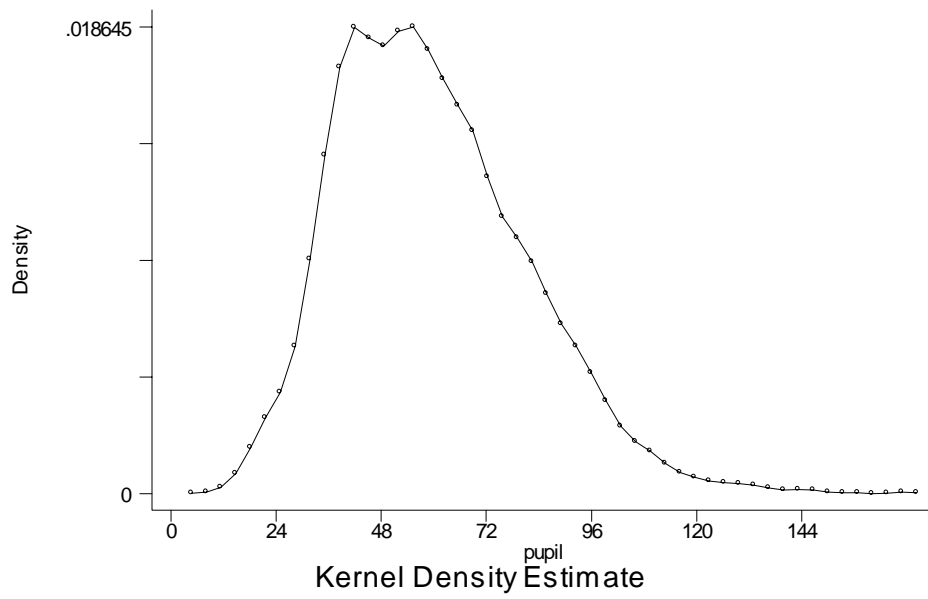


Figure 5. Average length of completed education (in years) and average expected class size (given by the instrument f), in enrollment intervals of 4.

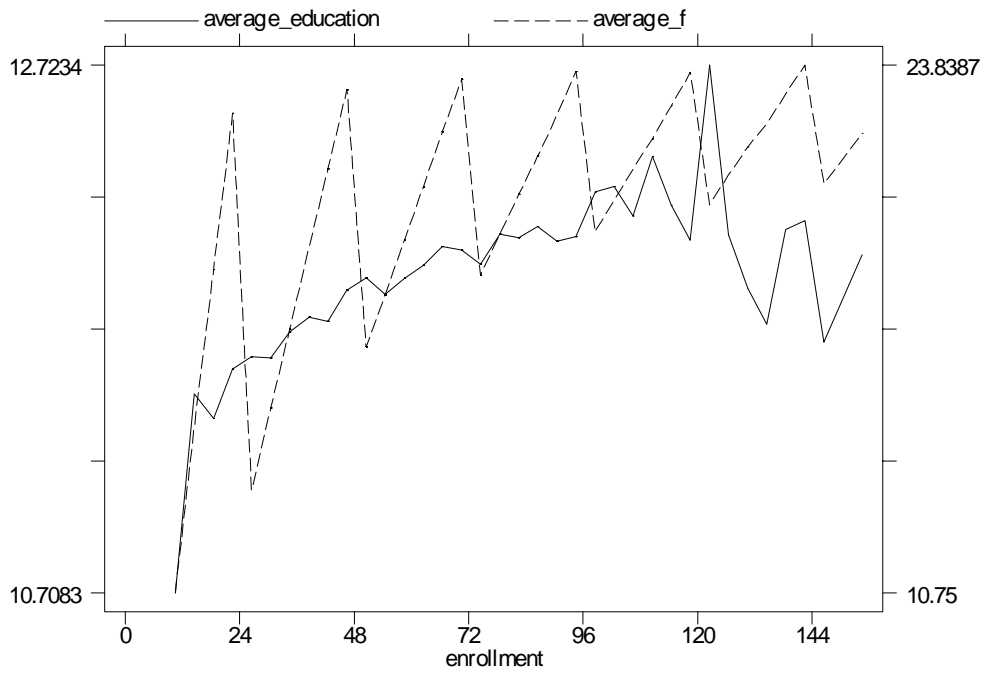


Figure 6. Average length of completed education (in years) and the average expected number of pupils per teacher hour (given by the instrument g), in enrollment intervals of 4.

