The Empirical Economics Letters, 4(2): (March 2005) *ISSN 1681 8997*

A Two-Level Model for Disentangling Student and School Productivity

Daniel Santín

Department of Applied Economics VI, Complutense University of Madrid
Campus de Somosaguas 28223, Pozuelo de Alarcón, Madrid, Spain

Email: dsantin@ccee.ucm.es

Abstract: The aim of this paper is to shed light on the estimation of school productivity. Conventional approaches do not consider explicitly that the educational production process takes place at least at two different levels, student and school. This paper proposes an alternative two-level model in order to provide an approximation for disentangling both effects. The model also permits calculation of school inputs elasticities and technical efficiency at school level. Finally, the paper provides an empirical application of the model using Spanish data from the Third International Mathematics and Science Study. Findings suggest that different school inputs matter for different students.

Keywords: Education economics, school productivity, artificial neural networks

JEL classification: I21, C45

1. Introduction

Since the early 1960s (Carroll, 1963), a large number of studies (*e.g*. Coleman, 1966; Summers and Wolfe, 1977; Hanushek, 1986, 1996, 1997; Pritchett and Filmer, 1999) has sought to explain the relationship between school inputs and student achievement. Unfortunately, the well-known “Does school matter?-Does money matter?” debate still remains open. Only a few general conclusions can be obtained from a review of this whole period. One evident result is that education takes place both at home and at school. However, the educational production function and the way that student, parents and school match with educational outputs continues to be unknown. This is a serious drawback for policy-makers taking decisions about allocating scarce public expenditure in schools. In fact, nowadays educational policies are decided without a valid theoretical framework. This means that it is impossible to know whether or not the additional resources invested in public education are being allocated to schools in the best productive way. In a largely cited article, Hanushek (1986, p.1148) summarizes this topic: "*It appears from the aggregate data that there is at best an ambiguous relationship and at worst a negative relationship between student performance and the inputs supplied by schools*" (italics in original).

A common framework for estimating the educational production function might take the following form (Levin, 1974; Hanushek, 1979):

$$A_i = f(B_{\mathrm{i}}, \mathrm{S}_{\mathrm{i}}, P_{\mathrm{i}}, I_{\mathrm{i}}) \quad (1)$$

where A_i equals the achievement of student i, B_i are student background characteristics, S_i are school inputs, P_i denotes the peer-group effect, and I_i are student innate abilities. As the f in equation 1 is unknown, economists assume typical econometrics requirements for a production function (Mas-Colell *et al*., 1995). Whether or not the straightforward results obtained by linear regression or an instrumental variable analysis could be interpreted as valid, depend on *a priori* assumptions about the educational production function. This paper owes a debt to recent literature provided mainly by Eide and Showalter, (1998), Figlio (1999) and Baker (2001) among others. These authors found strong evidence about the inconsistency of the use of a Cobb-Douglas specification for estimating the educational production function. Their critical analysis can be roughly summed up as follows: Using different flexible approaches for modelling the production function[1] we can enrich school productivity empirical results in order to better allocate education resources. Following this evidence, this work explores the possibilities of using artificial neural networks (ANNs) as a non-linear regression tool to approximate the educational production function at student level.

Another issue regarding the educational production function is its multilevel structure. In fact, equation 1 is frequently solved by econometric approaches that move all variables by aggregation or disaggregation to one single level of interest (e.g., student level, class level, school level and so on). Analyzing education at only one single level generates different problems (see Hox, 1995 for a review). Finally, most studies neither consider neither average test scores as school outputs can be misleading (Meyer, 2000) nor education that is a nested process.

---

[1] Eide and Showalter (1998) employ quantile regression approach, Figlio (1999) uses a *translog* functional form and Baker (2001) analyzes the educational production function at school level through different neural networks approaches.

This paper sheds additional light to educational research proposing a methodology for disentangling and measuring pupil and school effects in educational achievement. To fulfil with this purpose the paper is organized as follows. Second section provides a brief introduction to ANNs as a non-linear approach to match a multiple-inputs-multiple-outputs production function. Section 3 discusses the proposed two-level model for estimating the educational production function. Section 4 describes data set and variables and illustrates an empirical application along with a discussion of its results for policy makers. The final section of the paper offers conclusions and suggests directions for future research.

