The reasons behind the progression in PISA scores: An education production function approach using semi-parametric techniques.

April 2014

Ana Balcão Reis*,ᵗ Pedro Freitas*,ᵗ

Nova School of Business and Economics, Universidade Nova de Lisboa
Campus de Campolide 1099-0.32 Lisboa, Portugal

On December 2013, and for the fifth time since 2000, OECD published the results of the latest PISA survey. The existence of data for the last 12 years gives a view on how the students' performance has progressed during this time period.

Using PISA dataset, this paper proposes a methodology to analyze the evolution of the results by the perspective of an education production function, where the output - the performance in the PISA tests - is obtained using three types of inputs, related to the student, the student's parents and the school. We used non-parametric techniques, namely Data Envelopment Analysis (DEA), constructing a production frontier and deriving individual efficiency scores for each of the students who participated in the PISA tests. Focusing on relative efficiency measures and not in absolute scores it was possible to infer on the reasons behind the evolution of the PISA performance. This analysis was performed using data from Portugal, a country that has witnessed a sustained progression in PISA results since 2000 in reading, mathematics and science.

We obtain that the reasons behind the Portuguese PISA progression are neither due to a general higher level of efficiency nor to a higher ability to convert inputs into outputs. Thus the evolution in PISA is explained by a higher level of inputs, particularly related to parents' qualifications and economic comfort at home.

The DEA methodology is extended to measure the sensitivity of students with different economic backgrounds to school inputs and to analyze the impact on efficiency of two particular school policies, class size and school size.

Keywords: Education production function; Efficiency; Data envelopment analysis; PISA

JEL: I21, C61, C67

*E-mail: pedro.freitas@novasbe.pt; abr@novasbe.pt

ᵗ We acknowledge the feedback by Luís Catela Nunes and Carmo Seabra. We thank José Manuel Cordero-Ferrera for the help with Matlab coding. Pedro appreciates the comments from the participants in the research group of Nova Sbe, as well the many suggestions from his Master, Research Master and Phd colleagues namely, Sofia Oliveira, Maria Ana Matias and particularly João Vaz, to whom he thanks the constant support.

1 Introduction

PISA is a set of three tests in the fields of reading, mathematics and science organized by OECD since 2000. Presently there is available data regarding 5 PISA surveys, covering the last 12 years. This raises the question about what has been the evolution of each participating country through the years. The goal of this paper is to address this same issue, disentangling the drivers behind the evolution in PISA scores. To perform such an analysis, an education production function methodology is followed. However, instead of focusing in the evolution of absolute scores, we obtained relative measures of efficiency and compared them across time. To arrive to such efficiency measures, non-parametric techniques were explored, using Data Envelopment Analysis (DEA). In our specific case we have used data from Portugal, which has been a participating in PISA since 2000, showing a consistent, but not constant, evolution during this time period. [1] This first section motivates the paper; section 2 provides a theoretical framework and the methodology; section 3 presents the PISA data used; section 4 gives the results from the Data Envelopment Analysis (DEA) estimations and finally section 5 concludes.

2 Theoretical background and Methodology

This work follows closely the paper by Petra Todd and Kenneth Wolpin (2003), on the education production function where the authors give an exhaustive theoretical exposition on how to model the relation between socio-economic variables and the individual student achievement. Todd's and Wolpin's discussion follows a long literature on what are the determinants of the education production function, where Hanushek (1986, 2003) are reference contributions to this topic. Considering the data used in this work, provided by PISA, we followed a contemporaneous model, where the achievement of a certain student i which lives in the household j is given as:

$$T_{ij} = T(St_{ij}, F_{ij}, S_{ij}) + \varepsilon'_{ij} \tag{1}$$

The student's achievement is measured by the score in a particular test, T_{ij} and it is assumed to be a function of: 1. Student's characteristics, St_{ij}; 2. Family inputs, F_{ij} ; 3. School inputs, S_{ij}; 4. An additive error, ε'_{ij}, which includes all the unobserved characteristics of the individual. In this contemporaneous specification, the inputs are unchanged over time, and since no panel data is available, there is no indication on the past history of inputs.

We add to this specification of the education production function the notion of efficiency. Koopmans (1959) defined technical efficiency has the capacity of maximize the

[1] A graphical summary of the evolution of the Portuguese performance in PISA can be seen in Appendix 7.1.

output conditional on the level of the inputs. The implementability of this concept is brought by Farrell (1957), defining inefficiency as deviations from the frontier isoquant. Thus, the set of efficient points, or as in our case efficient students, are considered the ones who are on the frontier and the inefficient as the ones who are below the frontier.

Using the approaches described above we have derived the education production frontier, settling the three possible main reasons behind the evolution in Portuguese PISA results: 1. There is an overall improving in efficiency, meaning that more students are close to the frontier isoquant; 2. The Portuguese education frontier shifted, namely the ability of frontier students to transform inputs in outputs; 3. The Portuguese students' inputs increase, which consequently lead to a higher output. These three hypotheses are tested, concluding on which ones seem more plausible to explain the Portuguese case over the last decade.

Previous work tried to identify the reasons behind the evolution in PISA results. Some of these studies endured econometric decompositions, such as the Oaxaca-Blinder used by Coutinho (2011), where is assessed the evolution in the Portuguese results between 2003 and 2009, or by Osorio *et al.* (2011) where a similar decomposition is performed, but this time using data from Indonesia.

Here we have step away from these decompositions approaches adopting a frontier analysis. When dealing with frontier measurement, two paths can be followed: 1. A parametric estimation, based on econometric methodologies, like in Coutinho and Moreira (2011) to study the efficiency of the Portuguese secondary schools; 2. A non-parametric approach, whose most recognized method is the data envelopment analysis.

Given the puzzling results often obtained from standard parametric estimations, the approach trough non-parametric approaches has gained attention in the literature, as pointed by Worthington (2001). [2] This non-parametric procedure finds its roots in the founding paper of DEA by Charnes, Cooper and Rhodes (1978), where for the first time is presented this non-parametric approach, following the efficiency concept presented by Farrel (1957). The intensive use of DEA approaches to measure efficiency in education is found over the last decades in works such as the ones by Kirjavainen and Loikkanen (1998) on the efficiency of the Finnish secondary schools, Portela (2011, 2012) on the efficiency of Portuguese secondary schools or Afonso and St Aubyn (2005) on cross country efficiency of secondary education.

The wide majority of the literature applying DEA concepts to education focus on the measurement of school efficiency, taking each school as a DMU, which stands in DEA jargon as Decision Making Unit (DMU) (e.g the entity responsible for converting inputs in outputs).In

[2] The author points the freedom in the relation between inputs and outputs, particularly in education contexts when some of the usual axioms of productivity may breakdown (e.g cost minimization), or the relaxing on the restriction on data, as the main reasons for the attractiveness of non parametric methodologies.

this paper we wanted to look how each student takes advantage of the available inputs. Thus we extend the notion of DMU, centring our analysis at the student level. This approach raises several advantages, since it allows for an efficiency measurement for each individual student, constructing a DEA model with a significant larger sample than the ones normally used. This path of study is much less usual in the literature, with few exceptions such as Waldo (2005), where DEA techniques are applied to student individual data from the Swedish upper secondary schools.

