Estimating the causal impact of an intervention on efficiency in a dynamic setting

JOURNAL OF THE OPERATIONAL RESEARCH SOCIETY
https://doi.org/10.1080/01605682.2021.1979902

ORIGINAL ARTICLE

Estimating the causal impact of an intervention on efficiency in a dynamic setting

Anna Mergoni[a] and Kristof De Witte[a,b]

[a]KU Leuven, Leuven, Belgium; [b]Maastricht University, Maastricht, The Netherlands

ABSTRACT
This paper develops a novel methodology to estimate in a dynamic setting the causal impact of a policy on efficiency. Classical efficiency techniques evaluate multidimensional performance but ignore the endogeneity issues in policy evaluations. We develop an indicator which accounts for the dynamic performance of the observations and for the possible correlation between the treatment status and the efficiency score. Besides, we propose a decomposition of the indicator to disentangle the effect of the policy on the performance of the observations from the effect of the policy on the environmental harshness that the observations have to face. This innovative design allows us to introduce the notion of causality in efficiency studies and to shed light on the mechanisms underlying the inefficiency at the unit and policy level. In the application, the present study assesses the impact on the efficiency of a funding program that aims to foster educational equality.

ARTICLE HISTORY
Received 3 December 2020
Accepted 6 September 2021

KEYWORDS
Data envelopment analysis; Malmquist index; policy evaluation; causal inference; difference-in-differences; school resources

1. Introduction

The efficiency literature has a rich set of procedures to multidimensionally estimate the relative performance of observations. Contrary to traditional regression models, classical efficiency techniques integrate multidimensional inputs with multidimensional outputs. However, when it comes to policy evaluation, efficiency techniques falter as the relation they detect or highlight cannot be interpreted in a causal way. This contrasts with the classical policy evaluation techniques which are able to uncover causal mechanisms, although they are incapable in drawing a complete picture, as they evaluate one policy output at the time (for a review of these techniques see Abadie & Cattaneo, 2018). In the present paper, we blend the concepts of efficiency and effectiveness by developing a technique to evaluate in a causal way the dynamic effect of a policy in a multidimensional setting.

Non-parametric frontier estimations evaluate the efficiency of "decision making units" as the relative ability in transforming inputs into valuable outputs, given a reference set of units with similar characteristics and operating in similar environments. This definition makes efficiency estimation a well-established practice to evaluate the performance of public activities for at least two reasons: first, it is consistent with the idea that the production function can be retrieved directly from the data, without assuming a functional form. This is fundamental in the public context, where often the functional form of the production function is unknown. Second, it allows us to account for the multidimensional setting in which the observations operate. The same attractive features also create issues to interpret the findings in a causal way. To infer causality, the possible correlation between the assignment of the treatment and the potential outcome must be accounted for. Impact evaluation techniques analyze one output at a time, thus, they are based on identification strategies that control one correlation at a time. Instead, in the efficiency context the potential outcome, i.e. the efficiency score, involves several dimensions and this opens the possibility for additional correlations.

The current paper contributes to the policy evaluation literature by showing how non-parametric frontier methodologies can be used to assess the causal effect of a policy on the efficiency of the observations. As argued by Mergoni and De Witte (2021), there is a growing need for combining efficiency and program evaluation methods. Although these approaches were distinct in the past, they offer two complementary perspectives for the evaluation of policy measures. Currently, there are only few studies in the efficiency literature that aim to interpret the efficiency change of a policy action in terms of causal evidence. Moreover, previous literature typically considered a two-stage procedure where in a first step the efficiency scores are computed, and in a second one, they are used as dependent variable

CONTACT Anna Mergoni anna.mergoni@kuleuven.be Faculty of Economics and Business, KU Leuven, Naamsestraat 69, B-3000 Leuven, Belgium.

for classical econometric estimation (see for example Hurtado González & Herrero Chacón, 2014; Luca & Modrego, 2021; Varghese et al., 2011). However, by doing so, the issue of endogeneity in the computation of the efficiency scores, i.e. the possible correlation between the efficiency and the treatment variable (Orme & Smith, 1996; Santín & Sicilia, 2017), is not properly addressed. To overcome this issue, we propose a unified approach which accounts, already in the first stage, for this possible correlation. In particular, we combine performance evaluation with policy evaluation techniques by adapting the Difference-in-Differences identification strategy (Wooldridge, 2016) to the efficiency literature. As highlighted by Mergoni and De Witte (2021), a number of studies recently relied on a difference-in-differences identification strategy to infer causality in the context of efficiency estimation (see Baláž et al., 2020; He et al., 2020; Hitt & Tambe, 2016; Lin & Zhu, 2019; Ma et al., 2020; Pan et al., 2019), however, all of them relied on the contested separability condition.

Instead, we develop an indicator which accounts for the dynamic performance of the observations (relying on insights from Malmquist indices, see Färe et al., 1994; Pastor & Lovell, 2005) and for the possible correlation between the treatment status and the efficiency score (by relying on insights from program efficiency, see Charnes et al., 1981). Besides, we propose a decomposition of the indicator to disentangle the effect of the policy on the performance of the observations from the effect of the policy on the environmental harshness that the observations have to face. This innovative design allows us to introduce the notion of causality in efficiency studies and to shed light on the mechanisms underlying the inefficiency at the unit and policy level.[1]

To give insights on the practical utility of the proposed methodology, in our application we investigate the causal effects of the "Equal Education Opportunity program", a funding program offered by the Flemish government, on schools' productivity. The proposed method allows us to exploit the panel structure of the data and to investigate the effects of the policy from a dynamic and causal perspective. Our findings show that short and long term effects can be contradictory and, therefore, a dynamic perspective is fundamental to uncover the mechanisms through which the policy operates.