2. Artificial Neural Networks

The model proposed in this paper uses ANNs in its first stage.[2] The choice of an ANNs algorithm instead of another non-linear methodology responds to the relative success reached in educational applications (Gorr *et al*., 1994; Baker and Richards, 1999; Baker, 2001) as well as in other economics fields (Zapranis and Refenes, 1999).

The most used neural network architecture in the economics literature is the Multilayer Perceptron (MLP). In general terms, a MLP can be defined like a group of processing elements known as *neurons* organized in at least three layers, input, hidden(s) and output. These neurons are all connected in one direction by weights. The target of a MLP is learning to match input to output vectors through the interactions among neurons. This implies learning the function

$$\Re^n \xrightarrow{\phi} \Re^m$$
$$\phi(X;W) = Y$$

Through the sample $\{X(\text{p}), Y(\text{p})\}$, p = 1,2,...,N where $X(\text{p}) \in \Re^{\text{n}}$ is the input vector and $Y(\text{p}) \in \Re^{\text{m}}$ is the output vector. This is carried out by adjusting the matrix of weights (W) of given interconnections among the neurons according to some learning algorithm. MLP uses a supervised learning algorithm proposed by Rumelhart *et al*. (1986) called *backpropagation*. This learning is guided by specifying the desired response of the network, the observed output, for each training input pattern and its comparison with the

[2] The aim of this section is rather to provide a brief introduction to ANNs and not to do an extensive survey. Nevertheless, for a complete revision of mathematical and statistical properties of ANNs see for example Bishop (1996).

actual output computed by the network in order to adjust the weights. These adjustments have the purpose of minimizing the difference between desired and actual outputs.

Normally the implementation of the backpropagation algorithm implies to split the sample into three data sets. A training set is used to seek the parameters able to match inputs with outputs. A validation set is used at the same time to control for model complexity and for stopping learning when no gains are obtained in this sample with further optimization. Finally, after neural training, new observations never seen before by the neural net (test data set) are presented to the network to obtain an unbiased measure of the so-called generalization capability.

It has been shown that ANNs are universal approximators of functions (Funahashi, 1989; Hornik *et al*., 1989) and their derivates (Hornik *et al*., 1990). Scarselli and Chung (1998) provide an actual and complete review of this property.

3. The Two Level Model

This model is partially based on Meyer (1997). Let equation 2 be a first level student educational production function.

$$A_{is} = f(B_{is}) + \psi_{is} \qquad (2)$$

where B_{is} is a vector of background characteristics of the student *i* at school *s* and ψ_{is} denotes the school effect over the achievement that will be measured in the second level. Last component also captures school inefficiency and random noise.[3] Since the *f* in this model is completely unknown, the production function will be fitted using an OLS approach together with a standard *backpropagation* MLP. This comparison will allow detect whether or not exists non-linearities at student level. Afterwards, we will have a production function model able to predict, at the beginning of entering school, student performance according with background.

Students attend different schools and after the educational process each student will show a gap between real and predicted output that will be explained at school level by ψ_{is} factor. Before modelling school level equation we need to construct school outputs. A

[3] The database used in this paper does not contain information about student’s innate ability. It is assumed that this component is normally distributed among students independently of other inputs and captured in the error term.

school will be considered as productive only in those students obtaining better results than expected. These school indicators will be named as *value-added outputs* (eq. 3).

$$Y_{js} = g_j(A_{is}, \hat{A}_{is}) \qquad (3)$$

where Y_{js} is the value-added output j at school s, and g_j is a decision rule to construct this so called value-added outputs. The empirical application contained in this paper manages three different value-added outputs. First value-added output is the percentage of students at each school that perform better than expected (VA Output 1). Second value-added output is the percentage of at-risk-students (those *a priori* located at low quartile of the conditional predicted test score distribution) leaving out from this status (VA Output 2) after the school period. Last value-added output is the percentage of students with predicted test scores below high level (3rd quartile) and performing into this level after the educational process (VA Output 3).

Once we have defined value-added outputs we will research what school inputs explain then. To do this we can estimate the parameters contained in equation 4.

$$Y_{js} = \beta_{j0} + \beta_{j1}S_{js} + \beta_{j2}P_{js} + \varepsilon_j \qquad (4)$$

where ε_j is the error term. For the sake of simplicity in this paper technical efficiency could be estimated correcting the intercept in the following way $\hat{\beta}_0^* = \hat{\beta}_0 + \max\{\varepsilon_j\}$ in order to do that OLS estimation becomes a deterministic production frontier (COLS methodology proposed be Greene, 1980). Main drawback in econometric models is that they assume a parametric structure when in this case is unknown. However value-added school outputs measure different dimensions of performance allowing for variations in school inputs importance according with different students. Together with this, econometric results permit computing elasticities and statistical significance.

