LLAKES
CENTRE FOR LEARNING AND LIFE CHANCES IN KNOWLEDGE ECONOMIES AND SOCIETIES

IOE LONDON
Leading education and social research
Institute of Education University of London

Endogeneity Problems in Multilevel Estimation of Education Production Functions: an Analysis Using PISA Data

Saïd Hanchane and Tarek Mostafa

LLAKES Research Paper 14

Centre for Learning and Life Chances in Knowledge Economies and Societies

LLAKES is an ESRC-funded Research Centre - grant reference RES-594-28-0001.

To request printed copies of this paper or other LLAKES Research Papers,
please contact the LLAKES Centre - llakescentre@ioe.ac.uk.

To view this and other LLAKES Research Papers online, please visit www.llakes.org.uk.

ISSN 2042-5929

Endogeneity Problems in Multilevel Estimation of Education Production Functions: an Analysis Using PISA Data.

Saïd Hanchane and Tarek Mostafa*

Abstract

This paper explores endogeneity problems in multilevel estimation of education production functions. The focus is on level 2 endogeneity which arises from correlations between student characteristics and omitted school variables.

We first develop a theoretical model in order to show that school and peer characteristics are the by-product of student background. This theoretical framework helps the identification of the hypotheses we would like to test within the empirical part. From an econometric point of view, the correlations between student and school characteristics imply that the omission of some variables may generate endogeneity bias. Therefore, in the second section of the paper, an estimation approach based on the Mundlak (1978) technique is developed in order to tackle bias and to generate consistent estimates.

The entire analysis is undertaken in a comparative context between three countries: Germany, Finland and the UK. Each one of them represents a particular system. For instance, Finland is known for its extreme comprehensiveness, Germany for early selection and the UK for its liberalism. These countries are used to illustrate the theory and to prove that the level of bias arising from omitted variables varies according to the characteristics of education systems.

* Saïd Hanchane: Director of the National Authority of Assessment at the Moroccan Higher Education Council. Email: shanchane@cse.ma. Tarek Mostafa (corresponding author): Researcher at the Institute of Education - University of London, Centre for Learning and Life Chances in Knowledge Economies and Societies (LLAKES) 20 Bedford Way, London WC1H 0AL. Room 706. England, United Kingdom. Office phone: +44 (0)2073315115. Email: T.Mostafa@ioe.ac.uk. The authors are grateful to Martin Weale for his comments.

Contents

Tables

Introduction.

Multilevel estimation of education production functions is plagued by the problem of endogeneity resulting from omitted variables. Typically, endogeneity arises when unobserved variables affecting the outcome of education are correlated with independent variables included in the model. In this paper, we are concerned with level 2 endogeneity arising from correlations between included student characteristics and omitted school variables.

The correlations between student and school characteristics are mainly the result of stratification. For instance, unprivileged households are likely to live in relatively poor communities, due to the functioning of the housing market. These communities are populated with other households of similar type. Under these circumstances, the social mix of the schools operating in these neighbourhoods consists mainly of unprivileged students. Further, some school characteristics such as funding, teacher quality and availability may also be related to the status of the school and its location. Thus, it is possible to deduce that school characteristics are a by-product of students' social status. However, it should be noted that the strength of stratification varies between education systems. For instance, early selection in Germany exacerbates stratification and strengthens the relation between student and school status, while comprehensiveness in Finland does exactly the opposite.

The objective of the paper is to develop a multilevel estimation approach robust to endogeneity that allows us to overcome the omitted variable bias. The study is carried out in the context of three countries: Germany represents German speaking countries (known for early selection), Finland represents the Nordic countries (known for their comprehensiveness), and the UK for the English speaking ones (known for the liberal management of education).

The paper is organized as follows: In the first section, a theoretical model is built upon the work of Epple and Romano (1998). In this model, an economy populated with individuals differentiated by income, aptitudes and social status is considered. This economy has an arbitrarily fixed number of public, private and mixed finance schools and school quality depends on funding, aptitude and social peer effects. The latter are considered to be non-linear in their means. Note that linearity in means was criticized in Hoxby and Weingarth (2005). Schools maximize their profits under several quality constraints. This maximization transforms a continuum of student characteristics into a continuum of admission prices or tuition. In order to be admitted into a private school a student has to pay tuition fees that cover his marginal cost. In contrast, in public schools this marginal cost is covered entirely by public funds, and in mixed finance schools both tuition fees and public funds are used to cover the marginal cost of admitting a new student. In comparison with Epple and Romano (1998), non-linear peer effects, mixed finance schools and school funding are introduced. This theoretical model provides a convenient framework for the econometric analysis. On the one hand, it identifies the link between student and school characteristics and on the other, it justifies the multilevel nature of the empirical analysis.

In the second section, the education production function (EPF), identified in the theory, is assessed. This EPF utilises student, school and peer characteristics to explain variations in performance scores in a multilevel framework. In this context, the omission of some school variables leads to level 2 endogeneity bias. Therefore, various specifications of the model are tested and an endogeneity robust estimation approach is developed. This approach is based on the Mundlak (1978) technique developed for panel data.

In the third section of the paper, the estimation is carried out and the results are presented. The empirical illustration is done using PISA 2003 data for three countries: Germany, Finland and the UK. First, we present the results on the Hausman test and we show that the model omitting peer characteristics suffers from endogeneity. Further, the model which controls for the three vectors of variables: student, school and peer characteristics, is identified as the most robust. Secondly, we show that the level of bias differs between countries according to their contextual framework and according to their level of stratification. Thirdly, we prove that social peer effects are non-linear in their means. Finally, we compare our results to those published in PISA's 'Learning for Tomorrow's world' report, with the intention of affirming that the omission of key variables leads to bias and overestimation.

It should be noted that the theoretical literature on stratification is recent and dates back to the early 1970s with the founding articles of Barzel (1973) and Stiglitz (1974). The major developments occurred in the 1990s, when two distinct bodies of literature emerged. The first studied spatial stratification between jurisdictions and neighbourhoods. It includes Westhoff (1977), Rose-Ackerman (1979), De Bartolome (1990), Epple, Filimon, and Romer (1993), Nechyba (1997), Fernandez and Rogerson (1996), and Epple and Platt (1998). The second studied educational stratification between public and private schools. It includes Arnott and Rowse (1987), Epple and Romano (1998, 2006) and Nechyba (2003). The empirical literature includes a variety of studies that assess the determinants of achievements, such as peer effects, students' ethnicity and immigrant status, students' socioeconomic background, and school and teacher characteristics. Hanushek and Welch (2006) provide a good coverage of the studies of interest. Further, the recent econometric literature dealing with endogeneity problems includes Wooldridge (2002), Skrondal and Rabe Hesketh (2004), Fielding (2004), Snijders and Berkhof (2006), Kim and Frees (2006) and Grilli and Rampichini (2006).

Section One: The Theoretical Framework.

In this model, we consider an economy populated with a continuum of households differentiated by income, social status and student aptitudes. Social status is defined as a proxy for factors such as social class, cultural possessions, and parental education. All these factors are represented through a scalar k. Social status is taken into account to shed light on how school quality is affected by the social mix of enrolled students. Similarly, student aptitudes are defined to proxy factors such as motivation and interest for learning. Each household has parents and one student and forms a single decision making unit. Henceforth, "student" and "household" are used interchangeably.

A student i has an income y_i, an aptitude b_i and a social status k_i. Note that i may designate a particular student or a type of students with the same combination of y, b and k. These students attend a finite number of schools. A school is designated by an index j (with $j =$ $1,2,3, \dots, j$).

Income, aptitudes, and social status are distributed in the population according to the density function $f(b, y, k)$ which is positive and continuous on its support $S = (0, b_{\max}) \times (0, y_{\max}) \times (0, k_{\max})$. Correlations between these three endowments are not considered for simplicity.

Student utility is assumed to be a function of private consumption and school quality. It is noted as $U(c, q)$, where c is consumption and q is school quality. U is increasing, strictly quasi-concave and twice continuously differentiable. Students can attend only one school and they cannot supplement education elsewhere. Educational achievements are given by the education production function $a_i = a(b_i, q_j)$; a is continuous and increasing in both arguments. Achievements depend on students' aptitudes and on the quality of their school. In other words, the access to higher quality schools is translated into higher achievements. In terms of inequalities, unequal access to education generates unequal outcomes.

School quality is determined by expenditure per pupil, aptitude peer effects, social status peer effects and the dispersion of social status. Quality is increasing in all its arguments. It should be noted that we can reasonably assume that schools and policy makers appreciate social diversity within schools. However, one may think that parents with much social status prefer students of the same type and hence quality should be decreasing in the dispersion of social status. This is not the case in this model, since the problem is solved in two ways. On the one hand, schools maximize rents under a quality constraint which contains the dispersion of social status; so it is up to the schools to say whether school quality is increasing or not in this dispersion. On the other hand, high social status households who value social homogeneity may choose socially homogenous schools by paying higher tuition fees. A similar example would be desegregation in the US. Authorities may impose ethnic diversity into white majority schools (something that may enhance social cohesion, see Green et al 2009); however, white students can move to private non-diverse schools by paying more fees. In my model, authorities impose social diversity and then households choose a school according to their preferences (utility maximization).[2]

Aptitude peer effects are defined to be school average aptitudes, social status peer effects are defined to be school average social status, and the dispersion of social status is its within-school variance.

Three types of schools are considered: free public schools financed entirely by public funds, mixed finance schools financed by public funds and tuition fees paid by students and private schools financed solely by tuition fees.[3] In this economy, all households pay taxes, even if

[2] In this model, communities and geographical stratification are not considered for simplicity and since PISA does not include territorial data. Mostafa (2009 a) provides a theoretical framework for spatial stratification.

[3] Mixed finance schools represent government-dependent private schools controlled by non-government organizations or with governing boards not selected by a government agency that receive a considerable part of their core funding from government agencies.

their children do not attend public or mixed finance schools. The funds allocated to education are collected through a proportional income tax, and the number of students is larger than the number of schools.

The proportion of students of type (b,y,k) in school j is given by $\alpha_j(b,y,k)$, and the number of students in school j is given by l_j .with:

$$l_j = \iiint_S \alpha_j(b,y,k)\, f(b,y,k)\, db\, dy\, dk \tag{1}$$

And $\alpha_j(b,y,k) \in [0,1]$

1. Schools.

The production cost of education depends on the number of students in a school; it is given by $Co(l_j) = V(l_j) + F_j = n_1 l_j + n_2 l_j^2 + F_j$. $V' > 0$, $V'' > 0$, F is a fixed cost for school j and n_1 and n_2 are positive constants. Technical differences between schools are not included in the model.[4] The absence of economies of scale in the production of education is translated into a large number of schools catering for different types of students. [5] Furthermore, schools have perfect information on students' income, aptitudes and social status.

Schools are assumed to maximize profits under a quality constraint. Funding is provided by three sources: government subsidies, tuition paid by students and other earnings.

$$R_j = \iiint_S E_{ij}\alpha_j(b,y,k) f(b,y,k)\, db\, dy\, dk + \iiint_S p_{ij}\alpha_j(b,y,k) f(b,y,k)\, db\, dy\, dk + G_j$$

R_j : School resources or revenues.

E_{ij} : Government subsidies for student i attending school j .

p_{ij} : Tuition paid by student i in school j .

G_j : Other earnings.

[4] This assumption was used in Epple and Romano 1998 (p.38).
[5] Ferris and West (2004) provide evidence that large schools suffer from external and invisible costs "such as social problems that prevent the existence of economies of scale." This is reflected through the positive sign of n_1 and n_2.

