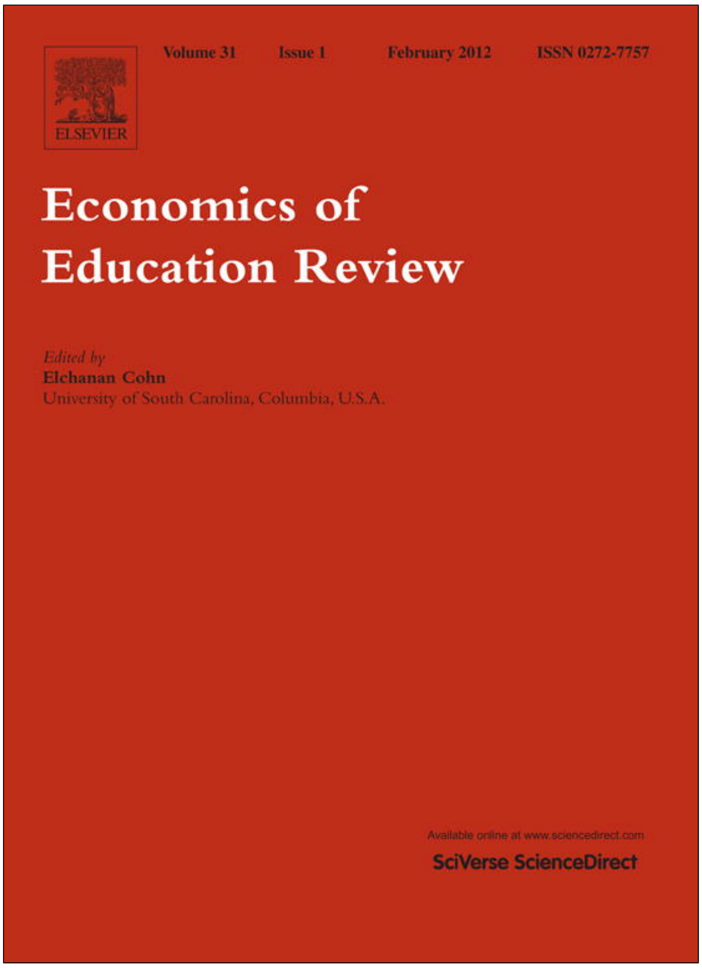
Volume 31 Issue 1 February 2012 ISSN 0272-7757
ELSEVIER
Economics of
Education Review
Edited by
Elchanan Cohn
University of South Carolina, Columbia, U.S.A.
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Economics of Education Review 31 (2012) 1136–1142

Contents lists available at SciVerse ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Peer effects and measurement error: The impact of sampling variation in school survey data (evidence from PISA)

John Micklewright[a,*], Sylke V. Schnepf[b], Pedro N. Silva[c]

[a] Institute of Education, University of London, United Kingdom
[b] Southampton Statistical Sciences Research Institute and School of Social Sciences, University of Southampton, United Kingdom
[c] Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Received 15 March 2010
Received in revised form 25 July 2012
Accepted 29 July 2012

JEL classification:
C21
C81
I21

Keywords:
Peer effects
Measurement error
School surveys
Sampling variation
PISA

ABSTRACT

Investigation of peer effects on achievement with sample survey data on schools may mean that only a random sample of the population of peers is observed for each individual. This generates measurement error in peer variables similar in form to the textbook case of errors-in-variables, resulting in the estimated peer group effects in an OLS regression model being biased towards zero. We investigate the problem using survey data for England from the Programme for International Student Assessment (PISA) linked to administrative micro-data recording information for each PISA sample member's entire year cohort. We calculate a peer group measure based on these complete data and compare its use with a variable based on peers in just the PISA sample. We also use a Monte Carlo experiment to show how the extent of the attenuation bias rises as peer sample size falls. On average, the estimated peer effect is biased downwards by about one third when drawing a sample of peers of the size implied by the PISA survey design.

1. Introduction

This paper addresses an aspect of peer group measurement that often arises in analyses based on sample survey data. If the survey's design means that only a random sample of peers is observed for each individual, rather than *all* peers, then any summary statistic of peer attributes that is based on the survey data and used as an explanatory variable in regression analysis is subject to sampling variation. This generates measurement error similar in form to the textbook case of errors-in-variables. As a result, the absolute value of the estimated peer group coefficient in an OLS regression is biased towards zero.

The problem has been recognised, for example by Ammermueller and Pischke (2009) for whom sampling variation is one source of error in peer group measurement. (See as well Sojourner, 2011.) There is also a parallel literature in statistics, little referenced by economists, that is concerned with multilevel models applied to survey data with a hierarchical structure when measures of variables at a higher level are formed by averaging the characteristics of units at a lower level (Kravdal, 2006; Woodhouse, Yang, Goldstein, & Rasbash, 1996). These papers have warned of the consequences of sampling variation in peer averages, but have been unable to conclude categorically about the extent of bias in any particular empirical setting. As Ammermueller and Pischke (2009) note, the bias will depend inter alia on the relative sizes of the within- and between-group variation in the individual characteristics. The bias is greatest when the former dominates – sampling from relatively heterogeneous groups can result in large sampling error.