4. Empirical results

4.1 Data

This study uses the Third International Mathematics and Science Study (TIMSS) database conducted in 1995 by the International Association for the Evaluation of Educational Achievement (IEA). The students who participated in TIMSS completed questionnaires about their home and school experiences related to learning mathematics and sciences. Teachers and school administrators completed questionnaires about instructional practices too.

The model presented above is used to analyze 3700 Spanish student's mathematics test results at eighth grade. A wide range of student, school, teacher and peer variables were included in the analysis. [4] Table 1 displays descriptive information for variables used at student and school levels. Student's variables include sex, age, language at home, resources at home, parental education, people at home and other activities. School variables include both, school and teachers characteristics. In particular the database collects different activities at classroom and teacher habits during the course.

Table 1: Descriptive Statistics (3700 Students and 142 Schools)

Student variables	Mean	Standard Deviation	School variables	Mean	Standard Deviation
Output			*Outputs*		
MATHSCORE	490,96	73,95	Average Result	485,76	32,86
Background			VA OUTPUT 1	48,15	17,61
SEX (1=boy)	0,48	0,50	VA OUTPUT 2	32,68	24,48
CASTILIAN	0,76	0,43	VA OUTPUT 3	22,27	15,01
CATALAN	0,01	0,09	*School*		
GALICIAN	0,15	0,36	YEARWEEKS	35,30	3,67
BORNINSPAIN	0,03	0,16	REPEATERS	0,61	0,49
LANGTESTHOME	0,79	0,41	SCHOOLPLAN	0,81	0,99
MOTHER	0,03	0,17	TEACHFOR+5Y	67,59	23,11
FATHER	0,11	0,31	ABSENT	3,18	2,42
BROTHER	0,34	0,47	ADMINISTRATION	1,31	2,24
SISTER	0,38	0,48	REGMEETINGS	0,04	0,20
OTHER(S)	0,90	0,29	PRIVATE	0,32	0,47
FAMILY	4,79	1,35	RATIO	27,13	7,33
BOOKS	3,50	1,15	PRESCHOOL	94,86	15,87
COMPUTER	0,57	0,50	DISAVBACKG	25,07	25,70

[4] The original variable name in TIMSS was replaced for another more illustrative one. Some of the variables were recoded from the original codification. However, the original student and school variables names are available from the author upon request. For an extensive description of original names and codes see Gonzalez and Smith (1997).

Table 1 continued

STUDYDESK	0,06	0,24	RURAL	0,47	0,50
MATHSBOOKS	0,34	0,47	*Teacher*		
EDUCVIDEOS	0,67	0,47	MATHS75%	10,60	18,14
EDUCSOFTWARE	0,74	0,44	SCHELOTHERS	3,42	3,56
STUDENT'S AGE	14,25	0,67	DISCRESPONSE	2,80	0,63
MPRIMARY	0,48	0,50	WORKINGROUPS	1,75	0,56
MSOMESEC	0,11	0,31	TIMETEXTBOOK	2,35	0,95
MSECONDARY	0,07	0,26	REASONING	1,88	0,54
MVOCATIONAL	0,04	0,20	HELPANSWER	2,02	0,55
MSOMEUNIVERS	0,03	0,16	EDUCATION	2,77	0,85
MUNIVERSITY	0,10	0,29	STDSUPERVISION	1,05	1,34
FPRIMARY	0,39	0,49	CORRECTANSWER	2,14	0,67
FSOMESEC	0,12	0,32	CORRECTOWN	3,46	0,88
FSECONDARY	0,09	0,28	HMWKDISCUSS	2,27	0,86
FVOCATIONAL	0,05	0,23	SCHEDMATHS	11,16	4,92
FSOMEUNIVERS	0,03	0,17	HMWKSTARTCLASS	66,82	21,20
FUNIVERSITY	0,12	0,33	COPYNOTES	59,67	23,51
PAIDLESS1H	0,04	0,21	TEACHERGUIDE	2,78	0,71
PAIDBETW1-2H	0,07	0,25	SPECIFICTEXT	3,36	0,85
PAIDBETW3-5	0,03	0,16	GRADE 6	0,61	0,49
PAIDMORE5H	0,03	0,18	TEACHERAGE	2,04	0,78
			TEACHERSEX	0,63	0,48

4.2. Results

To estimate first level equation a number of 35 variables were used. Note that many of these variables are simply different categories of one single variable. Table 2 provides the estimates for the OLS model including p-values and variance inflation factors (VIF) to control for multicollinearity among variables. [5]

[5] According with Baker (2001, pp. 87) it is assumed a standard threshold for multicollinearity of VIF>5.