For efficiency measures calculation it was pursed a DEA output oriented model with variable returns to scale based on the following linear programming problem:

$$\underset{\lambda,\theta}{Max}\ \theta_0$$

subject to:

$$\begin{aligned} &\theta_0 y_0 \leq \sum_{j=1}^{n} y_j \lambda_j \\ &x_0 \geq \sum_{j=1}^{n} x_j \lambda_j \\ &\sum_{J=1}^{n} \lambda_j = 1 \\ &\lambda_j \geq 0 \end{aligned} \tag{2}$$

The problem presented above is solved for j DMU's, having each one the outputs y_0 and the inputs x_0. This formulation stands for an output oriented model, where efficiency is read as the maximum amount of output that each DMU is able to generate with the inputs given. An efficient DMU is the one who is on the frontier which happens when $\theta_0 = 1$, meaning that the higher the value of θ_0, the less efficient is the respective DMU. The attractiveness of the DEA approach is that each unit is seen as a linear combination of the most efficient units closest to it, and the weight given to each one of this efficient units is given by the term λ_j. The assumption that the sum of all the λ_j equals one imposes variable returns to scale to our problem, as desired.

3 Data

We have focused in the PISA dataset. PISA has the virtue of having a rich set of variables related to the students, the students' parents and the schools. Despite the diversified information that can be collected from the PISA dataset, when analysing it for several years particular care must be taken. The initial test, 2000, was excluded from our analysis due to the difference between the number of individuals who performed the reading and mathematics test. We remain with 4 PISA samples (2003, 2006, 2009 and 2012). When dealing with PISA, particular attention must be given to existence of missing data. The percentage of missing data by variable seems relatively low, but when combining several variables for empirical estimation, the sample can be significantly reduced (e.g considering the variables chosen to the analysis, the samples are reduced for values between 50% in 2012 and 85% in 2009). [3] In order to address this same problem, we have opted for the implementation of data imputation techniques. Data imputation methodologies can be divided in single and multi imputation. This last one was excluded since we need a single imputation for each missing observation, to build the DEA method followed after. It was applied an Expected Maximization algorithm to fill part of the missing data, allowing that the final imputed samples include at least 90% of the observations in the original sample. [4]

4. Results

Translating this general DEA framework to the specific case of our data, it is assumed a multiple output for each DMU, given by the score achieved in the reading and mathematics test.[5] On inputs choice we have looked for a methodology that decreased the level of discretion. In DEA original formulations no particular methods were advanced to determine what inputs may or may not be relevant in the construction of the production frontier. Previous work, such as Nataraja and Johnson (2011) addressed this same issue. We have adopted a regression based methodology for the inputs choice, where estimation analysis was performed using PISA data and following the general theoretical formulation presented in equation (1). When choosing the variables to be included in our regressions, we took advantage of the indexes released by PISA, which aggregate several answers given by students and schools in the questionnaires. These indexes reveal to be important to define proxies, for example for the student's family economic situation. [6]

[3] Sample size without missing answers: 3462 (2003); 4014 (2006), 5535 (2009); 2831 (2012).
[4] Sample size after imputation: 4252 (2003); 4805(2006); 5984(2009);5178 (2012).
[5] These scores correspond to the average of the 5 plausible values released by PISA in both fields, Reading and Mathematics
[6] Todd and Wolpin (2003) advice for the use of proxies. However its use reveals to be necessary in order to control for the student social and economic background.

Table 4.1[7]

Type of inputs	PISA variable	Mean	Standard Deviation	Min	Max
Student's variables, **(St_{ij})**	Grade[8]	9.34	0.93	7	11
	Track of studies (Measured as the percentage of the curriculum that corresponds to the standard academic track, using the courses syllabus provided by the Ministry of Education)	92.08	20.58	23	100
	Female (dummy=1)	0.5139	0.49	0	1
	Age	15.79	0.29	15	16.25
Family variables, **(F_{ij})**	FISCED (Father's position in the international standard classification of education, scale 0-6)	2.52	2.03	0	6
	MISCED (Mother's position in the international standard classification of education, scale 0-6)	2.49	2.04	0	6
	Parents pt1 (student whose one of the parents were born in Portugal, dummy=1)	0.12	0.33	0	1
	Parents pt2 (student whose both parents were born in Portugal, dummy=1)	0.83	0.37	0	1
	CULTPOS (Index for level of cultural possessions at home)	2.02	0.92	0.49	4.10
	HEDRES (Index for investment in educational resources)	5.19	0.99	0.68	6.33
	HOMEPOS (Index for level of home possession at home)	7.17	0.98	0.58	11.21

[7] Average and standard deviations calculated based on the imputed sample.
[8] *CEF (Cursos de Educação e Formação)*, which are the vocational track of the lower secondary education are not coded in the variable grade. In the observation of the students' scores it was observed a distribution similar to the 7th grade. Thus this the grade was imputed to these observations in the sample.

School variables, (S_{ij})	Percentage of non-governmental financing of the school (%)	16.75	22.5	0	100
	Schools in small towns (dummy=1)	0.31	0.46	0	1
	Schools in towns (dummy=1)	0.4	0.49	0	1
	Schools in cities (dummy=1)	0.16	0.36	0	1
	Schools in large cities (dummy=1)	0.05	0.22	0	1
	Percentage of girls in schools	50.46	5.24	0.6	100
	School size	978.31	474.14	73	2750
	Students/teacher ratio	9.33	4.08	0.885	68
	SCMATEDU (Index for level of quality of school resources)	2.89	0.87	0.26	5.2
	Class size (For 2006 it corresponds to the average class size per school; and for the other years to the number of students reported by the students in their classes)	22.52	4.66	1	53
	Percentage of certified teachers	89.01	18.46	0.6	100
	RESPRES (Index for school autonomy in resources management)	1.99	1.27	0.897	8
	RESPCURR (Index for school autonomy in curriculum)	2.08	1.54	0.66	6
	TSCHORT (Index for teacher shortage)	1.26	0.6	0.79	5.59

OLS regressions were computed, implementing for the variance estimation the Balanced Repeated Replication using the 80 population weights provided in the datasets and a Fay's adjustment of 0.5, as recommended by PISA.

The estimation was run separately for the reading and mathematics scores considering different cases: 1. Just the students without missing answers (2003-2012); 2. The imputed databases without control for year adjustments (2003-2012); 3. The imputed databases controlled for year dummies (2003-2012).