The remainder of the paper is organized as follows. First, we review the relevant developments in non-parametric frontier literature. Second, we introduce the ingredients from the econometric and the efficiency toolbox (namely the Difference-in-Differences techniques, Program Efficiency and the Malmquist Index) and we present in Section 3.4 our unified approach. Third, the empirical application is presented. A final section concludes the paper.

2. Connections between malmquist and program efficiency

The non-parametric frontier measurements, and in particular Data Envelopment Analysis (DEA; Charnes et al., 1978), have been subject to a continuous evolution since their introduction. The development of DEA techniques has taken several paths, but all of them go towards the same direction: making the method more flexible, more reliable and more informative. For example, non-parametric frontier approaches have been refined to account for the return to scale (Banker et al., 1984; Banker & Thrall, 1992), to drop the convexity assumptions (Deprins et al., 1984; Tulkens, 1993), to consider non discretionary inputs (Ruggiero, 1998), the presence of outliers (Wilson, 1995) and the influence of exogenous variables (Cazals et al., 2002). As the models do not require price information, they are attractive to evaluate the performance of complex decision making units such as schools, hospitals, airports and any other (public) services which involve the use of multiple resources for the production of multiple goods or services and for which a clear production function is unknown.

A promising research avenue within the DEA literature is situated in the comparison of the efficiency scores of units operating in different environments. The comparison task is fundamental, as it allows researchers to assess the efficiency of (public) programs or to detect dynamic patterns. However, this research avenue is also challenging as the efficiency scores cannot be compared if different reference sets are used. Charnes et al. (1981) were the first to commit to this comparison task. The focal idea of their approach was that to allow "across-envelope" efficiency comparisons, an "inter-envelope" function is needed. This idea lead to the decomposition of the overall efficiency score as the product of managerial efficiency, computed by means of "intra-envelope" functions, and program efficiency, computed by means of the "across-envelope" function. The definition of program efficiency as the ratio between the managerial efficiency and the overall efficiency has been closely followed by the metafrontier analysis of Battese et al. (2004) and O'Donnell et al. (2008). Specifically, in the metafrontier context, the efficiency scores are decomposed into an element that measures the distance to the metafrontier (i.e. the frontier of an unrestricted technology, which correspond to the "across-envelope" function) and into an element that measures

the distance between the group frontier and the meta-frontier (which correspond to the "intra-envelope" function).

Similarly, Färe and Grosskopf (1992) developed a technique for inter-temporal comparisons, known under the name of the Malmquist Index. The central idea is that to allow comparisons there is the necessity of exploiting a common reference set. The comparison of the performances in different time periods allow one to measure the change in efficiency and to decompose it in a change in technical efficiency and a shift in technology (Färe et al., 1994).

The Malmquist Index and the Program Efficiency share some similarities. Pastor and Lovell (2005) introduce in the Malmquist literature a concept of the global frontier that resembles the idea of the across-envelope function of Charnes et al. (1981). Another meeting point is the work of Camanho and Dyson (2006) that suggests enriching the concept of program efficiency with the Malmquist index to develop a measure of comparison within group performances. Similarly, Asmild and Tam (2007) propose the construction of a global malmquist index for calculating differences between frontiers from different groups rather than different time periods and for allowing the estimation in case of unbalanced panels. Their index starts from the idea that productivity change is global effect. More recently, Aparicio et al. (2017) extended the approach of Camanho and Dyson (2006) to measure how the performance gap among groups changes over time. A recent review has been provided by Piot-Lepetit and Tchakoute Tchuigoua (2021).

From a policy evaluation perspective, it is important to compare the efficiency among units operating in different environments. However, non-parametric frontier techniques do not allow for a causal interpretation of the estimates. Despite the Malmquist Index and Program Efficiency have been developed to discover patterns underlining the performance/efficiency, they are not able to explain in a causal way these patterns. This lack can be motivated by the fact that the identification strategies developed in the econometric literature to ensure a causal interpretation cannot be directly transferred into the non-parametric frontier apparatus. As efficiency scores are multidimensional and computed by means of an optimising function of input and output vectors, the endogeneity is defined as the correlation of one of the inputs or outputs with the efficiency (Peyrache & Coelli, 2009). Therefore, a policy intervention consisting in a change of the level of one or more inputs or outputs potentially endogenize not only the single input or output affected, but the whole efficiency score.

With this respect, there have been some attempts to consider the issue of endogeneity in efficiency (see for example Cazals et al. (2016); Cordero et al., 2015; Orme & Smith, 1996; Peyrache & Coelli, 2009; Santín & Sicilia, 2017; Simar et al., 2016), but they have been driven by bias correction motivations more than by causal inference purposes. The first who explicitly addressed the question of causal inference was D'Inverno et al. (2021). They do so by exploiting the identification strategy of the Regression Discontinuity Design (Angrist & Pischke, 2008) and the idea of the Program Efficiency (Charnes et al., 1981). In particular, the use of tailored local frontiers allowed them to control for endogeneity and interpret in a causal way the difference among treated and control units.