Table 2: OLS results at Student Level

Variable	Estimate	Std. Errors	P-Value	VIF
Intercept	746,83	26,58	0,00	-
SEX	11,04	2,29	0,00	1,05
CASTILIAN	18,09	4,54	0,00	3,02
CATALAN	19,71	12,84	0,12	1,14
GALICIAN	4,48	5,08	0,38	2,71
BORNINSPAIN	-9,41	7,05	0,18	1,04
LANGTESTHOME	3,60	3,47	0,30	1,59
MOTHER	1,54	6,75	0,82	1,10
FATHER	3,65	3,77	0,33	1,13
BROTHER	2,61	2,53	0,30	1,16
SISTER	2,88	2,46	0,24	1,14
OTHER(S)	9,82	3,91	0,01	1,06
FAMILY	-1,14	0,93	0,22	1,26
BOOKS	12,93	1,11	0,00	1,30
COMPUTER	-6,58	2,98	0,03	1,76
STUDYDESK	-12,97	4,75	0,01	1,04
MATHSBOOKS	6,89	2,46	0,01	1,09
EDUCVIDEOS	6,14	2,50	0,01	1,11
EDUCSOFTWARE	-3,46	3,36	0,30	1,74
STUDENT’S AGE	-23,63	1,73	0,00	1,09
MPRIMARY	2,87	4,14	0,49	3,44
MSOMESEC	6,38	5,12	0,21	2,08
MSECONDARY	13,96	5,71	0,01	1,79
MVOCATIONAL	9,13	6,71	0,17	1,46
MSOMEUNIVERS	10,32	8,11	0,20	1,31
MUNIVERSITY	15,66	5,73	0,01	2,30
FPRIMARY	2,11	4,02	0,60	3,10
FSOMESEC	8,78	4,87	0,07	1,95
FSECONDARY	11,47	5,26	0,03	1,75
FVOCATIONAL	0,79	6,11	0,90	1,56
FSOMEUNIVERS	8,59	7,66	0,26	1,29
FUNIVERSITY	6,79	5,19	0,19	2,37
PAIDLESS1H	-11,49	5,45	0,04	1,03
PAIDBETW1-2H	-7,09	4,52	0,12	1,03
PAIDBETW3-5	10,57	7,01	0,13	1,02
PAIDMORE5H	-20,11	6,40	0,00	1,04
R-squared	0,167	Adj. R-squared		0,160

The estimated SEX and CASTILIAN coefficients are positive indicating that boys completing the mathematics exam in Castilian perform better than his counterparts. The effect of the variable -members of the family living with the student- shows that a student who lives with other family members different from parents, brothers and sisters performs worse. The resources at home variables are significant too. Number of books at home, having a computer, having an own student desk and the possession of maths books and educational videos are all significant. Student's age is negatively related with results due to the effect of students repeating grade. On the other hand, students whose mother completed secondary and university studies perform better. This effect is also found for students who have a father with some secondary education finished and with secondary completed. Last, a high negative significant coefficient is found for students who employ dairy time in a paid job.

Together with the OLS analysis a conventional MLP backpropagation-type algorithm was used to evaluate how this approach fits data.[6] In order to compare the prediction accuracy for each model, Pearson's correlation coefficient and coefficient of determination were calculated for both. To avoid overfitting problems only data test prediction results are valid and unbiased in order to be compared with OLS. Results are shown in table 3.

Table 3: Comparison of Prediction Accuracy for Backpropagation and OLS Procedures

	CORR*	R^2**	MIN***	MAX***
Total	0,4339	0,1883	0,021	262,873
Training	0,4402	0,1938	0,021	261,005
Test	0,4405	0,1941	0,231	195,316
OLS	0,4092	0,1674	0,073	260,853

Note: * implies coefficient of correlation between real and prediction results; ** coefficient of correlation squared; *** minimum and maximum residual terms.

