Peer Effects in European Primary Schools: Evidence from PIRLS

Andreas Ammermüller and Jörn-Steffen Pischke

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 06-027

Peer Effects in European Primary Schools: Evidence from PIRLS

Andreas Ammermüller and Jörn-Steffen Pischke

Nontechnical summary

Learning in schools takes place in a group setting, and the composition of the group possibly affects individual outcomes. There has been a lot of interest in these types of social interactions in economics recently, and in peer effects in school in particular. We revisit this issue in this paper, drawing on a previously unexploited data set in this context, the Progress in International Reading Literacy Study (PIRLS) for fourth graders. Our analysis covers six European countries, Germany, France, Iceland, the Netherlands, Norway, and Sweden.

One of the main challenges in the literature on peer effects is the feature that schools and class rooms are not formed randomly. School and class composition typically reflects neighborhood characteristics, and therefore the family background of students. The estimated peer effect may capture unobserved aspects of an individual student's performance if this problem is ignored. We estimate peer effects across classes within primary schools and argue that classes within schools are formed randomly with respect to family background.

We find that a one standard deviation change in our student background measure of peer composition leads to a 0.11 standard deviation change in reading test scores of fourth graders across our sample of six European countries. This is slightly larger than most previous estimates in the literature. The estimated peer effects are highest in France and The Netherlands and lowest in Norway, Iceland and Germany. For Sweden, the estimated effects are also high but might be driven by within-school selection. There is no strong evidence on non-linear peer effects.

We argue that there is little evidence for systematic sorting into class rooms within schools, and for different classes receiving different observable instructional resources. Hence, comparing students in different classes within schools should be an effective way of dealing with any selection at the school level. Surprisingly, we find that this selection does not seem to be very important once we take measurement error issues into account. The discussion in this literature seems dominated with solving the selection issues, while little attention is being paid to the measurement error issues, which we find to be important in our data.

Peer Effects in European Primary Schools:
Evidence from PIRLS

Andreas Ammermueller[*]
ZEW, Mannheim

Jörn-Steffen Pischke
LSE, London

April 2006

Abstract

We estimate peer effects for fourth graders in six European countries. The identification relies on variation across classes within schools. We argue that classes within primary schools are formed roughly randomly with respect to family background. Similar to previous studies, we find sizeable estimates of peer effects in standard OLS specifications. The size of the estimate is much reduced within schools. This could be explained either by selection into schools or by measurement error in the peer background variable. When we correct for measurement error we find within school estimates close to the original OLS estimates. Our results suggest that the peer effect is modestly large, measurement error is important in our survey data, and selection plays little role in biasing peer effects estimates. We find no significant evidence of non-linear peer effects.

1 Introduction

Learning in schools takes place in a group setting, and the composition of the group possibly affects individual outcomes. There has been a lot of interest in these types of social interactions in economics recently, and in peer effects in school in particular. We revisit this issue in this paper, drawing on a previously unexploited data set in this context, the Progress in International Reading Literacy Study (PIRLS) for fourth graders. Our analysis covers six European countries, Germany, France, Iceland, the Netherlands, Norway, and Sweden.

Peer effects are of interest for a variety of reasons. From the policy perspective, it is important to know how students should be grouped in schools and class rooms in order to maximize achievement, or to achieve certain equity goals. Tracking policies, the explicit grouping of students into classes by ability is widespread in many countries. Policies increasing school choice and school competition are being discussed or implemented in many countries. These policies will influence student outcomes through peer effects, if there is increased sorting of students as a result of these policies. Understanding peer effects is therefore an important ingredient in evaluating many education policies.

One of the main challenges in the literature on peer effects is the feature that schools and class rooms are not formed randomly. School and class composition typically reflects neighborhood characteristics, and therefore the family background of students. The estimated peer effect may capture unobserved aspects of an individual student's performance if this problem is ignored. We exploit the fact that the PIRLS data sample multiple class rooms within a single school. This allows us to estimate peer effects within schools. Since we study students in primary schools, there is no explicit tracking in any of the countries in our sample. We argue that classes within schools are in fact formed more or less randomly with respect to family background characteristics (other than immigrant status). The variation in our peer variable therefore most likely reflects that there will be small differences in composition when multiple groups are formed out of a small population (in essence the absence of the law of large numbers). Hence, our research design allows for a relatively credible identification of peer effects on student test scores.

The existing literature, of which some recent studies are summarized in Table 1, has used a wide variety of approaches to identify peer effects. The papers closest in spirit to ours are the ones by Hoxby (2000) for the US and Gould et al. (2005) for Israel. These papers similarly rely on differences in the compositions of individual classes within a school, which come about by chance. Hanushek et al. (2003) and McEwan (2003) also use within school variation to identify peer effects. However, it is more difficult to believe that differences in class composition are random in their cases. We will compare our methodology in detail to the existing literature in the next section. A number of recent studies have also used explicit random assignment to classes or schools, or other natural experiments. However, none of these studies is for European countries.

An important issue in our context is the fact that our peer measure is self-reported by the students' parents and that the sample does not include all students in a class room. These features will lead to measurement error in both the individual and peer level variables in the regression, and hence to biased estimates. Moreover, the size of the bias will differ in the OLS and within school estimates. We address these issues by using alternative variables for family background as instruments, and adjusting our estimates for the sampling error in our measure of the peer variable.

On average across countries, we find that a one standard deviation change in our measure of peer composition leads to a 0.11 standard deviation change in reading test scores, and this estimate is marginally significant. The size of this effect is slightly larger than most estimates reported in the literature. However, the confidence interval for our measurement error corrected estimates is fairly large.

The remainder of the paper is structured as follows. The next section discusses our methodology and compares our approach to the existing literature. Section 3 describes the PIRLS data. In section 4, we investigate the issue of whether classes are formed randomly within schools. Section 5 presents our preliminary estimates of peer effects, and section 6 discusses measurement error problems and how we deal with them. Section 7 converts our estimated peer effects to effect sizes and compares the results with the literature. A final section concludes.

2 Empirical framework and existing literature

A peer effects study typically starts from a specification of an education production function like

$$y_{ics} = \alpha + \beta X_{ics} + \gamma S_{cs} + \delta \overline{X}_{(-i)cs} + \phi \overline{y}_{(-i)cs} + \mu_{cs} + \varepsilon_{ics} \tag{1}$$

where y_{ics} is a student outcome, like a test score, for student i in class room c and school s, X_{ics} are student or family characteristics, like sex, family background, etc., S_{cs} are school or class level characteristics, like class size, teacher experience, characteristics of the municipality, etc., $\overline{X}_{(-i)cs}$ are the average characteristics of the peers of student i, and $\overline{y}_{(-i)cs}$ is the average outcome of the peers. In addition, μ_{cs} and ε_{ics} are a class level and an individual level error term.

In the language of Manski (1993), the coefficient δ reflects exogenous or contextual effects, ϕ reflects endogenous effects, and μ_{cs} reflects correlated effects. Exogenous effects arise when individuals learn more because the group of peers is more favorable in terms of their predetermined characteristics. Correlated effects arise when the group of peers is subject to a common influence, which is not modeled directly. These effects will give rise to a bias if they are correlated with peer group composition. For example, consider a remedial class room with relatively poorly performing children. This class room may be assigned a particularly able teacher but the exceptional characteristics of this teacher are not observable. Removing the potential bias from contextual effects is one of the main challenges in the peer effects literature, and we will discuss this issue shortly. Endogenous effects arise when individuals learn more because peers are learning more.

Even if the requisite data are available, eq. (1) cannot be estimated directly because of the well known reflection problem (see Manski, 1993). Student i influences his or her peers, and the peers influence student i in turn through the coefficient ϕ, if a symmetric equation to (1) holds for every student in the class. This implies that ε_{ics} will be mechanically correlated with $\overline{y}_{(-i)cs}$, leading to a classic simultaneity problem. It is very difficult to overcome the reflection problem without severe restrictions on the model. The literature has therefore basically resorted to estimating the reduced form equation

$$y_{ics} = \frac{\alpha}{1-\phi} + \beta X_{ics} + \gamma S_{cs} + \frac{\delta + \beta\phi}{1-\phi} \overline{X}_{(-i)cs} + \mu_{cs} + \varepsilon_{ics} \qquad (2)$$

$$= \alpha' + \beta X_{ics} + \gamma S_{cs} + \lambda \overline{X}_{(-i)cs} + \mu_{cs} + \varepsilon_{ics}$$

which is the approach we will follow as well. In eq. (2), it is not possible to distinguish endogenous and exogenous effects anymore; either one would give rise to a positive coefficient λ on the peer variable $\overline{X}_{(-i)cs}$. While distinguishing endogenous and exogenous effects may be of interest from a theoretical perspective, for policy purposes it will often not matter which give rise to peer effects. Hence, identifying the reduced form parameter is still of substantial interest.

Now consider the problem arising from the presence of correlated effects in eqs. (1) and (2). If some relevant school or class room characteristics are not controlled, the estimated peer effect λ will be biased. Random assignment of students and teachers to class rooms solves this problem, because random assignment breaks the link between peer characteristics and extraneous effects on the class, like unobserved teacher quality. Boozer and Cacciola (2001) and Graham (2004) exploit the random assignment in the Tennessee STAR experiment on class size. Cullen, Jacob, and Levitt (2003) use lotteries at oversubscribed Chicago public schools. However, their paper does not focus on the issue of peer effects.

True random assignment variation is rare in an education context, and unavailable in many countries. Hence, researchers have to resort to other strategies utilizing the existing data. In this paper, as in a variety of related studies, we use variation within schools in order to identify the peer effect. This means that we estimate

$$y_{ics} = \alpha_s + \beta X_{ics} + \gamma S_{cs} + \lambda \overline{X}_{(-i)cs} + \mu_{cs} + \varepsilon_{ics} \qquad (3)$$

instead of eq. (2). α_s is a set of dummy variables for each school. Alternatively, we also introduce peer variables at the school level directly into equation (2). Both approaches lead to very similar results.

The idea behind these strategies is the observation that different schools draw students from different neighborhoods, and hence family backgrounds. Hence, the unobserved characteristics μ_{cs} will be systematically related to $\overline{X}_{(-i)cs}$ at the school level. However, students are not generally grouped into classes on the basis of ability or family background in primary school. Although some countries, like Germany, track students into a rigid system of separate schools at the secondary level, there is no system wide tracking at the primary level. In fact, classes in primary schools with multiple class rooms at the same grade level are typically formed more or less on a random basis. In this case, $\overline{X}_{(-i)cs}$ will be uncorrelated with the class level shocks μ_{cs} conditional on a set of school fixed effects, or the characteristics of school peers. The bias from correlated effects is thus removed and λ can be estimated consistently.

In order to make this argument more precise, consider the following simple model generating student characteristics:

$$X_{ics} = \eta_{cs} + v_{ics} \tag{4}$$

I.e. student characteristics consist of a common class room level mean η_{cs} and an idiosyncratic, mean zero student level component v_{ics}, which is uncorrelated with η_{cs}. The peer mean is

$$\overline{X}_{(-i)cs} = \eta_{cs} + \overline{v}_{(-i)cs} \tag{5}$$

Correlated effects arise whenever $\text{cov}(\eta_{cs}, \mu_{cs}) \neq 0$. Estimates of both β and λ will therefore be biased in the estimation of equation (2). Our basic identifying assumption is $\eta_{cs} = \eta_s$, i.e. the systematic component of the student background characteristic arises only at the school level but not at the class level. Random assignment of students and resources to classes within schools would ensure that this condition is met. Hence, peer characteristics within schools are $\overline{X}_{(-i)cs} = \overline{v}_{(-i)cs}$, i.e. variation in the peer measure comes only from the fact that $\overline{v}_{(-i)cs} \neq 0$ in small groups. A necessary condition for the within school estimation to work is, of course, that there is sufficient variance in peer composition of a class room within a school. We will demonstrate that this is true in our data in section 4.

Our identification strategy is most closely related to that of McEwan (2003). He studies peer effects for 8th graders in Chile. However, random assignment to classes within schools is much less likely to happen at the secondary level because schools in many countries, including Chile, track students to at least some degree. If there is tracking on the basis of (unobserved) ability, estimates of λ are still confounded by correlated effects.

Gould et al. (2005), Hanushek et al. (2003), and Hoxby (2000) also use within school variation to identify peer effects. The Gould et al. and Hoxby studies are most similar in spirit to ours. We use comparisons across class rooms within the same grade for the same cohort of students. Hoxby uses comparisons between classes in the same grade across adjacent cohorts and years. Hence, she identifies peer effects from variation arising from the composition of subsequent cohorts. For example, one cohort may have more girls and the next cohort fewer for purely random reasons. Gould et al. also use data on multiple cohorts in the same grade. They condition on the student composition of the grade across years. Effectively, like Hoxby, they therefore exploit year to year variation in the composition of a cohort of students. However, these studies tend to focus on different peer group measures than ours. Hoxby looks at gender and race composition of the class room and performance by opposite gender and race groups, while Gould et al. look at the share of immigrants.

Hanushek et al. (2003) focus on a peer measure more similar to ours. They also control for school by grade effects like Hoxby and Gould et al. However, they track the same cohort of students over time, rather than different cohorts, and they control for student fixed effects. This means that they effectively only consider changes in the peer group, which come about through student mobility, because classes are typically kept together in primary schools. This means that an individual's peer group only changes if the individual herself or some peers move in or out of the classroom. But student mobility might well be related to unobserved time and grade specific shocks. For example, if a class is about to be assigned a particularly bad teacher, the parents of the best students may move their children to other schools. Peer quality in the class would fall. Student achievement may also subsequently fall in the class room but this might simply be due to the poor teacher, not the poor peers. Hence, controlling for student fixed effects may lead to an upward bias in the estimates. In fact, Hanushek et al. find an increase in the peer coefficient when they control for individual student effects compared to a similar specification without individual effects.