3. Methodology

To estimate the dynamic effect of a policy on efficiency, we combine insights from program evaluation with insights from efficiency analysis. We exploit the idea that it is possible to tackle the correlation among the inputs or outputs and the efficiency scores, and, by properly choosing the local frontiers, to account for the endogeneity. In particular, relying on a Difference-in-Differences (DiD) identification strategy, we divide the full set of observations in four categories (two for the treated units: at time t and time $t+1$ and, symmetrically, two for the non-treated units) and construct four local frontiers. The obtained local efficiency scores are then combined with the overall scores, which are constructed using an overall frontier enveloping all observations. The methodology is detailed in four steps. Section 3.1 discusses the relevant ideas behind the DiD, Section 3.2 presents the notion of program efficiency, followed by the notion of the Malmquist index in Section 3.3. Finally, Section 3.4 introduces our novel unified approach.

3.1. Difference-in-differences

The causal effect of an intervention is defined in theoretical terms as the average difference in the potential outcome in case of treatment and non-treatment. In the real world, it is not possible to estimate the causal effect of a policy at the unit level as for the same unit only the potential outcome in case of treatment or in case of non-treatment is observable. However, if the Stable Unit Treatment Value Assumption holds, which requires no interference between the units and no variation in the treatment, and if the assignment of the policy respects certain conditions, it is possible to estimate the

average treatment effect of the policy, τ (Imbens & Rubin, 2015).

The stricter the conditions that the policy fulfills, the simpler the estimation technique. In the simplest case, the assignment of the treatment is random and τ can be estimated as the difference between the average outcome for treated and non-treated units, as proposed by Rubin (1974). Unfortunately, due to ethical issues, policy makers can commonly not randomly assign a policy to individuals. The non-random assignment of the intervention possibly creates a correlation between the treatment and the potential outcomes (in our case the program efficiency scores) and, therefore, impedes a direct estimation of the causal effect. The phenomenon is known as selection bias and arises each time treated and non-treated units differ on (un)observed pre-treatment characteristics.

If the policy is exogenous along the time dimension and if the differences among treated and non-treated would stay constant over time in the absence of treatment, the Difference-in-Differences (DiD) technique allows for causal estimation (see Angrist & Pischke, 2008, 2014 for further details). This condition is known as parallel trend and is the key assumption for the identification strategy of the DiD estimator. Instead of comparing levels, the DiD estimator compares changes. So, instead of searching a counterfactual for the treated units, in the DiD setting a counterfactual for the changes in the treated units is searched. When the parallel assumption is fulfilled, it is possible to use as counterfactual the changes in the non-treated units. The DiD estimator of the causal effect is defined as follows:

$$\begin{aligned}\tau_{DID} = &(E[y(1)|t>T]-E[y(1)|t<T]) \\ &-(E[y(0)|t>T]-E[y(0)|t<T])\end{aligned} \tag{1}$$

where T is the time of the treatment $y(1)$ is the potential outcome in case of treatment and $y(0)$ the potential outcome in case of non treatment.

3.2. Program efficiency

In our analysis, the outcome variable is not directly observed, but it is an efficiency score $\theta_i(x_i, y_i)$ computed for each unit i by means of non-parametric frontier estimations, given a vector of inputs x_i and a vector of output y_i. Comparing the efficiency score of units belonging to different programs (or under different policy treatments) is not straightforward. This is due to two main reasons: first, the standard efficiency scores are assessed in relative terms, i.e. given a reference set of units with similar characteristics and operating in similar environments; second, the standard efficiency scores are not able to disentangle between managerial ability (i.e. the ability of the manager of taking decision regarding the level of inputs and the transformation process), technical advancement (i.e. the advancement in the level of technology involved in the transformation process) and other context specific factors (i.e. variables which influence the production process, but that are not under the control of the manager). Therefore, a direct comparison could lead to misleading conclusions as the influence of the environment is not accounted for.

The concept of Program Efficiency has been introduced by Charnes et al. (1981) to account for "good" units operating in unfavourable contexts and "bad" units operating in favourable contexts. For each unit i, the program efficiency is defined as the ratio of the global efficiency $\theta_i^{global}(x_i, y_i)$, which is the efficiency score of unit i with respect to all the other units, over the local efficiency $\theta_i^{local}(x_i, y_i)$, which is the efficiency score of unit i with respect to the units operating in the same program of unit i.

$$PE_i = \frac{\theta_i^{global}(x_i, y_i)}{\theta_i^{local}(x_i, y_i)} \tag{2}$$

The idea is to make comparable the efficiency scores by computing them in the same relative scale. This is achieved by the use of a global reference set to compute the global efficiency scores $\theta_i^{global}(x_i, y_i)$. Besides, the scores must account for the context specific performances, suggesting the introduction of the rescaling factor $\theta_i^{local}(x_i, y_i)$. It should be noted that Charnes et al. (1981) introduced the concept of Program Efficiency relying on the use of classical DEA methodology, i.e. input-oriented and constant return to scale DEA, to compute the global and the local efficiency scores. However, the idea of the ratio of an overall score over a local score can be extended in a straightforward way to any other appropriate non-parametric model introduced in the literature of frontier estimation (see for example the variable return to scale model by Banker et al. (1984) or the FDH model by Deprins et al. (1984)).

