

The Rise of Endogeneity in Multilevel Models: a Theoretical Assessment of the Role of Stratification.

Tarek Mostafa¹

Abstract

This paper studies the role of stratification in the rise of endogeneity bias in multilevel models. The theory is illustrated using educational stratification and its implication on the estimation of multilevel education production functions. Educational stratification results from the functioning of the education market; it transforms a continuum of student characteristics into a continuum of tuition fees. These fees enter students' utility functions and determine the school they attend and its quality. In other words, student characteristics are the major determinants of school quality and the two are correlated. In this paper, I analyse how these correlations arise and what are their implications for multilevel estimation of education production functions. The major problem posed by such correlations is cross-level endogeneity bias. The theory developed in this paper can be extended to any economic phenomenon that exhibits stratification or nesting of smaller units within larger units (e.g. employees within firms, residents within neighbourhoods, etc).

JEL Classification: I20, C02, D70.

Keywords: Stratification, Endogeneity Bias, Multilevel Models.

¹ Tarek Mostafa: Researcher at the Institute of Education - University of London. Centre for Longitudinal Studies (CLS - ESRC). 20 Bedford Way, London WC1H 0AL. England, United Kingdom. Office phone: +44 (0)2076126881. Email: T.Mostafa@ioe.ac.uk.

Introduction.

The efficiency of multilevel models is threatened by the existence of endogeneity bias resulting from omitted variables. Endogeneity bias arises when unobserved or omitted variables, that affect the outcome, are correlated with one or more of the included independent variables. The presence of such a bias in the estimation of panel data models has been recognized for a long time, and econometricians such as Mundlak (1978) and Maddala (1987) provided different solutions to tackle it. However, the study of endogeneity was not extended to other similar structures such as multilevel models. Note that panel models and multilevel models are similar since, in the former, time observations are nested within the same individual and in the latter individuals are nested within larger units (e.g. students nested within schools, residents within neighbourhoods, etc).

Multilevel models (also known as hierarchical models, nested models, or mixed models) are concerned with phenomena where observations are nested within larger units. They have been used in a variety of contexts such as longitudinal models (individuals observed at different points of time), organizational research (individuals nested within organizations), geographical research (residents within neighbourhoods or regions), educational research (students nested within schools), and in biometric, environmental and ecological studies.

In this paper, I am constructing a theoretical model that assesses the implications of stratification on the rise of endogeneity bias in the estimation of multilevel models. The theory is illustrated using education production functions where students are nested within schools. I chose this educational example because it is one of the most well known areas of application of multilevel modelling and because multilevel data is becoming more widely available (e.g. OECD's PISA, TIMSS and PIRLS, etc). In this context, endogeneity is caused by educational stratification which ties school characteristics to those of the student. In other words, students from a lower income and social strata are likely to be stratified into low-quality schools which are populated by students of similar type. This correlation between student and school characteristics leads to endogeneity if some of the school-level or student-level variables are omitted while being correlated with the included independent variables from the other level. Hanchane and Mostafa (2012). Note that even though the theory is illustrated using an educational example, it can be extended to any economic phenomenon that exhibits stratification or nesting of smaller units within larger units.

The theoretical model is built upon the work of Epple and Romano (1998). An economy populated with individuals differentiated by income, ability 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, ability 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 fees. 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.

The theoretical literature on stratification emerged in the 1970s with the founding articles of Barzel (1973) and Stiglitz (1974), and the major developments took place in the 1990s, when two distinct bodies of literature emerged. The first studied spatial stratification between jurisdictions and neighbourhoods. It includes 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). Further, the recent econometric literature studying endogeneity problems includes Wooldridge (2002), Skrondal and Rabe-Hesketh (2004), Fielding (2004), Snijders and Berkhof (2006), Kim and Frees (2006), Grilli and Rampichini (2006), Hanchane and Mostafa (2012).

The Theoretical Model.