6

Table 1 summarizes these and other recent studies on peer effects in schools. These previous literature finds peer effects which range from close to zero (Angrist and Lang, 2004, and Cullen, Jacob and Levitt, 2003) to relatively large (Hoxby, 2000).

3 Data and descriptive statistics

Thirty-five countries participated in the Progress in International Reading Literacy Study (PIRLS). This study was conducted by the International Association for the Evaluation of Educational Achievement (IEA) in 2001 and nine- and ten-year-olds in reading literacy. In the data, extensive information on home and school environment is available through student, parent, teacher and school questionnaires. With 150,000 students tested, PIRLS 2001 is the first in a planned 5-year cycle of international trend studies in reading literacy (Mullis et al., 2003).

The data are collected in a two-stage stratified sampling design. First, participating schools were chosen. Therefore, the schools are the primary sampling units and not the classes or students. Within each school, a sample of classes from the targeted grade was drawn. The targeted grade is the upper of the two grades with the most 9 year-olds at the time of testing. This is always the fourth grade in our sample of countries. Within each class, in principle, all students are sampled. In practice, the number of sampled students can be smaller than the actual class size, because of student non-participation. We use all European countries with a sufficient number of schools with at least two classes. These are France, Germany, Iceland, the Netherlands, Norway and Sweden.

Student performance is measured by test scores in reading literacy, which is the most important basic competency needed to acquire further skills and knowledge and to successfully participate in social life (Mullis et al., 2003). The test scores are plausible values that are drawn from an estimated proficiency distribution. Plausible values are imputed scores based on the students' answers to the test items (cf. Mislevy, 1991). The scores have then been standardized to an international mean of 500 and a standard deviation of 100, which facilitates the comparison across countries.

Table 2 provides information on mean reading scores and sample sizes in PIRLS at the student, class and school level. Students, classes and schools can be directly identified. Missing values of student background, class, and school variables are a serious problem in the data set. For parents' education, 36 percent of all values are missing. Instead of parents' education, we use the number of books at home as our indicator of family background. Among the variables reflecting family background, this is the one with best item response rate. In addition, this is an appealing variable in its own right. It is clearly correlated with parental income, education and origin. Table A3 in the appendix presents the weighted means of student background variables by categories of the books at home variable. The variable also reflects whether the parents value literary skills. Parents who own many books most likely will also promote reading among their children. In fact, Woessmann (2004) found the number of books to be the single most important predictor of reading skills among various family background variables in the Third International Math and Science Study (TIMSS) and Ammermueller (2005) in PIRLS and the Programme for International Student Assessment (PISA) data. As an alternative peer measure, we use an index of parents' highest educational attainment and report the results in the Appendix.

Table 2 demonstrates that the sample size, conditioning on non-missing student background and school variables, shrinks to about 40 to 75 percent of the original. Row 6 in the table gives the sizes of the samples we actually use. All figures from row 6 onward refer to the sample with no missing values. Reading scores in the selected samples are slightly higher than in the overall sample. Some sample schools have only one class. Our within school estimates will only be utilizing the schools with two or more classes. Information on the students, classes, and schools with more than one class can be found in the bottom rows of Table 2. Means and standard deviations of all the variables are reported in Table A1, while Table A2 presents the percentage of missing values for each variable in the original data. Both these tables are contained in the Appendix. The peer effects estimations have also been performed including all observations for which test scores are reported. Missing values have been replaced by zeros and dummy variables for missing values for each variable have been added to the regressions. The estimated peer effects are comparable to the evidence presented in section five and are available from the authors upon request.

The home questionnaire asked parents to report the number of books in their home in five categories: none or few books (0 – 10), enough to fill one shelf (11 – 25), enough to fill one

bookcase (26 – 100), enough to fill two bookcases (101 – 200), enough to fill three or more bookcases (more than 200).[1] In order to form a single measure of students' background, after some experimentation, we chose a simple index which assigns 1 to the lowest category (0 – 10), and 5 to the highest category (more than 200). The median parent reports 26 – 100 or 101 – 200 books, and the mean of the indices range from about 3.3 to 4, depending on the country (see Table 3 below and Table A1).

We generated peer variables as the class average of five student background variables: number of books at home, student's sex and age, whether at least one parent was born abroad, and whether a foreign language is spoken at home. There is an argument in the literature on peer effects whether class rooms or schools (or possibly even neighborhoods) are the more appropriate unit of peer interactions. Of course, peer interactions may occur at each of these levels, and it is an open question which is the most important. We focus on the class room level for the pragmatic reason that we want to analyze differences within schools. In the within school estimates, all peer interactions with students from other classes in the school will be absorbed into the school fixed effects. However, peer effects in the class room are clearly of interest, since classes are the basic unit where learning takes place. It is therefore natural to expect that a large fraction of total peer effects should arise at the class room level.

The peer averages are formed using information for all students who report a value for this specific variable in the data set, not just those students in the final sample. In Table 3, we decompose the total variance in these class averages into the parts of the variance within and between schools using the relationship

$$\frac{1}{N}\sum_{s=1}^{S}\sum_{c=1}^{C_s}\left(x_{cs} - \overline{x}\right)^2 = \frac{1}{N}\sum_{s=1}^{S}\sum_{c=1}^{C_s}\left(x_{cs} - \overline{x}_s\right)^2 + \frac{1}{N}\sum_{s=1}^{S}C_s\left(\overline{x}_s - \overline{x}\right)^2 \tag{6}$$

where x is the specific variable we are interested in, $s = 1, 2, …, S$ is a school indicator, $c = 1, 2, …, C_S$ is a class indicator, and there are C_S classes in school s. N is the total number of classes across all schools in our sample.[2]

---

[1] Using instead the number of books at home reported by students yields comparable results.
[2] For the variance decomposition to add to the total variance in an unbalanced panel, it is necessary to weight the between component by the number of classes in the sample. This is not what, for example, the Stata `xtsum` command calculates.

Table 3 presents the total, between and within school variance of the peer variables. The variation for the average reading test score is shown as well. It is obvious that most of the variance in all of these measures is between schools. Between 7 and 18 percent of the variance in the index for the number of books at home is within schools. The fraction is higher for the reading test scores. However, 70 percent or more of the test score variation is also between schools. This suggests that a large part of the variation in all these measures is accounted for by school effects. Nevertheless, there is also some non-negligible amount of variance left within schools.

4   Selection in class room formation

In this section we will discuss the assignment of students both between and within schools. We start by presenting some basic information on primary schooling in the countries we study. We then go on to present some evidence from the PIRLS data to shed light on the question whether classes are formed (more or less) randomly, and whether different class rooms systematically get different resources.

In all six countries in our sample, students attend a single track primary school from school enrolment to at least grade four, in which students have been tested in PIRLS.[3] While students are assigned to various school types after grade four in Germany, they stay on for at least two more years in primary school in most other countries (France, Iceland, the Netherlands, and Sweden) or go on to a single tracked secondary school (Norway). School choice for primary school is free in some countries (Germany and the Netherlands) and depends on the place of living in the other countries. However, parents have some means to influence the choice of schools also in these countries. In practice, most parents choose the nearest school for convenience in all countries (or live near the school of their choice). The heads of the school are responsible for the assignment of students to classes within schools. Most countries have legal rules on maximum class size and some school systems provide extra resources for schools with a high share of migrant students. The final responsibility lies with the heads of school, however. Grouping of students seems to happen in some cases based on the migration background of students. We have found no direct evidence that primary

---

[3] The information on the schooling systems is taken from Eurybase, the database in the information network on education in Europe, http://www.eurydice.org.

schools in any of the sample countries use any systematic ability grouping or sorting by family background in grade four.[4]

The PIRLS data asked in the schools questionnaire whether the school forms sample classes on the basis of ability. The last row in Table 2 reports the fraction of students in schools that report some ability grouping at the class level. This fraction is very low except in France and the Netherlands, where it reaches in the order of 30 percent. While we do not find much evidence that the classes in these tracked schools look very different from classes in other schools, we also show results excluding these schools which report some tracking.

We investigate two separate and distinct questions about class room formation with the PIRLS data. The first question is whether class rooms which differ in composition, for random or non-random reasons, receive different resources. The second question is whether the data are consistent with classes being formed randomly. In order to shed light on the first question, we ran a set of regressions of the peer variables described in the previous section on class room, teacher, and school characteristics. The observable characteristics of class rooms and schools which we use are class size, teacher gender, education, and experience, size of the town or city, instruction hours per day, and the degree of shortages of staff, materials, or rooms that the school reports.

Table 4 illustrates these regressions for Germany. For books at home, the only significant variable is shortage of rooms in the school, while none of the class level variables are individually significant. The other peer characteristics are also typically unrelated to class room and school characteristics, with coefficients usually being small and insignificant. The only exception is immigrants and non-native speakers, who tend to be clustered in more urban schools. Classes with more immigrants also tend to have better educated teachers.

Table 5 shows p-values for the corresponding F-tests on the joint significance of the variables. The first three rows of this table refer to simple OLS regressions of the type run in Table 4, which include both class room and school characteristics. F-tests for the class room level variables, the school level variables, and both groups together are reported. The fourth row reports the F-tests for the class level variables from a regression including school fixed effects. For our family background variable of interest, the number of books at home, the

---

[4] We consulted researchers in each of the sample countries and they also confirmed this impression.

class and school variables are typically insignificant, with a few exceptions. School level variables seem to matter in the Netherlands and Norway. This may simply indicate that students from more favorable backgrounds are not randomly allocated to schools. There is no correlation with class level variables, however. So this should not affect our conclusions when we look at within school results. There is some evidence for class level variables being related to family background in Iceland and in Sweden. In the case of Iceland, this only appears in the fixed effects regressions. It turns out that this correlation is solely driven by a single class room with a teacher with 20 years of experience (while all other teachers in Iceland have 10 or fewer years of experience). We discount this result as spurious. In the case of Sweden this seems indeed to indicate a non-random allocation of class room resources to classes with students from different background, even within schools. In particular, there is evidence that class size increases with average background of students in a class. The coefficients for the other class and school variables are not significant.

We also find some evidence that class rooms differ for students by age (in Germany, Iceland, and Norway) and by student sex (in Iceland, the Netherlands, and Norway). It also seems fairly clear that classes are different for immigrant students in all our sample countries. The higher the share of immigrant students in a class, the lower is teacher's education in Germany and Norway. Instead, teacher's education increases with the share of immigrants in France. In Sweden, there is weak evidence for an allocation of immigrant students to larger classes.

In order to test whether class rooms are formed randomly with respect to a particular student characteristic, we perform a series of Pearson Chi² tests. If classes are formed randomly, the student characteristic under study and the class the student is assigned to should be statistically independent. For example, consider student sex. The Pearson Chi² test asks whether there are more females in a particular class than is consistent with independence, given the number of students in the school. Formally, for each school the test statistic is given by

$$P = \sum_c \sum_j \frac{\left(n_{cj} - \hat{n}_{cj}\right)^2}{\hat{n}_{cj}} \qquad (7)$$

where n_{cj} is the number of students with characteristic j in class room c, $c = 1, \ldots, C$, $j = 1, \ldots, J$. Define

$$n_{c\bullet} = \sum_j n_{cj} \qquad\qquad n_{\bullet j} = \sum_c n_{cj} \qquad\qquad \hat{n}_{cj} = \frac{n_{c\bullet} n_{\bullet j}}{\sum_c \sum_j n_{cj}}$$

where $\hat{n}_{cj}$ is the predicted number of students with characteristic j in class room c when characteristic and class room are independent. Then, under the null hypothesis of independence, $P \sim \chi^2$ with $(C - 1)(J - 1)$ degrees of freedom.

We further assume that the S schools in a country are independent. In this case, we can simply add up the S test statistics to get an aggregate test statistic with $\left[\sum (C_s - 1)\right](J - 1)$ degrees of freedom (see, e.g. DeGroot, 1984, p.384). Obviously, the test can only be carried out on the sub-sample of schools with two or more class rooms. These test statistics are shown in the fifth row of each panel in Table 5. We found in a small Monte Carlo experiment that the test generally performs well but rejects somewhat too often under the null hypotheses in samples of our size.

Figure 1 shows the histograms of the school specific p-values for this test for Germany. Under the null hypothesis of random assignment to classrooms, the p-values should be (roughly) uniformly distributed. A policy of deliberate balancing of students' characteristics across classrooms (e.g. making sure that the same fraction of girls are in each class) would lead to a left-skewed distribution, while an assignment of similar students to the same classroom would lead to a right-skewed distribution (cf. Graham, 2004). Figure 1 suggests that students are roughly randomly assigned to classes on the basis of family background and sex. However, there is clustering of immigrant and non-native language students in certain classes, as low p-values predominate for these tests.

Table 5 confirms these results. The p-values of the aggregate tests for books at home and sex are 0.24 and 0.12, respectively. However, the p-values are only 0.016 for foreign born, and 0.050 for non-German speaking children. We also find evidence of non-random assignment of immigrant children for the Netherlands and Sweden. In addition, there seems to be some evidence that children of a similar family background are grouped together in Sweden.