3.3. Malmquist index

Although the Program Efficiency allows us to compare the performance of units belonging to different programs, comparing the changes in the efficiency scores is not straightforward. Färe and Grosskopf (1992) and Färe et al. (1994) exploited the fact that the inverse of a Shephard distance function is associated with a Farrell measure of technical efficiency (Pastor et al., 2020) and introduced the idea of the Malmquist index in the efficiency literature. This allowed the measurement of intertemporal changes in efficiency and the evaluation of the dynamic performances of the units.

For the task of comparing the dynamic performance of units belonging to different programs, the "global Malmquist productivity index" proposed by Pastor and Lovell (2005) is of particular interest. The use of a global reference set, known as intertemporal frontier (Berg et al., 1992; Samuelson & Swamy, 1974; Tulkens & Eeckaut, 1995), provides the circularity of the index. This characteristic ensures that given the comparison between a unit *a* and a unit *b*, and the comparison between the unit *b* and a unit *c*, the comparison of *a* and *c* can be assessed through *b*. Besides, the use of a constant return to scale model for efficiency (suggested by Grifell-Tatjé & Lovell, 1995) guarantees that two units that have the same ratio of outputs to inputs are evaluated as equally productive, regardless of the scale of production (also known as homogeneity). Circularity and homogeneity are fundamental to obtain sensible comparisons. Exploiting the relation between the Shepard distance function and the DEA efficiency score, we formulate the "global Malmquist productivity index" of Pastor and Lovell (2005) as:

$$M_i^G(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{\theta_{i,t+1}^G(x_i^{t+1}, y_i^{t+1})}{\theta_{i,t}^G(x_i^t, y_i^t)} \tag{3}$$

where $\theta_{i,t}^G(x_i^t, y_i^t)$ is the CRS DEA efficiency score of unit i at time t and it is computed using as reference set all the observation in the sample, at time t and $t+1$ and $\theta_{i,t+1}^G(x_i^t, y_i^t)$ is the CRS DEA efficiency score of unit i at time $t+1$, computed using as reference set all the observation in the sample, at time t and $t+1$.

3.4. Our unified approach

Our estimator combines insights from the DiD, program efficiency and Malmquist index in a two steps procedure. The first step constructs the *Malmquist Program Efficiency Index* as the combination of the program efficiency, which allows us to compare the efficiency scores of units belonging to different programs, and the Malmquist index, which allows us to compare changes in efficiency. We substitute the efficiency scores that typically enter the construction of the Malmquist Index with the program efficiency, as defined in Equation 2. This allows us to adapt the Malmquist Index proposed by Pastor and Lovell (2005) to a policy evaluation context. The *Malmquist Program Efficiency Index* is defined as follows:

$$MPE_i(x_i^t, y_i^t, x_i^{t+1}, y_i^{t+1}) = \frac{PE(x_i^{t+1}, y_i^{t+1})}{PE(x_i^t, y_i^t)} \tag{4}$$

In this framework, four local frontiers enter the analysis (two for the units in the treated group, one for time t and one for time $t+1$, and two for the units in the control group, one for time t and one for time $t+1$) which are compared by means of an overall global frontier encompassing all of them.

We decompose the *Malmquist Program Efficiency Index* to enrich its informative power. This decomposition accounts for the possible presence of factors influencing the production process, the so-called *environmental* variables, in line with the approach of Johnson and Ruggiero (2014). Note that the DiD specification controls for the time-invariant environmental factors, although there might be some variables changing over time.

In the first step we also decompose the program efficiency to account for the environmental variables z. The program efficiency is decomposed in a conditional program efficiency, $PE(x_i^{t+1}, y_i^{t+1}, z_i^{t+1})$, which accounts for the contribution of the program to the efficiency, *ceteris paribus*, and in an environmental harshness index $EPE(x_i^{t+1}, y_i^{t+1}, z_i^{t+1})$, which accounts for the harshness of the environment in terms of exploiting the program intervention. In particular, the conditional program efficiency $PE(x_i^{t+1}, y_i^{t+1}, z_i^{t+1})$ is computed by substituting $\theta(x_i^{t+1}, y_i^{t+1})$, the classical DEA score of equation 2, with $\theta(x_i^{t+1}, y_i^{t+1}, z_i^{t+1})$, the conditional efficiency score proposed by Daraio and Simar (2007). The definition of $EPE(x_i^{t+1}, y_i^{t+1}, z_i^{t+1})$ can be retrieved from the equation below:

$$PE(x_i^{t+1}, y_i^{t+1}) = PE(x_i^{t+1}, y_i^{t+1}, z_i^{t+1}) \times EPE(x_i^{t+1}, y_i^{t+1}, z_i^{t+1}) \tag{5}$$

By plugging in equation (5) in the *Malmquist Program Efficiency Index*, defined in equation (4), we obtain the following decomposition:

$$MPE_i = \frac{PE(x_i^{t+1}, y_i^{t+1}, z_i^{t+1})}{PE(x_i^t, y_i^t, z_i^t)} \times \frac{EPE(x_i^{t+1}, y_i^{t+1}, z_i^{t+1})}{EPE(x_i^t, y_i^t, z_i^t)} \tag{6}$$

where the first term accounts for the *ceteris paribus* change in the program efficiency and the second term represents the change in the environmental harshness. We refer to the latter as the *Malmquist Environmental Efficiency* (MEE). This decomposition allows us to complement the analysis of the change in the program efficiency (by means of the MPE) with the analysis of the change of the environmental harshness (by means of the MEE) due to interventions at the policy level.