Table 4.2

	5 Plausible values reading/mathematics (2003-2012) Weighted sample					
	(1)		(2)		(3)	
	Reading	**Maths**	**Reading**	**Maths**	**Reading**	**Maths**
Student's variables, St_{ij}						
Female	22.93***	-24.67***	22.59***	-25.31***	22.50***	-25.42***
Age	-17.05***	-14.70***	-17.36***	-14.89***	-15.77***	-13.03***
Grade	52.71***	53.85***	52.32***	53.23***	52.28***	53.13***
Track of studies	0.15***	0.082**	0.15***	0.06	0.17***	0.08**
Family variables, F_{ij}						
Mother – High school	6.62***	8.10***	5.91***	7.23***	6.3***	7.81***
Mother – More than high school	4.91*	10.74***	7.53***	12.55***	8.06***	13.24***
Father – High school	6.08***	7.25***	7.47***	8.55***	67.13***	7.11***
Father – More than high school	7.99***	12.09***	8.64***	12.77***	7.85***	11.84***
Parents pt 1	14.9*	20.32***	17.29**	19.87***	15.72***	17.75***
Parents pt 2	13.01***	19.48***	16.22***	20.15***	15.51***	19.3***
CULTPOSS	7.43***	4.24***	7.55***	4.25***	7.13***	3.66***
HEDRES	1.09	0.57	1.02	0.53	2.08***	1.99***
HOMEPOS	6.40***	8.47***	5.47***	8.08***	5***	7.38***
School variables, S_{ij}						
% Non government financing	0.17***	0.23****	0.15***	0.22***	0.12***	0.19***
Small town	2.54	-1.74	3.87	0.08	2.8	-1.4
Town	3.61	-1.60	4.08	-0.63	3.67	-1.4
City	11.98***	8.01**	11.41***	6.84	11.3***	6.44
Large city	16.25***	2.52	18.91**	3.02	18.92**	3.15
Prop. girls	0.42*	0.3	0.35*	0.3*	0.52**	0.49***
School size	0.07**	0.002	0.007***	0.002	0.07***	0.02
SCMATEDU	2.41**	3.41***	3.06***	3.85***	1.81	2.23***
Std/Teacher ratio	-0.009	0.14	0.06	0.19	-0.0	0.11
Class size	0.20**	0.48***	0.18	0.52***	0.23	0.63***
Prop.Certified teachers	0.14**	0.13**	0.11**	0.10**	0.06	0.04
RESPRES	3.65***	1.91	2.82**	1.34	5.43***	3.54*
RESPCURR	-5.26***	-6.19	-4.99**	-6.08***	-1.3	-1.42
Teacher shortage	-0.79	-0.75	-1.00*	-0.61	-0.96	-0.73
Constant	96.06***	69.84**	122.78***	82.93**	55.4	18.31
Year dummies	No	No	No	No	Yes	Yes
R^2	0.476	0.470	0.478	0.47	0.484	0.48

Statistically significant at *10%, *5%, ***1%

From the table above it is denoted some variability in the results, particularly between the reading and the mathematics score. We decided to include in the frontier estimation the inputs that are significant at least at 10% level in one of the two test scores. Following Bessent and Bessent (1979) it was also imposed that the relationship between inputs and outputs must be such that an increase in inputs is translated in an increase in outputs (e.g. $\beta_i \geq 0$). The inputs measurements must have non-zero elements, and when such cases occurs an infinitesimal value (0.0001) is assumed.[9]

The inputs were classified in two different categories. By one side we had to define those inputs that correspond to the environment where the student lives, and thus are not a choice of the student. In DEA literature these inputs are labelled as non-discretionary, z_j, and in our case correspond to: gender, age, mother's and father's level of education, parents' nationality, Index of Home possessions (HOMEPOS) and geographic location. All the other inputs, x_j, the discretionary, correspond to the ones that are indeed included in DEA linear programming problem. This choice implies that we are considering that the individuals can opt for the grade and track of studies they are enrolled in as well the level of investment in education, reflected by the index cultural possessions (CULTPOSS) or the index of educational resources (HEDRES). It was also assumed the student have information about the schooling system and can choose the school with the characteristics preferred.

5.1 DEA efficiency scores - Results

Three different DEA were run for the samples (1), (2) and (3), considering as outputs the scores in reading and mathematics and as inputs the ones that result from the regression based choice performed above. This heterogeneity reveals as an advantage in our estimation bringing more robustness to the efficiency measures found. In the following table is shown the quartile distribution on the efficiency scores, θ_0, derived:

[9] Some of the indexes in the PISA dataset present negative values, since they are constructed in such a way that the OECD average is 0. DEA standard measurements just assume positive values, therefore positive amounts were added to negative values, following the method by Pastor (1994) and Lovell (1995). Alternatively, other specifications on handling DEA with negative data can be used, as in Portela, Thanassoulis and Simpson (2004).

Table 4.1.1

(1)/(2)/(3)	1st quartile	Median	3th quartile	Mean
2003	1.11/1.12/1.12	1.20/1.22/1.22	1.31/1.33/1.33	1.23/1.24/1.24
2006	1.11/1.13/1.13	1.22/1.22/1.22	1.33/1.34/1.34	1.24/1.25/25
2009	1.11/1.12/1.11	1.22/1.22/1.21	1.33/1.34/1.32	1.23/1.24/1.23
2012	1.12/1.14/1.13	1.22/1.25/1.23	1.34/1.38/1.36	1.24/1.27/1.26

Observing the table above it is noticeable how the efficiency levels of the Portuguese students seem to be constant during the last decade. The results are robust not just over quartiles but when imputed and non imputed data are considered. To study these results in a more decomposed way the evolution of the efficiency levels were studied by school type (private vs public) and by grade:

Table 4.1.2

(1)/(2)/(3)	Private schools	Public schools
2003	1.18/1.18/1.18	1.23/1.24/1.24
2006	1.201.20/1.20	1.25/1.26/1.25
2009	1.19/1.18/1.18	1.24/1.24/1.23
2012	1.19/1.20/1.19	1.25/1.28/1.27

Table 4.1.3

(1)/(2)/(3)	7th grade	8th grade	9th grade	10th grade	11th grade
2003	1.35/1.35/1.34	1.31/1.32/1.32	1.29/1.3/1.3	1.19/1.2/1.2	1.05/1.08/1.08
2006	1.34/1.36/1.35	1.33/1.33/1.32	1.29/1.3/1.29	1.19/1.2/1.2	1.11/1.13/1.12
2009	1.24/1.27/1.24	1.36/1.38/1.36	1.3/1.31/1.29	1.2/1.2/1.18	1.06/1.06/1.05
2012	1.23/1.27/1.24	1.32/1.38/1.36	1.3/1.33/1.29	1.21/1.23/1.18	1.1/1.13/1.06

When desegregating the results it is denoted that the stagnation of the efficiency results seems to be happening across the private and public schools and across the majority of the grades. However the level of efficiency is higher in the private schools as well in those grades which correspond to students who never failed during their academic track (10th and 11th grade).[10]

In all of the three tables above we observe that whatever are the criteria adopted, the efficiency scores remain rather constant. However from 2003 to 2012, we observe a global improvement in the Portuguese scores in PISA tests. Then, we can conclude that the global

[10] To test for the possibility that the efficiency scores are being driven by top achieving students, the three DEA were run once again but just for those students below the percentile 95 both in reading and mathematics tests. The results show that the inefficiency levels are lower, but the stability of the results remain as in the tables presented.

positive evolution in absolute scores is not followed by an improvement in relative efficiency scores.