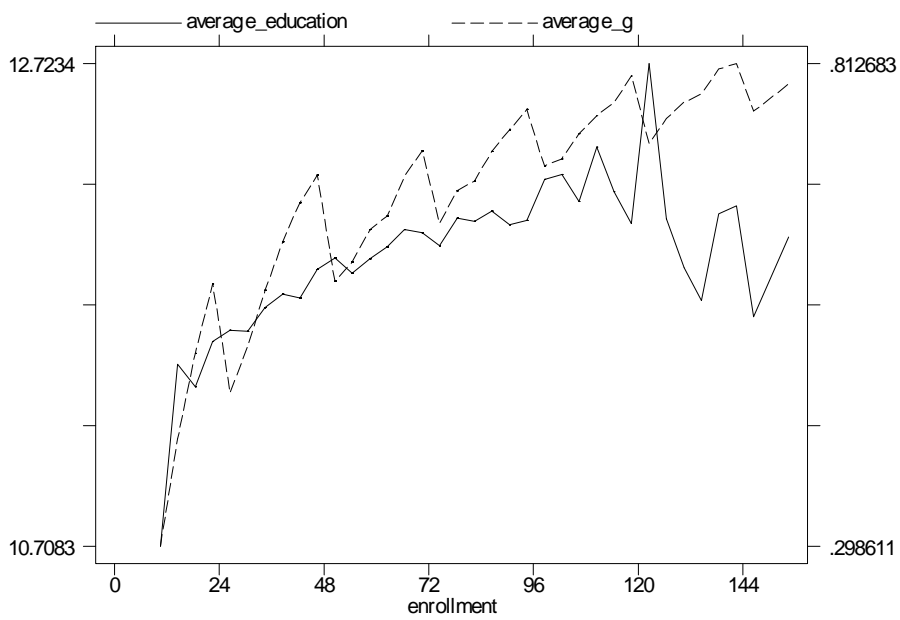


Figure 7. Average length of completed education (in years) and average expected class size (given by the instrument f), residuals from regressions on the full set of control variables, in enrollment intervals of 4.

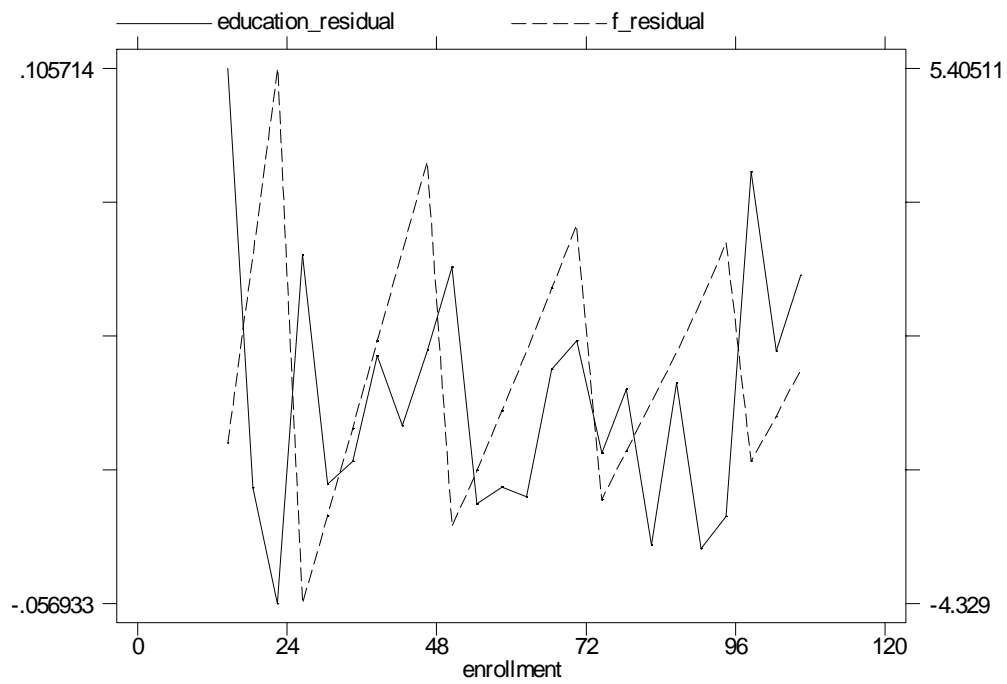


Figure 8. Average length of completed education (in years) and the average expected number of pupils per teacher hour (given by the instrument g), residuals from regressions on the full set of control variables, in enrollment intervals of 4.