The second step consists in combining the efficiency and the econometrics analysis (i.e. of the *MPE* and of the *DiD*). By exploiting the identification strategy of the DiD, and assuming that the SUTVA and the parallel trends assumptions hold, we are able to define the *Average Treatment Effect on Efficiency* (ATEE) as follows:

$$ATEE = \overline{MPE}_{TREATED} - \overline{MPE}_{CONTROL} \tag{7}$$

where $\overline{MPE}_{TREATED}$ and $\overline{MPE}_{CONTROL}$ are, respectively, the average *Malmquist Program Efficiency Index* among the treated and among the control units. The difference in differences of the DiD is accounted for in the ATEE as the terms of Equation 7, which accounts explicitly for the difference among the treated and the control and, hence, for the time trend within each group. Considering the proposed decomposition of the *MPE*, an *Environmental Average Treatment Effect on Efficiency* (EATEE) can also be defined to supplement the ATEE:

$$EATEE = \overline{MEE}_{TREATED} - \overline{MEE}_{CONTROL}. \quad (8)$$

The defined EATEE accounts for the effect of the policy on the environmental variables which contribute to the program efficiency. So, it allows us to distinguish between the direct effect of the policy on the *ceteris paribus* Program Efficiency and the effect of the policy on the environmental harshness.

4. Empirical application

Non-parametric efficiency techniques have been widely used for the evaluation of public services. Many methodological advances have been illustrated in the field of education (see, for example, Beasley, 1995; Cordero et al., 2017; Mayston, 2003; Thanassoulis et al., 2011 and for a review De Witte & Lòpez-Torres, 2017). Following this strand of literature, in the empirical application we implement the proposed methodology to provide novel evidence on the effect of increasing and reducing school resources. In particular, we evaluate the impact on efficiency of the "Equal Education Opportunity" (EEO) Program introduced by the Flemish Minister of Education in 2002.

4.1. The Flemish context

In Flanders, the majority of students follow education organized by private providers, the second largest group attends public education organized by the Flemish community, while the remaining students participate in the public education run by municipal or provincial authorities. Besides, Flemish secondary education is characterized by a hierarchical tracking system. In the first two years of secondary education, students choose between two curricula, according to their abilities and following the suggestion of their teachers and parents. After two years, the students who completed the more demanding curriculum can choose according to their preferences between a general track (ASO), an artistic track (KSO) a technical track (TSO) or a vocational track (BSO). While the other students are strongly recommended to enroll in the technical or in the vocational track (Nusche et al., 2015).

Despite persistently ranking in the top quartile of the distribution in international education rankings, the educational performances of Flemish students are declining over time. Besides, the Flemish education system is characterized by a high level of inequalities and a significant number of students who leave education without a degree (OECD., 2017, The Flemish Ministry of Education & Training, 2020a, 2020b). This situation can be partly explained by the high level of immigration, by the ineffective policies for education and by the rigid hierarchical tracking system (De Witte & Hindriks, 2018).

4.2. The equal education opportunity program

The EEO program has been implemented by the Flemish Government starting from 2002, with the aim of improving the schools outcomes of schools and reducing the educational gap of students with socioeconomic disadvantageous characteristics. The program consists of additional lump sum funds.

Although all schools and grades are eligible for the program, funding is only assigned to the schools with a certain share of disadvantaged students, and in cycles of three years. The amount of funding is computed according to the number and the kind of disadvantaged students. On average, the EEO funding amounts to about 15% of the school's budget. The students are classified as disadvantaged if they have at least one of the following characteristics: (1) the student receives an educational grant (proxy for the family income); (2) the student's mother does not have a secondary education degree (proxy for parental educational background); (3) the student lives outside the family; (4) one of the parents is part of the travelling population; (5) the student does not speak Dutch at home. Despite the schools being able to freely allocate the resources according to the specific necessity of the schools, in most of the cases the additional funds are used to provide additional teaching hours targeted for disadvantaged students.

The cyclical assignment of the extra funds leaves space for the investigation of two complementary questions: First, do additional public funds improve schools' efficiency? Second, does the reduction of public funds harm schools' efficiency? Both questions can be answered due to the change in schools' "treatment" status. As shown in Table 1, during the "2011–2013 cycle", among the 488 schools who received the extra funds, 455 schools already received them in the previous cycle, while for 33 schools it was the first time. Similarly, among the

Table 1. Changes in the treatment status for secondary Flemish schools.

	"2011-2013 cycle"	
"2008-2010 cycle"	extra-funds	no extra-funds
extra-funds	455	39
no extra-funds	33	122

161 schools who did not receive the funds during the "2011–2013 cycle", 122 did not receive them also during the previous cycle, but 39 schools saw a sudden decrease in the available resources starting from 2011.

4.3. Data

Our data consists of a 2010–2013 panel of all students in the first two years of secondary education. In addition, an extended database contains information from 2002 to 2013 for the outcomes of interest. This extended database is utilized to verify the assumptions for identifying causal effects (see Section 4.4).

To address our two complementary questions two samples have been individuated. A first sample involves the schools that started to be treated in the "2011–2013 cycle" of the policy, or that remain untreated. A second sample focuses on the schools that were treated in the "2008–2010 cycle" and did not receive the extra funds in the "2011–2013 cycle", and their counterfactual: the schools who remained treated.