Additionally, it was observed that, at the individual level, the correlation between the efficiency scores and the average scores in reading and mathematics is negative. Such fact means that higher the score in PISA test, lower the value of θ_0, and then more efficient is the individual.[11]

5.2 *Second stage DEA*

As pointed at the beginning of section 4 we divided the inputs in discretionary and non discretionary. The second stage DEA models intend to infer on how the efficiency measures presented in sub-section 4.1 are more or less affected by the environment the where student lives.

$$\theta_j = f(z_j, \mu_j) \tag{3}$$

Many of these analyses initially considered tobit censured models, since the levels of efficiency of each DMU are bounded to one. These routines are common in the literature, however the methodology to guaranteed the quality of the statistical inference has been subject to attention, particular due to the work by Simar and Wilson (2005) which propose two algorithm specifications to improve the results from a standard tobit censored approach. This method intends to correct for the expectable correlation between the non-discretionary inputs, z_j, and the error term μ_j. The estimations presented followed the first algorithm proposed by the authors[12], however and given the dimension of our sample the number of bootstraps was reduced to 1000. From now one, given the coherence in the estimation results between the non-imputed and the imputed samples, just these last ones (2 and 3) are considered:

Table 4.2.1

	θ_0, **DEA efficiency scores, (2003-2012)** Weighted sample	
	(2)	**(3)**
Female	-0.023*** [-0.032;-0.015]	-0.026*** [-0.036;-0.018]
Age	-0.0012 [-0.015;0.12]	0.005*** [-0.01;0.018]
Mother – High school	-0.032*** [-0.045;-0.021]	-0.032*** [-0.046;-0.021]

[11] A graphical representation of this same relation can be found in Appendix 7.2

[12] A detailed step by step exposition of the algorithm can be seen in Appendix 7.3. The second algorithm proposed by Simar and Wilson was not applied due to the large size of the samples used.

Mother – More than high school	-0.030*** [-0.046;-0.015]	-0.031*** [-0.049;-0.018]
Father – High school	-0.025*** [-0.036;-0.012]	-0.018*** [-0.032;-0.001]
Father – More than high school	-0.017*** [-0.033;0]	-0.08*** [-0.043;-0.018]
Parents pt 1	-.069*** [-0.059;-0.013]	-0.081*** [-0.017;-0.019]
Parents pt 2	-0.090*** [-0.105,-0.063]	-0.010*** [-0.117;-0.074]
HOMEPOS	-0.024*** [-0.03;-0.02]	-0.013*** [-0.02;-0.01]
Small town	-0.026*** [-0.044;-0.008]	-0.023*** [-0.038;-0.005]
Town	-0.018*** [-0.035;0]	-0.016*** [-0.031;0.003]
City	-0.065*** [-0.084;-0.048]	-0.068*** [-0.085;-0.048]
Large city	-0.079 [-0.107;-0.049]	-0.08*** [-0.108;-0.053
Constant	1.53*** [1.327;1.765]	1.36*** [1.149;1.617]
σ_ε	0.207*** [0.203;0.211]	0.20** [0.203;0.212]

Statistically significant at *10%, *5%, ***1%. In brackets, 95% Confidence interval based on Simar and Wilson (2007) 1st algorithm.

From the results presented above we note the high significance that the majority of the non discretionary inputs seem to have on the efficiency scores. It is observed that girls, from high qualified Portuguese parents who live in wealthier families and in cities tend to have better levels of efficiency.

In previous 2 stage methodologies, such in Afonso and St Aubyn (2005), the results from the regressions above are used to correct the initial efficiency scores. We have decided not to follow this reasoning, following Cordero, Pedraja and Santín-Gonzalez (2009), which show, by Monte Carlo simulations, that this kind of approaches behave poorly. [13]

5.3 Malmquist Index – Comparing PISA across the years.

Caves, Christiansen and Diewert (1981) proposed the concept of Malmquist index, allowing for efficiency evaluations of the same DMU across different points in time. This measurement of efficiency evolution across time is particularly appealing for our analysis, since

[13] Quoting the authors: “two stage models obtains the worst results due to its own structure which is focus on identifying external variables that really have influence on the results of production rather than worrying about how to construct a boundary frontier to take them into account”.

we are measuring efficiency in PISA considering several years. However, contrary to standard Malmquist index applications, the data used in this work is not panel data. PISA survey intends to be an accurate representation of the student population, and in each time different samples (with different sizes) of students between 15 and 16 years old are considered. This fact makes us use a modified version of the Malmquist index presented by Camacho and Dyson (2006) [CDMI]. [14] In their formulation the authors intended to compare the efficiency performance across different groups. In our specific case, the samples of each of the years in analysis were considered as different groups and CDMI's were computed for 2003-2006, 2006-2009 and 2009-2012. The Malmquist index has also the interest of allowing to desegregate between two possible sub-components: 1. Catch up component, standing for the measurement of the movement of the DMU'S to the frontier (*efficiency change)*; 2. Frontiers shifts, namely measuring how from one period to the other the technology allows that with the same inputs higher level of outputs are produced *(technical efficiency).* Camacho and Dyson index[15] allows to disentangle between what part of the Malmquist index corresponds to an efficiency change and to what corresponds to technical change:

$$CDMI = \frac{\left(\prod_{j=1}^{Nt} D^t(x_j^t, y_j^t)\right)^{\frac{1}{Nt}}}{\left(\prod_{i=1}^{Nt-1} D^{t-1}(x_j^{t-1}, y_j^{t-1})\right)^{\frac{1}{Nt-1}}} \cdot \left[\frac{\left(\prod_{j=1}^{Nt} D^{t-1}(x_j^t, y_j^t)\right)^{\frac{1}{Nt}}}{\left(\prod_{j=1}^{Nt} D^t(x_j^t, y_j^t)\right)^{\frac{1}{Nt}}} \frac{\left(\prod_{j=1}^{Nt-1} D^{t-1}(x_j^{t-1}, y_j^{t-1})\right)^{\frac{1}{Nt-1}}}{\left(\prod_{j=1}^{Nt-1} D^t(x_j^{t-1}, y_j^{t-1})\right)^{\frac{1}{Nt-1}}} \right]^{1/2} \quad (4)$$

Where D stands for the distance function which can be read as the inverse of the efficiency score previous derived: θ_0. The term outside the squares brackets represents the efficiency change factor, measuring the efficiency gap between groups (in our particular case different PISA samples in different years). The expressions in brackets, represents the change in the frontier between different PISA samples (years). Values close to 1 stand for non evolution in the different components of the CDMI:

Table 4.3.1

(2)/(3)	**Malmquist index**	**Efficiency change**	**Technical change**
2003-2006	≅1/≅1	≅1/≅1	≅1/≅1
2006-2009	≅1/≅1	0.99/≅1	≅1/≅1
2009-2012	≅1/≅1	≅1/≅1	≅1/≅1

[14] CDMI – Camacho and Dyson Malmquist index, following the nomenclature by Crespo-Cebada, Chaparro and Santín (2009).