In this model, we consider an economy populated with a continuum of households differentiated by income, social status and student ability. 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. 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 ability 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, ability, 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 $S = (0, b_{\max}) \times (0, y_{\max}) \times (0, k_{\max})$.

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’ abilities and on the quality of their school. In other words, the access to higher quality schools is translated into higher achievements. Note that this achievements function can be empirically modelled as a multilevel equation with two levels: students and schools.

School quality is determined by expenditure per pupil, ability 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 and 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 reflecting economic reality would be educational desegregation in the US. Authorities may

impose ethnic diversity into white majority schools; 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

Ability peer effects are defined to be school average abilities, 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 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:

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

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

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

² In this model, communities and geographical stratification are not considered for simplicity. Mostafa (2009a) 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.

⁴ This assumption was used in Epple and Romano 1998 (p.38).

large number of schools catering for different types of students.⁵ Furthermore, schools have perfect information on students' income, ability 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 fees paid by student i in school j .

G_j : Other earnings.

The sum of subsidies for a particular school is equal to $\sum E_{ij} = E_j$. At equilibrium, the sum of government subsidies to all schools is equal to tax revenues in the economy $\sum E_j = tY$ with Y being aggregate income (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 unrealistic. 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 model, I am assuming the exogeneity of the tax rate.⁶

School quality is given by:

$$q_j = q_j \left[\frac{R_j}{l_j}, \theta_j, O_j, \sigma_j^2 \right] \quad (2)$$

with: $\lim_{R_j \rightarrow 0} q_j = 0$, $\lim_{\theta_j \rightarrow 0} q_j = 0$, $\lim_{O_j \rightarrow 0} q_j = 0$, $\lim_{\sigma_j^2 \rightarrow 0} q_j = 0$.

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 .

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

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 $\frac{R_j}{l_j}$.

Ability peer effects are given by average abilities in a school:

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

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

$$O_j = \frac{1}{l_j} \iiint_s k_i \alpha_j(b, y, k) f(b, y, k) db dy dk \quad (4)$$

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

$$\sigma_j^2 = \frac{1}{l_j} \iiint_s (k_i - O_j)^2 \alpha_j(b, y, k) f(b, y, k) db dy dk \quad (5)$$

Ability 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, $R_j = Co(l_j)$.

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 abilities, 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

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

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:

$$l_j = \iiint_s \alpha_j(b, y, k) f(b, y, k) dbdydk \quad (1)$$

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

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

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

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

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

$$2- O_j \iiint_s \alpha_j(b, y, k) f(b, y, k) dbdydk - \iiint_s k_i \alpha_j(b, y, k) f(b, y, k) dbdydk = 0$$

$$3- \sigma_j^2 \iiint_s \alpha_j(b, y, k) f(b, y, k) dbdydk - \iiint_s (k_i - O_j)^2 \alpha_j(b, y, k) f(b, y, k) dbdydk = 0$$

The Lagrangian function is then written in the following form:

$$\begin{aligned} \Phi = & R_j - n_1 l_j - n_2 l_j^2 - F - \mu'_j [\theta_j \iiint_s \alpha_j(b, y, k) f(b, y, k) dbdydk - \iiint_s b_i \alpha_j(b, y, k) f(b, y, k) dbdydk] \\ & - \mu''_j [O_j \iiint_s \alpha_j(b, y, k) f(b, y, k) dbdydk - \iiint_s k_i \alpha_j(b, y, k) f(b, y, k) dbdydk] \\ & - \mu'''_j \left[\sigma_j^2 \iiint_s \alpha_j(b, y, k) f(b, y, k) dbdydk - \iiint_s (k_i - O_j)^2 \alpha_j(b, y, k) f(b, y, k) dbdydk \right] \end{aligned}$$

with $\mu'_j, \mu''_j,$ and μ'''_j the Lagrangian multipliers.

