WORKSHOP ON EFFICIENCY IN EDUCATION

Lancaster University Management School

The Work Foundation, London, September 19-20

"Assessing European primary school performance through a conditional nonparametric model"

Jose M. Cordero, *University of Extremadura, Spain*
Daniel Santín, *Complutense University of Madrid, Spain*
Rosa Simancas, *University of Extremadura, Spain*

OUTLINE

- ✓ Motivation
- ✓ Methodology
- ✓ Dataset and variables
- ✓ Results
- ✓ Concluding remarks

Motivation

- Comparative or cross-country studies have a long tradition within the educational research literature.
- They have become more popular since the development of international large-scale studies such as PISA, TIMSS or PIRLS, which facilitate the comparison by providing homogenous data.
- Most studies are focused on establishing relationships between school-related variables or students´ background and student achievement using the entire world as a laboratory (Bray & Thomas, 1995).
- These studies mainly use an econometric approach to estimate an educational production function with multiple covariates.
- Other empirical analysis use a similar approach to study specific aspects such as the differences among public and private schools or the influence of tracking.

Previous research

	PISA dataset	TIMSS-PIRLS datasets
Production function	Ammermuller (2005) Marks et al. (2006) Woessman et al. (2009) Hanushek & Woessman (2011) Le Donné (2014)	Ammermuller (2005) Woessman (2003)
Public vs. private	Vandenberghe & Robin (2004) Dronkers & Roberts (2008) Woessman et al. (2009)	
Accountability (central exams)	Woessman (2005) Fuchs & Woessman (2007) Bol et al. (2014)	Bishop (1997) Woessman (2003) Hanushek & Woessman (2010)
Tracking	Hanushek & Woessman (2006) Brunello & Checchi (2007) Raitano & Vona (2011) Bol et al. (2014)	Hanushek & Woessman (2006) Schuetz et al. (2008) Dupriez et al. (2008)
Instructional time	Lavy (2012) Rivkin & Schiman (2013)	

Motivation

- Most studies do not take into account the potential existence of inefficiency in the performance of schools.
- In times of economics crisis, improving efficiency has become a top priority in educational systems around the world
- All countries are interested in identifying which factors should be increased or reduced to improve efficiency in the education sector
- Many studies have attempted to measure efficiency of students, schools or districts within one specific country or state.
- However, few studies have used frontier methods to assess the performance of educational systems using a cross-country approach.

Educational efficiency at country level

	PISA data	TIMSS data
Student level	De Jorge and Santín (2010) Deutsch et al. (2013)	
School level	Wilson (2005)	
Country level	Afonso and StAubyn (2006) Thieme et al. (2012) Aristovnik (2013, 2014)	Gimenez et al. (2007)

Methodology

- Most of them use DEA with the exception of Deutsch et al. (2013) → COLS
- Dealing with external variables → Two-stage approach (only in some cases)
- None of them estimate efficiency scores considering at the same time external variables

How will we proceed?

- We use PIRLS data about 4th grade students in 16 European countries.
- We use order-*m* partial frontiers to measure efficiency of schools operating in different countries and a metafrontier approach to decompose overall inefficiency between school and country effects.
- Subsequently, we use a conditional nonparametric model to test the potential influence of different variables on inefficiency and obtain efficiency scores including data about external factors at school and country levels.
- Some previous studies have used these models to measure efficiency in education: De Witte *et al*. (2010); Cherchye et al. (2010); Haelermans & De Witte (2012); De Witte & Kortalainen (2013); Thieme *et al*. (2013); Cordero et al. (2014).
- However, all these studies are focused on a specific country instead of considering a cross-country approach.
- Our contributions:
  - Using these methods in a cross-country analysis
  - Considering both school and national factors affecting inefficiency

Methodology

- We use a fully nonparametric approach → FDH (Deprins et al., 1984)
- We apply the robust order-m approach (Cazals et al., 2002; Simar, 2003)

$$\lambda_m = E\left[\min_{i=1,..,m} \left\{ \max_{j=1,..,p} \left(\frac{x_i^j}{x^j} \right) \right\} \middle| y_i \geq y \right]$$

- This procedure is repeated B times resulting in multiple measures ($\hat{\lambda}^1_{mi},...,\hat{\lambda}^B_{mi}$) from which the final order-*m* efficiency score for each DMU is computed as the simple mean ($\hat{\lambda}_{mi}$)
- It does not include all the observations, thus it is less sensitive to the presence of extreme values or noise in the data
- When $\hat{\lambda}_{mi} < 1$ the evaluated observation can be labelled as super-efficient, since the order-*m* frontier exhibits lower levels of outputs than the unit under analysis