[15] Note that originally, CDMI is used by Camacho and Dyson (2006) under constant returns to scale. Like Portela, Camacho and Keshvari (2012) we use it a variable returns to scale DEA.

The results show a high stability of the Malmquist indexes and in both of its components, the *technical and efficient* components. These conclusions sustain that from 2003 to 2012 it was not registered significant approaches of the inefficient students to the efficient ones (represented on the frontier), neither were noticed considerable evolutions in the frontier (standing for the ability to convert the same inputs in higher amounts of output).

5.4 Inputs change

The efficiency scores derived trough a DEA approach show that neither efficiency changes nor frontier movements are in the core of the reasons to explain the Portuguese PISA results since 2003. The other reason behind such progression may lay on increases in the input levels. Since the education production function assumes positive marginal returns, this increase in inputs could be the source for the higher scores registered in the PISA tests.

From all the inputs pointed at the begging of section 4, we have focused in those ones which proved to be strongly significant in all of the three regressions performed and whose impact is positive both in the reading and mathematics test. Using the values and respective weights, we have computed the following input quantity index:

$$\text{Input quantity index} = \frac{\sum_{i=0}^{Nt} w_i input_i}{\sum_{j=0}^{Nt-1} w_j input_j} \qquad (5)$$

The index is calculated for N_t students in year t and N_{t-1} students in year t-1. w_i and w_j correspond to the respective final weights attributed by PISA in each year.

If the index is higher than 1, we have an increase in the total amount of input from one year to the other. The opposite conclusion can be inferred if this index reveals to be lower than one.

Table 4.4.1[16]

	Input quantity index		
	2003-2006	**2006-2009**	**2009-2012**
Grade	0.904	1.11	0.9372
Track of studies	0.90	1.05	0.9378
MISCED	0.85	1.39	0.872
FISCED	1.11	1.25	1.046
Parents pt 1	1.08	1.59	2.34

[16] The table presents the results on the imputed sample.

Parents pt 2	0.92	1.05	0.92
CULTPOSS	0.92	1.25	0.90
HOMEPOS	0.94	1.16	0.90
Non government financing (%)	0.93	1.02	1.10
SCMATEDU	0.83	1.20	1.06

From the table above, we can identify the evolution of the input quantity index from 2003 to 2012. It is observed that from 2003 to 2006 we have a general drop in the index, denoted by the decrease in the input quantity from one period to the other. The opposite movement is seen from 2006 to 2009, with all the selected inputs increasing. Finally from 2009 to 2012 we observe a mixed behaviour with some increases and some drops in index.

The evolutions just described must be compared with the own evolutions of the PISA scores during the period considered, evaluating in which way the input variations match, or not, the PISA scores evolution:

Table 4.4.2

	2003-2006	2006-2009	2009-2012
Δ Reading score (average 5 plausible values)[17]	-6.6	17.5	2.3
Δ Maths score (average 5 plausible values)	-0.9	20.5	4.2

It is denoted that from 2003 to 2006 the PISA scores fell, which was followed by a large increase from 2006 to 2009 and to a more modest evolution from 2009 to 2012. The pattern of this evolution seems to fit in the pattern of the input progression previously described.

The comparisons presented give important leads in the explanation of the Portuguese student results between 2003 and 2012. Combining the evolution of inputs with the facts concluded from the DEA computations, it is plausible to state that the reasons behind the evolution of the Portuguese PISA scores is more related to input variations than to improvement in students' efficiency.

5.5 More lessons from DEA

Given the non-parametric estimations previous described, we now build on previous results to study two particular issues related to school inputs: 1.We intend to capture the sensitivity of the individual efficiency scores previous derived to the inclusion, or not, of inputs related to the schools; 2. Given the stagnation of the efficiency through the last decade and in a

[17] These scores do not necessarily match the ones officially reported by PISA, since even after the imputation not all the PISA sample was used.

policy orientated perspective, we analyze the impact on efficiency of two particular school policies: school size and class size.

4.5.1 Sensitivity analysis to school inputs

In previous uses of DEA in educational contexts, as in Kirjavainen and Loikkanen (1998) or Montén and Tater (2010), when deriving the efficiency scores the authors assume several models, with different combinations of inputs and outputs. In this section we re-run the DEA using samples (2) and (3) without including school inputs and observe how the efficiency scores change. We consider two efficiency scores: θ_0 (all inputs) and θ_0 (no school inputs), and measure the difference between them as:

$$\Delta\theta_0 = \theta_0 \text{(all inputs)- } \theta_0 \text{(no school inputs)} \qquad (6)$$

If the difference between the two efficiency scores is negative it means that the efficiency score considering all the inputs is closer to one, and then the respective DMU is more efficient.

When deriving, $\Delta\theta_0$ we observe that for samples (2) and (3), it takes almost always negative values, as it is summarized in the distribution of its values by quartiles:

Table 4.5.1.1

	1st quartile	**Median**	**3th quartile**	**Mean**
(2)	-0.142	-0.083	-0.04	-0.10346
(3)	-0.131	-0.075	0.034	-0.0938

We also noted that these differences vary according to the score obtained by the student in the PISA tests, particularly lower the PISA score, higher is the difference between the two efficiency scores, $\Delta\theta_0$.[18]

The perception of such differences, raises the interest on how the difference between the two efficiency scores may be explained by the environmental inputs previously defined, z_j:

$$\Delta\theta_j = f(z_j, \mu_j) \qquad (7)$$

Then the regression (5) is run by a OLS, considering the individual final weights provided by PISA:

[18] A graphical representation of this same relation can be found in Appendix 7.4

Table 4.5.1.2

	θ_0, **DEA efficiency scores (2003-2012), (2003-2012)** Weighted sample	
	(2)	(3)
Female	0.016***	0.011***
Age	-0.005***	-0.001***
Mother – High school	-0.004	0.0001
Mother – More than high school	0.005***	0.009***
Father – High school	0.008***	0.007***
Father – More than high school	0.012***	0.013***
Parents pt 1	0.018***	0.006***
Parents pt 2	0.009***	0.008***
HOMEPOS	0.005***	0.002***
Small town	0.03***	0.006***
Town	0.02***	0.037***
City	0.02***	0.027***
Large city	-0.13***	0.024***
Constant	-0.13***	-0.12***
$\mathbf{R^2}$	0.0678	0.0532

Statistically significant at *10%, *5%, ***1%.