In contrast to Ammermueller and Pischke (2009) and other earlier papers, we are able to quantify the extent

* Corresponding author.
E-mail address: j.micklewright@ioe.ac.uk (J. Micklewright).

0272-7757/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.econedurev.2012.07.015

of the bias in peer group coefficient estimates obtained with school survey data since we have information on the population from which each sample of peers in the survey is drawn. We compare the regression estimate of the peer group parameter when the peer average is calculated with the survey sample of peers with the estimate obtained when the average is calculated for the population peer set. Using a Monte Carlo experiment, we are also able to show how the extent of the bias changes as the sample size of peers falls.

Our analysis uses data from England, collected in a major international school survey, the Programme for International Student Assessment (PISA), which measures the cognitive achievement of 15 year olds. International reports on PISA emphasise the estimated impact of peers on cognitive achievement e.g. OECD (2001, chap. 8) and OECD (2007, chap. 5). Subsequent papers have also estimated peer effects with the data e.g. Fertig (2003), Schindler-Rangvid (2003), Entorf and Lauk (2006), and Schneeweis and Winter-Ebmer (2007). But the potential for sampling variation to bias peer effect estimates in PISA has not been highlighted. Other things equal, there will be more attenuation bias in countries where schools are less socially segregated, that is where between-school variation in pupil characteristics is low. England is a middle-ranking country in this respect, with less segregation than high-ranking countries like Austria and Germany and more segregation than the low-ranking Nordic countries (Jenkins, Micklewright, & Schnepf, 2008).

Section 2 relates the classical measurement error problem to the PISA survey design. Section 3 describes our PISA data for England, which comprise the achieved sample in 2003 of responding schools and pupils together with data from administrative registers on all 15 year olds in the sampled schools. Section 4 presents results from regressions for cognitive achievement. The estimated coefficient on our measure of peer characteristics (the percentage of peers receiving free school meals available to low income families) is biased downwards in absolute terms by about half in our PISA sample. However, our Monte Carlo simulation shows that the coefficient would be expected to be biased downwards by about a third given the peer sample size implied by PISA's design. It also demonstrates that the expected absolute bias rises following a non-linear pattern as sample size falls. Section 5 concludes.

Our analysis is unlikely to reveal the 'true' impact of peers, even when we use the complete data on each individual's peer population. Our definition of the peer group (the year cohort at the individual's school) is very common but may be incorrect, our measure of peer characteristics (receipt of free school meals) may be inadequate, and we do not consider the selection of individuals into peer groups (schools in our analysis). (See e.g. Vigdor & Nechyba, 2007 for a review of measurement of the impact of peer characteristics.) However, attenuation bias of the type we analyse in this paper can be expected in any survey in which only a random sample of peers is observed, whatever the peer group definition and peer measure used and whether or not selection into peer groups is addressed.

2. Classical measurement error and the PISA sample design

In a regression model with one explanatory variable, classical measurement error in that variable leads to bias towards zero in the absolute value of the OLS estimate of the slope parameter – the 'iron law of econometrics' (Hausman, 2001). The size of this attenuation bias is determined by the relative magnitudes of the variances of the unobserved true variable x_i and the observed explanatory variable z_i. Let:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{1}$$

be the target regression model, where y_i is the response (measured without error) and ε_i is the error (disturbance) term.

Under a classical measurement error scenario, the observed values of the predictor variable z_i are related to the true unobserved values x_i as follows:

$$z_i = x_i + u_i \tag{2}$$

Therefore the researcher is forced to estimate:

$$y_i = \beta_0 + \beta_1 z_i + (-\beta_1 u_i + \varepsilon_i) \tag{3}$$

The OLS estimator of the slope coefficient for the observed data is given by:

$$\hat{\beta}_{1OLS} = \frac{(n-1)^{-1}\sum_i (z_i - \bar{z})(y_i - \bar{y})}{(n-1)^{-1}\sum_i (z_i - \bar{z})^2} = \frac{\mathrm{cov}(z_i, y_i)}{\mathrm{var}(z_i)} \tag{4}$$

Under the standard assumption that

$$(x_i, u_i, \varepsilon_i) \sim MN[(\mu_x, 0, 0); \mathrm{diag}(\sigma_x^2, \sigma_e^2, \sigma_u^2)]$$

are independent random vectors with a common Multivariate Normal distribution, it follows that, see e.g. Fuller (1987):

$$E(\hat{\beta}_{1OLS}) = \frac{VAR(x)}{VAR(z)}\beta_1 = \frac{\sigma_x^2}{\sigma_z^2}\beta_1 = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}\beta_1 \tag{5}$$

Under slightly weaker assumptions the following result holds for large samples:

$$\mathrm{plim}(\hat{\beta}_{1OLS}) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}\beta_1 \tag{6}$$

Thus measurement error implies that the composite error term in brackets in (3) is negatively correlated with the observed z_i, leading to bias in $\hat{\beta}_{1OLS}$, the OLS estimator of the slope β_1 in the target model (1).