Notice that according with these results ANNs overcomes OLS adjustment in almost three points in terms of coefficient of determination. A higher accurate fitting student

[6] The final MLP used 25 neurons in the hidden layer. Sample was divided in three subsamples assigning 80% data to training, 10% to validation and 10% to test the model.

results means a better specified model. It is also worth to note that ANN are a good tool to do an exploratory analysis for searching the existence of non-linear relationships between inputs and outputs before applying a conventional approach and avoiding possible functional form misspecifications (Lee *et al*., 1993).

The value-added ouputs described before in section 4.1 were constructed using ANN predictions and real test results to analyze what school inputs explain value-added outputs. Together with the so-defined value-added outputs the average school result is also computed to do a conventional estimation for the production function. The model is adjusted using a COLS approach for each output and results are showed in table 4. The initial hypothesis of the model is that different inputs matter different students.

Table 4: Comparison of COLS Results for Different Outputs

	VA Output 1		VA Output 2		VA Output 3		Average Result	
Independent Variables	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	-7,46	25,37	-34,79	40,38	-26,91	22,52	406,47*	42,93
YEARWEEKS	0,73***	0,38	0,65	0,59	0,27	0,34	1,32**	0,64
SCHELOTHERS	-0,84**	0,39	0,15	0,66	-0,22	0,35	-0,56	0,67
SCHCOOPPLAN	0,87	1,49	0,46	2,41	-0,14	1,32	3,52	2,53
DISCRESPONSE	4,38***	2,28	-0,23	3,61	1,41	2,03	5,60	3,86
WORKINGROUPS	-7,07*	2,44	-2,70	4,00	0,34	2,17	-4,89	4,13
TIMETEXTBOOK	2,45***	1,46	3,51	2,33	0,83	1,29	1,00	2,47
REASONING	3,51	2,53	9,37**	4,12	2,89	2,25	1,86	4,29
HELPANSWER	3,45	2,49	-4,16	3,98	2,91	2,21	-1,56	4,22
RURAL	1,34	2,87	8,18***	4,54	2,92	2,55	-7,92	4,86
TEACHFOR+5Y	0,05	0,06	0,06	0,10	-0,08	0,05	-0,10	0,10
EDUCATION	0,90	1,68	1,34	2,63	2,39	1,49	2,09	2,84
STDSUPERVISION	0,28	1,13	-0,43	1,81	0,55	1,00	-1,20	1,90
CORRECTANSWER	3,60***	2,12	0,60	3,41	2,58	1,88	0,01	3,59
CORRECTOWN	-0,31	1,57	-0,46	2,60	-1,28	1,40	-2,44	2,66
HMWKDISCUSS	1,38	1,69	1,83	2,69	1,37	1,50	1,30	2,86
ABSENT	-0,50	0,57	0,27	0,93	-0,33	0,51	-1,48	0,97
ADMINISTRATION	-0,04	0,62	-0,85	0,98	-0,42	0,55	-1,68	1,05

Table 4 continued

REGMEETINGS	-22,62*	6,69	-17,54***	10,47	-13,38**	5,94	-42,17*	11,32
SCHEDMATHS	0,81*	0,28	1,56*	0,46	0,74*	0,25	1,85*	0,48
HMWKSTARTCLASS	-0,21*	0,06	-0,18***	0,10	-0,15**	0,06	-0,48*	0,11
COPYNOTES	-0,10***	0,06	-0,05	0,10	-0,01	0,05	-0,24**	0,10
TEACHERGUIDE	-3,97***	2,03	-7,55**	3,26	-2,69	1,80	-4,77	3,43
SPECIFICTEXT	2,88***	1,67	2,05	2,62	4,58*	1,48	7,93*	2,82
GRADE 6	6,19**	2,74	7,59***	4,39	7,33*	2,44	11,33**	4,64
REPEATERS	-6,08**	2,91	-4,13	4,59	-4,18	2,58	-3,01	4,92
PRIVATE	1,59	3,44	6,58	5,54	5,11***	3,05	13,33**	5,81
RATIO	0,30	0,22	0,57	0,35	0,06	0,19	0,56	0,37
PRESCHOOL	0,12	0,08	0,06	0,13	0,16**	0,07	0,44*	0,14
MATHS75%	-0,17**	0,07	-0,05	0,12	-0,20*	0,07	-0,22***	0,13
DISAVBACKG	0,01	0,06	0,06	0,09	0,00	0,05	-0,01	0,10
TEACHERAGE	-2,71	1,86	2,95	2,97	-3,15***	1,65	-6,09***	3,14
TEACHERSEX	-4,66	2,96	-3,46	4,70	-1,69	2,62	4,40	5,00
R-squared	0,473		0,343		0,429		0,567	
Adj. R-squared	0,319		0,141		0,262		0,440	
Number of schools	142		137†		142		142	

