Solving Endogeneity Problems in Multilevel Estimation:

An Example Using Education Production Functions.

Saïd Hanchane and Tarek Mostafa[[1]](#footnote-1)

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.

Theses correlations are mainly the result of student stratification between schools. 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, an estimation approach based on the Mundlak (1978) technique is developed in order to tackle bias and to generate consistent estimates. Note that our analysis can be extended to any multilevel-structured data (students nested within schools, employees within firms, firms within regions, etc).

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.

JEL Classification: I20, C23.

Keywords: Multilevel analyses, endogeneity problems, Mundlak technique, PISA data.

**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 main motivation behind this analysis relates to the shortfalls of the findings published in the *‘Learning for Tomorrow’s World’*PISA 2003 report. 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. Furthermore, the multilevel nature of the data was neglected in most analyses. In this case, if heteroscedasticity or endogeneity arise, the regression coefficients will be inconsistent and the resulting inference will be distorted.[[2]](#footnote-2)

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 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 we discuss the implication of student stratification between schools on the estimation of education production functions (EPF). In the second, the EPF to be estimated is presented. It 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 developed and an endogeneity robust estimation approach is developed. This approach is based on the Mundlak (1978) technique developed for panel data and which we adapt for multilevel estimation.

In the third section of the paper, the estimation is carried out and the results are presented. The empirical analysis is done using the Program for International Student Assessment (PISA 2003) data. 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. Finally, we present the regression coefficients and the variance components.

**The Theoretical Framework.**

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. Note that Mostafa (2010) provides a concise theoretical model that describes how stratification arises in a general equilibrium model.

From an econometric point of view, stratification implies that the correlation between student and school characteristics might lead to endogeneity bias. In fact, the bias arises when the omitted variables are absorbed by the error term and the latter is correlated with the included ones. Moreover, the resemblance between students attending the same school warrants the use of multilevel modelling.

From this discussion it is possible to identify two 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.

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).

**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 was selected for several reasons. First, it was collected using the same sampling procedure across all countries which is very convenient for international comparisons. Secondly, the major domain of assessment in PISA 2003 is mathematics which is more universal than reading since it is not subject to cultural relativity. Thirdly, the PISA dataset provides a wide array of student and school variables that are needed for the analyses. The dependent variable is student performance scores in mathematics on the PISA 2003 standardized test. The independent variables are the following:

*Student Characteristics:*

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:*

DESCS: School average ESCS, depicting economic, social and cultural peer effects.

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:*

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.

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 (2009a and b).

**Table 1 about here.**

1. ***Endogeneity Problems in Multilevel Analyses.***

The general model to be estimated is the following:  with .

When is substituted out, the equation becomes.



This model is a random intercept multilevel model (i.e. different intercept or parallel slopes 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, and a random part , 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).  can be considered as comprising the unobserved school characteristics.

: are student performance scores in mathematics (student i attending school j). This is the dependent variable of the model.

: is a vector of student characteristics (student i attending school j).

: is a vector of peer effects (school aggregates of student characteristics).

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

: is the error term of the model.

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:

1. The independent variables at each level are not correlated with the random effects (error terms) on the other level, ,  and .
2. The level one independent variables are not correlated with level one error term. .
3. The level two independent variables are not correlated with level two error term. , and .
4. Each level one error term  is independent and normally distributed with a mean of 0 and a constant variance of . .
5. Each level two random effect (error term) is normally distributed with a mean of 0 and a constant variance . . These error terms are independent among the level two schools.
6. The error terms at level 1 and 2 are independent. .

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 is correlated with a level one independent variable. In this case, the assumption that  is violated, and some unobservable school characteristics relegated to the error term, are correlated with the observable student characteristics. 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 ( and ). 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  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 (or contextual 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, three different specifications of the aforementioned model are estimated. Model 1 omits peer effects, model 2 omits pure school characteristics, and model 3 considers the three vectors of variables. 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 a full explanation.

1. ***The Mundlak approach for panel data.***

It is useful to start with a description of our approach based on 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 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) to our multilevel data.

In his paper, Maddala (1987) gave an interesting example on how to solve endogeneity problems in a time panel data. The example he gave was based on the estimation of production functions in firms.[[3]](#footnote-3) The model he used is the following:

, with (i = 1, 2,…, N) and (t = 1, 2,…, T).

 is a subscript denoting a firm and t is a subscript denoting a time period. is the output, is the vector of inputs for firm i in period t.  is the regression coefficient,  is the firm specific unobserved inputs assumed to be constant over time. And finally,  is an error term assumed to be normally distributed with mean 0 and constant variance, .