Now assume that the target regression model includes additional explanatory variables, $\mathbf{t}_i$, free of measurement error:

$$y_i = \beta_0 + \beta_1 x_i + \boldsymbol{\beta}_2' \mathbf{t}_i + \varepsilon_i \tag{7}$$

The textbook result is that the OLS estimate of β_1 based on the observed covariate z_i is still biased towards zero. Moreover, the attenuation bias increases, relative to the case with a single explanatory variable, the greater is the correlation between x_i and $\mathbf{t}_i$ (e.g. Bound, Brown, & Mathiowetz, 2001, Eq. (5)). The OLS estimates of the coefficients in $\boldsymbol{\beta}_2$ are also biased but in unknown directions (e.g. Greene, 1993:

280–284), the measurement error in one variable contaminating the estimates of the other parameters in the model.

Suppose that y_i represents an individual's test score and x_i represents a measure of an individual's peer group, defined as the average value of a characteristic for all other persons in the individual's age cohort at school. This is a broad definition of peers, adopted in many studies out of necessity (e.g. Hanushek, Kain, Markman, & Rivkin, 2003), although authors often recognise that a narrower definition may be more suitable, such as the class or some other group within the school reflecting with whom the individual actually interacts. Many school surveys have a sampling design that results in x_i being measured with error since only a random sample of pupils is selected within each school for inclusion in the survey rather than all pupils.[1]

This problem is shared by PISA. The survey has a two-stage design. Schools are sampled with probability proportional to size and then 35 pupils aged 15 are randomly sampled within each school.[2] In England in 2003, the 35 students were sampled out of what is an average of about 170 students of this age per school. The mean characteristics of an individual's schoolmates that are observed in the PISA sample will be measured with error by z_i.[3] The error, $e_i = z_i - x_i$ is the result of sampling variation. Some of its properties resemble those of textbook 'classical' measurement error defined above, u_i. Critically, $COV(e_i, x_i)$ should be close to zero. On the other hand, $CORR(e_i, e_j)$ will be very high for students in the same school, although it should again be zero for students in different schools.[4] In the next section we investigate these features in practice in the PISA data.

3. The PISA data for 2003 in England and the measurement of peer variables

The 2003 PISA round in England resulted in data being collected from pupils at 159 responding schools. PISA tests 15 year olds on their competence in maths, science and reading. In 2003 maths was the 'major' subject to which the most time was devoted in the test instruments, while science and reading were 'minor' subjects, with less test time.

We have access to a version of the survey data that links schools and pupils to a Department for Education administrative register containing information for all 15 year old pupils in the country, the National Pupil Database (NPD).[5] The NPD provides us with one measure of pupils' socio-economic status, namely an indicator for whether they receive Free School Meals (FSM) – a state benefit for low income families. In the terminology of Manski (1993), peer receipt of FSM allows us to estimate 'contextual' peer group effects.

Receipt of FSM is the standard focus for research into social background in England's schools based on administrative data e.g. Burgess, McConnell, Propper, and Wilson (2004) and Goldstein and Noden (2003). A similar variable is used in US research on peer effects based on administrative registers and in that context has been summarised as 'likely to be a noisy measure of peer economic circumstances' (Hanushek et al., 2003: 537) that may 'proxy omitted or mismeasured factors that affect individual achievement, leading to biased results that quite generally exaggerate the importance of peers' (Hanushek et al., 2003: 530). The same is likely to be true in the UK: Hobbs and Vignoles (2009) demonstrate clearly that receipt of FSM is an imperfect proxy for low household income. Unfortunately, the NPD does not provide us with a good alternative measure of socio-economic status. However, our ambition is not to estimate the 'true' impact of peers. Rather it is to demonstrate the impact of measurement error bias resulting from survey design, albeit on the estimated parameter of an imperfect indicator of peer characteristics. This problem, generated by sampling variation, will be common to any estimate of peer effects based on a peer variable that is measured only for a random sample of peers, no matter how good that measure of peer characteristics is in principle (e.g. household income, lagged achievement, etc.) and whether or not the selection into the peer group is addressed. The measurement error that we focus on is that due to sampling error alone which we are able to isolate through comparison of results based on peer samples and peer population.[6]

We estimate regression models for the PISA maths test score with data on 3459 responding pupils in state schools for whom we have information on FSM receipt.[7] We exclude children at private schools for whom the information on FSM is not recorded (receipt is likely to be zero in this group) and a small number of respondents in state schools for whom the information is also missing (these two groups represent 5.9 percent and 1.5 percent respectively of all responding pupils whom we successfully linked to their NPD record). Among the responding state school pupils whom we analyse, 10.4 percent received FSM. We take the proportion of other 15 year olds in each individual's school who receive FSM as our measure of the peer group composition. The true value, x_i, is

[1] The sampling may be of pupils within a class, as in Henry and Rickman (2007), rather than the school year. The same problem of attenuation bias arises if the class is taken as the peer group (as in this case).