Recall that principals in a significant number of schools in France and the Netherlands report ability grouping in their schools. The p-values presented in Table 5 differ only slightly between these schools and those that report no tracking in France and the Netherlands. One exception is the evidence for sorting by age within the 19 schools that may be tracked in France.

Our results largely confirm that classes in the sample countries seem to be formed roughly randomly within schools. There is little evidence that students of different family backgrounds are more likely to be grouped in certain classes conditional on the school they attend, or that classes with different compositions receive different (observable) resources. This is comforting for our analysis. The only country, where this does not seem to be the case, is Sweden. Hence, the Swedish results may have to be taken with a grain more of salt. In addition, immigrant children, which are an important group in all of the sample countries, also seem to be non-randomly assigned and given different teaching resources.

5 Basic results on peer effects

We now turn to our results on peer effects. Table 6 reports results for Germany, which is the country with the biggest sample, and, in particular, with the largest number of schools with at least two classrooms. We report results for two versions of the peer variable on books at home and two versions of the individual level regressor on books at home. Since the variable takes on five categories, the most nonparametric way to use the variable is to use the fraction of peers in four of the categories. Alternatively, we use the 1 – 5 index we created from these five categories. In each case, the peer variable for student i used in the regressions is the leave-out mean for the classroom, omitting the value of the variable for student i from the calculation of the mean.

In the upper part of Table 6, four dummy variables are used as the individual level control variable, while the index is used in the lower part of the table. In each case, only the coefficients on the peer variable are shown in the table. Column 1 shows results when we only control for student level characteristics (gender, age, immigrant status, language at home, household size as well as the control for the number of books in the home of the student). Going from a classroom where all peers have 26-100 books to a classroom where all peers

have 101-200 books, a change of about a two standard deviations of the peer composition across classrooms, is associated with a 19 point increase in the test score. This is about 32 percent of the student level standard deviation in test scores. We find a very similar result in the specification with the single index measure for peers' books at home. This is not surprising since in practice most of the variation in the index variable occurs in the range between the 26-100 and the 101-200 books categories. Results are very similar, no matter how books at home are controlled for at the individual level.

In column (2) of Table 6, we add class and school level covariates to the regression. Adding these covariates does little to the results. This is not surprising, since we found in Tables 4 and 5 above that the class and school level covariates are basically uncorrelated with our measure of peer quality. In column (4) we add a school level measure of the peer group. This changes the results dramatically. The coefficient on the index variable drops from about 18 to zero and is insignificant. When the peer effects are measured by the fraction of peers in each category, the peer effects are no longer monotonic now. However, these effects are also imprecisely estimated. Finally, in columns (5) and (6), we add a full set of school fixed effects, leading to comparable results as for the school peer measures. The exclusion of schools that reported to form classes on the basis of student ability in column (6) hardly changes the results. Using the categories or the index for the individual books variable leads to very similar estimated peer effects. Using the peer's index of books at home seems a useful summary. Standard errors for the categorical measure of the peers' variable are often so high as to make it difficult to draw any conclusions from the results. For ease of interpretation, we will therefore only present results which use the index of books at home for both the individual level and the peer variable.

Table 7 summarizes the results for all the six countries. We find a relatively consistent pattern of results for all six countries in our sample. The size of the estimated peer effect is similar across the specifications with and without school and class level variables, and is in the order of 15 to 22 for moving peer quality to the next higher category. Only in Norway does the peer coefficient fall when school level covariates are added to the regression. Once we include school peer measures in column (3) or school fixed effects in column (4), the effect always falls, although the amount of the change is different across countries. In Germany, Iceland, and Norway the peer effect weakens the most in these specifications, while there is little change in France. Excluding schools that form classes based on student ability

predictably only changes the results in France and the Netherlands, the two countries with moderate shares of students in schools which form classes based on ability (see Table 2). Curiously, estimated peer effects are larger when the schools which report tracking are excluded for the Netherlands.

One reason for the high variation in the coefficients from the fixed effects models is that the standard errors of these estimates are reasonably large, so that the effects for each individual country cannot be estimated very precisely. If we believe that the peer effects are the same in each country then it makes sense to combine the estimates into a single estimate. The average of the six coefficients in the fixed effects specification in column (4), weighted by the inverse of the sampling variance, is 8.5. If the variation in country level estimates around this overall mean is only due to sampling variation, then the standard error for the meta-estimate is 3.3.[5] This estimate is much more precise than the country level estimates, and it is significant at the 5 percent level. One concern is with the results for Sweden, because we found some evidence for non-random assignment and targeted class room resources for Sweden above. The meta-estimate for the countries without Sweden is only slightly lower.

Our results show that standard OLS estimates of the peer effect may be biased upward substantially if the within school results are indeed reliable estimates of the true peer effect. This is true even after controlling for the typically available measures of class room and school characteristics. One reason why even the fixed effects estimates may be biased is the presence of immigrant children. We showed above that immigrant children are often not randomly assigned to classes within schools, and the classes with many immigrant children may get different resources. Since immigrants in these countries tend to be of lower SES (the index for books at home is on average 3.15 for immigrant families in the six countries but 3.56 for non-immigrant families), part of the peer effect may be explained by the non-random allocation of immigrants.

In order to probe this, we reran the regressions in Table 7 including the fraction of foreign born children in the class, and the fraction of children speaking a foreign language at home.

---

[5] The sampling variance of the mean is obtained as $v = \left[\sum v_c^{-1} \right]^{-1}$, where v_c is the sampling variance of the estimate for country c. One interpretation of this calculation is that the country average is the minimum distance estimate of the common peer effect across countries.

This attenuates the estimated peer effects at most very slightly.[6] We also experimented with regressions on the sub-sample of schools with few immigrant children. However, most sample countries have enough immigrants that there are relatively few such schools leading to small samples, and hence imprecise estimates. These results indicate that the effect of immigrant children in a class may be relatively well captured by our family background variable.

A further question is whether peer effects vary across students. This could give insights into the optimal assignment of students to classes. When students from a lower social background profit more from their peers' background than students from a high social background, more heterogeneous classes would benefit overall performance (Glewwe, 1997). To investigate this, we add interaction effects between the peer variable and the individual variable books at home to the regressions presented in Table 7. Since about half the students have more than 100 books at home, we interact the peer average with a dummy indicating whether the individual reports more than 100 books at home. The results are presented in Table 8. Peer effects seem to be stronger for students with a higher social background in France and the Netherlands, while they are stronger for students with a lower social background in Sweden and Norway. The meta-estimates are small and insignificant.

6   Measurement Error

Survey reports are subject to a lot of measurement error. In our case, measurement error in the books at home variable implies that there is measurement error in both the individual and the peer level regressor. In addition, the peer measure is not based on all students in a class because some students have not been sampled and others have not responded to the respective question. This problem will also arise in many studies based on administrative data, which frequently use lagged test scores as peer measure, since test taking may be incomplete or lagged scores cannot be matched to all students. Both these measurement problems will interact in leading to biased estimates of the peer effect in a non-standard way.

---

[6] Including measures from the school questionnaire on the share of economically disadvantaged students or the share of students leaving before the end of the academic year as alternative school level variables did not change the results.

In order to investigate the impact of measurement error in this setup we will return to the model we outlined in equations (2), (4), and (5) above. In order to focus on the variables of interest, consider a simplified version of equation (2) with only the individual level and the peer group regressor but no other variables:

$$y_{ics} = \beta X_{ics} + \lambda \overline{X}_{(-i)cs} + \varepsilon_{ics} \tag{8}$$

The student background variable X_{ics} is still given by equation (4) but this variable is not directly observed. Instead we observe

$$\tilde{X}_{ics} = X_{ics} + u_{ics} = \eta_{cs} + v_{ics} + u_{ics} \tag{9}$$

where u_{ics} is a classical measurement error. Our argument above has been that the correlated shocks η_{cs} only arise at the school level. Hence, we can think of our standard OLS results corresponding to those with $\sigma_\eta^2 > 0$ and the within school results to $\sigma_\eta^2 = 0$.

In this setup, the OLS estimate of λ will converge to

$$p\lim \hat{\lambda}_{OLS} = \beta \frac{(n_{cs}-1)\sigma_u^2 \sigma_\eta^2}{\left(\sigma_v^2 + \sigma_u^2\right)\left(n_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right)} + \lambda \frac{\dfrac{n_{cs}-1}{N_{cs}-1}\sigma_v^2\left(\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right) + (n_{cs}-1)\sigma_\eta^2\left(\sigma_v^2 + \sigma_u^2\right)}{\left(\sigma_v^2 + \sigma_u^2\right)\left(n_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right)} \tag{10}$$

as we show in the appendix. N_{cs} is the actual number of students in classroom c in school s and n_{cs} is the number of students in the sample. In order to understand the different sources of measurement error and the sign of the bias, it is instructive to look at some special cases. First, consider the case where all students in each class are sampled, so the only problem is classical measurement error. In this case

$$p\lim \hat{\lambda}_{OLS} = \beta \frac{(N_{cs}-1)\sigma_u^2 \sigma_\eta^2}{\left(\sigma_v^2 + \sigma_u^2\right)\left(N_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right)} + \lambda \frac{\sigma_v^2\left(N_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right) + (N_{cs}-1)\sigma_\eta^2 \sigma_u^2}{\sigma_v^2\left(N_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right) + \sigma_u^2\left(N_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right)} \tag{11}$$

It is easy to see in this formulation, that the second term implies an attenuation bias of λ if there is classical measurement error in X_{ics}. This measurement error will carry over to $\overline{X}_{(-i)cs}$, and lead to the standard attenuation. Since λ is likely positive, this will imply an underestimate of λ. Returning to equation (10), it becomes clear that the attenuation is greater, when some students in the class are not sampled. If $\sigma_\eta^2 > 0$ a second component of the bias arises, and this is captured by the first term in equation (11). The individual level regressor X_{ics} is also subject to error and hence will lead to an attenuation of the estimated $\hat{\beta}$. Since the peer variable contains $\overline{X}_{(-i)cs}$ information on η_{cs}, part of the signal in the individual level regressor will load on to the peer coefficient. This term is positive, and hence yields an upward bias.

Because of these two conflicting sources of bias it is impossible to tell what the net effect of the bias on $\hat{\lambda}_{OLS}$ is. The first term can dominate when β is sufficiently larger than λ. Hence measurement error may not lead to an underestimate of the peer effect in the standard OLS specification. Since the within school model corresponds to the case where $\sigma_\eta^2 = 0$, the first term in equation (10) and (11) vanishes, and the peer effect is now underestimated. Hence, measurement error alone may explain why we find lower peer effects in the fixed effects estimates in Table 7.

In order to address the measurement error problem, we rely on the standard remedy by using instruments for both X_{ics} and $\overline{X}_{(-i)cs}$. Recall that in our case the background variable X_{ics} is the parents' report of the number of books at home. The same question was asked of the students as well, so we use the students' report of the number of books at home as our instrument for the parents' report, and the peer mean of the students' report as instrument for the peer variable. A second plausible instrument is parents' education. However, books at home may not be a sufficient statistic for the way family background affects students' reading skills, and parents' education may affect students independently. Hence, we also re-estimate the IV models with only the student variable on books.

If the instrument is valid, the IV estimate of λ will converge to

$$p \lim \hat{\lambda}_{IV} = \lambda \left(\frac{n_{cs} - 1}{N_{cs} - 1} \right) \frac{N_{cs} \sigma_\eta^2 + \sigma_v^2}{n_{cs} \sigma_\eta^2 + \sigma_v^2}. \tag{12}$$

This is the same as the expression in equation (10) with $\sigma_u^2 = 0$, so IV solves the standard measurement error problem. It does not resolve the attenuation in the peer effect that arises due to the fact that we do not sample all the students in a class. When $\sigma_\eta^2 = 0$, equation (12) becomes

$$p \lim \hat{\lambda}_{IV} = \lambda \left(\frac{n_{cs} - 1}{N_{cs} - 1} \right). \tag{13}$$

This suggests that the within school IV estimate is simple to adjust for the sampling bias since N_{cs} and n_{cs} is known in our data. One problem we have ignored in these calculations is the fact that class sizes are not the same for each class. However, we find in a simple Monte Carlo experiment that using the adjustment in (13) works very well for the level of variation in class sizes present in our data. Our adjusted IV estimator will therefore be

$$\hat{\lambda}_{IVadj} = \hat{\lambda}_{IV} \left(\frac{\overline{N}_{cs} - 1}{\overline{n}_{cs} - 1} \right).$$

The first stages corresponding to our IV regressions indicate that both the relevant instruments for the individual level regressor and for the peer variable are always highly significant. The t-statistics on the students' report of books at home and parents' education are above 7 and typically above 10, and the corresponding F-statistics are also large.[7] This indicates that our IV models are not likely to suffer from any small sample bias.

Table 9 presents the results from the IV regressions. Both the individual and peers' index of the number of books at home from the home questionnaire are instrumented by the individual and peers' index of books at home from the student questionnaire and parents' highest educational level. Including only student level variables, the average peer effect across all

---

[7] The only exception is the Netherlands, where the instruments for the peer variable are not individually significant. However, the F-statistic for the joint significance of the instruments is above 10.

countries decreases slightly from 19 in the OLS regression in column (1) to 15 in the IV regression in column (2). Recall from equations (10) and (12) above that the bias in the OLS regression may be positive if $\sigma_\eta^2 > 0$.