Note: * implies coefficient estimates statistically significant at 99% level; ** coefficient estimates statistically significant at 95% level; *** coefficient estimates statistically significant at 90% level; that † 5 schools not were considered for this analysis because they did not have any *a priori* at risk student in their classes.

Table 4 reports the COLS coefficients for the three considered value-added outputs and for school average results. The results suggest that different school inputs matter for different students. Some remarkable results can be extracted from this analysis in terms of policy making. The estimated coefficients on REGMEETINGS and HMWKSTARTCLASS are negative and significant over all the outputs. Teachers that do not meet regularly to discuss instructional goals would obtain poorer results. Despite the traditional debate about instructional time the negative sign of HMWKSTARTCLASS points out that it is not a good activity to lose the class time starting homework at class. SCHEDMATHS and GRADE 6 present highly significant positive coefficients over all the outputs. The number of hours scheduled weekly for mathematics by the professor had a positive impact on mathematics results. However SCHELOTHERS provide the contrary result, time for preparing other subjects damages value-added output 1. GRADE 6 suggest that better results can be obtained if the teacher is full time at grade 8 and does not teach to other grades simultaneously during the year. A similar result reports the positive sign of MATHS75% for value-added outputs 1, 3 and average output. These outputs augment as far as it is increased the percentage of teachers with three quarters of more of their teaching in mathematics. From these results it seems to be more productive to concentrate teacher activity only in mathematics.

Together with these results several variables are only relevant for some outputs. Therefore, DISRESPONSE, WORKINGROUPS, TIMETEXTBOOK, CORRECTANSWER and REPEATERS have a significant coefficient for the so called value-added output 1. Except for the REPEATERS variable, suggesting that the presence of pupils repeating grade is negative for this output, the remaining variables refers to activities at class. These typical at class habits, say, discuss a response after a wrong answer, percentage of class time following a textbook and to call on another student to give the correct answer after a wrong one display positive coefficients. Nevertheless, students frequently working in pairs or small groups without assistance from the teacher WORKINGROUPS present a significant negative sign. The idea behind this result could be the same named before. Probably time class without teacher surveillance is lost by the students doing *other* activities quite different from study or solving mathematics problems.

REASONING reflects that asking frequently students to explain reasoning behind an idea matter for at risk students. COPYNOTES and TEACHERGUIDE suggest that "traditional learning" with teacher dictating the lesson to the whole class and with students copying notes is not a good methodology to obtain high output levels. On the other hand SPECIFICTEXT points out that it is recommendable that the teacher had a lot of influence to choose the specific mathematics textbook to be used. PRIVATE shows that private schools obtain significant better results for both, value added output 3 and average output. PREESCHOOL also obtains a significant positive result for the same outputs named before for pupils attended preschool. TEACHERAGE presents a negative and significant effect over value-added output 3 and average output. This variable is a proxy for teacher experience showing that young teachers do well in order to increase high performance students at class. Last, despite the high controversy about the variable

RATIO, the analysis does not provide any significant effect for number of student at class over all considered outputs.

5. Conclusions

A comprehensive review of economics of education literature shows that the process of transforming educational inputs in both, student and school levels, into test results is highly unknown. There exist also strong evidence about the presence of non-linear relationships between inputs and outputs. Despite this generalized result most of the studies continue applying conventional analysis at school level over average class results. Achievement is due to student and school but student's socioeconomic background predetermines potential results at school. In the empirical results presented above the presented two-level model provides a tool for researching educational production function in order to improve and rationalize educational policy decisions when allocating scarce resources.

This work leaves open different possibilities for further research. First, more work is needed in order to contrast different non-linear approaches at student level. Second, the measurement of technical efficiency at school level should incorporate both the multi-stage nature and the non-linear effects implicit in the educational process. Last, more research is still necessary comparing parametric and non-parametric efficiency techniques in education.