[2] Not all school surveys share this problem. For example, in the Trends in International Mathematics and Science Study (TIMSS) a whole class is randomly selected within each school. If the peer group is defined as the whole class rather than the whole cohort, then all peers are observed in this survey (provided survey response within each class is complete). Toma and Zimmer (2000) investigate peer effects with TIMSS data.

[3] The only exception, where x is observed, will be for small schools with 35 or less 15 year olds since in this case all students of this age in the school are sampled by PISA.

[4] In order to quantify the implications for the bias of this departure from the standard assumption of 'classical' error, we would need to develop multilevel models capable of capturing this dependence, which is not the purpose of the paper.

[5] We are able to link 97 percent of all PISA respondents to their records in the NPD (Micklewright et al., 2012); we exclude the other 3 percent from our analysis.

[6] We take the individual's year cohort as the peer group, but the problem of sampling error may well again arise with other peer group definitions, depending on the survey design – see footnotes 1 and 2.

[7] Results are very similar using either the science or reading test scores as an alternative. We use the average of the five 'plausible values' of the maths score estimated by the survey organisers for each respondent.

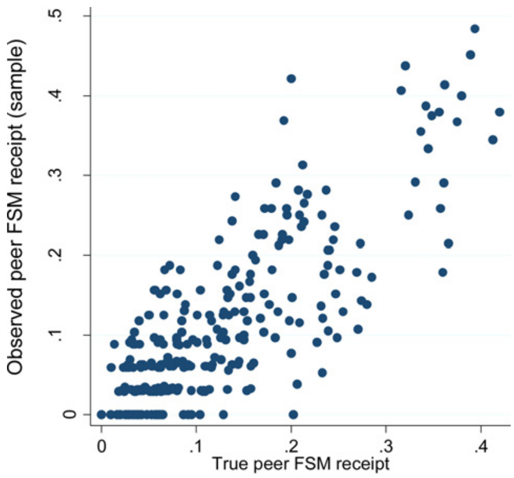

Fig. 1. 'Observed' (z_i) and 'true' (x_i) peer FSM variables. *Notes*: The observed peer FSM receipt relates to all sampled peers. The correlation coefficient is 0.82.

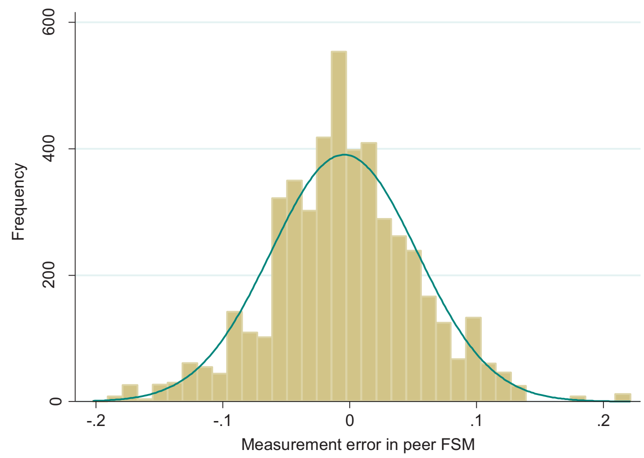

Fig. 2. Distribution of measurement error in peer FSM variable due to sampling (e_i). *Notes*: The graph shows $e_i = z_i - x_i$ with a normal distribution with the same mean and standard deviation superimposed. The observed peer FSM receipt, z_i, relates to all sampled peers and the true peer FSM receipt, x_i, to all members of the individual's year cohort in his or her school. The mean and standard deviation of e_i are −0.004 and 0.058 respectively.

measured by receipt of FSM among all 15 year olds in the individual's school, while the 'observed value in the PISA survey data, z_i, is measured by receipt of FSM among the 35 sampled students. In both cases the peer measure is obtained subtracting the FSM indicator value for the individual concerned from the count of all (or sampled) individuals receiving FSM in the individual's school. Measurement error e_i is given by z_i minus x_i.

A complication is introduced by non-response; 23 percent of sampled pupils in England in PISA 2003 declined to participate in the survey. This means that we can define the peer measure based on the survey data in two ways: (i) students sampled for PISA, and (ii) the subset of responding students. In the first case, z_i is indeed based on the 35 sampled students in each school, less the individual concerned. Here the measurement error e_i reflects only sampling error. In the second case, z_i is based on the other responding students in each individual's school. Here e_i is affected in addition by the pattern of response. The additional measurement error in the peer variable produced by the non-response would simply lead to further attenuation bias if the non-response were random. But the evidence shows this not to be the case: non-response is higher for pupils of lower ability (Micklewright, Schnepf, & Skinner, 2012). Hence the impact of the non-response on the estimate of the peer parameter is an empirical question to investigate with the data.