To implement our policy evaluation, two inputs and three outputs directly related with the EEO program have been considered to asses the schools' efficiency-performances. Specifically, schools' performance is evaluated in terms of a schools' ability in reducing the amount of operating grant per student[2] and teaching hours per student[3] for a constant level of educational outcomes, measured in terms of share of students who are able to obtain an A certificate[4], to progress in their academic path and with no problem of absenteeism. Besides, to account for the influence of the environment in which the school operates, we consider the exogenous variables that most differ among treated and control schools, namely, the share of disadvantaged students in the school, the school's size and the percentage of full time directors. The descriptive statistics for these variables are reported in the Appendix.

4.4. Underlying assumptions

To interpret the findings in a causal way, two assumptions need to hold, namely Stable Unit Treatment Value Assumption (SUTVA) and the parallel trends assumption.

The SUTVA requires that the potential outcomes are well defined (i.e. there is only one version of the treatment) and that the treatment of a unit does not interfere with the potential outcome of another unit. These conditions cannot be empirically verified, but they can be justified in the light of the available information regarding the policy under analysis. The EEO program provides additional funds to the schools with a certain share of disadvantaged students. To the best of our knowledge, no inter-school peer effects are reported in the literature, therefore, we can safely rely on the no interference assumption. Despite the amount of the extra funds varies according to school's characteristics such as the share of disadvantaged students and the size, we define as treatment the receipt of any positive amount. This prevents the existence of multiple versions of the treatment and ensures that the potential outcome is well defined.[5]

The parallel trends assumption requires that the potential outcome for the treated and the untreated would be parallel in case of no treatment. As in reality the treatment has been assigned, this assumption cannot be directly verified. However, scholars agree to use as a proxy for the potential post-treatment trends the pre-treatment trends. To discuss for the validity of the parallel trend we consider an extended database containing the relevant information from 2002 to 2015. In particular, to test formally the assumption, we check the significance of the interaction term of regression 9, as suggested, for example, by Autor et al. (2003), Table IV. The idea is that if the coefficient of the interaction term is not significantly different from zero for the period before the intervention, then the difference in the trend between the treated and the control group is also not significant, therefore, the assumption of parallel trends is fulfilled.

$$Y = \alpha + \beta Treatment + \gamma Time + \delta Treatment \times Time + \epsilon \tag{9}$$

Where *Treatment* is a dummy indicating whether the observation is in the control or in the treatment group and *Time* is a matrix of dummies indicating to which time period the observation is referred to. This regression is implemented separately for the sample relative to question 1 (i.e. the sample with the schools that started to be treated in the "2011–2013 cycle" of the policy or that remain untreated) and for the sample relative to question 2 (i.e. the sample with the schools that stop to receive the treatment in the "2011–2013 cycle" and those that remained treated). Besides, the tests have been implemented both for the share of students who obtained an A certificate and for the share of

students with no problem of absenteeism as both variables are included in the construction of the efficiency scores. Results from the regression are reported in Tables 8 and 9 of the Appendix and show that the parallel trends assumption reasonably holds in both samples for each outcome since the coefficient of the interaction terms between the dummies for years and the treatment dummy are not significant. Finally, we would like to notice that, despite the coefficient relative to the post-treatment year dummies (2011, 2012, 2013 and 2014) are also not significant, it is not implied that the policy itself has no effect on the efficiency performance of schools, since the efficiency measure is a composed index which accounts for the three outcome dimensions jointly.

4.5. Results and discussion

As the parallel trend and SUTVA assumptions are fulfilled and the data are sufficiently balanced we proceeded with the estimation of the average treatment effect of the policy on the efficiency of the schools (ATEE). It is worth to note that despite the parallel trends assumption holds, we have applied a nearest neighbour matching technique to select the most appropriate units for the control groups. The advantage is twofold: on the one hand, we reinforce the balance of our data, while on the other hand, we overcome the problem of the difference in the sample size for the treated and the control.[6]

The effect of an increase in funding on schools' efficiency is reported in Table 2. The negative average treatment effect on efficiency in 2011, $ATEE_{2011} = -0.28$, indicates that the schools which received the extra funds, have not been able to exploit these resources to increase the educational outcomes during the first year. A similar inefficiency at the policy level is detected when we consider the effect of a reduction of resources on the efficiency (see Table 3). We observe a positive short term effect $ATEE_{2011} = 0.61$, which signals that the schools with a reduction of funding have not reduced proportionally their educational outcomes.

These results partly contrast the findings of Bargagli-Stoffi et al. (2019) who investigated the heterogeneous effects of the EEO program and detected a positive impact on the students with young teachers or senior principals or with Ooghe (2011), who analyzed the impact of the program on the primary education, finding positive effects, especially for the disadvantaged students. However, the fact that additional resources are not associated with higher achievement rates is in line with the study of Leuven et al. (2007) and Clark et al. (2017), who analyzed a similar school funding policy for disadvantaged students in the Netherlands and in New Zealand, respectively. Besides, previous studies on the Flemish context confirm the possible waste of resources involved in the EEO program. In particular, D'Inverno et al. (2021), who focused on the effect of this program on the efficiency for the final four years of secondary education by means of a Regression Discontinuity Design based approach.

Table 2. The effects of an increase in the school resources.