The sum of subsidies for a particular school is equal to $\sum E_{ij} = E_j$. At equilibrium, the sum of government subsidies is equal to tax revenues in the economy $\sum E_j = tY$ (the budget is balanced). The tax rate is assumed to be exogenous. In some previous studies, tax rates were considered to be chosen through majority voting. However, the atomistic nature of the economy and the existence of political parties and complex political processes mean that majority voting over tax rates is simplistic. Furthermore, the presence of public and private schools prevents the existence of a majority voting equilibrium due to the non-single peakedness of individual preferences over tax rates. Hence, in order to avoid this problem and to simplify the theoretical framework, I am assuming the exogeneity of the tax rate.[6]

School quality is given by:

$$q_j = q_j\left[\frac{R_j}{l_j}, \theta_j, \mathrm{O}_j, \sigma_j^2\right] \tag{2}$$

With: $\lim_{R_j \to 0} q_j = 0$, $\lim_{\theta_j \to 0} q_j = 0$, $\lim_{\mathrm{O}_j \to 0} q_j = 0$, $\lim_{\sigma_j^2 \to 0} q_j = 0$.

q_j is increasing in $\frac{R_j}{l_j}$, θ_j, O_j and σ_j^2. where:

Expenditure per pupil is given by $\frac{R_j}{l_j}$.

Aptitude peer effects are given by average aptitudes in a school:

$$\theta_j = \frac{1}{l_j}\iiint_s b_i \alpha_j(b,y,k) f(b,y,k)\,dbdydk \tag{3}$$

Social status peer effects are given by average social status in a school:

$$\mathrm{O}_j = \frac{1}{l_j}\iiint_s k_i \alpha_j(b,y,k) f(b,y,k)\,dbdydk \tag{4}$$

The dispersion of social status in a school is given by its within-school variance:

$$\sigma_j^2 = \frac{1}{l_j}\iiint_s (k_i - \mathrm{O}_j)^2 \alpha_j(b,y,k) f(b,y,k)\,dbdydk \tag{5}$$

[6] Mostafa (2009 a) provides the conditions for the existence of a majority voting equilibrium when preferences are non-single peaked.

Aptitude peer effects, social status peer effects, the dispersion of social status and the number of enrolled students represent quality constraints under which school profit is maximized. Profit is equal to the difference between revenues and the cost of producing education $\pi_j = R_j - Co(l_j)$. At equilibrium, profits are equal to zero and no new entries on the education market are possible.[7] In this case, $R_j = Co(l_j)$. Since the sum of expenditure per pupil in a school is given by $\sum \frac{R_j}{l_j} = l_j \frac{R_j}{l_j} = R_j$, it is possible to write that at equilibrium $\sum \frac{R_j}{l_j} \equiv Co(l_j)$. In other words, the sum of expenditure per student is identical to the education production cost.

School profit maximization and price discrimination.

Schools maximize profit under several quality constraints. Even if public schools do not charge tuition, the authorities condition the level of subsidies according to student types. For private and mixed finance schools, the chosen level of quality and the level of public subsidies determine the tuition for each type of students. It should also be noted that private and mixed finance schools do not select students directly, since they admit any student who is able to pay the price corresponding to his type. In fact, displaying a prohibitive price is equivalent to refusing to admit a student. The quality constraints include average aptitudes, average social status, the dispersion of social status and the number of students attending a school. One should keep in mind that profit maximisation does not necessarily imply that schools are making an actual positive profit. Profit maximization is a mechanism that allows schools to optimize the allocation of resources (it is also used by public schools which are non-profit schools).

Schools maximize profit under several quality constraints:

$$\max \pi_j \equiv R_j - V(l_j) - F$$

subject to these constraints:

$$l_j = \iiint_s \alpha_j(b,y,k)\, f(b,y,k)\, db\, dy\, dk \qquad (1)$$

$$\theta_j = \frac{1}{l_j} \iiint_s b_i \alpha_j\left(b,y,k\right) f\left(b,y,k\right) db\, dy\, dk \qquad (3)$$

[7] As long as school profits are positive, new schools will enter the market. See Epple and Romano 1998 (p.39).

$$\mathrm{O}_j = \frac{1}{l_j}\iiint_s k_i \alpha_j(b,y,k) f(b,y,k) dbdydk \quad (4)$$

$$\sigma_j^2 = \frac{1}{l_j}\iiint_s (k_i - \mathrm{O}_j)^2 \alpha_j(b,y,k) f(b,y,k) dbdydk \quad (5)$$

These constraints can be transformed by replacing l_j by its value:

1- $\theta_j \iiint_s \alpha_j(b,y,k) f(b,y,k) dbdydk - \iiint_s b_i \alpha_j(b,y,k) f(b,y,k) dbdydk = 0$

2- $\mathrm{O}_j \iiint_s \alpha_j(b,y,k) f(b,y,k) dbdydk - \iiint_s k_i \alpha_j(b,y,k) f(b,y,k) dbdydk = 0$

3- $\sigma_j^2 \iiint_s \alpha_j(b,y,k) f(b,y,k) dbdydk - \iiint_s (k_i - \mathrm{O}_j)^2 \alpha_j(b,y,k) f(b,y,k) dbdydk = 0$

The Lagrangian function is then written in the following form:

$$\Phi = R_j - n_1 l_j - n_2 l_j^2 - F - \mu_j'[\theta_j \iiint_s \alpha_j(b,y,k) f(b,y,k) dbdydk - \iiint_s b_i \alpha_j(b,y,k) f(b,y,k) dbdydk]$$
$$- \mu_j''[\mathrm{O}_j \iiint_s \alpha_j(b,y,k) f(b,y,k) dbdydk - \iiint_s k_i \alpha_j(b,y,k) f(b,y,k) dbdydk]$$
$$- \mu_j'''\left[\sigma_j^2 \iiint_s \alpha_j(b,y,k) f(b,y,k) dbdydk - \iiint_s (k_i - \mathrm{O}_j)^2 \alpha_j(b,y,k) f(b,y,k) dbdydk\right]$$

with $\mu_j', \mu_j'', and \mu_j'''$ the Lagrangian multipliers.

Partial differentiation of the Lagrangian function over $\alpha_j(b,y,k)$ yields:

$$\frac{\partial \Phi}{\partial \alpha_j(b,y,k)} = R_j'^* - n_1 - 2n_2 l_j - \mu_j' (\theta_j - b_i) - \mu_j''(\mathrm{O}_j - k_i) - \mu_j'''\left[\sigma_j^2 - (k_i - \mathrm{O}_j)^2\right] = 0$$

The optimal level of resources per student $R_j'^*$ is the following:

$$MC = R_j'^* = n_1 + 2n_2 l_j + \mu_j' (\theta_j - b_i) + \mu_j''(\mathrm{O}_j - k_i) + \mu_j'''\left[\sigma_j^2 - (k_i - \mathrm{O}_j)^2\right] \quad (6)$$

Note that $R_j'^*$ represents the marginal cost (MC) of admitting a student of type (b,y,k), with $R_j'^* = E_{ij} + p_{ij}^*$. (After the partial differentiation of R_j over $\alpha_j(b,y,k)$).

$$\mu_j' = \frac{1}{l_j}\iiint_s \frac{\partial R'^*}{\partial \theta_j} \alpha_j(b,y,k) f(b,y,k) dbdydk$$

$$\mu_j'' = \frac{1}{l_j}\iiint_s \frac{\partial R'^*}{\partial \mathrm{O}_j} \alpha_j(b,y,k) f(b,y,k) dbdydk$$

$$\mu_j''' = \frac{1}{l_j}\iiint_s \frac{\partial R'^*}{\partial \sigma_j^2}\alpha_j(b,y,k)f(b,y,k)dbdydk$$

μ_j', μ_j'' and μ_j''' are positive and they vary between schools. μ_j' represents the change to resources per student deriving from a change in school average aptitudes θ_j. μ_j'' represents the change to resources per student deriving from a change in school average social status O_j. μ_j''' represents the change to resources per student deriving from a change in the within-school dispersion of social status σ_j^2.

At equilibrium, profit is equal to zero, $\pi_j = 0$. New entries on the education market are possible as long as $\pi_j > 0$. When $\pi_j = 0$, no new entries are possible.

In this equation, $R_j'^*$ represents the resources needed to cover the marginal cost of admitting a student. The first term $n_1 + 2n_2 l_j$ is the part resulting from the education production cost. It is positive and identical for all students attending the same school. The second, third and fourth terms reflect the externality of one's own aptitude and social status on the school and the cost resulting from them.

The second term, $\mu_j'(\theta_j - b_i)$, represents the impact of one's aptitude on average aptitudes. Students with above average aptitudes $\theta_j < b_i$ have a negative externality cost on the school. The reverse is true for $\theta_j > b_i$. This term is decreasing in b_i given a value of θ_j. $\lim_{b \to +\infty} \mu_j'(\theta_j - b_i) = -\infty$.

The third term, $\mu_j''(\mathrm{O}_j - k_i)$, represents the impact of one's social status on average social status. Students with above average social status $\mathrm{O}_j < k_i$ have a negative externality cost on the school. The reverse is true for $\mathrm{O}_j > k_i$. This term is decreasing in k_i given a value of O_j. $\lim_{k \to +\infty} \mu_j''(\mathrm{O}_j - k_i) = -\infty$.

The fourth term, $\mu_j'''\left[\sigma_j^2-\left(k_i-\mathrm{O}_j\right)^2\right]$, represents the cost of being too close to the average of social status. Those who are far from the mean (who create social diversity) represent a negative externality cost for the school. In other words, when k_i is far enough (higher or lower) from O_j, $\left(k_i-\mathrm{O}_j\right)^2$ is positive and high. If it is higher than σ_j^2, then the term $\mu_j'''\left[\sigma_j^2-\left(k_i-\mathrm{O}_j\right)^2\right]$ is negative. The reverse is true for k_i close enough to O_j. This term, given constant values of σ_j^2 and O_j, is concave in k_i and attains its maximum at $k_i=\mathrm{O}_j$. $\lim\limits_{k\to\pm\infty}\mu_j'''\left[\sigma_j^2-\left(k_i-\mathrm{O}_j\right)^2\right]=-\infty$.

Note that $R_j'^*$ might be negative, depending on the level of government subsidies and the position of individual aptitude and social status relative to the means. Equation (6) represents a compensation scheme; low aptitude students subsidize higher aptitude ones, low social status students subsidize high social status ones and students with social status close to its average subsidize those who are far from the average (those who create social diversity). Furthermore, this equation indicates that the access to educational quality is conditioned by student characteristics. In other words, those who have low combinations of (b,y,k) are the most disadvantaged since the market operates to their detriment (except that low k students might be rewarded through the forth term of the equation).

Equation (6) allows us to overcome the strict distinction between public and private schools. Different types of education finance can be considered: we can start with free admission public schools with $p_{ij}=0$ for all i and go through mixed finance schools where both p and E are positive and eventually reach purely private schools where $E_{ij}=0$.

When $R_j'^*$ is replaced by its value we obtain the following equations:

<u>For private schools, we have:</u>

$$p_{ij}^*=n_1+2n_2l_j+\mu_j'\ (\theta_j-b_i)+\mu_j''(\mathrm{O}_j-k_i)+\mu_j'''\left[\sigma_j^2-\left(k_i-\mathrm{O}_j\right)^2\right]$$

With $E_{ij}=0$.[8]

[8] Pricing is done according to the marginal cost of admitting a student.