Fig. 1 plots the observed z_i against the true x_i, where z_i in this case is defined in the first of the ways just described. The two measures are strongly correlated but there is also a fair degree of scatter around the 45° line reflecting the impact of sampling error. The extent of the sampling error, $e_i = z_i - x_i$, is shown more directly in Fig. 2. The error averages close to zero but ranges from about −0.2 to +0.2. The standard deviation of 0.058 may be compared with the mean of the true x_i, 0.118. The extent of the sampling error is sufficient for us to expect that a non-trivial degree of bias will arise from the use of the survey-based measure of the peer variable.

The properties of the observed e_i are not identical to those of u_i in the textbook measurement error set-up described earlier in Section 2. We have already noted that the correlation of sampling errors in peer measurement will be very high for students in the same school. In practice, it is also the case that we observe a correlation between e_i and true peer value x_i of −0.18, rather than the value of zero in the textbook case. (We easily reject the hypothesis that the correlation is zero e.g. at the 0.1 percent significance level.) Fig. 3 plots the two variables against each other. There is a bounding of both x_i and z_i from below by zero. While true x_i is only zero for one school, measured z_i is zero for about 10 percent of our pupil sample: sampling from schools with low levels of FSM can result in there being no peers in the PISA sample who are in receipt (recall that on average only 10 percent of pupils receive the benefit). In this case $e_i = -x_i$ and these are the observations on the line running from north-west to south-east at the left

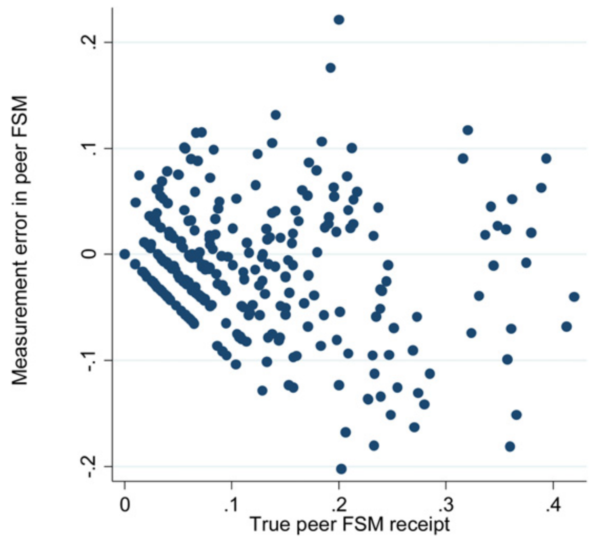

Fig. 3. Sampling error (e_i) in peer FSM measure and 'true' (x_i) peer FSM. *Notes*: See notes to Fig. 2.

Table 1
Fitted linear regression models for the PISA maths score with peer FSM receipt measured in three ways.

	1	2	3
True peer FSM receipt, *x*	−307.1		
	(15.6)		
Observed peer FSM receipt, *z* (sample)		−185.7	
		(14.5)	
Observed peer FSM receipt, *z* (responding pupils)			−172.4
			(14.6)
Constant	537.9	523.5	520.3
	(2.3)	(2.2)	(2.1)
Observations	3459	3459	3459
R-Squared	0.10	0.05	0.04

Note: The dependent variable is the PISA maths score (mean = 502.3, SD = 87.6). Standard errors are given in parentheses. Clustering in schools is allowed for when estimating the standard errors (see e.g. StataCorp, 2003). Results are based on unweighted data.

Table 2
Fitted linear regression models for the PISA maths score including peer FSM receipt measured in three ways.

	1	2	3
FSM in receipt (pupil)	−20.2	−25.8	−26.7
	(3.9)	(3.8)	(3.8)
True peer FSM receipt, *x*	−220.8		
	(31.1)		
Observed peer FSM receipt, *z* (sample)		−112.1	
		(25.8)	
Observed peer FSM receipt, *z* (responding pupils)			−103.1
			(24.5)
Female	−10.5	−10.9	−10.8
	(3.6)	(3.8)	(3.8)
Mother has secondary education	5.0	9.0	9.6
	(4.0)	(4.3)	(4.4)
Mother has tertiary education	15.1	16.2	16.2
	(3.4)	(3.4)	(3.4)
Missing value mother secondary education	−23.2	−21.2	−20.6
	(6.0)	(6.3)	(6.4)
Missing value mother tertiary education	62.8	69.6	71.1
	(18.0)	(20.4)	(20.8)
More than 100 books at home	44.9	48.2	48.8
	(3.1)	(3.4)	(3.4)
Missing value books	−16.0	−18.9	−19.8
	(13.8)	(14.8)	(15.1)
Constant	510.2	493.2	490.5
	(6.1)	(6.1)	(6.1)
Observations	3459	3459	3459
R-Squared	0.21	0.17	0.17

Note: The dependent variable is the PISA maths score (mean = 502.3, SD = 87.6). Standard errors are given in parentheses. Clustering in schools in allowed for when estimating the standard errors (see e.g. StataCorp, 2003). Results are based on unweighted data. The model specification in each column is the same, with column 1 including the variable for true peer FSM receipt, column 2 the variable for sampled peer FSM receipt, and column 3 the variable for responding peer FSM receipt.

hand side of the graph. But the negative correlation is still present with these observations excluded.

