Supplemental material to accompany: "Peer effects in European primary schools: Evidence from PIRLS"

Andreas Ammermueller and Jörn-Steffen Pischke

May 4, 2009

Our estimates of peer effects will be biased if students are sorted systematically into classrooms even within schools. Although we present some evidence which is suggestive of quasi-random assignment in most countries this possibility can never be fully ruled out. Here we discuss the consequences of such systematic sorting for our estimates. We consider the possibility that the true peer effect is zero. We then ask the question: would our test of random assignment in table 4 reject given plausible parameter values for a model of student achievement which are able to generate our estimation results?

We start with a model of peer effects similar to eq. (1) in the paper. We ignore covariates other than the background variable

$$x_{ic} = \eta_c + v_{ic}.$$

Also, we drop the distinction between classrooms and schools here, so the subindex c refers to classrooms. We will consider sorting across classrooms (and this would be within schools in our estimation setting). Student achievement is given by

$$y_{ic} = \beta_1 \eta_c + \beta_2 v_{ic} + \lambda w_{ic} + \mu_c + \varepsilon_{ic} \tag{1}$$

where $w_{ic} = \overline{x}_{(-i)c}$ is the peer leave-out mean. Notice that we have allowed separate effects on achievement of the sub-components of x_{ic}. There is no reason to assume that the effects of η_c and v_{ic} are the same, simply because we only observe x_{ic}. On the other hand, there is only a single coefficient λ on

w_{ic}, although w_{ic} contains the same two components η_c and the mean of v_{ic}. But the η_c component is just the same as before. Hence, β_1 is neither an indivdual effect nor a peer effect as the common classroom effect η_c does not allow this distinction. As a result, the model in (1) is the most general linear model for achievement using the variance components we have introduced.

We assume that v_{ic} and ε_{ic} are zero mean iid random variables. This implies that the model in (1) has 8 parameters: $\beta_1, \beta_2, \lambda, \sigma_\eta^2, \sigma_v^2, \sigma_\mu^2, \sigma_\varepsilon^2, \sigma_{\eta\mu}$. This can be thought of as a covariance structure model. Since only x_{ic} is observed rather than its components, this model generates the six moments $var(y)$, $var(x)$, $var(w)$, $cov(y,x)$, $cov(y,w)$, and $cov(x,w)$ and it is obvious that the model in this general form is not identified.[1]

We are not necessarily interested in the full model in (1). In particular, we would like to know whether a version of the model can generate our results under the restriction that $\lambda = 0$ but $\sigma_{\eta\mu} > 0$. The covariance $\sigma_{\eta\mu}$ governs the degree of sorting of students into classrooms. If $\sigma_{\eta\mu} = 0$, the conditions imposed in our paper hold as there are no correlated effects. The part of the classroom level variation in η_c and μ_c which is independent is not particularly interesting (it will just raise standard errors), so we make the further assumption that all classroom level variation is due to correlated effects, i.e. $\sigma_\eta^2 = \sigma_\mu^2 = \sigma_{\eta\mu}$, or in essence $\eta_c = \mu_c$. This restriction essentially assumes the maximum amount of sorting conditional on a value for σ_η^2. This seems like the specification that stacks things most in favor of finding spurious peer effects.

The restricted model is:

$$y_{ic} = \beta_1 \eta_c + \beta_2 v_{ic} + \eta_c + \varepsilon_{ic} = (\beta_1 + 1)\eta_c + \beta_2 v_{ic} + \varepsilon_{ic}. \tag{2}$$

This model has the five remaining parameters $\beta_1, \beta_2, \sigma_\eta^2, \sigma_v^2, \sigma_\varepsilon^2$. We have simulated this model with an arbitrary set of parameters which are able to generate the results in the paper. We then ask the question whether we can detect the classroom level component η_c in the background variable given our test in table 4.