Columns (3) and (4) report the corresponding results when school fixed effects and class level regressors are included. We know already that the standard fixed effects estimate is much smaller than the OLS estimate in column (1). Applying IV to the within estimates yields a much higher average peer effect of about 15. The magnitude of this coefficient is very similar to the pooled IV model in column (2). This is perfectly consistent with our results in equations (11) and (13). Column (5) applies the correction for sampling bias which raises the estimate slightly. Interestingly, this is actually slightly higher than the IV estimate in the pooled sample without school fixed effects. We should note that the average peer effect across countries masks large differences in the country level estimates. The coefficients in columns (4) and (5) basically only increase in France, and to a lesser degree in the Netherlands in Sweden, while they fall in the other countries. However, the country level IV results are very imprecisely estimated. Hence we focus on the average across countries. Finally, in column (6), we exclude tracked schools and the peer effect falls slightly but remains large. If we use only the children's report of books as instrument but not parental education, we obtain slightly larger IV estimates but these estimates are also less precise. Overall, these results suggest that the bias due to measurement error is important, particularly in the fixed effects results, while omitted variables bias actually does not play a large role.

7   Effect sizes

In order to facilitate comparisons with other studies, it is common in the literature to report effect sizes of the peer effects estimates. Effect sizes are typically calculated as $\sigma_{\bar{X}}\hat{\lambda}/\sigma_y$ where $\sigma_{\bar{X}}$ is the within country variation in the peer variable, and σ_y is the within country variation in the test scores. This quantifies the peer effect as the impact of a one standard deviation change in peer background in terms of individual level standard deviations of the outcome variable.

One complication with this measure in our context is that the standard deviation of the peer variable is not an unbiased measure of $\sigma_{\bar{X}}$ because of the measurement error. However, since we have both the parents' and the children's reports for books at home, the covariance of the two is a measure of the variance of the true variable if both reports are only subject to classical measurement error.

We report the effect size measure and the necessary ingredients in Table 10. As before, the effect sizes vary quite widely across countries, ranging from zero in Norway to almost 0.5 in France. The average effect size across countries is 0.11. This is larger than most of the estimates in the literature. The bulk of the reported effect sizes is in the range of 0.05 – 0.10. Our estimate is at the upper end of that range but well below the highest estimates reported in studies by Hoxby (2000) and McEwan (2003). One reason why our estimates are higher might be that we are careful about the measurement error in the peer effects variable. However, adjusting for measurement error lowers the estimate of $\sigma_{\bar{X}}$ and raises the estimate of $\hat{\lambda}$, so this cuts two ways. However, the upward adjustment in $\hat{\lambda}$ is much more important. Calculating the effect size on the basis of the estimates ignoring measurement error yields a value of only 0.06, almost half the size of our adjusted result. Hence, the treatment of measurement error may be rather important, particularly in studies based on survey data, like Schindler-Rangvid (2003) and Schneeweiss and Winter-Ebmer (2005).

8   Conclusion

Peer effects are potentially a major input into the process of educational production but are difficult to estimate empirically. We estimate peer effects across classes within primary schools and argue that classes within schools are formed randomly with respect to family background. We find that a one standard deviation change in our student background measure of peer composition leads to a 0.11 standard deviation change in reading test scores of fourth graders across our sample of six European countries. This is slightly larger than most previous estimates in the literature. The estimated peer effects are highest in France and The Netherlands and lowest in Norway, Iceland and Germany. For Sweden, the estimated effects are also high but might be driven by within-school selection.

We have argued that there is little evidence for systematic sorting into class rooms within schools, and for different classes receiving different observable instructional resources. Hence, comparing students in different classes within schools should be an effective way of dealing with any selection at the school level. Surprisingly, we find that this selection does not seem to be very important once we take measurement error issues into account. The discussion in this literature seems dominated with solving the selection issues, while little attention is being paid to the measurement error issues, which we find to be important in our data.

References

Ammermueller, Andreas (2005), Educational opportunities and the role of institutions, ZEW Discussion Paper No. 05-44.

Angrist, Joshua and Kevin Lang (2004), Does school intergration generate peer effects? Evidence from Boston's Metco Program, *American Economic Review* 94(5), 1613-1634.

Boozer, Michael and Stephen Cacciola (2001), Inside the 'Black Box' of Project STAR: Estimation of peer effects using experimental data, Economic Growth Center Center Discussion Paper No. 832, Yale University.

Cullen, Julie, Brian Jacob and Steven Levitt (2003), The effect of school choice on student outcomes: Evidence from randomized lotteries, NBER Working Paper No. 10113.

DeGroot, Morris (1984), *Probability and statistics*, 2nd edition, Reading, MA: Addison Wesley Publishing Company.

Gibbons, Stephen and Shqiponja Telhaj (2005), Peer effects and pupil attainment: Evidence from secondary school transition, London School of Economics, mimeo.

Glewwe, Paul (1997), Estimating the impact of peer group effects on socioeconomic outcomes: Does the distribution of peer group characteristics matter? *Economics of Education Review*.

Gould, Eric, Victor Lavy and Daniele Paserman (2005), Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence, Hebrew University, mimeo.

Graham, Bryan (2004), Identifying social interactions through excess variance contrasts, Harvard University, mimeo.

Hanushek, Eric, John Kain, Jacob Markman and Steven Rivkin (2003), Does peer ability affect student achievement?, *Journal of Applied Econometrics* 18 (5), 527-544.

Hoxby, Caroline (2000), The effects of class size on student achievement: New evidence from population variation, *Quarterly Journal of Economics* 115(4), 1239-1285.

Manski, Charles (1993), Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies* 60(3), 531-542.

McEwan, Patrick (2003), Peer effects on student achievement: Evidence from Chile, *Economics of Education Review* 22(2), 131-141.

Mislevy, Robert (1991), Randomization-based inference about latent variables from complex samples, *Psychometrika* 56, 177-196.

Mullis, Ina, Michael Martin, Eugene Gonzalez and Ann Kennedy, (2003), *PIRLS 2001 international report: IEA's study of reading literacy achievement in primary schools*, Chestnut Hill, Boston College.

Schindler-Rangvid, Beatrice (2003), Educational peer effects: Quantile regression evidence from denmark with PISA 2000 data, Kapitel 3 i "Do schools matter?", phd-afhandling nr. 2003:9 fra Handelshøjskolen i Århus.

Schneeweiss, Nicole and Rudolf Winter-Ebmer (2005), Peer effects in Austrian schools, Department of Economics, University of Linz, Working Paper No. 0502.

Woessmann, Ludger (2004), How equal are educational opportunities? Family background and student achievement in Europe and the United States, CESifo Working Paper No. 1162, Munich.

Figure 1: Distribution of school-specific p-values for Pearson Chi² test for Germany
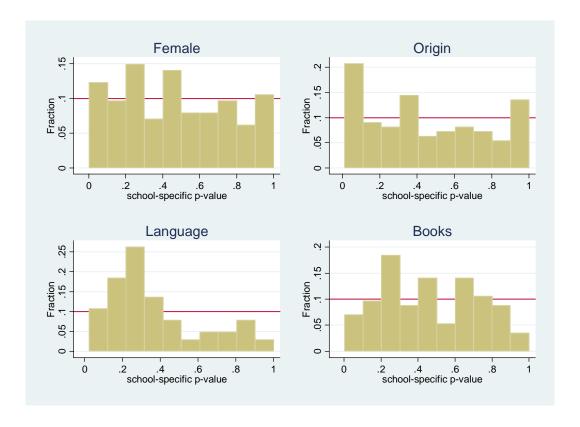
Table 1: Recent studies on peer effects

Studies	Context	Grade / student age	Outcome	Peer measure or treatment	Peer group	Identification	Magnitude of peer effect	Source of Magnitude	Non-linearities in peer effects
Hoxby (2000)	Public schools, Texas, US	3rd to 6th grade	Test scores (Reading, Math)	Test scores	Class	Cohort gender variation	1.s.d. → ~0.4 s.d.	Table 9, columns 3 and 4	No evidence
McEwan (2003)	Secondary schools, Chile	8th grade	Test scores (Spanish)	Mothers' education	Class	School fixed effects	1.s.d. → 0.27 s.d.	Table 3, column 4	Slightly concave
Hanushek et al. (2003)	Public schools, Texas, US	5th to 6th grade	Math test score gain	Math scores two years ago (also incl. prop. free lunch, std. dev of score)	School-grade	Student and school-by grade fixed effects	1.s.d. → 0.05.s.d.	Table 2, column 3 Table A1	No evidence
Cullen, Jacob and Levitt (2003)	Chicago public schools, US	9th and 10th grade	Test scores, and others	Winning a school voucher in lottery	-	Random assignment	Near zero and insignificant	Table 6, columns 1 and 2	No evidence
Schindler Rangvid (2004)	Secondary schools, Denmark	Age 15	PISA test scores (Reading)	Mothers' education	School-grade	Additional controls from register data	1 s.d. → 0.07 s.d.	Table 4 Table 1	Weak evidence for declining effects
Gibbons and Telhaj (2005)	State secondary schools, England	Age 14	Test scores (English, Math)	Attainment at age 11	School-grade	Random variation in transition from primary to secondary schools	1 s.d. → 0.08 s.d.	Tables 3 and 4, column 7	Slightly concave
Graham (2004)	STAR project, US	Kinder-garten to 3rd grade	Test scores (Math, Reading)	Test scores	Class	Random assignment and excess variance contrasts	50 percentile increase → 0.9-1.1 s.d.	Table 5, column 3	-
Schneeweiss and Winter-Ebmer (2005)	Secondary schools, Austria	Age 15	PISA test scores (Reading)	Socio-economic status	School-grade	School type fixed effects	1 s.d. → 0.06 s.d.	Table 2, column 1 Table 1	Weak evidence for declining effects

Builds on Table 1 in Gibbons and Tehlhaj (2005).

Table 2: Mean scores and sample sizes
(Standard deviations in parentheses)

	Germany	France	Iceland	Netherlands	Norway	Sweden
Reading score (all)	539.1	525.2	512.4	554.2	499.2	561.0
	(63.6)	(66.6)	(71.0)	(51.2)	(77.5)	(61.5)
Reading score (sample)	548.6	533.7	518.6	565.2	505.0	563.1
	(59.9)	(64.2)	(68.4)	(51.3)	(76.0)	(61.3)
Reading score (excl. tracked schools)	548.6	534.2	518.7	562.8	504.4	562.8
	(59.7)	(65.4)	(68.3)	(53.6)	(76.4)	(61.5)
Reading scores (tracked schools)	549.4	532.2	506.8	568.8	529.1	570.0
	(64.2)	(60.5)	(81.0)	(47.5)	(66.5)	(56.6)
No. of students (all)	7,633	3,538	3,676	4,112	3,459	6,044
No. of students (sample)	4,577	2,312	1,728	1,857	2,548	3,997
No. of students in schools with > 1 class	3,628	1,612	1,301	805	1,748	3,270
No. of schools	183	115	84	105	117	119
No. of schools with > 1 class	114	55	39	29	54	79
No. of classes	301	172	135	141	171	267
No. of classes in schools with > 1 class	232	112	90	65	108	227
Fraction of students in schools that apply tracking	0.067	0.278	0.006	0.328	0.035	0.046

Scores are weighted by students' sampling probability, standard deviations are in parentheses. Rows 6 and below refer to the sample used in the estimations. Last row reports the fraction of students in schools in which principals state that classes are formed by ability out of all students in schools for which principals reply to the question.

Table 3: Decomposition of variance in class level means

	Germany	France	Iceland	Netherlands	Norway	Sweden
Index of the number of books at home						
Mean	3.45	3.30	4.00	3.37	4.00	3.95
Total	.2401	.3138	.1480	.3922	.1542	.2643
Between	.2098	.2726	.1220	.3629	.1297	.2174
Within	.0303	.0412	.0259	.0293	.0245	.0469
Age						
Total	.0326	.0313	.0065	.0306	.0082	.0111
Between	.0250	.0183	.0050	.0212	.0060	.0060
Within	.0076	.0130	.0015	.0094	.0022	.0051
Female						
Total	.0145	.0226	.0212	.0156	.0145	.0158
Between	.0085	.0174	.0170	.0139	.0125	.0091
Within	.0061	.0052	.0043	.0017	.0020	.0067
Foreign parent						
Total	.0459	.0463	.0095	.0488	.0222	.0485
Between	.0404	.0413	.0069	.0451	.0189	.0386
Within	.0054	.0050	.0026	.0036	.0033	.0099
Foreign language at home						
Total	.0141	.0151	.0088	.0345	.0069	.0230
Between	.0112	.0128	.0058	.0330	.0052	.0167
Within	.0029	.0023	.0030	.0015	.0017	.0064
Reading test scores						
Total	1144.71	1223.61	751.93	896.62	1075.93	1123.78
Between	978.47	908.63	569.62	799.28	933.10	791.51
Within	166.24	314.97	182.31	97.34	142.83	332.27

Table 4: Regressions of peer variables on class room and school characteristics in Germany

Independent Variable	Books	Age	Female	Foreign parent	Foreign language
Class size	.002	-.011	.030	.010	.015
	(.091)	(.043)	(.017)	(.028)	(.014)
Class size squared*100	.03	.001	-.06	-.03	-.03
	(.2)	(.09)	(.04)	(.06)	(.03)
Female teacher	.109	.031	.006	-.008	-.009
	(.070)	(.024)	(.019)	(.030)	(.013)
Teacher education: ISCED 4	.001	.026	-.039	.006	-.018
	(.146)	(.050)	(.039)	(.036)	(.023)
Teacher education: ISCED 5+	.065	-.062	-.023	.173	.056
	(.114)	(.047)	(.034)	(.025)	(.015)
Teacher's experience in grade 4	.001	-.002	-.005	.005	.0001
	(.011)	(.003)	(.003)	(.004)	(.002)
Teacher's exp. Squared*100	.003	.008	.01	-.01	.00002
	(.03)	(.009)	(.008)	(.01)	(.006)
Size of town < 3000 inhab.	-.077	.141	.049	-.084	-.059
	(.265)	(.037)	(.074)	(.048)	(.039)
Size of town < 100,000 inhab.	.008	.086	-.029	.090	.044
	(.081)	(.026)	(.020)	(.030)	(.014)
Size of town < 500,000 inhab.	.045	.073	.004	.167	.084
	(.143)	(.047)	(.031)	(.044)	(.028)
Size of town > 500,000 inhab.	-.046	.115	.017	.230	.116
	(.154)	(.049)	(.035)	(.064)	(.033)
Instruction hours per day	.018	.005	-.022	.014	.012
	(.053)	(.018)	(.012)	(.022)	(.013)
Shortage of staff	-.020	.004	-.009	.034	.011
	(.055)	(.018)	(.010)	(.021)	(.011)
Shortage of material	.005	-.035	-.001	-.013	.005
	(.071)	(.022)	(.013)	(.025)	(.015)
Shortage of rooms	-.112	.041	.010	.004	.001
	(.057)	(.022)	(.012)	(.020)	(.013)

Regressions are weighted by the students' sampling probability. Number of observations is 4,577. Standard errors are clustered at the school level.