The element  can be treated as a fixed effect or as a random effect. If is considered to be fixed then one  should be estimated for each of the firms. In contrast, if is considered to be random, then it will be treated as a random component following a normal distribution with a mean of 0 and a constant variance (exactly like ). This random specification was used in Balestra and Nerlove (1966). The random effects model is similar to our multilevel specification.[[4]](#footnote-4) It should be noted that in fixed effects models, level 2 endogeneity problems do not exist since is treated as an intercept that has to be estimated for each firm. In contrast, in random effects models, level 2 endogeneity problems might exist since  is treated as random and since might be violated.

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

1. When the dataset contains a large number of observations, instead of estimating N values for  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).[[5]](#footnote-5)
2. The treatment of  as random allows us to measure firm-specific effects. 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.
3. Random effects models allow for the inclusion of time constant variables as independent variables. This is not possible in fixed effects models.

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 , with being a vector of time constant variables.

However, the drawback of random effects models is the possible existence of endogeneity bias. Mundlak (1978) studied the case where the s are correlated with thes.[[6]](#footnote-6) The author argued that this endogeneity problem will be solved if  is assumed to depend on the mean value of , such as , with a random part that has similar properties to . The equation becomes: .

Using the Fuller and Battese (1973) transformation of the model, the estimator of the $β$s from the random effects model is obtained through OLS estimation of the following semi-demeaned equation (demeaning consists of removing a fraction of the mean):



Then, the equation is expanded and  is added and subtracted from it:



Finally, the equation becomes:





We denote  and with 

Since and are independent (is orthogonal to ), it is possible to estimate each of ,  and  independently through OLS. The estimate of  is  (is the within-group estimator). The estimate of  is  and the estimate of  is . Similarly,  can be obtained by regressing the time constant variable  on the average of over time, which is. These estimates are robust and efficient.

1. ***Our adaptation of the Mundlak technique for multilevel analyses.***

Recall that the general multilevel model is the following:  with . In what follows we present our endogeneity-robust estimation approach for model 3. Note that the approach for models 1 and 2 is similar since these are simplified versions of model 3. All the estimations were carried out using LIMDEP and they were programmed step by step.

After replacing by() , model 3 becomes:



At this stage, the Hausman test is performed in order to compare the fixed effects model, containing only student characteristics , to the random effects model, containing ,and . Then the model is transformed in order to eliminate any possible endogeneity bias.

|  |
| --- |
| **Remark:**The within- and between-school variances are the ones on  and , respectively. The variance of  is  and the variance of  is .  and  are the estimates of  and , respectively. |

Using the Fuller and Battese (1973) argument the model is transformed in the following manner: 

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

We add and subtract  from the equation:



We expand the equation:



And finally it becomes:







With , ,

 and  with .

: is the estimate of the within-school variance.

: is the estimate of the between-school variance.

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

**Estimation:**

We assume that  and are independent and since  and  are orthogonal, the effects of the different components can be estimated separately.

* 1. We regress  on . is obtained as well as the variance components.
	2. We regress on  and .  and are obtained.
	3. We compute  with 

Then is computed as the average of ; hence .

* 1. We multiply  and by  to obtain:  and .
	2. We have, ,  and . It is possible to compute .

**Results.**

In this section, the estimation of the models is carried out using PISA 2009 data. 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 manner.

* 1. ***The Hausman Test.***

The Hausman test is a specification test developed by Jerry Hausman (1978). The test identifies the presence of level 2 endogeneity. The null hypothesis is that the school-level random effects  are not correlated with any of the students’ variables, in other words. 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. In the fixed effects model, is treated as a fixed effect and hence j values of it are estimated. This fixed effects model does not suffer from level 2 endogeneity since, as a fixed effect, cannot be correlated with . In contrast, in random effects models, is considered to be random and hence can be correlated with. The Hausman test compares the random and the fixed effects specifications, if the null hypothesis is rejected (i.e. ), we can conclude that the random effects model suffers from endogeneity and that the fixed effects specification is better.