Note that p_{ij}^{*} can be negative for some students (e.g. scholarship) but not for all, for the following reason: a private school cannot possibly offer scholarships for all its students. In order to offer a scholarship for one student, another has to pay a positive tuition. This can be seen through the sum of p_{ij}^{*} which is always positive.

$$\sum_{i=1}^{l_j} p_{ij}^{*} = \sum_{i=1}^{l_j}\left(n_1 + 2n_2 l_j\right) + \sum_{i=1}^{l_j}\left[\mu_j'\left(\theta_j - b_i\right)\right] + \sum_{i=1}^{l_j}\left[\mu_j''\left(\mathrm{O}_j - k_i\right)\right] + \sum_{i=1}^{l_j}\left\{\mu_j'''\left[\sigma_j^2 - \left(k_i - \mathrm{O}_j\right)^2\right]\right\} > 0 \text{ since}$$

$$\sum_{i=1}^{l_j}\left(n_1 + 2n_2 l_j\right) > 0 \text{ and}$$

$$\sum_{i=1}^{l_j}\left[\mu_j'\left(\theta_j - b_i\right)\right] + \sum_{i=1}^{l_j}\left[\mu_j''\left(\mathrm{O}_j - k_i\right)\right] + \sum_{i=1}^{l_j}\left\{\mu_j'''\left[\sigma_j^2 - \left(k_i - \mathrm{O}_j\right)^2\right]\right\} = 0$$

Mathematical details :

$$\sum_{i=1}^{l_j}\left[\mu_j'\left(\theta_j - b_i\right)\right] = \mu_j'\left(\theta_j - b_1\right) + \ldots + \mu_j'\left(\theta_j - b_{l_j}\right) = \mu_j'\left[\left(\theta_j - b_1\right) + \ldots + \left(\theta_j - b_{l_j}\right)\right]$$

$$= \mu_j'\left(l_j\theta_j - \iiint_s b_i \alpha_j(b,y,k)\, f(b,y,k)\, db\, dy\, dk\right) = \mu_j'\left(l_j\theta_j - \theta_j l_j\right) = 0$$

The same applies for $\sum_{i=1}^{l_j}\left[\mu_j''\left(\mathrm{O}_j - k_i\right)\right] = 0$ and $\sum_{i=1}^{l_j}\left\{\mu_j'''\left[\sigma_j^2 - \left(k_i - \mathrm{O}_j\right)^2\right]\right\} = 0$

For mixed finance schools, we have:

$$p_{ij}^{*} = -E_{ij} + n_1 + 2n_2 l_j + \mu_j'\ (\theta_j - b_i) + \mu_j''(\mathrm{O}_j - k_i) + \mu_j'''\left[\sigma_j^2 - \left(k_i - \mathrm{O}_j\right)^2\right]$$

For mixed finance schools, the level of subsidy per student E_{ij} is determined by authorities and not by optimization. Schools can only choose the level of tuition to apply. Pricing is done according to the level of quality, the type and number of enrolled students and the level of subsidies. Note that theoretically it is possible to charge negative tuition fees (scholarships) for all students, if E_{ij} is positive and very high. However, this is unrealistic, since authorities would not subsidize schools to the extent that they could offer scholarships to all students. A necessary condition is: $\sum_{i=1}^{l_j}(-E_{ij} + n_1 + 2n_2 l_j) > 0 \Rightarrow \sum_{i=1}^{l_j} p_{ij}^{*} > 0$.

For public schools, we have:

$$E_{ij} = n_1 + 2n_2 l_j + \mu'_j\ (\theta_j - b_i) + \mu''_j(\mathrm{O}_j - k_i) + \mu'''_j\left[\sigma_j^2 - \left(k_i - \mathrm{O}_j\right)^2\right]$$

With $p_{ij} = 0$ for all i.

Given the quality of a public school, student types determine the amount of local subsidies needed to cover the marginal cost of admitting them. Local authorities determine the level of subsidies according to the type of enrolled students while maintaining an open enrolment policy. In other words, all students are admitted regardless of their type. Note that E_{ij} can be negative for some students; however, $\sum_{i=1}^{l_j} E_{ij} > 0$ for the same aforementioned reasons.

2. Students.

Since public, mixed finance and private schools coexist, students have a large set of choices. They are price takers and they maximize their utility through school choice given their characteristics and school tuition.

Utility maximization is done under the following budget constraint: $c_i = (1-t)y_i - p_{ij}$, with t the tax rate. Note that the level of individual utility in the chosen school should at least be equal to the maximum utility that can be obtained elsewhere. The price taking assumption is given through the following property:

$$U\left[(1-t)y_i - p_{ij}, q_j\right] \geq MaxU\left[(1-t)y_i - p_{ij'}, q_{j'}\right]$$

with j and j' two schools with $j \neq j'$.

Utility maximization implies that students have to choose between consumption and tuition in order to attain the level of school quality that would maximize their utility function. Utility maximization yields the following indirect utility:

$$W_i(b_i, y_i, k_i, E_{ij}, \theta_j, \mathrm{O}_j, \sigma_j^2) = MaxU\left[(1-t)y_i - p_{ij}, q_j\right]$$

Describing students' feasible choice sets.

Consider two schools j and j'; the first offers a combination of tuition and quality $\left(p_{ij}, q_j\right)$ to student i, and the second offers a combination $\left(p_{ij'}, q_{j'}\right)$. If $p_{ij} > p_{ij'}$ and $q_j \leq q_{j'}$, then only school j' is part of the feasible choice set of student i, since no student would agree to pay a higher price for lower or equal quality. Hence, for each student, the feasible set of choices exhibits a hierarchy of tuition and quality levels. For j and j' to be part of the choice set, $p_{ij} > p_{ij'}$ and $q_j > q_{j'}$ should be verified. Furthermore, private schools should have higher quality than public schools since no one would agree to pay tuition fees when it is possible to obtain a higher quality free of charge. Note that the choice of a particular school from the set depends on utility maximization.

3. The Implications of the Theory.

First, school quality is a by-product of individual characteristics. The pricing function transforms a continuum of student characteristics into a continuum of tuition fees levels. Tuition fees affect utility through their impact on consumption and determine the educational quality that can be afforded. In other words, students are not randomly stratified into schools and stratification is determined by their aptitudes, social status and income. This stratification mechanism is summarized through the education production function (EPF) $a_i = a(b_i, q_j)$. This EPF not only tells us that achievements depend on student and school characteristics but it also says that school characteristics are determined by student type. Further, the theory implies that students stratified into the same school bear some resemblance. However, it should be noted that the theoretical framework only gives us suggestions about what to expect; therefore full understanding of the implications of stratification can only be achieved through empirical analyses.

Secondly, from an econometric point of view, the theory implies that student and school characteristics are expected to be correlated and that causality goes from the first to the second. This has major implications for the estimation. Omitting some school characteristics may generate endogeneity bias, since these variables will be absorbed by the error term and the latter will be correlated with the included individual characteristics. Moreover, the resemblance between students attending the same school warrants the use of multilevel modelling.

Thirdly, the level of correlation between student and school characteristics differs across countries. Some have comprehensive schooling (Finland) and the correlation is expected to be weak, while others have early selection (Germany) and the correlation and bias are expected to be stronger. Thus, international comparisons are essential in understanding the functioning of the EPF. The comparative nature of the empirical analysis can also be seen as an implication of the theory.

Fourthly, inequalities in performance scores can no longer be considered as the mere impact of a student's social and economic background on his performance, since school characteristics are likely to be a source of inequality too. Hence, inequalities are channelled through students' own characteristics and through stratification-determined school characteristics.

From this discussion it is possible to identify a number of hypotheses that we would like to test in the empirical part of the analysis.

First, the omission of some school characteristics (peer effects or pure school variables) may generate an endogeneity bias, thus it is essential to test various specifications of the empirical model in order to identify the most robust one. Secondly, countries with differing education systems may be affected differently by this bias due to different levels of stratification and to variations in the strength of correlation between student and school variables. Thirdly, it would also be interesting to test whether social peer effects are linear in their means and whether achievements depend on the distribution of peers (the dispersion of social status). Finally, we would like to compare our findings with some results obtained using the same data and published in the OECD's PISA reports. The objective would be to affirm that the omission of key variables leads to bias and overestimation.

In the next section, a full multilevel approach is developed in order to estimate the effect of student and school characteristics on performance scores. We assess endogeneity problems, using a Hausman test, by contrasting three different models that omit certain variables. The models are estimated for three countries: Germany, the UK and Finland. Each one of these represents a particular schooling system. Germany is an early selection system where student background is expected to be highly correlated with school characteristics and peer effects. In contrast, Finland is known for its extreme comprehensiveness, where student and school characteristics are not expected to be strongly correlated. The UK has a system characterised by the liberal organisation of education, and is expected to be middle ranking in comparison with the other two countries.[9] The results in the different models are generated through an estimation procedure based on the Mundlak technique (1978) developed for panel data and adapted for multilevel regressions in this paper.

It is worth noting that in the empirical analysis, we are not reconstructing stratification as we did in the theoretical model. Instead, we are using an already stratified sample of students in order to assess the consequences of stratification on estimation procedures. In other words, we assume that the first step – the rise of stratification – has happened at an earlier stage and that we can only analyze its implications. The theory in this paper is used to provide the structure for our analysis and to identify the hypotheses that we want to explore.

Section Two: Multilevel Modelling and Endogeneity Problems.

1. Data, Variables and Countries.

In the empirical analysis, we are using PISA 2003 for three countries: Germany, the UK, and Finland. This data source was selected for several reasons. First, the data is collected using the same sampling procedure across all countries which is very convenient for international comparisons. Secondly, PISA is age based and the sampled students are aged between 15 and 3 months and 16 and two months. This coverage helps measuring the extent to which knowledge was acquired till the age of 15-16 independently of the structure of the education system. Thirdly, the major domain of assessment in PISA 2003 is mathematics which tends to be more universal than reading since it is not subject to cultural relativity. Finally, the PISA dataset provides a wide array of student and school variables that are needed for the analyses.

[9] Mostafa (2009 a and b) provide a comparative analysis between these three different countries.

In the EPF $a_i = a(b_i, q_j)$, student aptitudes b_i are replaced with various student characteristics and school quality q_j is replaced with several proxies of quality and peer effects.

The dependent variable is student performance scores in mathematics on the PISA 2003 standardized test. The independent variables are the following:

Student characteristics (reflecting student aptitudes b_i):

ESCS: Economic, social and cultural status.
COMPHOME: An indicator of computer facilities at home.
INTMAT: An indicator of interest in mathematics.
ANXMAT: An indicator of anxiety in mathematics.
DISCLIM: An indicator of the perception of discipline in a school.
ETR: A dummy variable taking the value of one if a student is a first generation immigrant or a non-native. Henceforth, this category is simply called “non-natives”. Note that ETR is not a measure of ethnicity.

Peer effects, school aggregates of individual characteristics (reflecting θ_j, O_j and σ_j^2 in the theory):

DESCS: School average ESCS, depicting economic, social and cultural peer effects.
VARESCS: The within-school dispersion of ESCS, reflecting nonlinearities in peer effects (the impact of the social diversity within a school).
DCOMPH: School average COMPHOME, depicting the possession of computer facilities peer effects.
DINTMAT: School average INTMAT, depicting peer effects resulting from a generalized interest and enjoyment of mathematics within a school.
DANXMAT: School average ANXMAT, depicting peer effects resulting from a generalized feeling of anxiety and helplessness in mathematics.
DDISCL: School average DISCLIM, depicting the impact of a generalized perception of discipline in a school.
DETR: The percentage of non-natives or first generation immigrants in a school.