Table 5: P-values for F-tests on class room and school characteristics and for Chi² test for independence of peer variable and class assignment

	Germany	France	Iceland	Netherlands	Norway	Sweden
Index of the number of books at home						
Class	0.4532	0.2886	0.7094	0.4051	0.0995	0.0673
School	0.7095	0.1016	0.0595	0.0005	0.0431	0.3348
Class+school	0.7766	0.0588	0.3538	0.0000	0.0286	0.2448
Class (SFE)	0.4595	0.2552	0.0123	0.3370	0.5675	0.0000
Pearson Chi²	0.2415	0.3813	0.7964	0.7512	0.0893	0.0364
Age						
Class	0.0002	0.5581	0.0704	0.0249	0.1587	0.3492
School	0.0050	0.0137	0.3197	0.0015	0.4364	0.4234
Class+school	0.0000	0.0394	0.1603	0.0000	0.2139	0.4130
Class (SFE)	0.0017	0.2672	0.0021	0.0046	0.0000	0.9300
Pearson Chi²	0.0694	0.2402	0.1452	0.0992	0.0467	0.6247
Female						
Class	0.1270	0.7582	0.0006	0.0286	0.0089	0.9612
School	0.5170	0.3604	0.5958	0.0000	0.8831	0.3424
Class+school	0.4082	0.5402	0.0062	0.0000	0.0902	0.6536
Class (SFE)	0.5677	0.2838	0.0000	0.1467	0.0036	0.1589
Pearson Chi²	0.1240	0.4615	0.9608	0.6011	0.8827	0.9657
Foreign parent						
Class	0.0000	0.1350	0.1339	0.0254	0.5828	0.2174
School	0.0000	0.0004	0.7760	0.0000	0.0000	0.0000
Class+school	0.0000	0.0011	0.3009	0.0000	0.0000	0.0000
Class (SFE)	0.0003	0.0327	0.0000	0.0024	0.0000	0.1171
Pearson Chi²	0.0157	0.3129	0.3201	0.0444	0.1335	0.0046
Foreign language at home						
Class	0.0052	0.6215	0.0173	0.1830	0.3365	0.1542
School	0.0004	0.0098	0.6573	0.1222	0.1026	0.0000
Class+school	0.0000	0.0717	0.0476	0.0000	0.0867	0.0000
Class (SFE)	0.0000	0.4776	0.0001	0.0000	0.0029	0.0000
Pearson Chi²	0.0495	0.6920	0.1861	0.4217	0.4860	0.0009

The first four rows in each panel report p-values of Wald tests for the joint significance of the respective group of variables. The first three rows refer to regressions of peer variables on class and school variables. The fourth row refers to regressions of peer variables on class variables controlling for school fixed effects. Running the regressions at the class level or including the student level variables sex, age, origin and language in the regressions for the first three rows yields comparable results. The fifth row reports the p-value for Pearson Chi² tests of independence between the peer characteristic and class room assignment within each school using the individual level data. See text for details.

Table 6: Regressions for reading test score on peer composition for Germany
(Standard errors in parentheses)

Independent Variable	(1)	(2)	(3)	(4)	(5)
Individual books variable: five categories					
Peers' index of books at home	17.93	17.62	-.93	5.92	6.82
	(3.04)	(3.03)	(6.03)	(6.15)	(6.30)
Fraction peers 11-25 books at home	-20.15	-4.42	-31.63	-49.91	-68.77
	(23.44)	(21.11)	(22.87)	(42.75)	(44.67)
Fraction peers 26-100 books at home	21.41	28.10	-17.32	-22.90	-26.36
	(18.08)	(16.09)	(21.24)	(32.38)	(33.20)
Fraction peers 101-200 books at home	40.48	47.49	-10.31	-11.15	-16.21
	(18.88)	(17.24)	(24.85)	(39.44)	(40.55)
Fraction peers 200+ books at home	54.86	62.18	-28.67	-15.95	-21.31
	(17.69)	(16.16)	(30.35)	(36.86)	(37.36)
Individual books variable: index					
Peers' index of books at home	17.97	17.66	-.98	5.93	6.80
	(3.04)	(3.03)	(6.02)	(6.11)	(6.27)
Fraction peers 11-25 books at home	-21.42	-6.36	-27.71	-32.49	-45.14
	(22.83)	(20.63)	(21.74)	(33.48)	(35.74)
Fraction peers 26-100 books at home	20.37	26.54	-13.03	-7.50	-6.89
	(17.31)	(15.39)	(19.10)	(23.62)	(25.02)
Fraction peers 101-200 books at home	39.68	46.04	-5.19	5.40	5.37
	(18.16)	(16.51)	(22.38)	(28.67)	(30.17)
Fraction peers 200+ books at home	53.60	60.21	-21.18	-.37	-2.18
	(16.40)	(14.99)	(26.37)	(27.64)	(28.48)
Student level variables	✓	✓	✓	✓	✓
Class level variables		✓	✓	✓	✓
School level variables		✓	✓		
School peer variables			✓		
School fixed effects				✓	✓
Exclude tracked schools					✓

Weighted least squares regressions using students' sampling probability. Number of observations is 4,577. Standard errors are robust to clustering at the school level. Student level variables are student's sex and age, parents' origin, language spoken at home, number of books at home (four categories in the upper part and index in the lower part of the table) and number of persons living in household. Class level variables are class size, class size squared, teacher's sex, education, experience and experience squared. School level variables are community size, average daily instruction hours, shortage of staff, teaching material and buildings. School peer variables are school averages of index of number of books at home and parents' origin. Tracked schools are those for which principals state that fourth-grade classes are formed on the basis of ability.

Table 7: Regressions for reading test score on peer composition
(Standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)
Germany	17.97	17.66	1.52	5.93	6.80
	(3.04)	(3.03)	(5.29)	(6.12)	(6.27)
France	22.23	22.84	22.11	19.12	14.08
	(3.05)	(2.91)	(7.67)	(8.32)	(10.78)
Iceland	18.08	19.97	9.73	6.46	3.28
	(5.77)	(5.04)	(11.07)	(11.28)	(11.37)
The Netherlands	17.58	19.70	7.80	11.17	20.06
	(4.30)	(4.37)	(8.01)	(9.29)	(10.57)
Norway	15.46	9.84	-1.12	-6.93	-6.56
	(7.33)	(7.42)	(10.47)	(9.44)	(9.57)
Sweden	18.98	18.04	12.59	12.07	11.73
	(3.84)	(4.10)	(6.78)	(6.97)	(7.20)
Average across countries	**19.17**	**19.40**	**8.33**	**8.50**	**8.08**
	(1.61)	**(1.59)**	**(3.05)**	**(3.29)**	**(3.52)**
Student level variables	✓	✓	✓	✓	✓
Class level variables		✓	✓	✓	✓
School level variables		✓	✓		
School peer variables			✓		
School fixed effects				✓	✓
Exclude tracked schools					✓

Weighted least squares regressions using students' sampling probability. Peers' index of books at home is independent variable. Standard errors are robust to clustering at the school level. Student level variables are student's sex and age, parents' origin, language spoken at home, index of number of books at home and number of persons living in household. Class level variables are class size, class size squared, teacher's sex, education, experience and experience squared. School level variables are community size, average daily instruction hours, shortage of staff, teaching material and buildings. School peer variables are school averages of index of number of books at home and parents' origin. Tracked schools are those for which principals state that fourth-grade classes are formed on the basis of ability.

Table 8: Regressions for reading test score on peer composition and interactions with individual family background
(Standard errors in parentheses)

	(1)		(2)		(3)		(4)	
	Peer	Inter	Peer	Inter	Peer	Inter	Peer	Inter
Germany	17.58	.82	1.27	.66	5.69	.65	6.58	.63
	(3.03)	(1.04)	(5.28)	(1.02)	(6.11)	(1.04)	(6.26)	(1.08)
France	20.94	2.46	20.16	2.75	17.78	1.85	11.84	2.95
	(3.12)	(1.43)	(7.59)	(1.42)	(8.27)	(1.40)	(10.40)	(1.68)
Iceland	17.32	1.43	9.02	1.36	5.54	1.58	2.60	1.19
	(5.82)	(1.66)	(11.14)	(1.65)	(11.46)	(1.66)	(11.61)	(1.63)
The Netherlands	17.74	-.34	7.42	.92	10.68	1.15	18.45	4.38
	(4.49)	(1.36)	(8.14)	(1.23)	(9.46)	(1.28)	(10.71)	(1.35)
Norway	15.35	.16	-1.03	-.17	-6.58	-.71	-6.24	-.61
	(7.58)	(1.76)	(10.54)	(1.75)	(9.37)	(1.71)	(9.51)	(1.75)
Sweden	19.91	-1.35	13.81	-1.40	13.81	-2.01	13.84	-2.43
	(3.94)	(1.14)	(6.80)	(1.13)	(6.98)	(1.13)	(7.21)	(1.16)
Average across countries	**18.81**	**0.38**	**8.13**	**0.54**	**8.49**	**0.28**	**8.06**	**0.76**
	(1.64)	**(0.54)**	**(3.06)**	**(0.53)**	**(3.29)**	**(0.53)**	**(3.51)**	**(0.56)**
Student level variables	✓	✓	✓	✓	✓	✓	✓	✓
Class level variables			✓	✓	✓	✓	✓	✓
School level variables			✓	✓				
School peer variables			✓	✓				
School fixed effects					✓	✓	✓	✓
Exclude tracked schools							✓	✓

Weighted least squares regressions using students' sampling probability. Peers' index of books at home (columns "Peer") and interaction term of peers' index and individual level dummy variable for > 100 books at home (columns "Inter") are jointly included. Standard errors are robust to clustering at the school level. Student level variables are student's sex and age, parents' origin, language spoken at home, index of the number of books at home and number of persons living in household. Class level variables are class size, class size squared, teacher's sex, education, experience and experience squared. School level variables are community size, average daily instruction hours, shortage of staff, teaching material and buildings. School peer variables are school averages of index of number of books at home and parents' origin. Tracked schools are those for which principals state that fourth-grade classes are formed on the basis of ability.

Table 9: IV regressions for reading test score on peer composition
(Standard errors in parentheses)

	(1)	(2)	(3)	(4)	(5)	(6)
Germany	17.97	12.60	5.93	6.84	8.62	4.46
	(3.04)	(4.01)	(6.12)	(10.64)	(13.40)	(12.56)
France	22.23	16.51	19.12	47.75	64.99	63.63
	(3.05)	(3.64)	(8.32)	(17.84)	(24.28)	(27.08)
Iceland	18.08	-.93	6.46	7.27	9.57	12.96
	(5.77)	(8.39)	(11.28)	(21.21)	(27.91)	(28.43)
The Netherlands	17.58	13.47	11.17	24.80	41.53	163.64
	(4.30)	(5.48)	(9.29)	(31.38)	(52.54)	(139.90)
Norway	15.46	20.00	-6.93	.18	.22	.005
	(7.33)	(10.86)	(9.44)	(16.49)	(20.01)	(19.79)
Sweden	18.98	22.00	12.07	19.69	26.21	23.16
	(3.84)	(4.66)	(6.97)	(12.48)	(16.61)	(16.51)
Average across countries	**19.17**	**15.22**	**8.50**	**14.97**	**18.70**	**14.75**
	(1.61)	**(2.04)**	**(3.29)**	**(6.29)**	**(8.16)**	**(8.11)**
Student level variables	✓	✓	✓	✓	✓	✓
Class level variables			✓	✓	✓	✓
School fixed effects			✓	✓	✓	✓
IV books individual		✓		✓	✓	✓
IV books peer measure		✓		✓	✓	✓
IV parents educ. indiv.		✓		✓	✓	✓
IV parents educ. peers		✓		✓	✓	✓
Corrected for sampling bias					✓	✓
Exclude tracked schools						✓

Weighted least squares and instrumental variable regressions using students' sampling probability. Peers' index of books at home is independent variable. Standard errors are robust to clustering at the school level. Student level variables are student's sex and age, parents' origin, language spoken at home, index of number of books at home and number of persons living in household. Class level variables are class size, class size squared, teacher's sex, education, experience and experience squared. School level variables are community size, average daily instruction hours, shortage of staff, teaching material and buildings. The individual's and peers' index of the number of books at home from the home questionnaire are instrumented by the individual's and peers' index of books at home from the student questionnaire and individual's and peers' index of parents' highest educational level. A dummy for missing observations for the parental education variable has been added to not further restrict the sample size. The correction factor for sampling bias in columns (5) and (6) is $(N-1)/(n-1)$. Tracked schools are those for which principals state that fourth-grade classes are formed on the basis of ability.