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

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

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

$$MC = R_j^* = n_1 + 2n_2l_j + \mu'_j (\theta_j - b_i) + \mu''_j(O_j - k_i) + \mu'''_j[\sigma_j^2 - (k_i - O_j)^2] \quad (6)$$

Note that R_j^* represents the marginal cost (MC) of admitting a student of type (b, y, k) .

Furthermore, $R_j^* = E_{ij} + p_{ij}^*$ after the partial differentiation of

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

The Lagrangian multipliers are:

$$\mu'_j = \frac{1}{l_j} \iiint_s \frac{\partial R^*}{\partial \theta_j} \alpha_j(b, y, k) f(b, y, k) db dy dk$$

$$\mu''_j = \frac{1}{l_j} \iiint_s \frac{\partial R^*}{\partial O_j} \alpha_j(b, y, k) f(b, y, k) db dy dk$$

$$\mu'''_j = \frac{1}{l_j} \iiint_s \frac{\partial R^*}{\partial \sigma_j^2} \alpha_j(b, y, k) f(b, y, k) db dy dk$$

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 abilities 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 . The higher the multipliers are the stronger stratification is, since the higher the impact of individual characteristics on the marginal cost is. Note that if the multipliers are equal to zero, then the marginal cost of admitting a student will only depend on the number of students attending a school. In other words, student characteristics will not have any effect on the marginal cost and stratification according to ability and social status will not exist.

In equation (6), R_j^* represents the resources needed to cover the marginal cost of admitting a student. The first term $n_1 + 2n_2l_j$ is the part resulting from the education production cost (reflecting the impact of the number of students in a school). 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 ability 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 ability on school average abilities. Students with above average abilities $\theta_j < b_i$ 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 .

$$\lim_{b \rightarrow +\infty} \mu'_j(\theta_j - b_i) = -\infty.$$

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

The fourth term, $\mu'''_j[\sigma_j^2 - (k_i - O_j)^2]$, represents the cost of being too close to the average of social status. Those who are far from the mean (i.e. they 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 , $(k_i - O_j)^2$ is positive and high. If it is higher than σ_j^2 , then the term $\mu'''_j[\sigma_j^2 - (k_i - O_j)^2]$ 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 $k_i = O_j$.

$$\lim_{k \rightarrow \pm\infty} \mu'''_j[\sigma_j^2 - (k_i - O_j)^2] = -\infty.$$

Note that R_j^* might be negative, depending on the level of government subsidies and the position of individual ability and social status relative to the means. Equation (6) represents a compensation scheme; low ability students subsidize higher ability 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 (i.e. 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). Note that the pricing function is not necessarily meritocratic. In fact, the second term is meritocratic since it rewards higher ability students, but the third term is not because it rewards students with higher social status.

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:

For private schools, we have:

$$p_{ij}^* = n_1 + 2n_2l_j + \mu'_j(\theta_j - b_i) + \mu''_j(O_j - k_i) + \mu'''_j[\sigma_j^2 - (k_i - O_j)^2]$$

With $E_{ij} = 0$.⁸

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 fee. This can be seen through the sum of p_{ij}^* which is always positive.

$$\sum_{i=1}^{l_j} p_{ij}^* = \sum_{i=1}^{l_j} (n_1 + 2n_2l_j) + \sum_{i=1}^{l_j} [\mu'_j(\theta_j - b_i)] + \sum_{i=1}^{l_j} [\mu''_j(O_j - k_i)] + \sum_{i=1}^{l_j} \{\mu'''_j[\sigma_j^2 - (k_i - O_j)^2]\} > 0 \text{ since}$$

$$\sum_{i=1}^{l_j} (n_1 + 2n_2l_j) > 0 \text{ and}$$

$$\sum_{i=1}^{l_j} [\mu'_j(\theta_j - b_i)] + \sum_{i=1}^{l_j} [\mu''_j(O_j - k_i)] + \sum_{i=1}^{l_j} \{\mu'''_j[\sigma_j^2 - (k_i - O_j)^2]\} = 0$$