[1]In general, it would be possible also to look at the within classroom moments in addition to the variation across the whole sample. Since w_{ic} is a leave-out variable which varies by student, there is within class variation which in principle helps to identify parameters (see Lee, 2007). For example, this variation could be used to separately estimate the effects due to μ_c (which doesn't vary within classes) and w_{ic} (which does). In practice, this variation is typically not sufficient to be very helpful. As a result, we will abstract from the within variation here.

We proceed as follows. In the paper we estimate regressions of the form

$$y_{ic} = a + bx_{ic} + cw_{ic} + e_{ic}. \tag{3}$$

If the data are generated by (2) then this regression will yield

$$\text{plim } \widehat{b} = (1 + \beta_1)\alpha + \beta_2(1 - \alpha)$$

where

$$\alpha = \frac{\sigma_\eta^2}{\overline{N}\sigma_\eta^2 + \sigma_v^2}$$

N is the class size, and

$$\text{plim } \widehat{c} = (\overline{N} - 1)(1 + \beta_1 - \beta_2)\alpha.$$

Since β_1, β_2 and λ don't have a natural scale, we focus in the paper on the ratio between the individual and the peer effect estimate, $\widehat{c/b}$. Define

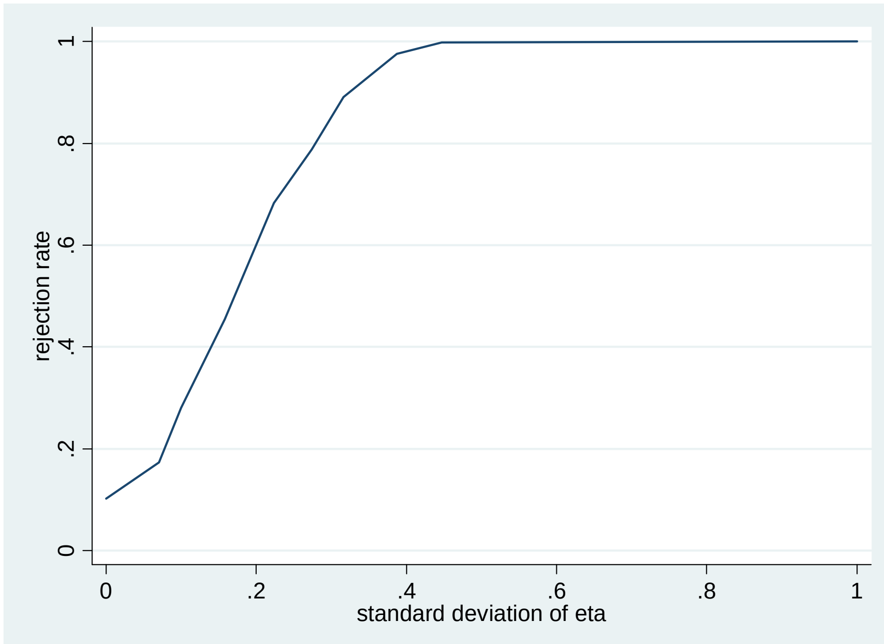
$$\kappa \equiv \frac{\text{plim } \widehat{c}}{\text{plim } \widehat{b}} = \frac{(\overline{N} - 1)(1 + \beta_1 - \beta_2)\alpha}{(1 + \beta_1)\alpha + \beta_2(1 - \alpha)}.$$

We want to choose parameters holding this ratio κ fixed. In practice, we choose the parameters $\beta_1, \sigma_\eta^2, \sigma_v^2, \sigma_\varepsilon^2$ and, solving for β_2, impose the constraint

$$\beta_2 = \frac{(1 + \beta_1)\alpha(\overline{N} - 1 - \kappa)}{(\overline{N} - 1)\alpha + \kappa(1 - \alpha)}.$$

Note that the regression estimates of b and c involve all the parameters (except σ_ε^2). The test for random assignment of students to classes, on the other hand, only depends on the two parameters σ_η^2 and σ_v^2. While we could infer the relative magnitude of these variances directly from the between and within classroom variance in x_{ic} in our data, we do not pursue this avenue. Estimating σ_η^2 and assessing whether $\sigma_\eta^2 > 0$ amounts to the same as testing for random assignment to classes. Instead, we therefore analyze the model for a series of prespecified values for σ_η^2.