Ignoring these differences, if we were to approximate u_i by e_i in Eq. (2) and the population variances by our sample values, we would conclude from the textbook formula (6) that the plim of the OLS estimate of the coefficient on the peer measure differs from the true value by a factor of 0.72 when measuring peer FSM receipt using the survey data (and when estimating a regression with a single explanatory variable).

4. Estimated bias in the peer group and other coefficients

In order to keep close to the textbook case, we begin with a simple linear regression of the PISA maths test score on the peer FSM measure (Table 1). The maths score has a mean of 502.3 and a standard deviation of 87.6.[8] Peer FSM in column 1 is measured by x_i, the proportion of all other 15 year olds in the individual's school who are receiving this benefit. There is therefore no measurement error bias from sampling affecting the estimate here, although we would expect omitted variable bias due to peer FSM proxying other factors influencing maths achievement – as yet there are no other variables included in the model. (There would also be other biases if there were any mismeasurement of FSM in NPD.) By contrast, the estimates of the peer effect reported in columns 2 and 3 suffer from attenuation since they are obtained by using measures of FSM receipt based respectively on peers drawn for the PISA sample and on the subset who respond to the survey. A comparison of these estimates with the figure in column 1 gives an estimate of the extent of attenuation bias when using only the sample of peers. The estimated coefficients in columns 2 and 3 are about 60 percent of that in column 1, indicating a larger problem in practice than would be suggested by the calculation we reported at the end of Section 3 based on the textbook formula.

Table 2 shows results from estimating models that include other explanatory variables: a dummy variable indicating own receipt of FSM, and dummy variables for gender, the level of the mother's education, and the number of books in the home. There is no measure of family income in PISA, so the FSM dummy obtained from the NPD is the only direct indicator of low income available to us. Mother's education is a well-recognised correlate of children's educational attainment, e.g. Haveman and Wolfe (1995). The association reflects both a direct impact on the quantity and quality of time and goods inputs in the child and an indirect impact coming through family income. It may also proxy unobserved parental ability that is passed on to the child through his or her gene endowment. The number of books in the home is estimated by the child and reported in categorical form. This is a standard variable collected in international surveys of children's learning and is often used to proxy family background. It is used as the main measure of both individual and peer characteristics in the analysis of peer effects by Ammermueller and Pischke (2009).

With other variables included in the model, it now makes more sense to consider the size of the estimated peer effect in column 1. A one standard deviation rise in peer FSM receipt, equal to 0.092, is associated with a fall of nearly

[8] These figures refer to unweighted data, as do all our calculations (including the Monte Carlo simulations). Point estimates and hence our estimates of attenuation bias are very similar if we use weighted data. The weights are those supplied with the data by the OECD; they adjust for different sampling probabilities, the level and pattern of school response, and the level of pupil response (Micklewright & Schnepf, 2006).

0.25 of a standard deviation of the maths score. This is large, above the range of most peer effect sizes measured in this standardised way that are reported by Ammermueller and Pischke (2009: 340) from their review of the literature, and about 1.5 times the size of the average effect these authors find for the six European countries in their study of primary school children. In line with our discussion of the case of multiple regression in Section 2, the inclusion of the other variables in the model leads to an increase in attenuation bias in the peer effect estimates in columns 2 and 3 – the estimated peer FSM coefficients are now only about half that of the column 1 estimate.

Measurement error in one of the explanatory variables in a regression model also biases estimates of the coefficients of the other variables – see discussion of Eq. (7) in Section 2 – and we see evidence of this from Table 2. Moving to the peer FSM measures based only on sampled or responding pupils in columns 2 and 3 leads to coefficient estimates for most variables that are biased upwards in absolute size, rather than attenuated as in the case of the coefficient on the peer measure itself. Coefficients on several variables rise in absolute size by an amount equal to about one to one and a half standard errors: the individual FSM receipt dummy, the mother's secondary education dummy (the coefficient almost doubles in this case) and the books dummy.[9]

How does bias change as sample size within each school falls? We investigate this question using a simple Monte Carlo experiment, focusing on the estimated coefficient of the peer FSM measure. We consider a range of sample sizes, from 200 down to 25. Consider the case of n equal to 100. We draw a random sample (without replacement) of 15 year olds of size 100 for each school in the NPD that contains PISA respondents (where the number of 15 year olds in the school is less than 100, we take all 15 year olds, i.e. the population). Suppose that in a particular school, 20 pupils receive FSM in the sample that we draw. Where a PISA respondent in the school receives FSM we assign a value of the peer FSM variable equal to 19/99 and where he or she does not receive FSM we assign a value equal to 20/99. Constructing the peer FSM variable in this way for each PISA respondent, we then estimate the multiple regression model shown in Table 2. We repeat this process 200 times and calculate the mean and standard deviation of the estimated peer FSM variable in the 200 regressions. This procedure is followed with n equal to 200, 175, 150, 125, 100, 75, 50, 35 (the intended PISA sample size), and 25.