	2011	2012	2013
$\bar{MPE}_{treated}$	0.99	1.16	1.34
$\bar{MPE}_{control}$	1.27	1.15	1.28
ATEE	−0.28	0.01	0.06
$E\bar{MPE}_{treated}$	0.61	0.67	0.78
$E\bar{MPE}_{control}$	0.78	0.8	0.86
EATEE	−0.17	−0.13	−0.08

Table 3. The effects of a decrease in school resources.

	2011	2012	2013
$\bar{MPE}_{treated}$	1.58	1	1.27
$\bar{MPE}_{control}$	0.97	1.03	0.98
ATEE	0.61	−0.03	0.29
$E\bar{MPE}_{treated}$	1.03	0.67	0.76
$E\bar{MPE}_{contrl}$	0.52	0.51	0.49
EATEE	0.51	0.16	0.27

To better understand the mechanism through which an increase in the resources can affect the school's efficiency, and possibly explain the conflicting evidence stemming from previous literature, we enrich the analysis by investigating the longer term effect and the effect of the policy on environmental harshness (EATEE). A positive impact of additional resources is detected in the subsequent years, $ATEE_{2012} = 0.01$ and $ATEE_{2013} = 0.06$ suggesting that schools experience a learning process and in the medium to long run are able to employ in a productive way the extra-funds. For reducing the funding, the long term effect provides more complex dynamics. A mid term negative effect, $ATEE_{2012} = -0.03$ is followed by a long term positive effect $ATEE_{2013} = 0.29$. This evolution suggests that, in the case of prolonged funds restriction, after a first moment in which the schools struggle to maintain the initial level of performance, they learn how to use efficiently the funds at their disposal.

From a policy perspective, this evidence shows that it is possible for schools to learn how to use in an efficient way additional resources, and also to adapt to funds restrictions, even if this requires a longer process. The results are of particular interest first, because the evidence on the long run effect of school resources is scarce; second, because it can partly explain the contradictory evidence offered by previous literature.

In addition, the decomposition of the program efficiency proposed in equation 5 allows us to detect the causal effect of the policy on the environmental harshness (see equation 8). Surprisingly, an increase

in the school resources has a negative effect on the environment where schools operate, while a restriction in the funding has a positive effect, suggesting that higher resources are associated with higher wastes. As the environmental harshness is measured in terms of share of disadvantaged students, school size and quality of school management (using as a proxy the full time or not full time status), the result can be explained by two mechanisms. First, it is possible that the school production function has decreasing return to scale, therefore, it is not sufficient to rise the school resources proportionally with school size or to the share of disadvantaged students. Second, it is possible that school directors do not have the proper skills to manage resources efficiently if the amount is too large.

5. Conclusion

The main objective of this paper was to develop a methodology to study in a dynamic perspective the causal effect of a policy on efficiency. The combination of efficiency and policy evaluation analysis is fundamental to detect inefficiency at the policy level and could help in tailoring more effective policy actions. However, despite efficiency measures being well established to evaluate the performance of complex (public) utilities, previous efficiency literature, and especially the non-parametric frontier literature, has overlooked the endogeneity issues, preventing a causal interpretation of the results.

The paper contribution expands the recent literature which attempts to endow the efficiency scores with a causal interpretation. In particular, the study shows how to exploit the temporal dimension of a panel dataset to examine causal relations. The main idea is to adapt the identification strategy of the Difference-in-Differences to the context of non-parametric frontier estimation. It is done so by considering as potential outcome the Malmquist Program Efficiency (MPE), which is constructed using the tools offered by the program efficiency and by the Malmquist index. Besides, the study proposes a decomposition of this index to disentangle between the direct effect of the policy on the efficiency of the units and the indirect effect that the policy exerts through the environment.

This innovative methodology has allowed us to shed new light on the effect of school funding. In particular, we investigated the effect of the "Equal Education Opportunities" program in Flemish secondary education. We observed that schools' efficiency is more related with schools' experience in using its resources than in the level of resources itself. From a policy perspective, this suggests that policy makers should consider designing subsidiary interventions to the EEO program, such as training sessions and other intervention to help schools in acquiring experience.

With respect to previous literature, our analysis offers a more complete picture thanks to our innovative methodology. First, it analyzes the effect of the policy in a causal way and using a multidimensional evaluation of school performance. Despite the effect of additional funds on specific school's outcomes has been deeply investigated, evidence on the effect of school policies on efficiency are scarce. Second, this study orchestrates different perspectives, considering both the effect of increasing and decreasing the resources and both a short and longer term period. To the best of our knowledge, this is the first study which investigates the effect of a resource reduction or longer term effect of funding on schools' performance. Third, the decomposition of the Malmquist index has provided further evidence on the mechanism through which the level of resources affect the school's efficiency.

Despite the steps forward in our understanding of the effect of funding on school's efficiency provided by this research, further investigations are needed to detect the possible heterogeneous effects of the policy and to investigate the equalizing role of the policy itself. Besides, from a methodological perspective, further research should focus on combining econometric tools to the efficiency framework. Recent studies, such as D'Inverno et al. (2021) and Mergoni et al. (mimeo), translated the idea of the identification strategy of the Regression Discontinuity Design in the realm of efficiency estimation, while in this paper we relied on the identification strategy of the Difference-in-Differences. Future research should continue in this direction by exploiting other econometric techniques, such as the instrumental variable approach, the fixed effect model, or the synthetic control method.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes

1. To foster further applications, the R code is available upon request.
2. The total budget allocated to meet operating costs, weighted for the number of students
3. The number of teaching hours provided by the schools, weighted for the number of students
4. This is the highest certificate that the students can receive at the end of the year and this allows them to proceed to the following academic year without restrictions.