Pure school characteristics (reflecting funding $\frac{R_j}{l_j}$ *and other school characteristics):*

Compweb: The proportion of computers connected to the web in a school.
Mactiv: The number of activities used to promote engagement with mathematics in a school.
Mstrel: An index measuring poor student teacher relations.
Tcshort: An index measuring principals' perception of potential factors hindering the recruitment of new teachers, and hence instruction.
Tcmorale: An index depicting principals' perception of teacher morale and commitment.
Teacbeha: An index depicting principals' perception of teacher-related factors hindering instruction or negatively affecting school climate.
Private: A dummy variable taking the value of one if a school is private (private dependent and independent schools are combined). Note that each of the selected countries, in fact, has only one of the two types of private schools. Thus, the two types have to be combined since estimation is not possible if the frequency of one of the types is close to zero. However, the interpretation of the results is made according to the predominant type.
Scmatedu: The quality of educational infrastructure in a school as perceived by the principal.
Academic: A dummy variable taking the value of one if a school selects its students according to their academic records.

The countries included in the analysis are: Germany, Finland and the UK. On the one hand, Germany is one of the few remaining countries in Western Europe to have selective schooling in the lower secondary phase, which starts around the age of 10. This early selection is the main source of social and ability stratification. In contrast, Finland is known for its extreme comprehensiveness with nine years of all-through schooling in the primary and lower secondary phases. Therefore, Finland is one of the least stratified education systems in the world. On the other hand, the UK has four distinct education systems in England, Wales, Scotland and Northern Ireland which vary in significant respects. Whilst the system in Scotland is fully comprehensive at the lower secondary stage, the other three systems retain selective grammar schools in varying degrees. The UK generally, is characterized by large territorial disparities, and an uneven spread of comprehensive schooling. Thus, it also has a stratified education system, even though stratification is more moderate than in Germany. For descriptive statistics on stratification see Mostafa (2009b).

Table 1. The number of sampled students and schools for each country is the following:

	Germany	Finland	UK
Number of students	4114	5728	9045
Number of schools	216	197	383

2. Endogeneity Problems in Multilevel Analyses.

The general model to be estimated is the following: $Y_{ij} = \beta_{0j} + \beta X_{ij} + \gamma_1 \overline{X}_{\bullet j} + \gamma_2 K_j + \varepsilon_{ij}$ with

$\beta_{0j} = c + V_j$

when β_{0j} is substituted out, the equation becomes.

$$Y_{ij} = c + \beta X_{ij} + \gamma_1 \overline{X}_{\bullet j} + \gamma_2 K_j + V_j + \varepsilon_{ij}$$

This model is a random intercept multilevel model. The intercept is divided into two elements: c is the overall intercept, which is constant for all schools and equal to the average of the intercepts β_{0j}, and a random part V_j, denoting school *j*'s departure from the overall intercept, which can also be seen as a unique effect of school *j* on the average intercept (Raudenbush and Bryk, 2002). V_j can be considered as comprising the unobserved school characteristics, and is assumed to have a zero mean and a variance of τ^2.

X_{ij} : is a vector of student characteristics (student i attending school j).

$\overline{X}_{\bullet j}$: is a vector of peer effects (school aggregates of student characteristics).

K_j : is a vector of pure school characteristics (e.g. funding, teacher morale…).

$V_j + \varepsilon_{ij}$: is the error term of the model. With $\varepsilon_{ij} \sim \mathrm{N}(0, \sigma^2)$.

In this multilevel model, the student level is called level one and the school level is called level two.

The assumptions on which this model relies are the following:

a) The independent variables at each level are not correlated with the random effects (error terms) on the other level, $\mathrm{cov}(X_{ij}, V_j) = 0$, $\mathrm{cov}(\overline{X}_{\bullet j}, \varepsilon_{ij}) = 0$ and $\mathrm{cov}(K_j, \varepsilon_{ij}) = 0$. In other terms, any unobservable student characteristics relegated to the error term should not be correlated with the observable school characteristics $\overline{X}_{\bullet j}$, and K_j. Similarly, any unobservable school characteristics relegated to the error terms should not be correlated with the observable student characteristics X_{ij}.

b) The level one independent variables are not correlated with level one error terms. $\text{cov}(X_{ij}, \varepsilon_{ij}) = 0$. In other words, any unobservable student characteristics relegated to the error term should not be correlated with the observable student characteristics X_{ij}.

c) The level two independent variables are not correlated with level two error term - $\text{cov}(\overline{X}_{\bullet j}, V_j) = 0$, and $\text{cov}(K_j, V_j) = 0$. In other words, any unobservable school characteristics relegated to the error term should not be correlated with the observable school characteristics $\overline{X}_{\bullet j}$, and K_j.

d) Each level one error term ε_{ij} is independent and normally distributed with a mean of 0 and a constant variance of σ^2. $\varepsilon_{ij} \sim \text{N}(0, \sigma^2)$.

e) Each level two random effect (error term) is normally distributed with a mean of 0 and a variance τ^2. $V_j \sim \text{N}(0, \tau^2)$. These error terms are independent among the level two schools.

f) The error terms at level 1 and 2 are independent. $\text{cov}(\varepsilon_{ij}, V_j) = 0$.

It should be noted that the homoscedasticity and normality assumptions (assumptions b, c, d, e, f) are tested using scatter plots of error terms and Q-Q plots respectively.

In this paper, the main concern is to test the cross-level assumption (assumption a), where the random effect on the intercept V_j is correlated with a level one independent variable X_{ij}. In this case, the assumption that $\text{cov}(X_{ij}, V_j) = 0$ is violated, and some unobservable school characteristics relegated to the error term, are correlated with the observable student characteristics X_{ij}. If this assumption is breached, the coefficient estimates might be biased. This problem is called the level 2 endogeneity problem (Grilli and Rampichini, 2006).

Other forms of endogeneity may arise when level 2 independent variables (school characteristics) are correlated with level 1 error terms ($\text{cov}(\overline{X}_{\bullet j}, \varepsilon_{ij}) \neq 0$ and $\text{cov}(K_j, \varepsilon_{ij}) \neq 0$). Or in other words, omitted student characteristics are correlated with the included school variables. In this paper we will only focus on the level 2 endogeneity problem.

In what follows, the endogeneity-robust Mundlak approach (1978) used for the estimation of panel data models is adapted for the estimation of multilevel models. It should be noted that multilevel data and panel data are very similar. In the former, we have a number of students nested within a number of schools, while in the latter we have a number of time periods nested within a number of individual units. Mundlak (1978) noted that a straightforward solution to solve endogeneity problems would be to include level 2 means $\overline{X}_{\bullet j}$ into the equation. Snijders and Berkhof (2006) also noted that the inclusion of such a variable permits the disentanglement of within- and between-clusters effects. In the case of PISA, this has an intuitive interpretation, since school averages represent different forms of peer effects within a school. One should note that in this type of models it is not possible to disentangle peer effects from selection effects, and probably there is no need to do so, since we are interested in knowing how the correlation between student and school characteristics generates endogeneity bias. For instance, In Germany students have been stratified at the age of 10; and since PISA assesses students at the age of 15, we can say that our "peer effects" represent the impact of 5 years of socialization as well as the initial effect of selection. Furthermore, one should not confuse our model with conventional models of peer effects where the objective is to identify the effect of peers at period t on the outcomes of $t+1$.

In order to assess endogeneity problems, four different specifications of the aforementioned model will be estimated. The first omits peer effects; the second omits pure school characteristics; the third considers the three vectors of variables, and finally the fourth introduces the within-school dispersion of economic social and cultural status (VARESCS) as an independent variable. Note that VARESCS is not included in models 1, 2 and 3; its inclusion is only intended to assess whether social peer effects are linear in their means. Further, models 1 and 2 are intended to show that the omission of key level 2 variables (school characteristics) might cause endogeneity problems. As mentioned earlier, all these models are estimated using a multilevel approach based on the Mundlak technique (1978). The approach consists of semi-demeaning the estimated equations. By doing so, it is possible to separate the within and between parts of the model and to estimate them separately. In what follows, we present the results. All algebraic details related to the estimation are relegated to the appendix.

Section Three: Estimation and Results.

In this section, the estimation of the four aforementioned models is carried out. Since the objective of the paper is to prove the existence of endogeneity when some school variables are omitted; we decided to limit the interpretation to the Hausman test and to the comparison between the different models. It should be noted that Mostafa (2009a) provides an interpretation of the regression coefficients in a cross-country comparative framework.

1. The Hausman Test.

The Hausman test is a specification test developed by Jerry Hausman. The test identifies the presence of level 2 endogeneity. The null hypothesis is that the random effects (on the intercept) are not correlated with any of the students' variables, $\mathrm{cov}(X_{ij}, V_j) = 0$. If the null hypothesis holds, then the estimates of the coefficients are both consistent and efficient. The Hausman test tests a fixed effects specification of the models against the random effects one. If the null hypothesis is rejected, we can conclude that the random effects model suffers from endogeneity and that the fixed effects specification is still better. Furthermore, after the transformation of the model according to the Mundlak approach, the estimates of the betas (the coefficients on level 1 'student' variables) are consistent regardless of whether the null hypothesis is valid.

Table 2. The results of the Hausman test.

	Germany	Finland	UK
Model 1	309.07	5.68	125.66
Model 2	0	0	0
Model 3	0	0	0
Model 4	0	0	0

As we can see, the Hausman test fails for the first model where the null hypothesis is rejected and holds for models 2, 3 and 4. A number of findings can be drawn:

a) Model one did not control for peer effects (school averages of the X_{ij} s). These school characteristics were relegated to the error term V_j and are correlated with student level variables. Thus, model 1 suffers from endogeneity and the null hypothesis on the

Hausman test is rejected. The fixed effects specification is preferred to the random effects one and the coefficients on the latter are biased.

b) The only country that is close to passing the Hausman test in model one is Finland. This indicates that the strength of the correlation between student characteristics and unobserved peer effects is low. This is perhaps due to the fact that schooling in Finland is very comprehensive and schools are homogenous. Hence, it is unlikely that student characteristics are highly correlated with those of the school (mainly peer characteristics). Finland is followed by the UK, then by Germany. The latter has the strongest value on the test which means that student characteristics and school peer effects are highly correlated. This is not surprising since early selection implies that student characteristics determine to a large extent those of their school. The UK is middle ranking.

c) The failure of the Hausman test is a strong indication that the specification in model 1 is not reliable. Even if the Mundlak transformation generates consistent estimates for the βs, the rest of the coefficients are still biased.

d) Models 2, 3 and 4 passed the Hausman test. The null hypothesis holds and there are no correlations left between students' variables and unobserved school characteristics relegated to the error term.

e) In model 2, pure school characteristics were omitted. However, the model still passed the Hausman test. This is an indication that endogeneity arises not from the correlation between student level variables and omitted pure school characteristics but from the correlation between the former and omitted peer effects (since model 1 did not pass the test).

f) Models 3 and 4 are the most complete; they controlled for peer effects and pure school characteristics and they passed the test. Therefore, these are the ones to be interpreted.

In conclusion, we can say that the Hausman test answers two questions that arose from the theory developed in section one. Firstly, when peer effects are omitted the model failed the Hausman test. But, when pure school characteristics are omitted, the model passed the test. Therefore, we can deduce that student characteristics are highly correlated with peer characteristics (b_i and θ_j in the theory). Thus, it is essential to control for peer effects in any estimation, since their omission generates endogeneity problems and biased results. Further, peer effects are more important than pure school characteristics, since the omission of the latter did not affect the viability of the model. Secondly, the Hausman test has proven that different countries have different levels of correlation between student variables and peer characteristics. In other words, the extent of the bias that may arise from level 2 endogeneity vary according to the context of each country. In fact, countries known for their comprehensiveness (Finland) are less affected by the bias than countries known for early selection (Germany) or for freedom of choice (the UK). This is due to the fact that comprehensiveness mitigates the impact of stratification by making schools more homogenous and choice less relevant. In contrast, early selection and liberalism in the organization of schooling exacerbate the impact of stratification by intensifying the role that student-related factors play in determining school characteristics.