Fig. 4 plots the results. With n equal to 200, the mean of the estimated coefficients is −218.8, very close to the value in column 1 of Table 2 (−220.8), and the two standard

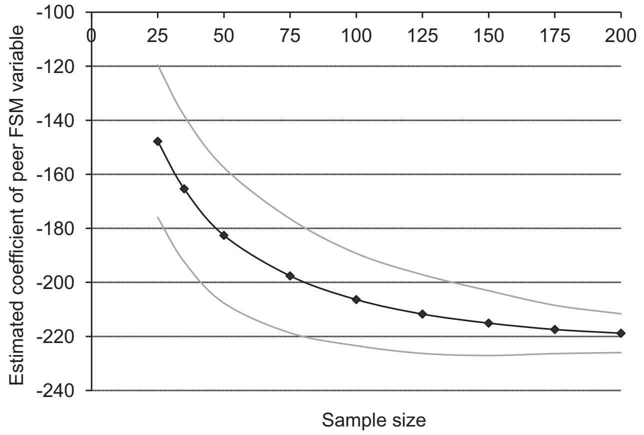

Fig. 4. Monte Carlo simulation of the effect of changing within-school sample size on the estimate of the peer group FSM coefficient. *Notes*: The series with symbols for each data point shows the mean value of the peer group FSM coefficient in regression models estimated for the same 3459 individuals and with the same model specification as in Table 2. To generate each point, we estimate the model having randomly drawn a sample of the size indicated of peers (defined as 15 year olds in the same school) from the NPD for each individual, repeating the process 200 times and averaging the estimated peer FSM coefficient (see text for more details). The two series without symbols show the values of the mean ± 2 standard deviations. Results are based on unweighted data.

deviation interval around the mean is small. In this case, the sample is sufficiently large (for some schools it is even a 100 percent sample) that there is negligible bias due to sampling variation in the peer measure. As sample size falls, attenuation bias increases and the mean estimated coefficient falls in absolute value. The change is non-linear and the mean coefficient rises sharply as sample size falls below 50. With n equal to 35, the intended PISA sample size, the mean estimated peer FSM coefficient is equal to −165.4. This is notably larger than the figure in column 2 of Table 2 (−112.1), and the latter is also outside the two standard deviation interval. This underlines the danger of drawing conclusions on the impact of sampling variation based on a single sample of peers.[10] We therefore conclude that with the sample size intended by the PISA sample design, the expected attenuation in the estimated peer FSM coefficient is about one third. By contrast, it is clear that with sample sizes of 100 or more pupils per school the attenuation effects are quite mild: 10 percent or less.

5. Conclusions

We have investigated attenuation bias in peer effect estimates that arise when information is available for just a random sample of peers rather than all peers, a situation that is not uncommon in school surveys. In our particular empirical setting of the PISA sample for England for 2003 and a peer variable measuring the proportion of children receiving an in-kind benefit for low income families, we

[9] We also estimated the model with other specifications, adding further survey variables to measure socio-economic background: dummies for the father's education and a more detailed set of dummies for the number of books in the home. We calculated an estimated attenuation ratio by dividing the coefficient on the peer FSM receipt when measured with sampled or responding peers by the coefficient on the 'true' FSM peer variable reported in column 1 of Table 2. The value was essentially unchanged from those implied by Table 2: around 0.5. We also experimented with dropping cases with missing values; results were essentially unchanged.

[10] The Monte Carlo experiment could have been designed to force the simulation results to agree with the point estimate obtained in Table 2 column 2 (by merely adding additional pupils to the sampled PISA peers for values of n greater than 35). We rejected this alternative as providing less general results.

were able to exploit linked administrative data on benefit receipt among all children in the same age cohort at each individual's school. We found substantial attenuation bias in the estimated peer effect when measuring peer receipt using just the peers present in the survey data. Biases were also present in estimates of other parameters. A Monte Carlo simulation provided more general results: a non-linear increase in bias as peer sample size fell and expected attenuation bias of about one third in the peer group coefficient with the sample size implied by PISA's survey design.

These results suggest that caution is needed when estimating peer effects with survey data of the type we have used here. The extent of attenuation bias will vary with the empirical setting.[11] As far as use of PISA data is concerned, there should be less attenuation bias in countries where schools are more socially segregated and hence where peer groups are more homogenous. In the case of England, a medium-ranking country for segregation in schools, the level of attenuation bias can be expected to be in the middle of the range.