Acknowledgements

This paper owes a debt to Aurelia Valiño for helpful comments. Financial support from the Spanish Institute for Fiscal Studies (IEF) is also gratefully acknowledged. All detected errors are my own.

References

Baker, B. D. and Richards, C. E., 1999, A comparison of conventional linear regression methods and neural networks for forecasting educational spending. Economics of Education Review, 18, 405-415.

Baker, B. D., 2001, Can flexible non-linear modelling tell us anything new about educational productivity? Economics of Education Review, 20, 81-92.

Bishop, C. M., 1996, Neural networks for pattern recognition, Oxford: Oxford University Press.

Carroll, J., 1963, A model of school learning, Teachers College Record 64, 723-733.

Coleman, J. S. *et al.,* 1966, Equality of educational opportunity (Washington, DC: U.S. GPO).

Eide, E. and Showalter, M. H., 1998, The effect of school quality on student performance: A quantile regression approach. Economics Letters 58, 345-350.

Figlio, D.N., 1999, Functional form and the estimated effects of school resources. Economics of Education Review 18, 241-252.

Funahashi, K., 1989, On the approximate realization of continuous mappings by neural networks, Neural Networks 2, 183-192.

Gonzalez, J. E. and Smith, T. A. (Eds)., 1997, User guide for the TIMSS international database. Primary and middle school years. (Population 1 and Population 2). (TIMSS International Study Center. Chestnut Hill, MA. Boston College).

Gorr, W. L., Nagin, D. and Szczypula, J., 1994, Comparative study of artificial neural network and statistical models for predicting student grade point averages, International Journal of Forecasting 10, 17-34.

Greene, W., 1980, On the estimation of a flexible frontier production model. Journal of Econometrics, 13(1), 101-115.

Hanushek, E. A., 1979, Conceptual and empirical issues in the estimation of educational production functions, Journal of Human Resources, 14, 351-388.

Hanushek, E. A., 1986, The economics of schooling, Journal of Economic Literature 24 (3), 1141-1171.

Hanushek, E. A., 1996, School resources and student performance. In: G. Burtless, Does money matter? Washington DC: The Brookings Institutions.

Hanushek, E.A., 1997, Assesing the effects of school resources on student performance: An update, Educational Evaluation and Policy Analysis. 19, 141-164.

Hornik, K., Stinchcombe, M. and White, H., 1989, Multilayer Feed-forward Networks are Universal Approximators, Neural Networks 2, 359-66.

Hornik, K., Stinchcombe, M. and White, H., 1990, Universal Approximation of an Unknown Mapping and its Derivatives using Multilayer Feed-forward Networks. Neural Networks 3, 551-60.

Hox, J. J., 1995, Applied multilevel analysis, TT-Publikaties, Amsterdam.

Lee, T., White, H. and Granger, C. W. J., 1993, Testing for neglected nonlinearity in time series models: a comparison of neural networks methods and alternative tests. Journal of Econometrics, 56 (3), 269-91.

Levin, H. M., 1974, Measuring efficiency in educational production, Public Finance Quaterly, 2, 3-24.

Mas-Colell, A., Whinston, M., and Green, J. (1995): Microeconomic Theory, Oxford University Press.

Meyer, R.H., 1997, Value-Added indicators of school performance: A primer. Economics of Education Review 16, 283-301.

Meyer, R. H., 2000, Value-added indicators: a powerful tool for evaluating science and mathematics programs and policies, National Institute for Science Education, Brief 3 (3)

Pritchett, L. and Filmer, D., 1999, What education production functions really show: a positive theory of education expenditures, Economics of Education Review 18, pp. 223-239.

Rumelhart, D. E., Hinton, G. and Williams, R. (1986), Learning internal representations by error propagation, in D. E. Rumelhart, J. L. Mc Clelland, and the PDP Research Group (Eds.) vol.I, Parallel Distributed Processing: Explorations on the Microstructure of Cognition, (MIT Press, Cambridge, MA).

Scarselli, F. and A. Chung, 1998, Universal approximation using feedforward neural networks: a survey of some existing methods, and some new results, Neural Networks, 11 (1), 15-37.

Summers, A. A. and Wolfe, B. L., 1977, Do schools make a difference? American Economic Review 67 (4), 639-652.

Zapranis, A. and Refenes, A-P., 1999, Principles of Neural Model Identification, Selection and Adequacy with Applications to Financial Econometrics, Springer.