Mathematical details :

$$\begin{aligned} \sum_{i=1}^{l_j} [\mu'_j(\theta_j - b_i)] &= \mu'_j(\theta_j - b_1) + \dots + \mu'_j(\theta_j - b_{l_j}) = \mu'_j[(\theta_j - b_1) + \dots + (\theta_j - b_{l_j})] \\ &= \mu'_j \left(l_j \theta_j - \iiint_s b_i \alpha_j(b, y, k) f(b, y, k) db dy dk \right) = \mu'_j(l_j \theta_j - \theta_j l_j) = 0 \end{aligned}$$

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

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

For mixed finance schools, we have:

$$p_{ij}^* = -E_{ij} + n_1 + 2n_2l_j + \mu_j'(\theta_j - b_i) + \mu_j''(O_j - k_i) + \mu_j'''[\sigma_j^2 - (k_i - O_j)^2]$$

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: $\sum_{i=1}^{l_j} (-E_{ij} + n_1 + 2n_2l_j) > 0 \Rightarrow \sum_{i=1}^{l_j} p_{ij}^* > 0.$

For public schools, we have:

$$E_{ij} = n_1 + 2n_2l_j + \mu_j'(\theta_j - b_i) + \mu_j''(O_j - k_i) + \mu_j'''[\sigma_j^2 - (k_i - O_j)^2]$$

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

Given the quality of a public school, student types determine the amount of public funds needed to cover the marginal cost of admitting them. Public authorities determine the level of funds according to the type of enrolled students while maintaining an open enrolment policy.

Note that E_{ij} can be negative for some students; however, $\sum_{i=1}^{l_j} E_{ij} > 0$ for the same aforementioned reasons.

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 fees.

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:

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

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:

$$W_i(b_i, y_i, k_i, E_{ij}, \theta_j, O_j, \sigma_j^2) = \text{Max}U[(1-t)y_i - p_{ij}, q_j]$$

This utility function indicates that student and school characteristics enter the function through the combination of price (tuition fees) and school quality. Furthermore, even though income is not part of the pricing function (equation 6), it affects utility maximisation through the budget constraint which determines which school is affordable.

Describing students' feasible choice sets:

Consider two schools j and j' ; the first offers a combination of tuition fees and quality (p_{ij}, q_j) to student i , and the second offers a combination $(p_{ij'}, q_{j'})$. 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 fees and quality levels. For j and j' to be part of the choice set, $p_{ij} > p_{ij'}$ and $q_j > q_{j'}$ must hold. 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 aforementioned utility maximization.

The Implications of the Theory.

First, school profit maximization transforms a continuum of student characteristics into a continuum of tuition fees through the pricing function (i.e. equation 6). Tuition fees enter the utility function through their impact on consumption and determine the educational quality that can be afforded (through student utility maximization). Hence, students are not randomly stratified into schools and stratification is determined by their ability, social status and income. This stratification mechanism is summarized through the pricing function.

$$MC = R_j^* = n_1 + 2n_2l_j + \mu'_j (\theta_j - b_i) + \mu''_j (O_j - k_i) + \mu'''_j [\sigma_j^2 - (k_i - O_j)^2]$$

This function tells us that the marginal cost of a student and hence the tuition fees he has to pay are determined by his type. Given school quality and tuition fees, utility maximization determines which school a student will attend. Therefore, one can conclude that students are sorted into schools according to their type and to schools' strategic behaviour on the education market. As mentioned before, the higher the Lagrangian multipliers are the higher the level of stratification is. In other words, student characteristics will have a strong bearing on students' marginal costs.