Table 10: Effect sizes

	(1) S. D. test score σ_y	(2) S. D. peer variable $\sigma_{\tilde{\bar{X}}}$	(3) S. D. peer var. adjusted $\sigma_{\bar{X}}$	(4) Peer effect $\hat{\lambda}$	(5) Effect size $\sigma_{\bar{X}}\hat{\lambda}/\sigma_y$
Germany	59.9	.462	.379	8.6	.055
France	64.2	.572	.480	65.0	.485
Iceland	68.4	.367	.278	9.6	.039
The Netherlands	51.3	.579	.420	41.5	.340
Norway	76.0	.370	.300	.2	.001
Sweden	61.3	.458	.420	26.2	.179
Average across countries	**63.5**	**.468**	**.379**	**18.7**	**.112**

Column (3) is the square root of the covariance between the peer variables index of books at home taken from the student and the home questionnaire. The estimates of the peer effects in column (4) are taken from column (5) in Table 8. The results in column (5) are calculated as (4)*(3)/(1)

Table A1: Weighted mean values
(standard deviations in parentheses)

	Germany	France	Iceland	Netherlands	Norway	Sweden
Student background						
Student female	.52	0.49	0.50	0.53	0.49	0.49
	(.50)	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Student's age	10.51	10.10	9.72	10.24	9.97	10.80
	(.47)	(0.49)	(0.28)	(0.46)	(0.33)	(0.32)
Foreign parent	.22	0.29	0.12	0.17	0.17	0.23
	(.42)	(0.46)	(0.32)	(0.37)	(0.38)	(0.42)
Foreign language at home	.09	0.12	0.11	0.13	0.08	0.09
	(.29)	(0.33)	(0.32)	(0.34)	(0.26)	(0.29)
Books at home (index 1 – 5)	3.49	3.32	3.99	3.36	4.03	3.91
	(1.19)	(1.25)	(1.01)	(1.29)	(1.04)	(1.12)
0 – 10 books (ref.)	.06	0.10	0.02	0.10	0.02	0.03
	(.24)	(0.30)	(0.13)	(0.31)	(0.13)	(0.18)
11 – 20 books	.12	0.14	0.05	0.15	0.05	0.07
	(.32)	(0.35)	(0.21)	(0.35)	(0.22)	(0.25)
21 – 100 books	.36	0.35	0.27	0.30	0.26	0.26
	(.48)	(0.48)	(0.44)	(0.46)	(0.44)	(0.44)
101 – 200 books	.19	0.18	0.26	0.19	0.23	0.23
	(.39)	(0.38)	(0.44)	(0.39)	(0.42)	(0.42)
More than 200 books	.27	0.24	0.41	0.26	0.44	0.41
	(.44)	(0.43)	(0.49)	(0.44)	(0.50)	(0.49)
Household size	3.54	3.62	3.78	3.51	3.49	3.56
	(1.50)	(1.45)	(1.35)	(1.15)	(1.17)	(1.26)
Class						
Class size	22.77	24.04	19.56	26.65	20.32	24.25
	(3.93)	(3.41)	(3.98)	(5.26)	(5.40)	(6.36)
Teacher female	.80	0.62	0.94	0.51	0.84	0.79
	(.40)	(0.48)	(0.23)	(0.50)	(0.36)	(0.41)
Teacher's education ISCED 3 (ref.)	.07	0.42	0.13	1.00	0.02	1.00
	(.25)	(0.49)	(0.34)	(0.00)	(0.14)	(0.00)
Teacher's education ISCED 4	.01	0.23	0.12	0	0.98	0
	(.12)	(0.42)	(0.33)		(0.14)	
Teacher's education ISCED 5+	.92	0.35	0.75	0	0	0
	(.28)	(0.48)	(0.43)			
Teacher's experience	8.26	9.06	2.84	6.25	3.17	6.78
	(7.25)	(8.86)	(2.73)	(5.81)	(2.35)	(6.54)

Table A1, continued: Weighted mean values
(standard deviations in parentheses)

	Germany	France	Iceland	Netherlands	Norway	Sweden
School						
Village (ref.)	.31	0.25	0.06	0.01	0.25	0.14
	(0.46)	(0.44)	(0.23)	(0.09)	(0.43)	(0.34)
< 3000 inhab.	.02	0.06	0.17	0.14	0.15	0.08
	(0.12)	(0.25)	(0.37)	(0.35)	(0.36)	(0.27)
< 100,000 inhab.	.47	0.51	0.45	0.68	0.43	0.64
	(0.50)	(0.50)	(0.50)	(0.47)	(0.50)	(0.48)
< 500,000 inhab.	.12	0.15	0.32	0.13	0.10	0.07
	(0.32)	(0.36)	(0.47)	(0.33)	(0.30)	(0.26)
> 500,000 inhab.	.09	0.02	0	0.04	0.07	0.07
	(0.28)	(0.13)		(0.20)	(0.25)	(0.26)
Instruction hours/day	3.76	5.05	4.21	5.02	3.47	4.39
	(0.65)	(0.35)	(0.54)	(0.16)	(0.77)	(0.77)
Shortage of staff	1.76	1.23	1.56	1.55	1.53	2.09
(index 1 – 4)	(0.81)	(0.45)	(0.81)	(0.71)	(0.91)	(1.04)
Shortage of material	1.30	1.50	1.70	1.51	1.53	1.55
(index 1 – 4)	(0.51)	(0.75)	(0.87)	(0.66)	(0.74)	(0.83)
Shortage of rooms	1.49	1.40	1.62	1.49	1.67	1.74
(index 1 – 4)	(0.69)	(0.76)	(0.94)	(0.76)	(0.86)	(0.92)

Mean values are weighted by the students' sampling probability.

Table A2: Percentage of missing values

	Germany	France	Iceland	Netherlands	Norway	Sweden
Reading score	0	0	0	0	0	0
Student female	0	0	0	0	0	0
Student's age	0.03	0.17	0.03	0.54	1.07	0.02
Foreign parent	9.60	7.80	6.80	2.48	2.46	3.16
Foreign lang. at home	4.40	3.00	3.51	1.53	2.20	2.38
Books	13.32	10.97	16.89	35.41	9.42	9.33
Household size	8.23	6.53	4.68	4.60	3.79	4.42
Class size	6.86	4.49	6.86	11.84	1.24	3.26
Teacher female	8.14	4.83	7.64	11.31	0.69	3.34
Teacher's education	9.45	3.87	11.29	100	0.69	100
Teacher's experience	14.45	4.49	12.40	14.42	3.32	6.72
Size of town	4.48	4.49	23.91	11.87	2.66	4.33
Instruction hours/day	7.91	6.02	18.47	10.00	6.27	11.15
Shortage of staff	4.48	7.55	20.08	10.34	2.11	5.15
Shortage of material	5.04	5.17	21.33	10.80	2.60	4.91
Shortage of rooms	4.48	6.22	20.13	10.09	2.89	3.87

Table A3: Weighted means of student background variables by the number of books

	Germany				France			
	0-20 books	21-100	101-200	> 200 books	0-20 books	21-100	101-200	> 200 books
Student female	.49	.55	.51	.50	.50	.49	.50	.49
Student's age	10.68	10.52	10.45	10.42	10.26	10.10	10.03	9.97
Foreign parent	.40	.23	.15	.14	.34	.29	.24	.29
Foreign language	.19	.09	.06	.05	.21	.11	.11	.07
Household size	3.60	3.51	3.49	3.57	3.85	3.59	3.46	3.58
No secondary	.07	.02	.01	.01	.19	.09	.03	.03
Lower secondary	.29	.13	.05	.03	.52	.37	.20	.08
Upper secondary	.23	.27	.24	.20	.17	.24	.27	.12
Post secondary	.36	.46	.41	.23	.09	.16	.21	.20
Tertiary	.05	.12	.30	.54	.04	.13	.28	.57
Income < 30,000$	.63	.45	.30	.17	-	-	-	-
30,000$ – 50,000$	.32	.41	.47	.43	-	-	-	-
> 50,000$	.06	.15	.23	.41	-	-	-	-
Number of obs.	835	1,641	849	1,252	528	782	412	590

	Iceland				The Netherlands			
	0-20 books	21-100	101-200	> 200 books	0-20 books	21-100	101-200	> 200 books
Student female	.44	.50	.48	.53	.53	.54	.53	.51
Student's age	9.76	9.71	9.74	9.71	10.33	10.23	10.19	10.17
Foreign parent	.25	.12	.10	.10	.25	.17	.11	.12
Foreign language	.15	.12	.11	.11	.18	.15	.13	.09
Household size	3.67	3.76	3.74	3.84	3.41	3.47	3.47	3.68
No secondary	.06	.02	.01	.01	.02	.01	.002	.002
Lower secondary	.22	.18	.16	.08	.69	.56	.36	.15
Upper secondary	.55	.54	.47	.27	.14	.18	.13	.09
Post secondary	.12	.11	.09	.11	.10	.22	.40	.45
Tertiary	.05	.15	.28	.53	.05	.03	.11	.31
Income < 30,000$	.34	.27	.20	.15	.48	.34	.25	.16
30,000$ – 50,000$	.49	.43	.41	.32	.37	.41	.39	.33
> 50,000$	.17	.30	.39	.54	.16	.25	.35	.51
Number of obs.	110	467	440	711	448	562	336	511

Table A3, continued: Weighted means of student background variables by the number of books

	Norway				Sweden			
	0-20 books	21-100	101-200	> 200 books	0-20 books	21-100	101-200	> 200 books
Student female	.47	.50	.45	.50	.49	.46	.52	.50
Student's age	9.99	9.97	9.97	9.97	10.80	10.81	10.80	10.80
Foreign parent	.31	.17	.14	.16	.45	.26	.20	.17
Foreign language	.21	.10	.06	.05	.24	.11	.07	.05
Household size	3.46	3.37	3.42	3.60	3.83	3.58	3.50	3.51
No secondary	.01	.004	0	.001	.03	.003	.004	.001
Lower secondary	.09	.05	.02	.01	.17	.08	.05	.02
Upper secondary	.73	.64	.47	.26	.53	.51	.39	.22
Post secondary	0	0	0	0	.15	.24	.30	.21
Tertiary	.18	.30	.51	.73	.12	.17	.26	.55
Income < 30,000$	.45	.28	.19	.12	.47	.29	.21	.14
30,000$ – 50,000$	.42	.47	.41	.31	.43	.52	.47	.38
> 50,000$	.13	.24	.41	.57	.10	.19	.32	.48
Number of obs.	178	670	565	1,135	359	967	865	1,806

Mean values are weighted by the students' sampling probability. Number of observations refers to variables included in Table A1. For the variables highest educational level attained by at least one parent and annual household income in US$ before taxes the number of observations is lower in most countries.

Table A4: Regressions for reading test score on alternative peer composition
(Standard errors in parentheses)

	(1)		(2)	
	Books	Parent's education	Books	Parent's education
Germany	17.97	12.12	5.93	-2.00
	(3.04)	(3.53)	(6.12)	(5.00)
France	22.23	17.38	19.12	3.67
	(3.05)	(2.75)	(8.32)	(7.40)
Iceland	18.08	14.21	6.46	-11.70
	(5.77)	(3.42)	(11.28)	(11.93)
The Netherlands	17.58	16.18	11.17	-17.53
	(4.30)	(3.57)	(9.29)	(9.48)
Norway	15.46	15.26	-6.93	-10.05
	(7.33)	(5.23)	(9.44)	(8.84)
Sweden	18.98	25.53	12.07	-2.77
	(3.84)	(4.31)	(6.97)	(7.40)
Average across countries	**19.17**	**16.47**	**8.50**	**-4.37**
	(1.61)	**(1.46)**	**(3.29)**	**(3.05)**
Student level variables	✓	✓	✓	✓
Class level variables			✓	✓
School fixed effects			✓	✓

Weighted least squares regressions using students' sampling probability. Peers' index of books at home and peers' index of parents' highest education are independent variables. Standard errors are robust to clustering at the school level. Student level variables are student's sex and age, parents' origin, language spoken at home, index of number of books at home and number of persons living in household. Class level variables are class size, class size squared, teacher's sex, education, experience and experience squared.

Appendix

We are interested in estimating equation (8) in the text

$$y_{ics} = \beta x_{ics} + \lambda w_{cs} + \epsilon_{ics} \tag{1}$$

where $w_{cs} = \overline{x}_{(-i)cs}$ is the peer effect. The background variable x_{ics} is given by

$$
\begin{aligned}
x_{ics} &= \eta_{cs} + v_{ics} \\
E(v_{ics}) &= 0
\end{aligned}
$$

and measurement error is classical so that the measured variable is

$$
\begin{aligned}
\widetilde{x}_{ics} &= x_{ics} + u_{ics} \\
E(u_{ic}) &= 0
\end{aligned}
$$

There are N_{cs} students in a class. Even though $E(v_{ic}) = 0$, because N_{cs} is relatively small, $\overline{v}_{(-i)cs}$ will generally be different from zero. It is useful to distinguish $\overline{v}_{(-i)cs}$ from η_{cs}, because η_{cs} will carry information about x_{ics}, while $\overline{v}_{(-i)cs}$ will not.