**Table 2 about here.**

As we can see, the Hausman test fails for the first model (in Germany and the UK) where the null hypothesis is rejected and holds for models 2, and 3. A number of conclusions can be drawn:

1. Model one did not control for peer effects (school averages of the s). These school characteristics were relegated to the error term  and are correlated with student level variables. Thus, model 1 suffers from endogeneity and the null hypothesis on the Hausman test is rejected in Germany and in the UK. The fixed effects specification is preferred to the random effects one and the coefficients on the latter are biased.
2. The only country that passed 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 extremely 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 the school. The UK is middle ranking.
3. 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.
4. Models 2 and 3 passed the Hausman test in all countries. The null hypothesis holds and there are no correlations left between students’ variables and unobserved school characteristics relegated to the error term.
5. In model 2, pure school characteristics were omitted. However, the model still passed the Hausman test. This is an indication that this model does not suffer from level two endogeneity. One should note that endogeneity arose when peer effects were omitted (model 1) and not when pure school characteristics were neglected (model 2). In other words, in model 1, the correlation between peer effects and student variables generated endogeneity; while in model 2 the correlation between pure school variables and student characteristics was too weak to cause endogeneity.
6. 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 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.
7. Model 3 is the most complete; it controlled for peer effects and pure school characteristics and it passed the test. Therefore, it can be considered to be the best.

In conclusion, 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 more correlated with peer characteristics than pure school variables. One may find this to be obvious since peer effects are averages of student variables and hence the two are expected to be correlated. However, in a school with 30 sampled students, the value on a student-level variable has a 1/30 weight in the corresponding peer effect (i.e. the average of the student variable). Hence the correlation between a student-level variable and its peer effect is not necessarily strong. The strength of the correlation is a factor mainly determined by the strength of educational stratification. The Hausman test indicated 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 varies according to country characteristics. In a stratified system, such as Germany, students are more likely to be sorted according to their type. Hence students will end up attending schools with similar peers and student variables are more likely to be correlated with peer effects. The reverse is true in countries such as Finland where education is very comprehensive. 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 management of schooling exacerbate the impact of stratification by intensifying the role that student-related factors play in determining school characteristics. In summary, we can say that peer effects are more important than pure school characteristics and it is essential to control for them in order to avoid level 2 endogeneity.

* 1. ***The Regression Coefficients.***

**Table 3 about here.**

**Table 4 about here.**

As noted before, the coefficients on student level variables ($βs$) are identical for all models. However, the coefficients on school level variables ($γs$) 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. Model 3 is the most complete, since it considered the three vectors of independent variables: student characteristics, pure school characteristics, and peer effects. This model is the benchmark against which models 1 and 2 are compared.

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 (i.e. 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 model 3, 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. 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.

In conclusion, we can say that when some school variables were omitted, the regression coefficients were overestimated and inconsistent. In contrast, in model 3, the random effects model was efficient in comparison with the fixed effects model. In other words, including the averages at the school level (level 2) of level one variables solves the endogeneity problem. Further, these averages have an intuitive meaning since they represent school contextual effects.

* 1. ***The Variance Components.***

**Table 5 about here.**

In table 5, we present the estimates of the variance components. The first finding to note is that the between school variance on $V\_{j}$ is sufficiently high to justify the use of a random intercept multilevel models. In fact, in Germany in model 3, the between school variance accounts for 18% of the total variance. In Finland and in the UK the percentage is lower. This is intuitive because in both countries schools are more homogenous than in Germany and hence their specific effects are less likely to vary. Secondly, we can see that the estimate of the between school variance drops from model 1 to model 3. The drop is the most significant in Germany, followed by the UK, while in Finland the estimates are almost the same across all models. Recall that model 1 was the least reliable and Germany was the country most affected by endogeneity.

**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 peer effects 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 is weak. Finally, in relation to the results published in PISA 2003 report, this paper clearly shows that any future multilevel analyses of education production functions should simultaneously control 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.

**References**

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.

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.

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.

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.

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.

Mostafa, T. (2010) ***Endogeneity Problems in Multilevel Estimation of Education Production Functions: an Analysis Using PISA Data. LLAKES Research Paper No 14. 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.

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.

**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 |

**Table 2. The results of the Hausman test.**

|  |  |  |  |
| --- | --- | --- | --- |
|   | Germany | Finland | UK |
| Model 1 | 309.07(0.00) | 5.68(0.45) | 125.66(0.00) |
| Model 2 | 0.00(0.99) | 0.00(0.99) | 0.00(0.99) |
| Model 3 | 0.00(0.99) | 0.00(0.99) | 0.00(0.99) |

P values are in parentheses based on χ2 test.