From the table above, the positive coefficients in variable such as HOMEPOS, for example, indicate that individuals whose level of economic possessions at home is higher have a less negative value of $\Delta\theta_0$. This means that higher the economic condition of the student less is the impact in the efficiency scores from adding the school inputs.

Such sensitivity analysis, provides relevant insights on who are the students whose efficiency scores suffer a larger impact if we consider the inputs related to the school context.

5.5.2 School inputs – Class size and School size

In this last section we aim to analyze how the efficiency scores previous derived can contribute to the discussion of particular school policy issues. To perform this analysis we need to extend our DEA model to incorporate the concept of slacks. In section 4, the linear programming problem was solved using a one Phase DEA problem. In this unique phase, we approach the concept of radial efficiency, which is reflected by the efficiency score, θ_0, whose values higher than 1 suggest that the outputs can be expanded given the inputs available. However this definition does not cover another dimension of efficiency related with inputs excess and outputs shortfalls, normally designated in DEA literature as *slacks*:

$$s^{-} = x_0 - \sum_{j=1}^{n} x_j \lambda_j \qquad (8.1)$$

$$s^{+} = \sum_{j=1}^{n} y_j \lambda_j - \theta_0^{*} y_0 \qquad (8.2)$$

(8)

In our case we focus exclusively in the inputs slacks, given in equation 8.1. These inputs slacks stand for the amount that a certain input could be reduced without hurting the output of the individual, and are normally treated in DEA literature as a second source of inefficiency. In our case we step away from seeing these input slacks as an additional measure of inefficiency[19] and we use them as a way to identify in our PISA samples who are the students who have higher slacks in the inputs related to school size and class size. Thus, higher the slack value means that the student maybe enrolled in a larger or smaller class or in a larger or a smaller school that his performance remains constant.

We choose to study the slacks related to these two inputs, since they have been particularly important in the discussion of school policy. In the Portuguese case school size is a particular relevant question given that during the last years many school with less of 21 students have been closed, and school administration has been concentrated all over the country. [20] The class size input was chosen given it is for long a topic of discussion in what concerns school policy. [21]

To obtain these slack variables, it was necessary to run the linear programming problem, but now a two phase problem was considered, adding the slacks measurement to the previous specification in equation (2):

[19] This source is normally labelled as "non-radial efficiency" or "mix inefficiency".
[20] According to 2010 data from the Portuguese ministry of education, since 2005 3.200 schools were closed, mainly elementary schools in small villages in the country. In 2013, 67 large administrative bodies were created, whose number of students under their responsibility can amount to more than 3.500.
[21] According to OECD 2012 Education at a glance on Portugal between 2000 and 2010 the average class size decreased 34% in the country.

$$\underset{\lambda,\theta,s^-,s^+}{Max} \quad \theta_0 + \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right)$$

subject to:

$$\begin{aligned} &\theta_0 y_0 \leq \sum_{j=1}^{n} y_j \lambda_j + s_i^- \\ &x_0 \geq \sum_{j=1}^{n} x_j \lambda_j - s_r^+ \\ &\sum_{J=1}^{n} \lambda_j = 1 \\ &\lambda_j, s_i^-, s_r^+ \geq 0 \end{aligned} \tag{9}$$

This problem was solved for m inputs, s outputs and ε, the "non-Archimedean infinitesimal", being $\varepsilon > 0$.

Collected the slacks associated to the school size and class size inputs, the propose was to observe how the environmental variables, z_j, may affect them. The study of the slacks seems to receive less attention in DEA literature, with a particular remark for the work by Fried, Lovell, Schmidt and Yaisawarng (2002). Given that slacks are bounded at zero the first intention would pass by the use of a Tobit censored model. However, and once again, recalling the work from Simar and Wilson (2005), the authors point that the approach by them proposed can be applied to cases where slacks are the dependent variable. Again the 1st algorithm by Simar and Wilson was used to perform the following estimation: [22]

$$s_j^- = f(z_j, \mu_j) \tag{10}$$

The results from this estimation using both samples can be observed in the following table:

	Dependent variable: s_i^-, Input slacks, (2003-2012) Weighted sample			
	School size slack		**Class size slack**	
	(2)	**(3)**	**(2)**	**(3)**
Female	2.82 [-41.35;44.78]	-81.23*** [-145.73;-15.26]	0.40*** [0.259;0.549]	0.56*** [0.413;0.716]
Age	38.70*** [-35.81;113.2]	182.47*** [74.54;311.875]	-0.27*** [-0.523;-0.019]	-0.49*** [-0.757;-0.243]
Mother – High school	-41.20*** [-98.94;22.07]	-158.52*** [-271.15;-47.47]	-0.43*** [-0.68;-0.205]	-0.35*** [-0.615;-0.125]

[22] The detailed explanation of Simar and Wilson 1st Algorithm is in Appendix 7.3, with a note for its use in the slacks case

Mother – More than high school	-20.58*** [-82.87;30.49]	-135.96*** [-251.857;-15.46]	-0.27*** [-0.541;-0.011]	-0.35*** [-0.617;-0.07]
Father – High school	50.20*** [-23.021;124.099]	126.14*** [14.88;223.05]	0.26*** [0.036;0.492]	0.46*** [0.232;0.703]
Father – More than high school	51.19*** [-20.14;122.05]	244.12*** [104.592;350.02]	0.55*** [0.269;0.825]	0.84*** [0.558;1.126]
Parents pt 1	171.06*** [-54.243;224.12]	585.84*** [66.787;648.06]	0.052 [-0.466;0.477]	0.07*** [-0.466;0.493]
Parents pt 2	305.50*** [159.06;421.77]	759.77*** [296.08;733.28]	0.27*** [-0.143;0.690]	0.04 [0-0.382;0.467]
HOMEPOS	65.65*** [41.26;90.76]	22.54*** [-10.835;67.87]	0.08*** [-0.002;0.172]	0.14*** [-0.054;0.235]
Small town	-169.17*** [-265.28;-67.21]	-199.04*** [-290.3;-95.25]	0.38*** [0.094;0.719]	0.09*** [-0.208;0.425]
Town	245.12*** [150.21;343.94]	120.16*** [-37.38;258.59]	0.99*** [0.663;1.309]	0.45*** [0.11;0.774]
City	468.57*** [362.46;571.25]	570.02*** [406.21;721.65]	0.66*** [0.31;1.017]	0.34*** [-0.5;0.699]
Large city	136.45*** [-5.041;275.75]	-4.98 [-231.183;205.104]	-0.14*** [-0.65;0.33]	-0.4*** [-0.924;0.008]
Constant	-1779.73*** [-2987.7;-599.51]	-4810.81*** [-6900.92;-2977.69]	6.1** [2.27;10.044]	9.86*** [5.933;13.878]
$\boldsymbol{\sigma_\varepsilon}$	732.84*** [707.82;762.49]	931.47*** [844.90;990.85]	3.76*** [3.69;3.83]	3.90*** [3.826;3.977]

Statistically significant at *10%, *5%, ***1%. In brackets, 95% Confidence interval based on Simar and Wilson (2007) 1st algorithm.