Hence the variances of the true values of the variables are

$$
\begin{aligned}
var(x_{ics}) &= \sigma_\eta^2 + \sigma_v^2 \\
var(w_{cs}) &= var(\overline{x}_{(-i)cs}) \\
&= \sigma_\eta^2 + var(\overline{v}_{(-i)cs}) \\
&= \sigma_\eta^2 + \frac{\sigma_v^2}{N_{cs} - 1}
\end{aligned}
$$

We have a sample on $n_{cs} \leq N_{cs}$ students in the class. Hence the variance of the estimated variables $\widetilde{x}_{ics}$ and $\widetilde{\overline{x}}_{(-i)cs}$ in the sample is

$$
\begin{aligned}
var(\widetilde{x}_{ics}) &= \sigma_\eta^2 + \sigma_v^2 + \sigma_u^2 \\
var(\widetilde{w}_{cs}) &= \sigma_\eta^2 + \frac{\sigma_v^2}{n_{cs} - 1} + \frac{\sigma_u^2}{n_{cs} - 1}
\end{aligned}
$$

We will also need various covariances below. Denote by S_{-i} the set of students other than student i in the sample, and by P_{-i} the set of all other students in the classes under study. Then

$$
\begin{aligned}
cov(\widetilde{x}, \widetilde{w}) &= E\left[(\eta_{cs} + v_{ics} + u_{ics}) \left(\eta_{cs} + \frac{1}{n_{cs} - 1} \sum_{S_{-i}} (v_{ics} + u_{ics}) \right) \right] \\
&= \sigma_\eta^2
\end{aligned}
$$

43

$$cov(\widetilde{x}, w) = E\left[(\eta_{cs} + v_{ics} + u_{ics})\left(\eta_{cs} + \frac{1}{N_{cs}-1}\sum_{P_{-i}} v_{ics}\right)\right]$$
$$= \sigma_\eta^2$$

$$cov(x, \widetilde{w}) = E\left[(\eta_{cs} + v_{ics})\left(\eta_{cs} + \frac{1}{n_{cs}-1}\sum_{S_{-i}} (v_{ics} + u_{ics})\right)\right]$$
$$= \sigma_\eta^2$$

$$cov(\widetilde{w}, w) = E\left[\left(\eta_{cs} + \frac{1}{n_{cs}-1}\sum_{S_{-i}} (v_{ics} + u_{ics})\right)\left(\eta_{cs} + \frac{1}{N_{cs}-1}\sum_{P_{-i}} v_{ics}\right)\right]$$
$$= \sigma_\eta^2 + \frac{1}{(n_{cs}-1)(N_{cs}-1)}(n_{cs}-1)\sigma_v^2 = \sigma_\eta^2 + \frac{\sigma_v^2}{(N_{cs}-1)}$$

The estimates of β converges to

$$\text{plim } \widehat{\beta}_{OLS} = \frac{var(\widetilde{w})cov(y,\widetilde{x}) - cov(\widetilde{w},\widetilde{x})cov(y,\widetilde{w})}{var(\widetilde{x})var(\widetilde{w}) - cov(\widetilde{w},\widetilde{x})^2}$$

$$= \frac{var(\widetilde{w})(\beta\sigma_x^2 + cov(w,\widetilde{x})) - cov(\widetilde{w},\widetilde{x})(\beta cov(\widetilde{w},x) + \lambda cov(\widetilde{w},w))}{var(\widetilde{w})(\sigma_x^2 + \sigma_u^2) - cov(\widetilde{w},\widetilde{x})^2}$$

$$= \frac{\left(\sigma_\eta^2 + \frac{\sigma_v^2}{n_{cs}-1} + \frac{\sigma_u^2}{n_{cs}-1}\right)\left[\beta\sigma_x^2 + \lambda\sigma_\eta^2\right] - \sigma_\eta^2\left[\beta\sigma_\eta^2 + \lambda\left(\sigma_\eta^2 + \frac{\sigma_v^2}{N_{cs}-1}\right)\right]}{\left(\sigma_\eta^2 + \frac{\sigma_v^2}{n_{cs}-1} + \frac{\sigma_u^2}{n_{cs}-1}\right)\left(\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right) - \sigma_\eta^4}$$

$$= \frac{\beta\left[\sigma_x^2\left(\sigma_\eta^2 + \frac{\sigma_v^2}{n_{cs}-1} + \frac{\sigma_u^2}{n_{cs}-1}\right) - \sigma_\eta^4\right] + \lambda\left[\sigma_\eta^2\left(\sigma_\eta^2 + \frac{\sigma_v^2}{n_{cs}-1} + \frac{\sigma_u^2}{n_{cs}-1}\right) - \sigma_\eta^2\left(\sigma_\eta^2 + \frac{\sigma_v^2}{N_{cs}-1}\right)\right]}{\frac{1}{n_{cs}-1}\left(\sigma_v^2 + \sigma_u^2\right)\left(\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right) + \sigma_\eta^2\left(\sigma_v^2 + \sigma_u^2\right)}$$

$$= \frac{\beta\left[\left(\sigma_\eta^2 + \sigma_v^2\right)\left(\frac{\sigma_v^2}{n_{cs}-1} + \frac{\sigma_u^2}{n_{cs}-1}\right) + \sigma_\eta^2\left(\sigma_\eta^2 + \sigma_v^2\right) - \sigma_\eta^4\right] + \lambda\left[\sigma_\eta^2\left(\frac{\sigma_v^2}{n_{cs}-1} - \frac{\sigma_v^2}{N_{cs}-1} + \frac{\sigma_u^2}{n_{cs}-1}\right)\right]}{\frac{1}{n_{cs}-1}\left(\sigma_v^2 + \sigma_u^2\right)\left(\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right) + \frac{n_{cs}-1}{n_{cs}-1}\sigma_\eta^2\left(\sigma_v^2 + \sigma_u^2\right)}$$

$$= \frac{\beta\left[\left(\sigma_\eta^2 + \sigma_v^2\right)\left(\frac{\sigma_v^2}{n_{cs}-1} + \frac{\sigma_u^2}{n_{cs}-1}\right) + \sigma_\eta^2\sigma_v^2\right] + \lambda\left[\sigma_\eta^2\left(\frac{\sigma_v^2}{n_{cs}-1} - \frac{\sigma_v^2}{N_{cs}-1} + \frac{\sigma_u^2}{n_{cs}-1}\right)\right]}{\frac{1}{n_{cs}-1}\left(\sigma_v^2 + \sigma_u^2\right)^2 + \frac{n_{cs}}{n_{cs}-1}\sigma_\eta^2\left(\sigma_v^2 + \sigma_u^2\right)}$$

$$= \frac{\beta\left[\sigma_v^2\left(\frac{\sigma_v^2}{n_{cs}-1} + \frac{\sigma_u^2}{n_{cs}-1}\right) + \sigma_\eta^2\left(\frac{n_{cs}}{n_{cs}-1}\sigma_v^2 + \frac{\sigma_u^2}{n_{cs}-1}\right)\right] + \lambda\left[\sigma_\eta^2\left(\sigma_v^2\frac{N_{cs}-n_{cs}}{(n_{cs}-1)(N_{cs}-1)} + \frac{\sigma_u^2}{n_{cs}-1}\right)\right]}{\frac{1}{n_{cs}-1}\left(\sigma_v^2 + \sigma_u^2\right)^2 + \frac{n_{cs}}{n_{cs}-1}\sigma_\eta^2\left(\sigma_v^2 + \sigma_u^2\right)}$$

44

$$= \frac{\beta\left[\sigma_v^2\left(\sigma_v^2 + \sigma_u^2\right) + \sigma_\eta^2\left(n_{cs}\sigma_v^2 + \sigma_u^2\right)\right] + \lambda\left[\sigma_\eta^2\left(\frac{N_{cs}-n_{cs}}{N_{cs}-1}\sigma_v^2 + \sigma_u^2\right)\right]}{\left(\sigma_v^2 + \sigma_u^2\right)\left(n_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right)}$$

$$= \frac{\beta\left[\sigma_v^2\left(\sigma_v^2 + \sigma_u^2\right) + \sigma_\eta^2\left(n_{cs}\sigma_v^2 + \sigma_u^2\right)\right] + \lambda\left[\sigma_\eta^2\left(\frac{N_{cs}-n_{cs}}{N_{cs}-1}\sigma_v^2 + \sigma_u^2\right)\right]}{\left(\sigma_v^2 + \sigma_u^2\right)\left(n_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right)}$$

$$= \beta\frac{\sigma_v^2\left(\sigma_v^2 + \sigma_u^2\right) + \sigma_\eta^2\left(n_{cs}\sigma_v^2 + \sigma_u^2\right)}{\sigma_v^2\left(\sigma_v^2 + \sigma_u^2\right) + \sigma_\eta^2\left(n_{cs}\sigma_v^2 + \sigma_u^2\right) + \sigma_u^2\left((n_{cs}-1)\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right)}$$

$$+ \lambda\frac{\sigma_\eta^2\left(\frac{N_{cs}-n_{cs}}{N_{cs}-1}\sigma_v^2 + \sigma_u^2\right)}{\left(\sigma_v^2 + \sigma_u^2\right)\left[n_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right]}$$

This shows if $\lambda = 0$ or $\sigma_\eta^2 = 0$, then there is only classical attenuation bias in $\widehat{\beta}$. However, if $\lambda > 0$ and $\sigma_\eta^2 > 0$, then there is also a positive bias term. However, this term will tend to be small because of the factor $(N_{cs} - n_{cs})/(N_{cs} - 1)$, unless very few students are sampled per classroom.

However, we are primarily interested in the estimate of the peer effect:

$$\text{plim } \widehat{\lambda}_{OLS} = \frac{var(\widetilde{x})cov(y, \widetilde{w}) - cov(\widetilde{w}, \widetilde{x})cov(y, \widetilde{x})}{var(\widetilde{x})var(\widetilde{w}) - cov(\widetilde{w}, \widetilde{x})^2}$$

$$= \frac{var(\widetilde{x})\left[\beta cov(\widetilde{w}, x) + \lambda cov(\widetilde{w}, w)\right] - cov(\widetilde{w}, \widetilde{x})\left[\beta\sigma_x^2 + \lambda cov(w, \widetilde{x})\right]}{var(\widetilde{x})var(\widetilde{w}) - cov(\widetilde{w}, \widetilde{x})^2}$$

$$= \frac{\left(\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right)\left[\beta\sigma_\eta^2 + \lambda\left(\sigma_\eta^2 + \frac{\sigma_v^2}{N_{cs}-1}\right)\right] - \sigma_\eta^2\left[\beta\left(\sigma_\eta^2 + \sigma_v^2\right) + \lambda\sigma_\eta^2\right]}{\frac{1}{n_{cs}-1}\left(\sigma_v^2 + \sigma_u^2\right)\left[n_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right]}$$

$$= \frac{\beta\left[\left(\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right)\sigma_\eta^2 - \left(\sigma_\eta^2 + \sigma_v^2\right)\sigma_\eta^2\right] + \lambda\left[\left(\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right)\left(\sigma_\eta^2 + \frac{\sigma_v^2}{N_{cs}-1}\right) - \sigma_\eta^4\right]}{\frac{1}{n_{cs}-1}\left(\sigma_v^2 + \sigma_u^2\right)\left[n_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right]}$$

$$= \frac{\beta\left[\sigma_u^2\sigma_\eta^2\right] + \lambda\left[\sigma_\eta^2\frac{\sigma_v^2}{N_{cs}-1} + \left(\sigma_v^2 + \sigma_u^2\right)\left(\sigma_\eta^2 + \frac{\sigma_v^2}{N_{cs}-1}\right)\right]}{\frac{1}{n_{cs}-1}\left(\sigma_v^2 + \sigma_u^2\right)\left[n_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right]}$$

$$= \beta\frac{(n_{cs}-1)\sigma_u^2\sigma_\eta^2}{\left(\sigma_v^2 + \sigma_u^2\right)\left[n_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right]} + \lambda\frac{\frac{n_{cs}-1}{N_{cs}-1}\sigma_v^2\left(\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right) + (n_{cs}-1)\sigma_\eta^2\left(\sigma_v^2 + \sigma_u^2\right)}{\left(\sigma_v^2 + \sigma_u^2\right)\left[n_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right]}$$

Consider the following special cases. If there is no measurement error, $\sigma_u^2 = 0$:

$$\text{plim } \widehat{\lambda}_{OLS} = \lambda\frac{\frac{n_{cs}-1}{N_{cs}-1}\sigma_v^2\left(\sigma_\eta^2 + \sigma_v^2\right) + (n_{cs}-1)\sigma_\eta^2\sigma_v^2}{\sigma_v^2\left[n_{cs}\sigma_\eta^2 + \sigma_v^2\right]}$$

$$= \lambda\frac{\frac{n_{cs}-1}{N_{cs}-1}\left(\sigma_\eta^2\sigma_v^2 + \sigma_v^4\right) + (n_{cs}-1)\sigma_\eta^2\sigma_v^2}{\sigma_v^2\left[n_{cs}\sigma_\eta^2 + \sigma_v^2\right]}$$

$$= \lambda \frac{\left(\frac{n_{cs}-1}{N_{cs}-1} + n_{cs} - 1\right) \sigma_\eta^2 \sigma_v^2 + \frac{n_{cs}-1}{N_{cs}-1} \sigma_v^4}{\sigma_v^2 \left[n_{cs}\sigma_\eta^2 + \sigma_v^2\right]}$$

$$= \lambda \frac{\left(\frac{n_{cs}-1}{N_{cs}-1} + n_{cs} - 1\right) \sigma_\eta^2 \sigma_v^2 + \frac{n_{cs}-1}{N_{cs}-1} \sigma_v^4}{\sigma_v^2 \left[n_{cs}\sigma_\eta^2 + \sigma_v^2\right]}$$

$$= \lambda \left(\frac{n_{cs}-1}{N_{cs}-1}\right) \frac{N_{cs}\sigma_\eta^2 \sigma_v^2 + \sigma_v^4}{\sigma_v^2 \left[n_{cs}\sigma_\eta^2 + \sigma_v^2\right]}$$

$$= \lambda \left(\frac{n_{cs}-1}{N_{cs}-1}\right) \frac{\sigma_v^2 \left(N_{cs}\sigma_\eta^2 + \sigma_v^2\right)}{\sigma_v^2 \left[n_{cs}\sigma_\eta^2 + \sigma_v^2\right]}$$

$$= \lambda \left(\frac{n_{cs}-1}{N_{cs}-1}\right) \frac{N_{cs}\sigma_\eta^2 + \sigma_v^2}{n_{cs}\sigma_\eta^2 + \sigma_v^2}$$

In this case attenuation due to not sampling the whole class is most important when $\sigma_\eta^2 = 0$, but will vanish as σ_η^2 grows.