After setting σ_η^2 to a particular value, we draw the parameters $\beta_1, \sigma_v^2, \sigma_\varepsilon^2$ independently from a uniform distribution on the interval [0,10]. We then calculate the implied β_2, setting $\kappa = 0.75$. This is the minimum value for κ we find in the last two columns of table 7. We then draw the random



variables $\eta_c, v_{ic}, \varepsilon_{ic}$ from normal distributions with mean zero and the chosen variances $\sigma_\eta^2, \sigma_v^2, \sigma_\varepsilon^2$. Finally, we generate y_{ic} according to (2). We generate data on 370 schools, which is the number of schools in our data set with more than one class (pooling all countries). Three quarters of these schools have two classes, the remainder three classes, generating the observed average of 2.25 class rooms per school in our sample. Each class in our sample has 25 students, the average class size in our data.

We then run the regression (3) on the simulated data. We do this in order to verify that the regression produces values of $\widehat{c}/\,\widehat{b}$ close to the desired value of κ. If $\sigma_\eta^2 = 0$, $\alpha = 0$, and as a result plim $\widehat{c} = \kappa = 0$. But as soon as σ_η^2 is small and positive, the regressions hit the chosen value of κ quite closely. More importantly, we then apply the χ^2-test for independence of x_{ic} and classroom assignment to the simulated data. Since this test is based on a discrete x-variable, we discretize x_{ic} into five approximately equal sized bins. The test is in essence trying to assess whether $\sigma_\eta^2 > 0$. We repeat this process a 1,000 times for each value of σ_η^2 we have chosen, and we record the empirical rejection probabilities of the χ^2-test with nominal size of 5%. $\sigma_\eta^2 = 0$ corresponds to the null hypothesis of the test, i.e. classroom assignment is independent of x_{ic}. The rejection rate for $\sigma_\eta^2 = 0$ is therefore the actual size of the test in our simulated data, and the rejection rates for $\sigma_\eta^2 > 0$ trace out the power function of the test for these values of σ_η^2. The results are displayed in the figure above.

The figure demonstrates that the test has a size distortion under the null: the 5% rejection rate is around 0.13 for $\sigma_\eta^2 = 0$. In table 4 in the paper we found a p-value for the test of 0.036 for Sweden. The size distortion in our simulated data implies that this corresponds to an actual p-value of 0.103. The power of the test then rises quickly and becomes close to 1.0 for values of $\sigma_\eta^2 = 0.15$. Recall that we draw all other variances from a range of 0 to 10. For $\sigma_\eta^2 = 0.15$, the variance σ_η^2 will therefore be smaller than σ_v^2 in 98.5% of the simulations, and on average $\sigma_\eta^2 = 0.03\sigma_v^2$. This suggests that the test has very good power to detect even very small classroom components in the background variable.

References

Lee, Lung-fei (2007) "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics*, Volume 140, Issue 2, October 2007, Pages 333-374.

Derivation of the plims

Here we derive the plims for $\widehat{b}$ and $\widehat{c}$:

$$\begin{aligned}
\text{plim}\ \widehat{b} &= \frac{var(w)cov(y,x) - cov(w,x)cov(y,w)}{var(x)var(w) - cov(w,x)^2} \\
&= \frac{\left(\sigma_\eta^2 + \frac{\sigma_v^2}{\overline{N}-1}\right)\left[\left(\beta_1+1\right)\sigma_\eta^2 + \beta_2\sigma_v^2\right] - \sigma_\eta^2\left[\left(\beta_1+1\right)\sigma_\eta^2\right]}{\left(\sigma_\eta^2+\sigma_v^2\right)\left(\sigma_\eta^2 + \frac{\sigma_v^2}{\overline{N}-1}\right) - \sigma_\eta^4} \\
&= \frac{\left(\overline{N}-1\right)\beta_2\sigma_\eta^2\sigma_v^2 + \sigma_v^2\left[\left(\beta_1+1\right)\sigma_\eta^2 + \beta_2\sigma_v^2\right]}{\sigma_\eta^2\sigma_v^2 + \left(\overline{N}-1\right)\sigma_\eta^2\sigma_v^2 + \sigma_v^4} \\
&= \frac{\left(\beta_1+1\right)\sigma_\eta^2\sigma_v^2 + \beta_2\sigma_v^2\left[\left(\overline{N}-1\right)\sigma_\eta^2 + \sigma_v^2\right]}{\overline{N}\sigma_\eta^2\sigma_v^2 + \sigma_v^4} \\
&= \frac{\left(\beta_1+1\right)\sigma_\eta^2 + \beta_2\left[\left(\overline{N}-1\right)\sigma_\eta^2 + \sigma_v^2\right]}{\overline{N}\sigma_\eta^2 + \sigma_v^2} \\
&= (1+\beta_1)\alpha + \beta_2(1-\alpha)
\end{aligned}$$

where

$$\alpha = \frac{\sigma_\eta^2}{\overline{N}\sigma_\eta^2 + \sigma_v^2}$$

and

$$\begin{aligned}
\text{plim } \widehat{c} &= \frac{var(x)cov(y,w) - cov(w,x)cov(y,x)}{var(x)var(w) - cov(w,x)^2} \\
&= \frac{(\sigma_\eta^2 + \sigma_v^2)\left[(\beta_1 + 1)\,\sigma_\eta^2\right] - \sigma_\eta^2\left[(\beta_1 + 1)\,\sigma_\eta^2 + \beta_2\sigma_v^2\right]}{(\sigma_\eta^2 + \sigma_v^2)\left(\sigma_\eta^2 + \frac{\sigma_v^2}{\overline{N}-1}\right) - \sigma_\eta^4} \\
&= \frac{\left(\overline{N} - 1\right)\left[(\beta_1 + 1)\,\sigma_\eta^2\sigma_v^2 - \beta_2\sigma_\eta^2\sigma_v^2\right]}{\sigma_\eta^2\sigma_v^2 + \left(\overline{N} - 1\right)\sigma_\eta^2\sigma_v^2 + \sigma_v^4} \\
&= \frac{\left(\overline{N} - 1\right)[\beta_1 + 1 - \beta_2]\,\sigma_\eta^2}{\overline{N}\sigma_\eta^2 + \sigma_v^2} \\
&= (\overline{N} - 1)(1 + \beta_1 - \beta_2)\alpha.
\end{aligned}$$

Solving the ratio of the two for β_2 yields

$$\begin{aligned}
\kappa &= \frac{(\overline{N} - 1)(1 + \beta_1 - \beta_2)\alpha}{(1 + \beta_1)\alpha + \beta_2(1 - \alpha)} \\
\left[(1 + \beta_1)\alpha + \beta_2(1 - \alpha)\right]\kappa &= (\overline{N} - 1)(1 + \beta_1 - \beta_2)\alpha \\
(1 + \beta_1)\alpha\kappa + \beta_2(1 - \alpha)\kappa &= (\overline{N} - 1)(1 + \beta_1)\alpha - (\overline{N} - 1)\beta_2\alpha \\
\beta_2(1 - \alpha)\kappa + (\overline{N} - 1)\beta_2\alpha &= (\overline{N} - 1)(1 + \beta_1)\alpha - (1 + \beta_1)\alpha\kappa \\
\beta_2\left[(1 - \alpha)\kappa + (\overline{N} - 1)\alpha\right] &= \left[\overline{N} - 1 - \kappa\right](1 + \beta_1)\alpha \\
\beta_2 &= \frac{(1 + \beta_1)\alpha(\overline{N} - 1 - \kappa)}{(\overline{N} - 1)\alpha + \kappa(1 - \alpha)}.
\end{aligned}$$