**Table 3. The regression coefficients on student level variables.**

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Germany | Finland | UK |
|   | Coef. | Sig | SE | Coef. | Sig | SE | Coef. | Sig | SE |
| ESCS | 11.61 | \*\*\* | 1.24 | 26.85 | \*\*\* | 1.24 | 23.26 | \*\*\* | 0.99 |
| COMHOME | 1.28 |  | 1.22 | -2.72 | \*\* | 1.15 | 8.2 | \*\*\* | 0.99 |
| INTMAT | 4.69 | \*\*\* | 0.95 | 14.51 | \*\*\* | 1.05 | -1.16 |  | 0.87 |
| ANXMAT | -19.03 | \*\*\* | 0.9 | -31.96 | \*\*\* | 1.08 | -25.11 | \*\*\* | 0.87 |
| DISCLIM | 2.63 | \*\*\* | 0.82 | 1.41 |  | 1 | 12.33 | \*\*\* | 0.7 |
| ETR | -27.92 | \*\*\* | 2.92 | -63.49 | \*\*\* | 6.37 | -5.18 | \* | 2.9 |

**Table 4. The regression coefficients on school level variables.**

|  |  |
| --- | --- |
| Variables | Germany |
|   | Model 1 | Model 2 | Model 3 |
|   | Coef. |  Sig | SE | Coef. |  Sig | SE | Coef. |  Sig | SE |
| DESCS | . | . | . | 66.15 | \*\*\* | 1.63 | 60.38 | \*\*\* | 1.81 |
| DCOMPH | . | . | . | 26.08 | \*\*\* | 2.78 | 28.58 | \*\*\* | 2.94 |
| DINTMAT | . | . | . | -23.7 | \*\*\* | 2.19 | -24.31 | \*\*\* | 2.08 |
| DANXMAT | . | . | . | -14.14 | \*\*\* | 1.97 | -16.54 | \*\*\* | 1.93 |
| DDISCL | . | . | . | 28.19 | \*\*\* | 1.32 | 25.03 | \*\*\* | 1.47 |
| DETR | . | . | . | 10.57 | \*\*\* | 3.41 | 6.38 | \*\*\* | 3.31 |
| COMPWEB | 38.95 | \*\*\* | 2.98 | . | . | . | 10.97 | \*\*\* | 1.79 |
| MACTIV | 29.8 | \*\*\* | 1.53 | . | . | . | 1.26 |  | 0.8 |
| MSTREL | -137.52 | \*\*\* | 24.01 | . | . | . | 12.95 |  | 11.57 |
| TCSHORT | -17.86 | \*\*\* | 1.19 | . | . | . | -7.63 | \*\*\* | 0.54 |
| TCMORALE | -0.04 | \*\*\* | 0 | . | . | . | 0.03 | \*\*\* | 0 |
| TEACBEHA | -12.64 | \*\*\* | 1.41 | . | . | . | -4.3 | \*\*\* | 0.79 |
| PRIVATE | 34.07 | \*\*\* | 2.1 | . | . | . | -11.36 | \*\*\* | 2.05 |
| SCMATEDU | 0 |  | 0.01 | . | . | . | -0.07 | \*\*\* | 0 |
| ACADEMIC | 25.91 | \*\*\* | 2.02 | . | . | . | 7.29 | \*\*\* | 0.98 |
| INTERCEPT | 459.18 |   | 3.44 | . | . | . | 466.33 |   | 2.15 |

|  |  |
| --- | --- |
| Variables | Finland |
|   | Model 1 | Model 2 | Model 3 |
|   | Coef. | Sig | SE | Coef. | Sig | SE | Coef. | Sig | SE |
| DESCS | . | . | . | 2.44 | \*\*\* | 1.12 | 2.82 | \*\*\* | 0.95 |
| DCOMPH | . | . | . | 3.96 |  | 1.41 | 1.76 |  | 1.43 |
| DINTMAT | . | . | . | -2.84 | \*\*\* | 1.47 | -6.01 | \*\*\* | 1.37 |
| DANXMAT | . | . | . | 6.14 | \*\*\* | 1.7 | 2.87 | \*\*\* | 1.58 |
| DDISCL | . | . | . | 3.11 | \*\*\* | 0.8 | -0.1 |  | 0.82 |
| DETR | . | . | . | 44.12 | \*\*\* | 3.31 | 55.58 | \*\* | 3.14 |
| COMPWEB | 13 | \*\*\* | 4.02 | . | . | . | 4.83 | \*\* | 2.44 |
| MACTIV | 2.74 | \* | 1.46 | . | . | . | 1.44 | \*\*\* | 0.47 |
| MSTREL | -131.34 | \*\*\* | 21.28 | . | . | . | -124.76 | \*\*\* | 8.02 |
| TCSHORT | 1.06 |  | 1.05 | . | . | . | -0.35 |  | 0.37 |
| TCMORALE | 5.47 | \*\*\* | 0.96 | . | . | . | 1.78 | \*\*\* | 0.32 |
| TEACBEHA | -2.45 | \*\*\* | 0.78 | . | . | . | -1.4 | \*\*\* | 0.43 |
| PRIVATE | -15 | \*\*\* | 3.7 | . | . | . | -18.42 | \*\*\* | 1.46 |
| SCMATEDU | 0.16 |  | 0.81 | . | . | . | 0.24 |  | 0.37 |
| ACADEMIC | 12.64 | \*\*\* | 3.51 | . | . | . | 11.15 | \*\*\* | 1.01 |
| INTERCEPT | 532.41 |   | 5.17 | . | . | . | 528.74 |   | 2.44 |