2. The Regression Coefficients.

(***) stands for significance at the level of 1%, (**) for significance at the level of 5% and (*) for significance at the level of 10%.

Table 3. The regression coefficients on student level variables.

Variables	Germany		Finland		UK	
	Coefficient	Std Error	Coefficient	Std Error	Coefficient	Std Error
ESCS	11.61(***)	1.24	26.85(***)	1.24	23.26 (***)	0.99
COMHOME	1.28	1.22	-2.72(**)	1.15	8.20 (***)	0.99
INTMAT	4.69(***)	0.95	14.51(***)	1.05	-1.16	0.87
ANXMAT	-19.03(***)	0.90	-31.96(***)	1.08	-25.11(***)	0.87
DISCLIM	2.63(***)	0.82	1.41	1.00	12.33(***)	0.70
ETR	-27.92(***)	2.92	-63.49(***)	6.37	-5.18 (*)	2.90

Table 4. The regression coefficients on school level variables.

Variables	Germany											
	Model 1			Model 2			Model 3			Model 4		
	Coef.		SE	Coef.		SE	Coef.		SE	Coef.		SE
DESCS	.	.	.	66.15	***	1.63	60.38	***	1.81	61.16	***	1.80
DCOMPH	.	.	.	26.08	***	2.78	28.58	***	2.94	24.95	***	2.98
DINTMAT	.	.	.	-23.70	***	2.19	-24.31	***	2.08	-23.98	***	2.08
DANXMAT	.	.	.	-14.14	***	1.97	-16.54	***	1.93	-16.36	***	1.89
DDISCL	.	.	.	28.19	***	1.32	25.03	***	1.47	25.27	***	1.47
DETR	.	.	.	10.57	***	3.41	6.38	***	3.31	6.09	***	3.38

Compweb	38.95	***	2.98	.	.	.	10.97	***	1.79	8.85	***	1.78
Mactiv	29.80	***	1.53	.	.	.	1.26		0.80	0.52		0.81
Mstrel	-137.52	***	24.01	.	.	.	12.95		11.57	6.06		11.45
Tcshort	-17.86	***	1.19	.	.	.	-7.63	***	0.54	-7.54	***	0.53
Tcmorale	-0.04	***	0.00	.	.	.	0.03	***	0.00	0.04	***	0.00
Teacbeha	-12.64	***	1.41	.	.	.	-4.30	***	0.79	-4.15	***	0.76
Private	34.07	***	2.10	.	.	.	-11.36	***	2.05	-12.56	***	2.13
Scmatedu	0.00		0.01	.	.	.	-0.07	***	0.00	-0.07	***	0.00
Academic	25.91	***	2.02	.	.	.	7.29	***	0.98	7.61	***	0.96
Varescs	.	.	.	.	.	.	.	.	.	16.50	***	1.60
Intercept	459.18		3.44	.	.	.	466.33		2.15	461.20		2.15

Variables	**Finland**											
	Model 1			Model 2			Model 3			Model 4		
	Coef.		SE	Coef.		SE	Coef.		SE	Coef.		SE
DESCS	.	.	.	2.44	***	1.12	2.82	***	0.95	2.82	***	0.95
DCOMPH	.	.	.	3.96		1.41	1.76		1.43	1.60		1.43
DINTMAT	.	.	.	-2.84	***	1.47	-6.01	***	1.37	-6.10	***	1.37
DANXMAT	.	.	.	6.14	***	1.70	2.87	***	1.58	2.38	***	1.58
DDISCL	.	.	.	3.11	***	0.80	-0.10		0.82	-0.43		0.82
DETR	.	.	.	44.12	***	3.31	55.58	**	3.14	56.65	**	3.15
Compweb	13.00	***	4.02	.	.	.	4.83	**	2.44	5.20	**	2.43
Mactiv	2.74	*	1.46	.	.	.	1.44	***	0.47	1.56	***	0.47
Mstrel	-131.34	***	21.28	.	.	.	-124.76	***	8.02	-122.70	***	7.99
Tcshort	1.06		1.05	.	.	.	-0.35		0.37	-0.42		0.37
Tcmorale	5.47	***	0.96	.	.	.	1.78	***	0.32	1.76	***	0.31
Teacbeha	-2.45	***	0.78	.	.	.	-1.40	***	0.43	-1.47	***	0.43
Private	-15.00	***	3.70	.	.	.	-18.42	***	1.46	-18.37	***	1.46
Scmatedu	0.16		0.81	.	.	.	0.24		0.37	0.08		0.37
Academic	12.64	***	3.51	.	.	.	11.15	***	1.01	11.13	***	1.02
Varescs	.	.	.	.	.	.	.	.	.	-4.38	***	1.53
Intercept	532.41		5.17	.	.	.	528.74		2.44	530.51		2.61

Variables	**UK**											
	Model 1			Model 2			Model 3			Model 4		
	Coef.		SE	Coef.		SE	Coef.		SE	Coef.		SE
DESCS	.	.	.	51.40	***	0.84	42.82	***	1.09	41.64	***	1.06
DCOMPH	.	.	.	-15.78	***	1.40	-6.33		1.58	-3.78	***	1.53
DINTMAT	.	.	.	-8.88	***	1.12	-8.27	***	1.16	-8.62	***	1.16

DANXMAT	.	.	.	-13.83	***	1.19	-12.61	***	1.14	-10.00	***	1.14
DDISCL	.	.	.	11.39	***	0.69	9.38	***	0.72	10.08	***	0.70
DETR	.	.	.	-35.14	***	2.40	-35.83	***	2.36	-40.18	***	2.37
Compweb	18.46	***	3.09	.	.	.	17.70	***	1.17	18.63	***	1.13
Mactiv	-2.36	***	0.59	.	.	.	-2.97	***	0.24	-3.17	***	0.23
Mstrel	-282.90	***	26.28	.	.	.	-76.85	***	7.78	-92.12	***	7.43
Tcshort	-6.76	***	0.49	.	.	.	-3.79	***	0.31	-2.91	***	0.31
Tcmorale	0.51		0.71	.	.	.	-2.15	***	0.34	-1.61	***	0.33
Teacbeha	6.83	***	0.75	.	.	.	-1.28	***	0.34	-2.22	***	0.33
Private	71.68	***	2.33	.	.	.	14.53	***	1.46	22.75	***	1.55
Scmatedu	-0.96	*	0.56	.	.	.	2.56	***	0.27	2.34	***	0.25
Academic	11.37	***	1.82	.	.	.	4.51	***	0.97	3.13	***	0.99
Varescs	.	.	.	.	.	.	.	.	.	21.90	***	1.09
Intercept	505.68		3.93	.	.	.	492.80		1.38	478.27		1.53

As noted before, the coefficients on student level variables (betas) are identical for all models. However, the coefficients on school level variables (gammas) differ between models. Model 1 failed the Hausman test and thus, is considered to be unreliable. Model 2 passed the Hausman test but is still incomplete since pure school characteristics were omitted. Models 3 and 4 are the most complete. Model 3 considered the three vectors of independent variables: student characteristics, pure school characteristics, and peer effects, and model 4 added the within-school dispersion of student ESCS as an independent variable. These two models are considered to be the benchmark against which model 1 and 2 are compared. Note that, since the variance of ESCS was added as an independent variable in model 4, the coefficients of models 3 and 4 can be slightly different.

One should bear in mind that the Mundlak estimation procedure only solves endogeneity problems that arise from a correlation between included student variables and omitted school characteristics (e.g. cross-level endogeneity or level 2 endogeneity). Thus, even if model 2 passed the Hausman test, it may still suffer from endogeneity bias resulting from the correlation between included school peer effects and omitted pure school characteristics (e.g. same-level endogeneity).

From the regression results, it is possible to see that the coefficients for model 1 are systematically overestimated when compared with those of models 3 and 4, and some have different sign and significance levels. Thus, the results from model 1 are clearly inconsistent.

In contrast, the results from model 2 are relatively close to those of models 3 and 4. This finding confirms the results on the Hausman test. Model 1 suffers from level-2 endogeneity bias while model 2 is slightly inconsistent due to the omission of pure school variables.

The other major result concerns the regression coefficient on the within school dispersion of ESCS (Varescs) which measures within-school social diversity, and reflects nonlinearities in peer effects.

The linear in means models assume that a single student whose ESCS raises that of a school by one point has the same effect as several students whose combined ESCS raises that of a school by one point. As noted by Hoxby and Weingarth (2005), if peer effects were linear in means then neither economists nor policy makers would care much about them, because regardless of how peers are arranged, society would have the same average level of outcomes.[10]

The inclusion of Varescs in model 4 identifies the impact of social diversity on performance scores. Firstly, it provides an empirical proof for the nonlinearity assumption made in the theory. In fact, if Varescs has a statistically and economically significant effect, then it is possible to say that social peer effects are non-linear in means and that the distribution of ESCS within a school has an important impact on achievements. Secondly, it provides a full image on the functioning of peer effects under different education systems.

The first interesting finding concerns the statistical significance of the coefficient on Varescs. The coefficient is significant across all countries. This significance indicates that peer effects are non-linear in means and that the inclusion of ESCS and its average is not sufficient. This empirical finding confirms my theoretical assumptions.

The signs on the Varescs coefficient differ between countries. For Germany and the UK, the sign is positive, while it is negative for Finland. A positive sign indicates that higher levels of social diversity lead to higher levels of performance scores. The reverse is true when the sign is negative. The values of the coefficients also differ between countries. An increase of one unit in the dispersion of ESCS leads to an increase of 16 points in performance scores in

[10] The different models of Epple and Romano (1998, 2006) used a linear in means specification of peer effects.

Germany and 22 points in the UK. In contrast, an increase of one unit in Varescs leads to a decrease of 4 points in performance scores in Finland.

In order to understand these findings, the context of each country should be considered. For instance, German and British schools have wide between-school disparities in their social intakes. Some schools are socially diverse while others are elitist. Elitist schools in the UK are mostly private; while in Germany they are the general education ones (note that vocational schools are usually attended by lower social class students). The presence of important between-school disparities in social intakes is reflected through the statistical and economic significance of the coefficient on Varescs. The positive sign indicates that an increase of within-school social diversity has a favourable effect on performance scores. In other words, educational policy should be concerned with fighting social elitism and segregation. In contrast, the sign of the coefficient in Finland is negative, indicating that a further increase in social diversity leads to a decrease in performance scores. However, the impact is quantitatively weak (only a 4 point decrease for a one unit increase in Varescs). The negative sign can be explained by the fact that Finnish schools already have wide social intakes. Hence, any further increase in social diversity has a slightly negative impact on achievements (this is a form of the focus model of peer effect, where too much disparities in students' characteristics might lead to negative effects on their performance). One should note that in our theory we assumed that educational quality is increasing in the dispersion of social status and that educational performance is increasing in quality. Hence, Germany and the UK verify the theory (performance scores are increasing in Varescs) while Finland does not (performance scores are decreasing in Varescs). This is expected, since countries know for their differences cannot possibly verify the same theory.

3. Comparison with the Results from PISA's 'Learning for Tomorrow's World' Report.

In this subsection, the results obtained from model 3 are compared with those presented in the PISA 2003 report "Learning for tomorrow's world". Note that the PISA report did not assess the impact of all variables on performance scores. In fact, for some variables the analysis consisted of an interpretation of descriptive statistics without any estimation, and even when estimations were carried out, the models did not simultaneously control for a large array of variables. Hence, the results tend to diverge on a number of issues.

Table 5. The PISA 2003 report's results (relation between different variables and performance scores in mathematics).