The table present that individuals whose level of Home possessions is higher, live in families where at least one of the parents is Portuguese and where fathers are more qualified tend to have higher levels of slacks, both in what concerns the school size and the class size. From the coefficient in the variable regarding small towns is observable that students who live in smaller locations are less indifferent to a higher school size. Also to note the different signs of the coefficients for mother's and father's qualifications, suggesting different impacts from the parents' level of schooling.

The use of the results presented may give interesting policy recommendations, identifying who are the students that may be more or less indifferent to variations in these two specific school inputs.

5. Final Remarks

Every three years when PISA results are released they constitute a theme of discussion about how the students' performance and schooling systems have been changing over time.

Our intention was to contribute to this debate proposing a methodology to disentangle such evolution, proposing three possible drivers for the progression in PISA scores. We have departed from a production function framework, where the score obtained by the student in the PISA tests are seen as the output, which is a function of several socio-economic inputs. This production function is then evaluated through an efficiency perspective, identifying the students' distribution around the efficient frontier. The analysis was performed through a non-parametric methodology, using Data envelopment analysis.

This methodology was applied to PISA data on Portugal since 2003, concluding that the higher scores obtained by the Portuguese students are neither due to an higher individual level of efficiency or to a change in the frontier, indicating no progression in the way Portuguese students transform the inputs they are given into better academic performance. After detailed analysis of the inputs evolution, it was observed that a higher level of inputs, namely parent's qualification, level of cultural and home possessions and the quality of the school resources are on the core of the evolution registered in the Portuguese PISA scores. This simple fact does not take away the merit from the Portuguese evolution in PISA, but it raises new questions about how we should use the PISA results to measure the true evolution of the capacity of the students to generate higher school results. From a policy evaluation perspective, if by one side the investment in school resources appears to contribute to a better school performance, by the other side the stagnant average efficiency scores show that the schooling system has not generated more efficient students during the period analyzed.

DEA techniques were also used to take more conclusions about the overall behaviour of the Portuguese educational system. Particularly, the school inputs seem to be able to boost the efficiency of those students whose socio-economic conditions are more challenging. We focused as well in two particular school polices (school size and class size). The results provide some indications on how different students are affect by these two school recourses, which can be an important reference for the policy maker when addressing the distribution and allocation of these resources.

6. References

Afonso, A, St. Aubyn, M. "Cross country Efficiency of Secondary Education. A semi-parametric Anlysis with Non-discretionary inputs", ECB Working Paper, 2005

Aderson P., Peterseon, N., "A Procedure for Ranking Efficient units in Data Envelopment Analysis", Management Society, 1993

Besset A, Besset W. "Determining the comparative efficiency of schools through data envelopment analysis", Center of Cybernetic Studies, 1979

Camacho, A., Dyson, R. "Data Envelopment Analysis and Malmquist indices for measuring group performance", Journal of Productivity Analysis, 2006

Charnes, A. Cooper, W., Rhodes,E. "Measuring the efficiency of decision Making units", European Journal of Operational Research, 1978

Cooper, W., Seinford, L., Tone, K., " Data Envelopment Analysis: A comprehensive Text with Models, Applications, References and DEA-Solver Software", Springer, 2nd Edition, 2007

Cebada-Crespo, Eva, F., Santín-Gonazález, D., "Comparing school ownership performance using a pseudo-panel database: a Malmquist-type index approach", 2009

Cordero-Ferrera, J., Pedraja-Chaparro, F., Santín-Gonazález, D., " Alternative Approaches to Include Exogenous Variables in DEA Measures: A comparison using Monte Carlo", Computers & Operations Research, 2009

Coutinho Pereira, M., "An analysis of the Portuguese Students' Performance in the OECD Programme for International Student Assessment (PISA)", Economic Bulletin, Banco de Portugal, 2011

Coutinho Pereira, M., "An analysis of the Portuguese Students' Performance in the OECD Programme for International Student Assessment (PISA)", Economic Bulletin, Banco de Portugal, 2011

Färe, R., Grosskopf, S., Zhang, Z. "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries", The American Economic Review, 1994

Farrell, M. "The measurement of productivity efficiency", Journal of Royal Statistics Society" 1957

Fried, H., Lovell, C., Schmidet S., Yasaiawarng S., "Accounting for environmental Effects and Statistical Noise in Data Envelopment analysis", Journal of Productivity Analysis, 2002

Hanushek, E, "The economics of schooling: production and efficiency in public schools", Journal of Economics Literature", 1986

Hanushek, E, "The failure of input-based schooling policies", Economic Journal, 2003

Johnson, A.,Ruggiero, J, "Non parametric Measurement of Productivity and Efficiency in Education", Non parametric Measurement of Productivity and Efficiency in Education, 2011

Karjavainen, T., Loikkanen, H., "Efficiency Differences of Finnish Senior Secondary Schools: An application o DEA and Tobit Analysis", Economics of Education Review, 1998

Lovell, C., Pastor, J. "Units invariant and translation invariant DEA models", Operations Research Letters, 1995

Mancebón-Torrubia, M., Calero, J., Choi, A., Ximénez, D., "Efficiency of public and publicly-subsidized high schools in Spain, Evidence from PISA 2006", MPRA Paper, 2010

Montén. A., Thater, C., "Determinants of child care provision", IFO working paper, 2010

Nataraja, N, Johnson, A., "Guidelines for Using Variable Selection Techniques in Data Envolpment Analysis", European Journal of Operational Research, 2011

Pastor, J. "How to discount environmental effects in DEA: an application to bank branches" Working paper N. 011/94, Depto. De Estadistica e Investigacion Operativa, Universidad de Alicante, Spain, 1994

Portela, M., Thanassoulis, E., "Malmquist-type indices in the presence of negative data: An application to bank branches", Journal of Banking and Finance, 2010

Portela, M., Thanassoulis, E., Simpson G. " Negative Data in DEA: A directional Distance Approach Applied to Bank branches", Journal of the Operational Research Society, 2004

Portela, M., Camacho, A., Keshvari, A, " Assessing the evolution of school performance and value added: trends over four years", Journal of Productivity Analysis, 2012

Sena, V. "The Frontier Approach to the Measurement of Productivity and Technical Efficiency", 2003

Simar, L., Wilson, P., "Estimation and inference in two-stage, semi-parametric models of production process", Journal of Econometrics, 2005