Secondly, the education production function (EPF) $a_i = a(b_i, q_j)$ not only tells us that achievements depend on student ability and school quality but it also says that school quality is determined by student type (since student type determines the school a student attends as explained before). Hence b_i and q_j are correlated. And if ability is assumed to be correlated with income and social status, then the correlation between b_i and q_j might even be stronger. Further, if the level of stratification is high (high values on the Lagrangian multipliers), then students stratified into the same school will bear some resemblance. This happens because students of the same type are likely to have similar choice sets and utility maximization is likely to lead them to the same schools. In fact, students with identical abilities, income and social status will have identical choice sets and identical tuition fees, (school quality held constant).

Thirdly, from an econometric point of view, the theory implies that student and school characteristics are correlated and that causality goes from the first to the second. In fact, student endowments in ability, income and social status determine the quality of their school through utility maximization. This has major implications for any econometric estimation. Omitting some student or 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 (independent variables). Moreover, the resemblance between students attending the same school warrants the use of multilevel modelling.

Fourthly, this model can be extended to any multilevel phenomenon where level 1 units are expected to be stratified. For instance, the implications of this model can be extended to urban economics where households are stratified between neighbourhoods which differ in terms of their housing prices. In this case, stratification is determined by household characteristics and the functioning of the real-estate market. The model can also be extended to the analysis of industrial clusters where firms are stratified into geographical zones. Furthermore, the implications can be extended to the analysis of labour productivity where individuals are stratified between firms based on their characteristics and on the functioning of the labour market.

Finally, in this theoretical model I used a single economy populated by a continuum of households. However, this model can be extended to multiple economies. In these economies, the strength of educational stratification needs not be the same across all of them. In fact, in my model the Lagrangian multipliers define the strength of the variation in the marginal cost resulting from variations in school quality after admitting a student of a particular type. These multipliers are positive, and the higher they are the stronger stratification is, since the higher the impact of student characteristics on the marginal cost will be. Note that if the multipliers are equal to zero, then the marginal cost of admitting a student will only be determined by the number of students in a school. In other words, student characteristics will not have any effect and stratification according to ability and social status will not exist. This is very important for any empirical work based on this model. Some countries have comprehensive schooling (Finland) and stratification levels are weak, while others are more liberal (UK) and levels of stratification are expected to be relatively higher. Our theoretical model allows for such variations in stratification levels.

Conclusion.

The objective of this paper was to develop a theoretical model to study the empirical and econometric implications of stratification. For this I used the example of educational stratification of students between schools.

The model clearly shows that the omission of some school characteristics generates endogeneity bias in multilevel estimations because these characteristics are the by-product of students' attributes. The reverse is also true. The omission of student characteristics leads to endogeneity because these are correlated with the included school-level variables. Secondly, countries with different education systems may be affected differently by this bias due to varying levels of stratification and hence to variations in the strength of correlation between student and school variables. Finally, non-linear social peer effects were introduced in order to model the impact of social diversity within a school. The hypotheses identified in this model can be tested empirically, especially that multilevel data accounting for student and school characteristics are available (e.g. PISA, TIMSS, PIRLS). Further, the theory developed in this paper can be extended to any economic phenomenon that exhibits stratification or nesting of smaller units within larger units (e.g. employees within firms, residents within neighbourhoods, etc).

References.

- Arnott, R. and Rowse, J. (1987) 'Peer Group Effects and Educational Attainment,' *Journal of Public Economics*, 32, pp. 287–306.
- Barzel, Y. (1973) 'Private Schools and Public School Finance,' *The Journal of Political Economy*. 81, 1, pp. 174-186.
- 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.

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.

Hanchane, S. and Mostafa, T. (2012) Solving Endogeneity Problems in Multilevel Estimation: An Example Using Education Production Functions. *Journal of Applied Statistics*, 39, 5, pp. 1101-1114.

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.

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

Mostafa, T. (2009b) Decomposing Inequalities in Performance Scores: The Role of Student Background, Peer Effects and School Characteristics. *International Review of Education*, 56, 5, pp. 567-589.

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.

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.

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