Variables	Germany	Finland	UK
	Coefficient	Coefficient	Coefficient
ESCS	**17**	**33**	**31**
INTMAT	**10.2**	**30.5**	**13.6**
DESCS	**90**	**-2**	**58**
DDISCL	**18.6**	**10.4**	**24.7**
Teacbeha	-3.4	1.7	**20.3**
Scmatedu	11	0.2	**13**
Tcmorale	7.4	**5**	**13.4**
Private	**-66**	5	**-87**
Private after controlling for ESCS and DESCS	**14**	**16**	1

Numbers shown in bold stand for significant regression coefficients.

The analyses carried within the PISA report followed several separate axes. Chapter 3 assessed the distribution and impact on achievements of a number of student characteristics, including students' attitudes, students' learning strategies and students' beliefs about themselves. Chapter 4 assessed the impact of student and school ESCS on performance scores along with the impact of immigration background. Finally, chapter 5 assessed the distribution and impact of some school variables, including school climate, school policies and practices and school resources.

A major trait of the analyses undertaken in the PISA report is the use of a very limited number of variables in each regression. In fact, most regressions were bivariate, and the resulting coefficients can be described as correlation coefficients. This technique certainly suffers from the omitted variable bias. As mentioned before, when an important variable is omitted, and when this variable is correlated with one or more of the included independent variables, the model may suffer from heteroscedasticity and endogeneity. In other words, some of the included independent variables will be correlated with the error terms of the model.

If heteroscedasticity or endogeneity exist, the coefficients will be inconsistent and the resulting inference will be distorted. Further, we can reasonably expect that the regression coefficients will be overestimated. More intuitively, when important variables are not accounted for, the omission may artificially inflate the effect of included ones.

Moreover, in comparative studies, some researchers assume that if the estimation bias is identical across countries, then it is no longer a problem. However, in reality, there is no explicit empirical or theoretical evidence to support this claim. In fact, the Hausman test carried out earlier shows that the bias that may arise from the omission of some variables (e.g. peer effects) differs in magnitude depending on the characteristics of each education system. For instance, countries with limited social stratification, such as Finland, are weakly affected by endogeneity bias since the correlation between student and school variables is weak. The reverse is true for Germany and the UK. Hence, bias is unlikely to be identical across countries, and consequently the interpretation of the results is distorted.

A solution for bias problems is to control for all variables of interest, especially those that could be correlated with key independent variables included in the model. This has not been done in the PISA report. For instance, when student level ESCS was controlled for along with school average ESCS, other school characteristics that may be correlated with ESCS and DESCS were omitted. In our model, we controlled for a wide range of variables, including student and school characteristics as well as different forms of peer effects.

In bivariate regression analyses (used in the PISA report), the omission of all variables except one leads to an artificial inflation of its effect. In comparison to our findings, the PISA results are systematically overestimated. In addition to this, it is possible to see that some coefficients seem to be counter intuitive. For instance, DESCS has a negative effect on performance scores in Finland. Similarly, in Germany the coefficient on DESCS is very high: an increase of one unit in DESCS leads to an increase of 90 points in achievements. In other words, an increase of one unit in average ESCS causes previously low performing students to become high achievers. Similarly, the coefficients on INTMAT, DDISCL and Tcmorale in the PISA report are in general overestimated, even if the extent of the bias seems to vary between countries. Note that, for some countries, the significance and sign of these coefficients are also different from ours.

Another interesting finding concerns the coefficient on Scmatedu; the results are overestimated for all countries except for Finland, where they seem to be identical to those obtained in model 3. This is an indication that the absence of stratification, in the comprehensive Finnish system, reduces the level of correlations between Scmatedu and the variables relegated to the error term (e.g. other student and school characteristics), and thus reduces the estimation bias. As a consequence, the coefficient on Scmatedu for Finland tends to be consistent and very close to the one obtained in model 3.

The results for private schooling are interesting, too. When ‘Private’ was included separately, some of the coefficients were negative. However, when ESCS and DESCS were controlled for, all coefficients became positive even for Finland. In our model, the coefficient on ‘private’ has a negative sign for all countries except the UK. These results are more intuitive and more realistic, especially for Finland, where private schools are attended by students who fail to be integrated in the public system, and the UK, where most private schools are elitist and expected to provide better results.

Moreover, the negative sign on the coefficient for private schooling indicates that, when peer effects, school and student characteristics are controlled for, private schools do not have an advantage that stems from the fact that they are privately managed or funded. This finding has important consequences for educational policy. In fact, private schools do not have higher achievements because they are privately funded and managed, but because they have better inputs in terms of peer quality, funding and school climate. The only exception to this rule is the UK. When the different inputs are controlled for in the regression analysis, the apparent advantage of private schools turns into a negative effect (in Germany and Finland).

In conclusion, we can say that the techniques used by PISA (simple bivariate regressions) are insufficient to decompose inequalities and to indentify their sources. By contrast, the multilevel approach, developed in the remit of this paper, allows us to overcome endogeneity and omitted variable bias, and to provide better and more consistent results on which educational policy can be based.

Conclusion.

The research developed in this paper sheds light on the mechanisms of stratification and their implications for estimation strategies. It explores level 2 endogeneity problems in multilevel modelling of education production functions which arise from correlations between student characteristics and omitted school variables.

Our findings show that the omission of key student level variables leads to level 2 endogeneity bias. This bias can be dealt with through a transformation of the model according to the Mundlak approach (1978). Further, the bias resulting from omitted variables varies across countries according to the characteristics of each education system. Hence, it is no longer possible to claim that bias is identical across countries and that the results are affected in the same way. In fact, comprehensive education systems are less likely to be affected by level 2 endogeneity bias than stratified ones, since the correlation between student and school characteristics are weak. In addition to this, the paper confirms that social peer effects are non-linear in their means indicating that the distribution of peers within schools also affects their performance scores. Finally, the paper compares the results obtained through endogeneity robust multilevel regressions to those published by the OECD in the 2003 PISA report. The conclusion is that regression coefficients will always be overestimated unless the model simultaneously controls for student, school and peer characteristics. In this paper, our multilevel approach allows us to overcome endogeneity bias, and to provide better and more consistent results on which educational policy can rely.

Appendix

In this appendix, the multilevel estimation approach is developed. First, we start with a presentation of the Mundlak technique (1978) for panel data, and then we adapt it for multilevel estimation.

The Mundlak approach for panel data.

Before proceeding with the specification of the models, it is useful to start with the formulation developed in Maddala (1987). In his paper, the author reviewed some estimation issues that arise when the dependent variable is continuous in a panel data set. Note that a panel data has a number of cross-sectional units (individuals…) observed at several points of time. In other words, different time observations are nested within individual units. In the context of PISA, we have a similar structure. Students are nested within schools. Hence, it is possible to adapt the endogeneity robust estimation procedure developed by Mundlak (1978) in a panel data context to the PISA multilevel data. In a paper published in the *Journal of human resources* Maddala (1987) gave an interesting example on the estimation of production functions in firms; his model is the following: $Y_{it} = \beta X_{it} + \alpha_i + \varepsilon_{it}$, with (i = 1, 2,…, N) and (t = 1, 2,…, T).

i is a subscript denoting a firm and t is a subscript denoting a time period. Y_{it} is the output, X_{it} is the vector of inputs for firm i in period t. β is the regression coefficient, α_i is the firm specific unobserved inputs assumed to be constant over time. And finally, ε_{it} is an error term assumed to be normally distributed with mean 0 and constant variance, $\varepsilon_{it} \sim \mathrm{N}(0,\sigma^2)$.

The element α_i can be treated as a fixed effect, and hence one α_i should be estimated for each of the firms. In contrast, α_i can also be treated as a random variable (exactly like ε_{it}) as in Balestra and Nerlove (1966). When α_i is treated as random, the model is called a random effects model. The random effects model is similar to our multilevel specification.[11] It should be noted that in fixed effects models, level 2 endogeneity problems do not exist since α_i is treated as an intercept. In contrast, in random effects models, level 2 endogeneity problems might exist since α_i is treated as random and since $\mathrm{cov}(X_{it},\alpha_i)=0$ might not be verified.

The estimators of the regression coefficients are obtained in the following manner. First, $\overline{Y}_i = \frac{1}{T}\sum Y_{it}$ denotes the within-firm output average over time, and $\overline{Y} = \frac{1}{N}\sum \overline{Y}_i$ denotes the population average (global output average). It is possible to decompose the total sum of squares (TSS) as follows. $TSS = \sum\left(Y_{it}-\overline{Y}\right)^2 = \sum\left(Y_{it}-\overline{Y}_i\right)^2 + \sum\left(\overline{Y}_i-\overline{Y}\right)^2 = W_{YY} + B_{YY}$.

W_{YY} measures the within firm variations and B_{YY} measures the between firms variations. Using a similar decomposition of the variance and covariance, we obtain the estimates of β: $\hat{\beta} = W_{XX}^{-1}W_{XY}$. With $W_{XY} = \sum\left(X_{it}-\overline{X}_i\right)\left(Y_{it}-\overline{Y}_i\right)$ for the fixed effects model.

And $\hat{\beta}_{GLS} = (W_{XX}+\Theta B_{XX})^{-1}(W_{XY}-\Theta B_{XY})$ as the estimator of β in the random effects model.

With $\Theta = \dfrac{\sigma^2}{\sigma^2 + T\sigma_\alpha^2}$. (See Maddala 1971, p. 308-309).

Fuller and Battese (1973), noted that the GLS estimation of the betas used in Maddala (1971) is similar to the OLS estimation with the transformed data: $Y_{it} - \lambda\overline{Y}_i$ and $X_{it} - \lambda\overline{X}_j$ where $\lambda = 1-\sqrt{\Theta}$. This transformation is worth noting for three reasons:

[11] The subscript i is for the level two units (firms) and t is for the level one units (time observations). This should not be confused with the notation in our multilevel model, where i is the subscript for the level 1 units (students) and j is the one for the level 2 units (schools).

a) It has been used by Mundlak (1978) to solve the level 2 endogeneity problem.
b) This transformation rearranges the model in a form that is easily estimated with OLS.
c) When $\lambda = 1$, the model is identical to the fixed effects one.

Maddala gave two reasons for which the use of random effects models is more appropriate when the data shows some nesting features.

a) When the dataset contains a large number of observations, instead of estimating N values for α_i with fixed effects models, it is possible to estimate only the mean and variance with random effects models. This saves a lot of degrees of freedom (Maddala 1987, p. 309).[12]
b) The treatment of α_i as random allows us to measure firm-specific effects that we are ignorant about. In other terms, we are able to estimate the departures from the overall intercept for each firm. These departures reflect the effects of firm unobservable factors.
c) Another important reason is that if we want to make inferences about the actual set of cross-section units included in the dataset, we should treat α_i as fixed. However, if we want to make inferences about the population from which these cross section units came, α_i should be treated as random. Usually the latter is the case (Maddala 1987, p. 309).

In the example utilized by Maddala (1987), it is also possible to add time constant variables. These are similar to our student constant variables, which are school characteristics. His model becomes $Y_{it} = \beta X_{it} + \gamma Z_i + \alpha_i + \varepsilon_{it}$. With Z_i being a vector of time constant variables.

Mundlak (1978) studied the case where the α_i's are correlated with the X_{it}'s. This is similar to our level 2 endogeneity problem, where the X_{ij}'s might be correlated with the V_j's. The author argued that this endogeneity problem will be solved if α_i is assumed to depend on the

[12] In the case of PISA, if the β_{0j} is treated as fixed and is not decomposed into an overall intercept and a random part, then we have to estimate a β_0 for each school (this will cause the loss of j degrees of freedom). However, when β_{0j} is decomposed in the following manner $\beta_{0j} = c + V_j$, only the constant overall intercept c and the random parts are estimated, thus saving some degrees of freedom.

mean value of X_{ij}, such as $\alpha_i = \pi\overline{X}_i + w_i$. With w_i a random part that has similar properties to α_i. The equation becomes: $Y_{it} = \pi\overline{X}_i + \beta X_{it} + \gamma Z_i + w_i + \varepsilon_{it}$.