If there is only measurement error but the whole class is sampled, $N_{cs} = n_{cs}$:

$$\text{plim } \widehat{\lambda}_{OLS} = \beta \frac{(N_{cs} - 1)\sigma_u^2 \sigma_\eta^2}{(\sigma_v^2 + \sigma_u^2)\left[N_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right]} + \lambda \frac{\sigma_v^2 \left(\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right) + (N_{cs} - 1)\sigma_\eta^2 \left(\sigma_v^2 + \sigma_u^2\right)}{(\sigma_v^2 + \sigma_u^2)\left[N_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right]}$$

$$= \beta \frac{(N_{cs} - 1)\sigma_u^2 \sigma_\eta^2}{(\sigma_v^2 + \sigma_u^2)\left[N_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right]} + \lambda \frac{\sigma_v^2 \left(N_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right) + (N_{cs} - 1)\sigma_\eta^2 \sigma_u^2}{\sigma_v^2 \left(N_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right) + \sigma_u^2 \left(N_{cs}\sigma_\eta^2 + \sigma_v^2 + \sigma_u^2\right)}$$

In this case there is attenuation in $\widehat{\lambda}$ if $\beta = 0$. Attenuation is greatest if $\sigma_\eta^2 = 0$, but becomes less important when σ_η^2 grows. If $\beta > 0$, this raises the estimated value of $\widehat{\lambda}$, and more so when σ_η^2 is bigger. The positive bias due to the β-term may outweigh the attenuation for large enough σ_η^2. If the whole class is not sampled, this introduces additional attenuation.

We now turn to the instrumental variables estimator. The instruments

$$z_{1ics} = x_{ics} + u_{1ics}$$
$$z_{2cs} = \overline{z}_{1cs} = \eta_{cs} + \overline{v}_{cs} + \overline{u}_{1cs}$$

are based on an independent measurement of x_{ics}, i.e. we assume $cov(u_{ics}, u_{1ics}) = 0$. Then

$$cov(z_1, \widetilde{x}) = E\left[(\eta_{cs} + v_{ics} + u_{1ics})(\eta_{cs} + v_{ics} + u_{ics})\right]$$
$$= \sigma_\eta^2 + \sigma_v^2 = cov(z_1, x)$$

$$cov(z_1, \widetilde{w}) = E\left[(\eta_{cs} + v_{ics} + u_{1ics})\left(\eta_{cs} + \frac{1}{n_{cs} - 1}\sum_{S_{-i}}(v_{ics} + u_{ics})\right)\right]$$
$$= \sigma_\eta^2 = cov(z_1, w)$$

46

$$
\begin{aligned}
cov(z_2, \widetilde{x}) &= E\left[\left(\eta_{cs} + \frac{1}{n_{cs}-1}\sum_{S_{-i}}(v_{ics} + u_{1ics})\right)(\eta_{cs} + v_{ics} + u_{ics})\right] \\
&= \sigma_\eta^2 = cov(z_2, x)
\end{aligned}
$$

$$
\begin{aligned}
cov(z_2, \widetilde{w}) &= E\left[\left(\eta_{cs} + \frac{1}{n_{cs}-1}\sum_{S_{-i}}(v_{ics} + u_{1ics})\right)\left(\eta_{cs} + \frac{1}{n_{cs}-1}\sum_{S_{-i}}(v_{ics} + u_{ics})\right)\right] \\
&= \sigma_\eta^2 + \frac{1}{(n_{cs}-1)}\sigma_v^2
\end{aligned}
$$

$$
\begin{aligned}
cov(z_2, w) &= E\left[\left(\eta_{cs} + \frac{1}{n_{cs}-1}\sum_{S_{-i}}(v_{ics} + u_{1ics})\right)\left(\eta_{cs} + \frac{1}{N_{cs}-1}\sum_{P_{-i}}v_{ics}\right)\right] \\
&= \sigma_\eta^2 + \frac{1}{(n_{cs}-1)(N_{cs}-1)}(n_{cs}-1)\sigma_v^2 = \sigma_\eta^2 + \frac{\sigma_v^2}{(N_{cs}-1)}
\end{aligned}
$$

Hence

$$
\begin{aligned}
\text{plim}\widehat{\beta}_{IV} &= \frac{cov(z_2,\widetilde{w})cov(y,z_1) - cov(\widetilde{w},z_1)cov(y,z_2)}{cov(z_1,\widetilde{x})cov(z_2,\widetilde{w}) - cov(z_1,\widetilde{w})cov(z_2,\widetilde{x})} \\[2mm]
&= \frac{cov(z_2,\widetilde{w})\left[\beta cov(x,z_1) + \lambda cov(w,z_1)\right] - cov(\widetilde{w},z_1)\left[\beta cov(x,z_2) + \lambda cov(w,z_2)\right]}{cov(z_1,\widetilde{x})cov(z_2,\widetilde{w}) - cov(z_1,\widetilde{w})cov(z_2,\widetilde{x})} \\[2mm]
&= \frac{\left(\sigma_\eta^2 + \frac{1}{(n_{cs}-1)}\sigma_v^2\right)\left[\beta\left(\sigma_\eta^2 + \sigma_v^2\right) + \lambda\sigma_\eta^2\right] - \sigma_\eta^2\left[\beta\sigma_\eta^2 + \lambda\left(\sigma_\eta^2 + \frac{1}{(N_{cs}-1)}\sigma_v^2\right)\right]}{\left(\sigma_\eta^2 + \sigma_v^2\right)\left(\sigma_\eta^2 + \frac{1}{(n_{cs}-1)}\sigma_v^2\right) - \sigma_\eta^4} \\[2mm]
&= \frac{\beta\left[\left(\sigma_\eta^2 + \sigma_v^2\right)\left(\sigma_\eta^2 + \frac{1}{(n_{cs}-1)}\sigma_v^2\right) - \sigma_\eta^4\right] + \lambda\left[\sigma_\eta^2\left(\sigma_\eta^2 + \frac{1}{(n_{cs}-1)}\sigma_v^2\right) - \sigma_\eta^2\left(\sigma_\eta^2 + \frac{1}{(N_{cs}-1)}\sigma_v^2\right)\right]}{\left(\sigma_\eta^2 + \sigma_v^2\right)\left(\sigma_\eta^2 + \frac{1}{(n_{cs}-1)}\sigma_v^2\right) - \sigma_\eta^4} \\[2mm]
&= \beta + \lambda\frac{\left[\sigma_\eta^2\frac{1}{(n_{cs}-1)}\sigma_v^2 - \sigma_\eta^2\frac{1}{(N_{cs}-1)}\sigma_v^2\right]}{\left(\sigma_\eta^2 + \sigma_v^2\right)\left(\sigma_\eta^2 + \frac{1}{(n_{cs}-1)}\sigma_v^2\right) - \sigma_\eta^4} \\[2mm]
&= \beta + \lambda\frac{\sigma_\eta^2\sigma_v^2\frac{N_{cs}-n_{cs}}{(n_{cs}-1)(N_{cs}-1)}}{\left(\sigma_\eta^2 + \sigma_v^2\right)\left(\sigma_\eta^2 + \frac{1}{(n_{cs}-1)}\sigma_v^2\right) - \sigma_\eta^4} \\[2mm]
&= \beta + \lambda\left(\frac{N_{cs}-n_{cs}}{N_{cs}-1}\right)\frac{\sigma_\eta^2\sigma_v^2}{\left(\sigma_\eta^2 + \sigma_v^2\right)\left((n_{cs}-1)\sigma_\eta^2 + \sigma_v^2\right) - (n_{cs}-1)\sigma_\eta^4} \\[2mm]
&= \beta + \lambda\left(\frac{N_{cs}-n_{cs}}{N_{cs}-1}\right)\frac{\sigma_\eta^2\sigma_v^2}{\sigma_v^2\left(n_{cs}\sigma_\eta^2 + \sigma_v^2\right)} \\[2mm]
&= \beta + \lambda\left(\frac{N_{cs}-n_{cs}}{N_{cs}-1}\right)\frac{\sigma_\eta^2}{n_{cs}\sigma_\eta^2 + \sigma_v^2}
\end{aligned}
$$

47

so the IV estimator of β is consistent if $\sigma_\eta^2 = 0$.

$$
\begin{aligned}
\text{plim } \widehat{\lambda}_{IV} &= \frac{cov(z_1,\widetilde{x})cov(y,z_2) - cov(\widetilde{x},z_2)cov(y,z_1)}{cov(z_1,\widetilde{x})cov(z_2,\widetilde{w}) - cov(z_1,\widetilde{w})cov(z_2,\widetilde{x})} \\
&= \frac{cov(z_1,\widetilde{x})\left[\beta cov(x,z_2) + \lambda cov(w,z_2)\right] - cov(\widetilde{x},z_2)\left[\beta cov(x,z_1) + \lambda cov(w,z_1)\right]}{cov(z_1,\widetilde{x})cov(z_2,\widetilde{w}) - cov(z_1,\widetilde{w})cov(z_2,\widetilde{x})} \\
&= \frac{\left(\sigma_\eta^2 + \sigma_v^2\right)\left[\beta\sigma_\eta^2 + \lambda\left(\sigma_\eta^2 + \frac{1}{(N_{cs}-1)}\sigma_v^2\right)\right] - \sigma_\eta^2\left[\beta\left(\sigma_\eta^2 + \sigma_v^2\right) + \lambda\sigma_\eta^2\right]}{\left(\sigma_\eta^2 + \sigma_v^2\right)\left(\sigma_\eta^2 + \frac{1}{(n_{cs}-1)}\sigma_v^2\right) - \sigma_\eta^4} \\
&= \frac{\beta\left[\sigma_\eta^2\left(\sigma_\eta^2 + \sigma_v^2\right) - \sigma_\eta^2\left(\sigma_\eta^2 + \sigma_v^2\right)\right] + \lambda\left[\left(\sigma_\eta^2 + \sigma_v^2\right)\left(\sigma_\eta^2 + \frac{1}{(N_{cs}-1)}\sigma_v^2\right) - \sigma_\eta^4\right]}{\left(\sigma_\eta^2 + \sigma_v^2\right)\left(\sigma_\eta^2 + \frac{1}{(n_{cs}-1)}\sigma_v^2\right) - \sigma_\eta^4} \\
&= \lambda\left(\frac{n_{cs}-1}{N_{cs}-1}\right)\frac{\left(\sigma_\eta^2 + \sigma_v^2\right)\left((N_{cs}-1)\sigma_\eta^2 + \sigma_v^2\right) - (N_{cs}-1)\sigma_\eta^4}{\left(\sigma_\eta^2 + \sigma_v^2\right)\left((n_{cs}-1)\sigma_\eta^2 + \sigma_v^2\right) - (n_{cs}-1)\sigma_\eta^4} \\
&= \lambda\left(\frac{n_{cs}-1}{N_{cs}-1}\right)\frac{\sigma_v^2\left(N_{cs}\sigma_\eta^2 + \sigma_v^2\right)}{\sigma_v^2\left(n_{cs}\sigma_\eta^2 + \sigma_v^2\right)} \\
&= \lambda\left(\frac{n_{cs}-1}{N_{cs}-1}\right)\frac{N_{cs}\sigma_\eta^2 + \sigma_v^2}{n_{cs}\sigma_\eta^2 + \sigma_v^2}
\end{aligned}
$$

This is the same as in the OLS case when $\sigma_u^2 = 0$. When $\sigma_\eta^2 = 0$ this simplifies to

$$
\text{plim } \widehat{\lambda}_{IV} = \lambda\left(\frac{n_{cs}-1}{N_{cs}-1}\right)
$$

so that the IV coefficient can easily be adjusted.