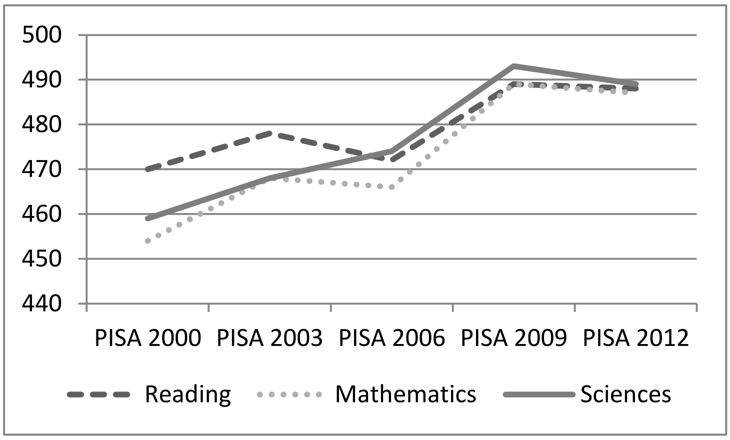
Todd, P., Kenneth, W., "On the specification and estimation of the production function for cognitive achievement", The Economics Journal, 2003

Waldo, S., "On the use of student data in efficiency analysis - technical efficiency in Swedish upper secondary school", Economics of Education Review, 2005

Worthington, Andrew, "An empirical Survey of Frontier Efficiency Measurement Techniques in Education", Education Economics, 2001

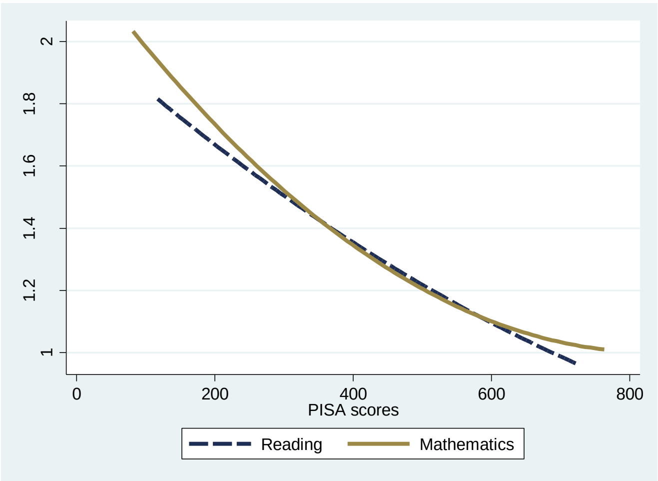
7. Appendix

7.1

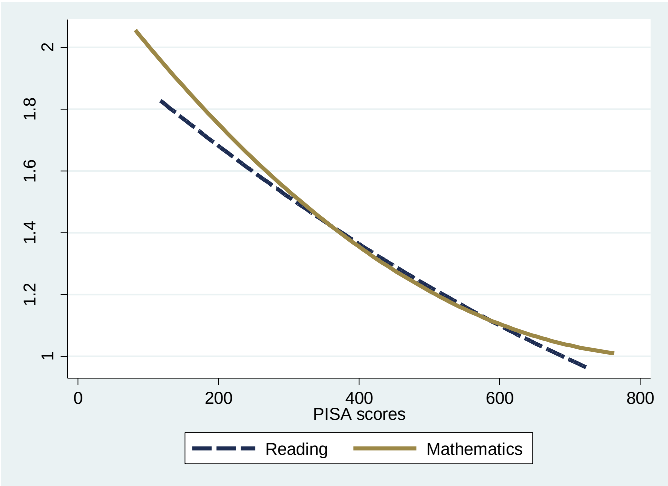


Portuguese average PISA scores 2003-2012

7.2



Quadratic fitting between efficiency scores and PISA scores in mathematics and reading, sample (2)



Quadratic fitting between efficiency scores and PISA scores in mathematics and reading, sample (3)

7.3

1) Estimation of $\hat{\theta}_j$ for all DMU'S considering the original data.

2) By maximum likelihood obtain estimator $\hat{\beta}$ and $\hat{\sigma}_\varepsilon$ from the truncated regression model, $\hat{\theta}_j = f(z_j, \mu_j)$.

3) Loop over the next three steps for L bootstrap estimates for β and σ:

   3.1 For each of the j DMU's draw μ_j from N(0, $\hat{\sigma}_\varepsilon^2$) distributed with left truncation at 1-$z_j\hat{\beta}$

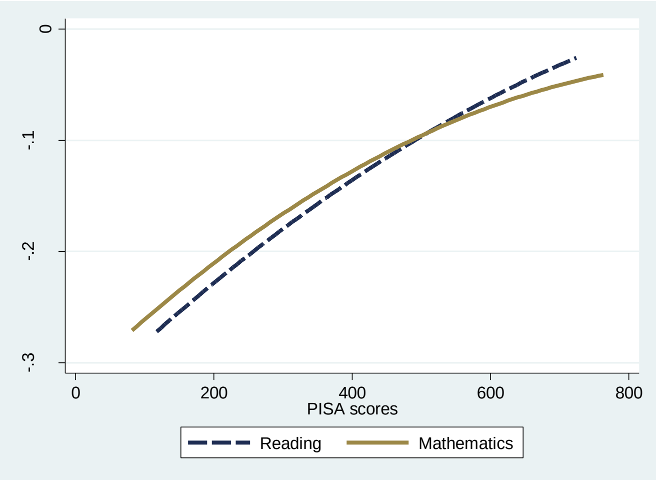
3.2 Compute $\theta_j^* = z_j \hat{\beta} + \mu_j$ for each DMU j

3.3 Coming back to the maximum likelihood truncated regression to estimate $\theta_j^* = f(z_j, \mu_j)$, providing the bootstrap estimates, $\hat{\beta}^*$ and $\hat{\sigma}_\varepsilon^*$

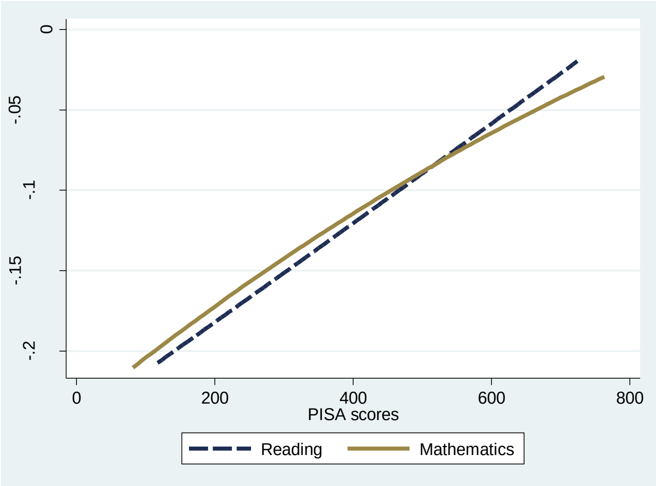
4) Using the bootstrap estimates and the original $\hat{\beta}$ and $\hat{\sigma}_\varepsilon$ to construct confidence intervals for β and σ.

Note: When using slacks as the depend variable, instead of, $\hat{\theta}_j$, the left truncation point is $z_j \hat{\beta}$.

7.4



Quadratic fitting between $\Delta\theta_0$ and PISA scores in reading and mathematics, sample (2)



Quadratic fitting between $\Delta\theta_0$ and PISA scores in reading and mathematics, sample (3)