Using the Fuller and Battese transformation of the model, the estimator of the beta from the random effects model is obtained through OLS estimation of the following equation:

$$Y_{it} - \lambda\overline{Y}_i = \pi\left(\overline{X}_i - \lambda\overline{X}_i\right) + \beta\left(X_{it} - \lambda\overline{X}_i\right) + \gamma\left(Z_i - \lambda Z_i\right) + v_{it}$$

Then, the equation is developed and $\beta\overline{X}_i$ is added and subtracted from it:

$$Y_{it} - \lambda\overline{Y}_i = \pi\overline{X}_i - \lambda\pi\overline{X}_i + \beta X_{it} - \lambda\beta\overline{X}_i + \beta\overline{X}_i - \beta\overline{X}_i + \gamma\left(Z_i - \lambda Z_i\right) + v_{it}$$

Finally, the equation becomes:

$$Y_{it} - \lambda\overline{Y}_i = (\pi + \beta)\overline{X}_i - \lambda(\pi + \beta)\overline{X}_i + \beta\left(X_{it} - \overline{X}_i\right) + \gamma(1 - \lambda)Z_i + v_{it}$$

$$Y_{it} - \lambda\overline{Y}_i = (1 - \lambda)(\pi + \beta)\overline{X}_i + \beta\left(X_{it} - \overline{X}_i\right) + \gamma(1 - \lambda)Z_i + v_{it}$$

We denote $\delta = (1 - \lambda)(\pi + \beta)$ and $\lambda = 1 - \sqrt{\Theta}$ with $\Theta = \dfrac{\sigma^2}{\sigma^2 + T\sigma_\alpha^2}$

Since $(X_{it} - \overline{X}_i)$ and $\overline{X}_i$ are independent ($\overline{X}_i$ is orthogonal to $(X_{it} - \overline{X}_i)$), it is possible to estimate each of δ, β and γ independently through OLS. The estimate of beta is $\hat{\beta} = W_{XX}^{-1}W_{XY}$ ($\hat{\beta}$ is the within-group estimator). As it is possible to see, the estimate of the betas in the random effects model is identical to the aforementioned fixed effects estimator. The estimate of delta is $\hat{\delta} = \left(\sum \overline{X}_i\overline{X}_i\right)^{-1}\left(\sum \overline{Y}_i\overline{X}_i\right)(1 - \lambda)$ and the estimate of pi is $\hat{\pi} = \left(\sum \overline{X}_i\overline{X}_i\right)^{-1}\left(\sum \overline{Y}_i\overline{X}_i\right) - \hat{\beta}$. Similarly, $\hat{\gamma}$ can be obtained by regressing the time constant variable Z_i on the average of Y_{it} over time, which is $\overline{Y}_i$. These estimates are robust and efficient.

The models to be estimated.

Recall that the general model is the following: $Y_{ij} = \beta_{0j} + \beta X_{ij} + \gamma_1\overline{X}_{\bullet j} + \gamma_2 K_j + \varepsilon_{ij}$ with $\beta_{0j} = c + V_j$. All the estimations were carried out using LIMDEP and they were programmed step by step.

Model 1.

In the first specification of the model, school peer effects are dropped from the equation. The model becomes:

$Y_{ij} = c + \beta X_{ij} + \gamma_2 K_j + V_j + \varepsilon_{ij}$

The Hausman test is performed in order to compare the fixed effects model, containing only student characteristics, to the random effects model, containing both X_{ij} and K_j. Note that, in fixed effects models, V_j is not treaded as random and j of the β_{0j} are estimated. Hence, level 2 endogeneity problems would not arise. In contrast, in random effects models V_j is treated as random, and only an overall intercept c (the average intercept) and j school departures are estimated (the dispersion of these departures is denoted as the between school variance). In random effects models, level 2 endogeneity problems might arise (in model 1, the omitted school peer effects are absorbed by V_j and might be correlated with X_{ij}). Further, in the fixed effects models, the parameters γ_2 cannot be estimated since a school variable is constant for students attending the same school. Hence, the variables K_j cannot be included in its estimation.

The Hausman test.

The Hausman test is a specification test named after Jerry Hausman; it was developed in his article of 1978 and it tests for the presence of level 2 endogeneity. In other words, the null hypothesis is that the random effects V_j are not correlated with any of the observable students' variables. If the null hypothesis is correct, then the estimates of the coefficients are both consistent and efficient. It should also be said that after the transformation of the model, according to the Mundlak approach, the estimates of the betas (the within-group estimator) are consistent regardless of whether the null hypothesis is valid (see Maddala 1987, page 311).

Using the Fuller and Battese (1973) argument, the model is transformed in the following manner: $Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta(X_{ij} - \lambda \overline{X}_{\bullet j}) + \gamma_2(K_j - \lambda K_j) + w_{ij}$

The estimator of β is obtained through OLS estimation of this equation.

We add and subtract $\beta \overline{X}_{\bullet j}$ from the equation:

$Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta(X_{ij} - \lambda \overline{X}_{\bullet j}) + \gamma_2(K_j - \lambda K_j) + \beta \overline{X}_{\bullet j} - \beta \overline{X}_{\bullet j} + w_{ij}$

The equation is developed:

$$Y_{ij} - \lambda\overline{Y}_{\bullet j} = c + \beta X_{ij} - \lambda\beta\overline{X}_{\bullet j} + \gamma_2 K_j - \lambda\gamma_2 K_j + \beta\overline{X}_{\bullet j} - \beta\overline{X}_{\bullet j} + w_{ij}$$

And finally it becomes:

$$Y_{ij} - \lambda\overline{Y}_{\bullet j} = c + \beta(X_{ij} - \overline{X}_{\bullet j}) + (1-\lambda)\beta\overline{X}_{\bullet j} + (1-\lambda)\gamma_2 K_j + w_{ij}$$

$$Y_{ij} - \lambda\overline{Y}_{\bullet j} = c + \beta(X_{ij} - \overline{X}_{\bullet j}) + (1-\lambda)\beta\overline{X}_{\bullet j} + \delta_2 K_j + w_{ij}$$

With $\delta_2 = (1-\lambda)\gamma_2$ and $\lambda = \overline{\lambda}_j$, with $\lambda_j = 1 - \dfrac{\sigma_w}{\sqrt{\sigma_w^2 + n_j\sigma_b^2}}$.

σ_w^2: is the within school variance.

σ_b^2: is the between school variance.

n_j: is the number of observations in each school for an unbalanced data set (e.g. the number of students).

Remark:

The within- and between-school variances are the ones on ε_{ij} and V_j, respectively. The variance of ε_{ij} is σ^2 and the variance of V_j is τ_0^2. σ_w^2 and σ_b^2 are the estimates of σ^2 and τ_0^2, respectively.

Estimation:

We assume that $(X_{ij} - \overline{X}_{\bullet j})$ and K_j are independent, so their effects can be estimated separately.

1. We regress $(Y_{ij} - \overline{Y}_{\bullet j})$ on $(X_{ij} - \overline{X}_{\bullet j})$. $\hat{\beta}$ is obtained as well as the variance components.
2. We regress $\overline{Y}_{\bullet j}$ on $\overline{X}_{\bullet j}$, and K_j. $\hat{\gamma}_2$ is obtained.
3. We compute λ_j with $\lambda_j = 1 - \dfrac{\sigma_w}{\sqrt{\sigma_w^2 + n_j\sigma_b^2}}$

   Then λ is computed as the national average of λ_j; hence $\lambda = \overline{\lambda}_j$.
4. We multiply $\hat{\gamma}_2$ by $(1-\lambda)$ to obtain $\hat{\delta}_2$.

The regression can be fitted through OLS as suggested by Fuller and Battese (1973).

Model 2.

In the second specification of the model, pure school characteristics K_j are dropped from the education production function. The equation becomes: $Y_{ij} = c + \beta X_{ij} + \gamma_1 \overline{X}_{\bullet j} + V_j + \varepsilon_{ij}$

As in model 1, the Hausman test is performed in order to compare the fixed effects model, containing only student characteristics, to the random effects model, containing both X_{ij} and $\overline{X}_{\bullet j}$.

Using the Fuller and Battese (1973) argument the model is transformed in the following manner: $Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta(X_{ij} - \lambda \overline{X}_{\bullet j}) + \gamma_1(\overline{X}_{\bullet j} - \lambda \overline{X}_{\bullet j}) + w_{ij}$

The estimator of beta is obtained through OLS estimation of this equation.

We add and subtract $\beta \overline{X}_{\bullet j}$ from the equation:

$$Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta(X_{ij} - \lambda \overline{X}_{\bullet j}) + \gamma_1(\overline{X}_{\bullet j} - \lambda \overline{X}_{\bullet j}) + \beta \overline{X}_{\bullet j} - \beta \overline{X}_{\bullet j} + w_{ij}$$

We develop the equation:

$$Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta X_{ij} - \lambda \beta \overline{X}_{\bullet j} + \gamma_1 \overline{X}_{\bullet j} - \lambda \gamma_1 \overline{X}_{\bullet j} + \beta \overline{X}_{\bullet j} - \beta \overline{X}_{\bullet j} + w_{ij}$$

And finally it becomes:

$$Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta(X_{ij} - \overline{X}_{\bullet j}) + (\gamma_1 - \lambda \gamma_1 - \lambda \beta + \beta) \overline{X}_{\bullet j} + w_{ij}$$

$$Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta(X_{ij} - \overline{X}_{\bullet j}) + (1 - \lambda)(\gamma_1 + \beta) \overline{X}_{\bullet j} + w_{ij}$$

$$Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta(X_{ij} - \overline{X}_{\bullet j}) + \delta_1 \overline{X}_{\bullet j} + w_{ij}$$

With $\delta_1 = (1 - \lambda)(\gamma_1 + \beta)$, and $\lambda = \overline{\lambda}_j$ with $\lambda_j = 1 - \dfrac{\sigma_w}{\sqrt{\sigma_w^2 + n_j \sigma_b^2}}$,

σ_w^2: is the within-school variance.

σ_b^2: is the between-school variance.

n_j: is the number of observations in each school for an unbalanced data set (the number of students).

Estimation:

Since $(X_{ij} - \overline{X}_{\bullet j})$ and $\overline{X}_{\bullet j}$ are orthogonal, the effects of the two components can be estimated separately.

1. We regress $(Y_{ij} - \overline{Y}_{\bullet j})$ on $(X_{ij} - \overline{X}_{\bullet j})$. $\hat{\beta}$ is obtained as well as the variance components.
2. We regress $\overline{Y}_{\bullet j}$ on $\overline{X}_{\bullet j}$. $(\gamma_1 \hat{+} \beta)$ is obtained.
3. We compute λ_j with $\lambda_j = 1 - \dfrac{\sigma_w}{\sqrt{\sigma_w^2 + n_j \sigma_b^2}}$

   Then λ is computed as the national average of λ_j; hence $\lambda = \overline{\lambda}_j$.
4. We multiply $(\gamma_1 \hat{+} \beta)$ by $(1 - \lambda)$ to obtain $\hat{\delta}_1 = (1 - \lambda)(\gamma_1 \hat{+} \beta)$.
5. We have $\hat{\beta}$, $(\gamma_1 \hat{+} \beta)$, λ, and $\hat{\delta}_1$. It is possible to compute $\hat{\gamma}_1$.