Methodology

- Efficiency scores estimated using the order-m approach might be explained by the school environment, but there might also exist specific features in each country that can affect the results
- Decomposition → Metafrontier approach developed by Battese & Rao (2002), Battese *et al.* (2004) and O`Donnell *et al.* (2008)
- A metafrontier is defined as the boundary of the unrestricted technology set that envelops each of the separate local frontiers (one for each country)
- This technique allows us to decompose the performance of each school into a part attributable to the country and a part attributable to the school (Silva-Portela and Thanassoulis, 2001):

$$OE = SCE \times CNE$$

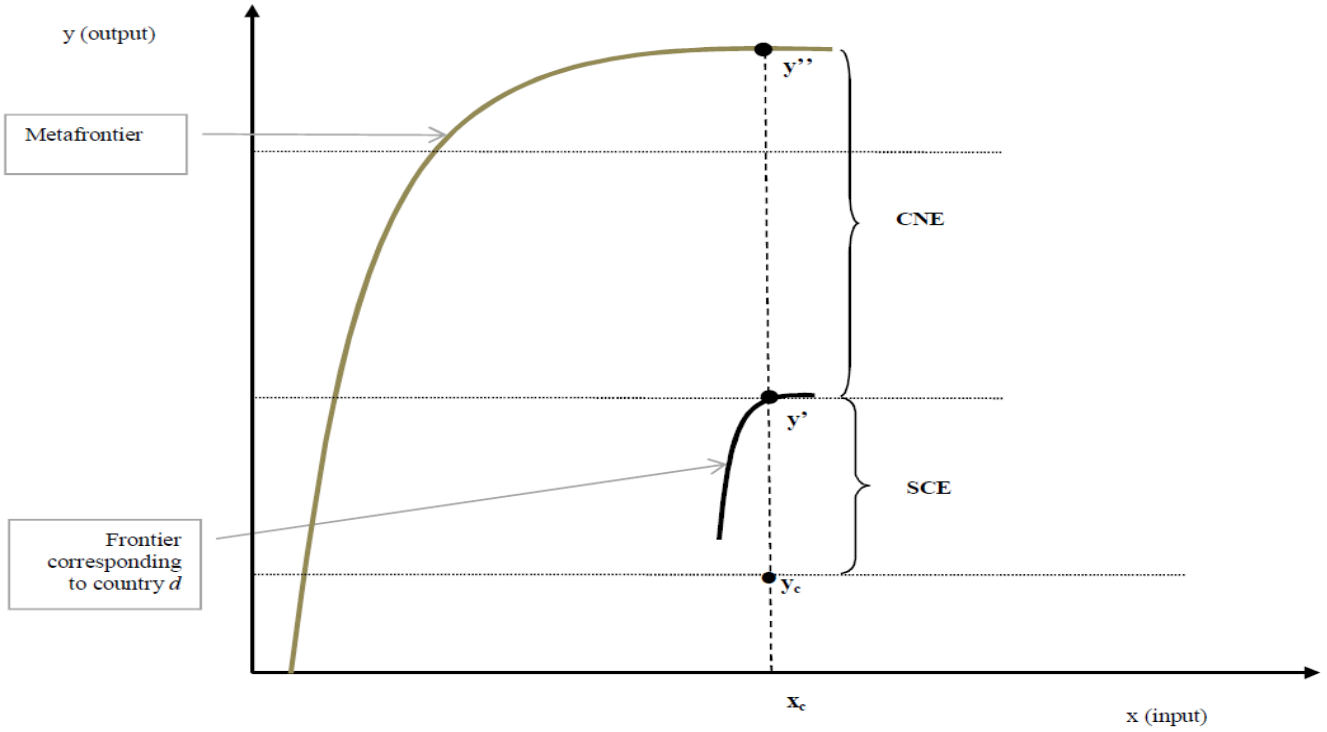
OE → Overall efficiency (distance to the metafrontier)

SCE_1 → School efficiency (distance to the local frontier)

CNE_1 → Country effect (distance between overall and local frontier)

METHODOLOGY

Figure 1. Metafrontier illustration (decomposition of school and country effect)



Conditional nonparametric approach

- ✓ To test the potential influence of factors affecting the performance of schools in different countries we use a conditional nonparametric approach (Cazals et al., 2002; Daraio & Simar, 2005; 2007)

$$H_{XY|Z}(x, y|z) = \Pr(X \leq x, Y \geq y|Z = z)$$

- ✓ This function represents the probability of a unit operating at level (x, y) being dominated by other units facing the same environmental conditions z.
- ✓ The conditional order-m estimator can be defined as

$$\hat{\lambda}_m(x, y|z) = \int_0^\infty \left[1 - (1 - \hat{S}_{Y|X,Z}(uy|x, z))^m\right] du$$

- ✓ The estimation of $S_Y(y|x, z)$ requires to use smoothing techniques for the exogenous variables in z (nonparametric kernel functions) to select the appropriate reference partners and an 'optimal bandwidth' h.
- ✓ This process is even more difficult if we want to test the influence of discrete variables in addition to continuous variables.