|  |  |
| --- | --- |
| Variables | The UK |
|   | Model 1 | Model 2 | Model 3 |
|   | Coef. | Sig | SE | Coef. | Sig | SE | Coef. | Sig | SE |
| DESCS | . | . | . | 51.4 | \*\*\* | 0.84 | 42.82 | \*\*\* | 1.09 |
| DCOMPH | . | . | . | -15.78 | \*\*\* | 1.4 | -6.33 |  | 1.58 |
| DINTMAT | . | . | . | -8.88 | \*\*\* | 1.12 | -8.27 | \*\*\* | 1.16 |
| DANXMAT | . | . | . | -13.83 | \*\*\* | 1.19 | -12.61 | \*\*\* | 1.14 |
| DDISCL | . | . | . | 11.39 | \*\*\* | 0.69 | 9.38 | \*\*\* | 0.72 |
| DETR | . | . | . | -35.14 | \*\*\* | 2.4 | -35.83 | \*\*\* | 2.36 |
| COMPWEB | 18.46 | \*\*\* | 3.09 | . | . | . | 17.7 | \*\*\* | 1.17 |
| MACTIV | -2.36 | \*\*\* | 0.59 | . | . | . | -2.97 | \*\*\* | 0.24 |
| MSTREL | -282.9 | \*\*\* | 26.28 | . | . | . | -76.85 | \*\*\* | 7.78 |
| TCSHORT | -6.76 | \*\*\* | 0.49 | . | . | . | -3.79 | \*\*\* | 0.31 |
| TCMORALE | 0.51 |  | 0.71 | . | . | . | -2.15 | \*\*\* | 0.34 |
| TEACBEHA | 6.83 | \*\*\* | 0.75 | . | . | . | -1.28 | \*\*\* | 0.34 |
| PRIVATE | 71.68 | \*\*\* | 2.33 | . | . | . | 14.53 | \*\*\* | 1.46 |
| SCMATEDU | -0.96 | \* | 0.56 | . | . | . | 2.56 | \*\*\* | 0.27 |
| ACADEMIC | 11.37 | \*\*\* | 1.82 | . | . | . | 4.51 | \*\*\* | 0.97 |
| INTERCEPT | 505.68 |   | 3.93 | . | . | . | 492.8 |   | 1.38 |

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

**Table 5. The variance components.**

|  |  |  |  |
| --- | --- | --- | --- |
| Variance components | Germany | Finland | UK |
|   | Model 1 | Model 2 | Model 3 | Model 1 | Model 2 | Model 3 | Model 1 | Model 2 | Model 3 |
| Total variance | 4924 | 3820 | 3747 | 4183 | 4210 | 4183 | 4911 | 4692 | 4643 |
| Within school variance | 3038 | 3038 | 3038 | 3973 | 3973 | 3973 | 4305 | 4305 | 4305 |
| Between school variance | 1885 | 782 | 709 | 210 | 237 | 210 | 606 | 387 | 338 |
| Between/Total | 0.383 | 0.205 | 0.189 | 0.050 | 0.056 | 0.050 | 0.123 | 0.082 | 0.073 |

1. 1 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 - ESRC). 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. [↑](#footnote-ref-1)
2. Note that in the PISA 2009 report, some of these shortfalls have been dealt with even though the problem of endogeneity was not spelt out. [↑](#footnote-ref-2)
3. Note that in our case we are estimating multilevel education production functions in schools. [↑](#footnote-ref-3)
4. 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). [↑](#footnote-ref-4)
5. In the case of PISA, if the is treated as fixed and is not decomposed into an overall intercept and a random part, then we have to estimate a  for each school (this will cause the loss of j degrees of freedom). However, when is decomposed in the following manner , only the constant overall intercept c and the random parts are estimated, thus saving some degrees of freedom. [↑](#footnote-ref-5)
6. This is similar to our level 2 endogeneity problem, where thes might be correlated with the s. [↑](#footnote-ref-6)