The estimation is fitted through OLS, as suggested in Fuller and Battese (1973).

Model 3.

In the third specification, the full model is estimated.

$Y_{ij} = c + \beta X_{ij} + \gamma_1 \overline{X}_{\bullet j} + \gamma_2 K_j + V_j + \varepsilon_{ij}$

The Hausman test is performed in order to compare the fixed effects model, containing only student characteristics, to the random effects model, containing X_{ij}, $\overline{X}_{\bullet j}$ and K_j.

Using the Fuller and Battese (1973) argument the model is transformed in the following manner: $Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta(X_{ij} - \lambda \overline{X}_{\bullet j}) + \gamma_1(\overline{X}_{\bullet j} - \lambda \overline{X}_{\bullet j}) + \gamma_2(K_j - \lambda K_j) + w_{ij}$

The estimator of beta is obtained through OLS estimation of this equation.

We add and subtract $\beta \overline{X}_{\bullet j}$ from the equation:

$Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta(X_{ij} - \lambda \overline{X}_{\bullet j}) + \gamma_1(\overline{X}_{\bullet j} - \lambda \overline{X}_{\bullet j}) + \gamma_2(K_j - \lambda K_j) + \beta \overline{X}_{\bullet j} - \beta \overline{X}_{\bullet j} + w_{ij}$

We develop the equation:

$Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta X_{ij} - \lambda \beta \overline{X}_{\bullet j} + \gamma_1 \overline{X}_{\bullet j} - \lambda \gamma_1 \overline{X}_{\bullet j} + \gamma_2 K_j - \lambda \gamma_2 K_j + \beta \overline{X}_{\bullet j} - \beta \overline{X}_{\bullet j} + w_{ij}$

And finally it becomes:

$$Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta(X_{ij} - \overline{X}_{\bullet j}) + (\gamma_1 - \lambda\gamma_1 - \lambda\beta + \beta)\overline{X}_{\bullet j} + (1-\lambda)\gamma_2 K_j + w_{ij}$$

$$Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta(X_{ij} - \overline{X}_{\bullet j}) + (1-\lambda)(\gamma_1 + \beta)\overline{X}_{\bullet j} + (1-\lambda)\gamma_2 K_j + w_{ij}$$

$$Y_{ij} - \lambda \overline{Y}_{\bullet j} = c + \beta(X_{ij} - \overline{X}_{\bullet j}) + \delta_1 \overline{X}_{\bullet j} + \delta_2 K_j + w_{ij}$$

With $\delta_1 = (1-\lambda)(\gamma_1 + \beta)$, $\delta_2 = (1-\lambda)\gamma_2$,

and $\lambda = \overline{\lambda_j}$ with $\lambda_j = 1 - \dfrac{\sigma_w}{\sqrt{\sigma_w^2 + n_j \sigma_b^2}}$.

σ_w^2: is the within-school variance.

σ_b^2: is the between-school variance.

n_j: is the number of observations in each school for an unbalanced data set (the number of students).

Estimation:

We assume that $(X_{ij} - \overline{X}_{\bullet j})$ and K_j are independent and since $(X_{ij} - \overline{X}_{\bullet j})$ and $\overline{X}_{\bullet j}$ are orthogonal, the effects of the different components can be estimated separately.

1. We regress $(Y_{ij} - \overline{Y}_{\bullet j})$ on $(X_{ij} - \overline{X}_{\bullet j})$. $\hat{\beta}$ is obtained as well as the variance components.
2. We regress $\overline{Y}_{\bullet j}$ on $\overline{X}_{\bullet j}$ and K_j. $(\widehat{\gamma_1 + \beta})$ and $\hat{\gamma}_2$ are obtained.
3. We compute λ_j with $\lambda_j = 1 - \dfrac{\sigma_w}{\sqrt{\sigma_w^2 + n_j \sigma_b^2}}$

   Then λ is computed as the national average of λ_j; hence $\lambda = \overline{\lambda_j}$.
4. We multiply $(\widehat{\gamma_1 + \beta})$ and $\hat{\gamma}_2$ by $(1-\lambda)$ to obtain: $\hat{\delta}_1 = (1-\lambda)(\widehat{\gamma_1 + \beta})$ and $\hat{\delta}_2 = (1-\lambda)\hat{\gamma}_2$.
5. We have $\hat{\beta}$, $(\widehat{\gamma_1 + \beta})$, λ and $\hat{\delta}_1$. It is possible to compute $\hat{\gamma}_1$.

Important remarks:

b) In models 1 and 2, different sets of school variables have been dropped, while in model 3 the full equation was estimated. The content of these sets of variables, (X_{ij} , $\overline{X}_{\bullet j}$ and K_j) is the same across the three models. The selection of each variable within each set has been done previously. We started with very simple models and added one variable after the other, while dropping non-significant ones. The result was three sets of variables controlling for student characteristics, peer effects and pure school characteristics. However, these steps are not included within the paper because of space constraints and since they represent the preliminary work in the development of the models.

c) Notice that step one of the estimation procedure enables the estimation of the betas "within effects," which measure the strength of the relation between student level variables and performance scores. Step two enables the measurement of the "between effects," which are the effects of school characteristics on school average performance. School characteristics include averages of student variables and pure school variables.

d) The transformation of the model according to the Mundlak approach is needed in order to ensure that the within and between parts of the model are independent and can be estimated separately.

Model 4.

Model 4 is identical to model 3, except that a new school level variable, which is the within-school dispersion of ESCS (VARESCS), is added. VARESCS can be considered as a complement to average ESCS (social peer effects). If this dispersion has a significant effect, it would be possible to say that ESCS peer effects are non-linear. This model provides an answer to our theoretical investigations, and affirms whether social diversity has a positive effect on performance scores. The model is estimated using the same aforementioned Mundlak approach.

References

Anderson, P. (1979) *Lineages of the Absolutist State*. Verso, London.

Arnott, R. and Rowse, J. (1987) ‘Peer Group Effects and Educational Attainment,’ *Journal of Public Economics*, 32, pp. 287–306.

Balestra, P. and Nerlove, M. (1966) ‘Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas,’ *Econometrica* 34, 3.

Barzel, Y. (1973) ‘Private Schools and Public School Finance,’ *The Journal of Political Economy*. 81, 1, pp. 174-186.

Boli, J. (1989) *New Citizens for a New Society: the Institutional Origins of Mass Schooling in Sweden*. Pergamon Press, Oxford.

De Bartolome, C. (1990) ‘Equilibrium and Inefficiency in a Community Model with Peer Group Effects,’ *Journal of Political Economy*, 98, pp. 110-133.

Epple D, Filimon R, and Romer T. 1993. ‘Existence of Voting and Housing Equilibrium in a System of Communities with Property Taxes,’ *Regional Science and Urban Economics*, 23, pp. 585-610.

Epple, D. and Platt, G. (1998) ‘Equilibrium and Local Redistribution in an Urban Economy when Households Differ in both Preferences and Income,’ *Journal of Urban Economics,* 43, pp. 23-51.

Epple, D. and Romano, R. (1998) ‘Competition between Public and Private Schools, Vouchers, and Peer Group Effects,’ *The American Economic Review*, 88, 1, pp. 33-62.

Epple, D. and Romano, R. (2006) ‘Admission, Tuition, and Financial Aid Policies in the Market for Higher Education,’ *Econometrica*, 74, 4, pp. 885-928.

Fernandez, R. and Rogerson, R. (1996) ‘Income Distribution, Communities, and the Quality of Public Education,’ *The Quarterly Journal of Economics*. 111, 1, pp. 135-164.

Ferris, J. and West, E. (2004) ‘Economies of Scale, School Violence and the Optimal Size of Schools,’ *Applied Economics*, 36, pp. 1677-1684.

Fielding, A. (2004) ‘The Role of the Hausman Test and whether Higher Level Effects should be treated as Random or Fixed,’ *Multilevel Modeling Newsletter*, 16, 2, pp. 3–9.

Fuller, W. and Battese, G. (1973) ‘Transformations for Estimation of Linear Models With Nested Error Structure,’ *Journal of the American Statistical Association*, 68, pp. 626-32.

Grilli, L. and Rampichini, C. (2006) *Model Building Issues in Multilevel Linear Models with Endogenous Covariates.* Working Paper, Dipartimento di Statistica, Università di Firenze, Florence.

Hanushek, E. and Welch, F. (2006) *Handbook of the Economics of Education*. Elsevier, London.

Hausman, J. (1978) ‘Specification Tests in Econometrics,’ *Econometrica*, 46, 6, pp. 1251-1271.

Hoxby, C. and Weingarth, G. (2005) *Taking Race out of the Equation: School Reassignment and the Structure of Peer Effect.* Working paper, Harvard University, Cambridge.

Kim, J. and Frees, E. (2006) ‘Omitted variables in multilevel models’ *Psychometrika*, in press.

Maddala, G. (1987) ‘Limited Dependent Variable Models Using Panel Data,’ *The Journal of Human Resources*, 22, 3 pp. 307-338.

Maddala, G. (1971) ‘The Use of Variance Components Models in Pooling Cross-Section and Time-Series Data,’ *Econometrica*, 39, 2, pp. 341-58.

Marsden, D. and Ryan, P. (1995) ‘Work, Labour Markets and Vocational Preparation: Anglo-German Comparisons of Training in Intermediate Skills’ in L. Bash and A. Green (eds) *Youth, Education and Work*. Kogan, London, pp. pp. 67-79.

Mostafa, T. (2009 a) *The Anatomy of Inequalities in Attainments: an International Investigation on Stratification and Choice*. PhD Thesis, Université de la Méditerranée.

Mostafa, T. (2009 b) *The Anatomy of Inequalities in Educational Achievements: An International Investigation of the Effects of Stratification.* LLAKES Research Paper No 3. Institute of Education, London.

Mundlak, Y. (1978) ‘On the Pooling of Time Series and Cross-Sectional Data,’ *Econometrica*, 46, pp. 69–86.

Nechyba, T. (1997) ‘Existence of Equilibrium and Stratification in Local and Hierarchical Tiebout Economies with Property Taxes and Voting,’ *Economic Theory*, 10, 2, pp. 277-304.

Nechyba, T. (2003) ‘Centralization, Fiscal Federalism, and Private School Attendance,’ *International Economic Review,* 44, 1, pp. 179-204.

OECD (2003) *Learning for Tomorrow’s World*. OECD, Paris.

OECD (2003) *Education at a glance.* OECD, Paris.

Raudenbusch, S. and Bryk, A. (2002) *Hierarchical Linear Models* (Second edition). Sage Publications, London.

Rose-Ackerman, S. (1979) ‘Market Models of Local Government: Exit, Voting, and the Land Market,’ *Journal of Urban Economics,* 6, pp. 319-337.

Skrondal, A. and Rabe-Hesketh, S. (2004) *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models.* Chapman & Hall/ CRC Press, Boca Raton, Florida.

Snijders, T. and Berkhof, J. (2008) ‘Diagnostic Checks for Multilevel Models.’ In J. Leeuw and E. Meijer (eds) *Handbook of Multilevel Analysis*. Springer, New York.

Stiglitz, J. (1974) ‘The Demand for Education in Public and Private School Systems,’ *Journal of Public Economics,* 3, 4, pp. 349-385.

Westhoff, F. (1977) ‘Existence of Equilibrium in Economies with a Local Public Good,’ *Journal of Economic Theory,* 14, pp. 84-112.

Wooldridge, J. (2002) *Econometric analysis of cross section and panel data*. The MIT Press, Cambridge.

For more information, please contact
llakescentre@ioe.ac.uk
LLAKES Centre
Institute of Education
20 Bedford Way
WC1H 0AL
London
UK