Methodology

- ✓ We use the model proposed by De Witte & Kortalainen (2013), which consists of multiplying different kernel functions (one for each type of variable) to obtain a generalized product kernel function.
- ✓ The conditional approach also allows us to evaluate the direction of the effect of exogenous variables on the production process by examining the scatter plot of the ratio between conditional and unconditional measures and its smoothed nonparametric regression line:

$$Q^z = \hat{\lambda}_m(x, y|z) / \hat{\lambda}_m(x, y)$$

  - Increasing line (favorable effect)
  - Decreasing line (unfavorable effect)

- ✓ Moreover, it is also possible to test the statistical significance of Z variables using nonparametric regression significance tests (Racine and Li, 2004)
- ✓ Those tests can be interpreted as the nonparametric equivalent of standard t-tests in OLS regression

Dataset

- ✓ Dataset: European countries participating in PIRLS 2011
- ✓ The sample of schools in each country ranges from 100 to 200
- ✓ The survey collects data from students, parents, teachers and principals.
- ✓ Some limitations:
  - In some countries data about home background is not provided (e.g. US, England, Ireland)
  - An index compiling students´ socioeconomic background (e.g. ESCS in PISA) is not available
  - Budgetary data about schools are not included

Variables

- ✓ The selection of input and output variables is the most difficult decision to make in this type of analysis
- ✓ Outputs → test scores
- ✓ There is a certain level of consensus about two input categories:
  - Characteristics of pupils (prior attainment or socioeconomic background)
  - School resources (capital, number of teachers, expenditure per pupil)

<u>OUR SELECTION</u>

Outputs

- Test scores in Reading (PIRLS)

Inputs:

- Index of students´ abilities before starting at school
- Ratio of computers per 100 students (proxy for capital resources)
- Ratio of teachers per 100 students (proxy for human resources)

Variables

- ✓ **School contextual variables (MODEL 1)**
  - Instructional time (continuous)
  - Disciplinary index (continuous)
  - Level of absenteeism (discrete ordered)
  - Level of parents´ involvement at home (discrete ordered)
  - Level of parents´ involvement in school (discrete ordered)
  - More than 50% of students from a disadvantaged background (dummy)
  - School placed in a rural area (dummy)

(*) Other factors such as the ownership (public or private), the role of innovations or the percentage of immigrants were not considered due to the lack of data about them in PIRLS

Variables

- ✓ **Institutional or country variables (MODEL 2)**
  - *Economic factors*
    - GDP per capita
    - Expenditure in education (% GDP)
  - *Cultural values* (source: pool data from *World Value Surveys*) regarding some qualities that are considered as the most important for children.
  - People had to choose up to five from the following list: 'independence', '**hard work'**, 'feeling of **responsibility'**, 'imagination', 'tolerance', 'thrift', '**perseverance'**, 'religious faith', 'unselfishness' and 'obedience'.
  - According to Heckman (2011) we choose the variables related with 'Factor conscientiousness' that is related with school performance (openess to experience, extraversion, Agreeableness and Emotional Stability complete 'The Big Five' dimensions related with personality).

RESULTS

- ✓ We adopt an output orientation, since we consider that schools attempt to maximize the attainment of their students.
- ✓ We select m=100 and B=200 for statistical inference
- ✓ Multiple steps in the estimation

1. UNCONDITIONAL MODEL

- Estimation of the unconditional order-m metafrontier.
- Estimation of multiple unconditional frontiers for each country.
- Decomposition of efficiency in school and country effects.