Acknowledgements

Micklewright's work was supported by ESRC grant RES-576-25-0014 'Administrative Data – Methods, Inference and Network', the ADMIN node of the National Centre for Research Methods. The paper builds on an appendix on the impact of sampling variation on estimates of peer effects in Micklewright and Schnepf (2006), work funded by the Department for Education. We are grateful to the Department for continued access to the PISA data linked to the National Pupil Database (views expressed in the paper should not be associated with the Department). Micklewright thanks the Institute for Research on Poverty at the University of Wisconsin-Madison, where he worked on the paper during a sabbatical visit. Participants at seminars at Southampton and the Institute of Education and at the ISA RC 04 conference in Barcelona made very helpful comments, as did the referees.

References

Ammermueller, A., & Pischke, J.-S. (2006). *Peer effects in European primary schools: Evidence from PIRLS*. ZEW Discussion Paper 06-027.

Ammermueller, A., & Pischke, J.-S. (2009). Peer effects in European primary schools: Evidence from the Progress in International Reading Literacy Study. *Journal of Labor Economics*, *27*(3), 315–348.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In J. Heckman, & E. Leamer (Eds.), *Handbook of econometrics*. Amsterdam: Elsevier.

Burgess, S., McConnell, B., Propper, C., & Wilson, D. (2004). *Sorting and choice in English secondary schools*. Working Paper 04/111. CMPO, University of Bristol.

Entorf, H., & Lauk, M. (2006). *Peer effects, social multipliers and migrants at school: An international comparison*. Discussion Paper 2182. Bonn: IZA.

Fertig, M. (2003). *Educational production, endogenous peer group formation and class composition – Evidence from the PISA 2000 study*. Discussion Paper 714. Bonn: IZA.

Fuller, W. A. (1987). *Measurement error models*. Hoboken, NJ: Wiley.

Goldstein, H., & Noden, P. (2003). Modelling social segregation. *Oxford Review of Education*, *29*(2), 225–237.

Greene, W. (1993). *Econometric analysis*. Upper Saddle River, NJ: Prentice Hall.

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics*, *18*, 527–544.

Hausman, J. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *Journal of Economic Perspectives*, *15*, 57–67.

Haveman, R., & Wolfe, B. (1995). The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature*, *33*(4), 1829–1878.

Henry, G. T., & Rickman, D. K. (2007). Do peers influence children's skill development in preschool? *Economics of Education Review*, *26*, 100–112.

Hobbs, G., & Vignoles, A. (2009). Is children's free school meal 'eligibility' a good proxy for family income? *British Educational Research Journal*, *36*(4), 673–690.

Jenkins, S. P., Micklewright, J., & Schnepf, S. V. (2008). Social segregation in secondary schools: How does England compare with other countries? *Oxford Review of Education*, *34*(1), 21–38.

Kravdal, O. (2006). A simulation-based assessment of the bias produced when using averages from small DHS clusters as contextual variables in multilevel models. *Demographic Research*, *15*(1), 1–20.

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

Micklewright, J., & Schnepf, S. V. (2006). *Response bias in England in PISA 2000 and 2003*. DfES Research Report 771. www.dcsf.gov.uk/research/data/uploadfiles/RR771.pdf

Micklewright, J., Schnepf, S. V., & Skinner, C. (2012). Non-response biases in surveys of school children: The case of the English PISA samples. *Journal of the Royal Statistical Society, Series A*, *175*(4), 915–938.

Neidell, M., & Waldfogel, J. (2008). *Cognitive and non-cognitive peer effects in early education*. NBER Working Paper 14277.

OECD. (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD.

OECD. (2007). *PISA 2006 – Science competencies for tomorrow's world. Vol. 1: Analysis*. Paris: OECD.

Schindler-Rangvid, B. (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.

Schneeweis, N., & Winter-Ebmer, R. (2007). Peer effects in Austrian schools. *Empirical Economics*, *32*, 387–409.

Silva, P. N., Micklewright, J., & Schnepf, S. V. (2012). *The impact of sampling variation on peer measures: A comment on a proposal to adjust estimates for measurement error*. Department of Quantitative Social Science Working Paper. Institute of Education, University of London.

Sojourner, A. (2011). *Identification of peer effects with missing peer data: Evidence from Project STAR*. IZA Discussion Paper 5432.

StataCorp. (2003). *Stata survey data – Reference manual – Release 8*. College Station, TX: Stata Press Publication.

Toma, R., & Zimmer, E. (2000). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management*, *19*(1), 75–92.

Vigdor, J., & Nechyba, T. (2007). Peer effects in North Carolina public schools. In L. Woessmann, & P. E. Peterson (Eds.), *Schools and the equal opportunity problem*. Cambridge, MA: The MIT Press.

Woodhouse, G., Yang, M., Goldstein, H., & Rasbash, J. (1996). Adjusting for measurement error in multilevel analysis. *Journal of the Royal Statistical Society, Series A*, *159*(2), 201–212.

[11] In Silva, Micklewright, and Schnepf (2012) we cast doubt on a simple adjustment factor for attenuation bias resulting from sampling error in the peer measure that is suggested in Neidell and Waldfogel (2008), who drew on Ammermueller and Pischke (2006, 2009).