2. CONDITIONAL MODEL

- Estimation including only school variables (Model 1)
- Estimation including both school and country variables (Model 2)
- Estimation of nonparametric regression significance tests and analysis of scatter plots

RESULTS

Countries	Overall Efficiency	School Effect		Country Effect	
	Mean	Mean	%	Mean	%
Finland	1.0673	1.0327	48.53%	1.0341	50.65%
Netherlands	1.0995	1.0456	45.80%	1.0515	51.77%
Czech Republic	1.1088	1.0494	45.40%	1.0574	52.71%
Hungary	1.1129	1.1088	96.41%	1.0042	3.74%
Germany	1.1210	1.0805	66.54%	1.0385	31.79%
Sweden	1.1255	1.0538	42.90%	1.0685	54.57%
Italy	1.1258	1.0968	76.98%	1.0276	21.96%
Bulgaria	1.1276	1.0837	65.60%	1.0415	32.54%
Lithuania	1.1393	1.0423	30.39%	1.0939	67.38%
Poland	1.1459	1.0773	48.44%	1.0643	47.08%
Slovenia	1.1546	1.0581	50.00%	1.0918	41.57%
France	1.1646	1.081	35.28%	1.0786	55.75%
Spain	1.1802	1.0902	44.95%	1.0832	43.61%
Norway	1.1807	1.0436	49.90%	1.1316	46.07%
Romania	1.1881	1.0784	23.16%	1.1033	69.97%
Georgia	1.2277	1.0742	34.41%	1.1433	45.38%
TOTAL	**1.1432**	**1.0707**	**48.15%**	**1.0687**	**45.77%**

RESULTS

Unconditional		Conditional Model 1		Conditional Model 2	
Finland	1.0673	Netherlands	1.0247	Norway	1.0211
Netherlands	1.0995	Hungary	1.0358	Netherlands	1.0247
Czech Republic	1.1088	Italy	1.0385	Romania	1.0307
Hungary	1.1129	Romania	1.0425	Lithuania	1.0312
Germany	1.1210	Finland	1.0435	Georgia	1.0339
Sweden	1.1255	Germany	1.0445	Finland	1.0345
Italy	1.1258	Lithuania	1.0461	Sweden	1.0359
Bulgaria	1.1276	Czech Republic	1.0474	Czech Republic	1.0363
Lithuania	1.1393	Bulgaria	1.0563	Hungary	1.0364
Poland	1.1459	Sweden	1.0600	Slovenia	1.0381
Slovenia	1.1546	Georgia	1.0636	Germany	1.0398
France	1.1646	Poland	1.0674	Italy	1.0402
Spain	1.1802	Norway	1.0679	Poland	1.0445
Norway	1.1807	France	1.0733	Bulgaria	1.0500
Romania	1.1881	Slovenia	1.0848	France	1.0580
Georgia	1.2277	Spain	1.0914	Spain	1.0634
TOTAL	**1.1432**	**TOTAL**	**1.0576**	**TOTAL**	**1.0405**

RESULTS

Estimation of nonparametric significance test

	Model 1		Model 2	
School variables	**p-value**	**Influence (scatter plot)**	**p-value**	**Influence (scatter plot)**
Disadvantage background	**0.01***	Unfavorable	1.00	Unfavorable
Rural area	**(<2e-16)****	Unfavorable	1.00	Unfavorable
Parents´ involvement at home	**(<2e-16)****	Favorable	1.00	Favorable
Parents´ involvement in school	**0.04***	Favorable	1.00	Favorable
Absenteeism	0.14	Unfavorable	1.00	Unfavorable
Disciplinary index	1.00	Favorable	1.00	Favorable
Instructional time	1.00	Favorable	1.00	Favorable
Country variables				
GDPpc			**(<2e-16)*****	Favorable
Expenditure in education			**(<2e-16)*****	Favorable
Hard work			**(<2e-16)*****	Favorable
Responsibility			**(<2e-16)*****	Favorable
Perseverance			**(<2e-16)*****	Favorable

CONCLUDING REMARKS

- The comparison of educational results in a cross-country perspective should take into account the resources employed and the specific characteristics of schools and educational systems
- The best performers (e.g. Finland, Czech Republic or Germany) might not be the most efficient educational systems.
- There are significant divergences across countries with regard to the importance of schools or country effects to explain inefficiency
- “Cultural values” and “economic variables” have more influence on explaining inefficiency than school factors.
- Some traditional school factors such as the ‘instructional time’ or ‘the level of absenteeism’ do not seem to have influence on efficiency

WORKSHOP ON EFFICIENCY IN EDUCATION

Lancaster University Management School

The Work Foundation, London, September 19-20

"Assessing European primary school performance through a conditional nonparametric model"

Jose M. Cordero, *University of Extremadura, Spain*
Daniel Santín, *Complutense University of Madrid, Spain*
Rosa Simancas, *University of Extremadura, Spain*