5. The "Positivity assumption" is ensured as every non-treated school has a positive probability of receiving the treatment and every treated school has a positive probability of stopping to receive the funds. Moreover, the "Exchangeability assumption" is verified by looking at the differences in means among the treated and the control group along the variables of interest available in the database. Results are reported in Appendix (see Tables 4–6).
6. As shown by De Witte and Marques (2009), differences in sample size might bias the efficiency estimates.

Acknowledgements

The authors would like to thank Giovanna D'Inverno, participants of the North American Productivity Workshop, participants of AEDE conference and seminar participants at the University of Leuven. Anna Mergoni acknowledges funding from Flemish Science Organisation through grant 11G5520N.

ORCID

Anna Mergoni http://orcid.org/0000-0002-9128-5341
Kristof De Witte http://orcid.org/0000-0003-0505-8642

References

Abadie, A., & Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, *10*(1), 465–503. https://doi.org/10.1146/annurev-economics-080217-053402

Angrist, J. D., & Pischke, J. S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Angrist, J. D., & Pischke, J. S. (2014). *Mastering'metrics: The path from cause to effect*. Princeton University Press.

Aparicio, J., Crespo-Cebada, E., Pedraja-Chaparro, F., & Santín, D. (2017). Comparing school ownership performance using a pseudo-panel database: A malmquist-type index approach. *European Journal of Operational Research*, *256*(2), 533–542. https://doi.org/10.1016/j.ejor.2016.06.030

Asmild, M., & Tam, F. (2007). Estimating global frontier shifts and global malmquist indices. *Journal of Productivity Analysis*, *27*(2), 137–148. https://doi.org/10.1007/s11123-006-0028-0

Autor, W. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, *118*(4), 1279–1333. https://doi.org/10.1162/003355303322552801

Baláž, V., Nežinský, E., Jeck, T., & Filčák, R. (2020). Energy and emission efficiency of the Slovak regions. *Sustainability*, *12*(7), 2611. https://doi.org/10.3390/su12072611

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*(9), 1078–1092. https://doi.org/10.1287/mnsc.30.9.1078

Banker, R. D., & Thrall, R. M. (1992). Estimation of returns to scale using data envelopment analysis. *European Journal of Operational Research*, *62*(1), 74–84. https://doi.org/10.1016/0377-2217(92)90178-C

Bargagli-Stoffi, F. J., De Witte, K., & Gnecco, G. (2019). Heterogeneous causal effects with imperfect compliance: A novel bayesian machine learning approach. arXiv Preprint arXiv:1905.12707.

Battese, G. E., Rao, D. S. P., & O'donnell, C. J. (2004). A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *Journal of Productivity Analysis*, *21*(1), 91–103. https://doi.org/10.1023/B:PROD.0000012454.06094.29

Beasley, J. E. (1995). Determining teaching and research efficiencies. *Journal of the Operational Research Society*, *46*(4), 441–452. https://doi.org/10.1057/jors.1995.63

Berg, S. A., Førsund, F. R., Jansen, E. S., Berg, S. A., & Forsund, F. R. (1992). Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980–89. *The Scandinavian Journal of Economics*, *94*(Supplement), S211–S228. https://doi.org/10.2307/3440261

Camanho, A. S., & Dyson, R. G. (2006). Data envelopment analysis and malmquist indices for measuring group performance. *Journal of Productivity Analysis*, *26*(1), 35–49. https://doi.org/10.1007/s11123-006-0004-8

Cazals, C., Fève, F., Florens, J. P., & Simar, L. (2016). Nonparametric instrumental variables estimation for efficiency frontier. *Journal of Econometrics*, *190*(2), 349–359. https://doi.org/10.1016/j.jeconom.2015.06.010

Cazals, C., Florens, J. P., & Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics*, *106*(1), 1–25. https://doi.org/10.1016/S0304-4076(01)00080-X

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Charnes, A., Cooper, W. W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. *Management Science*, *27*(6), 668–697. https://doi.org/10.1287/mnsc.27.6.668

Clark, J., Das, S. R., & Menclova, A. (2017). Evaluating the effectiveness of school funding and targeting different measures of student disadvantage: Evidence from New Zealand. *Economic Record*, *93*(303), 576–599. https://doi.org/10.1111/1475-4932.12354

Cordero, J. M., Santín, D., & Sicilia, G. (2015). Testing the accuracy of dea estimates under endogeneity through a Monte Carlo simulation. *European Journal of Operational Research*, *244*(2), 511–518. https://doi.org/10.1016/j.ejor.2015.01.015

Cordero, J. M., Santìn, D., & Simancas, R. (2017). Assessing European primary school performance through a conditional nonparametric model. *Journal of the Operational Research Society*, *68*(4), 364–376. https://doi.org/10.1057/jors.2015.42

D'Inverno, G., Smet, M., & De Witte, K. (2021). Impact evaluation in a multi-input multi-output setting: Evidence on the effect of additional resources for schools. *European Journal of Operational Research*, *290*(3), 1111–1124. https://doi.org/10.1016/j.ejor.2020.08.042

Daraio, C., & Simar, L. (2007). *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications*. Springer Science & Business Media.

De Witte, K., & Hindriks, J. (2018). *De (her) vormende school*. Itinera Institute.

De Witte, K., & Lòpez-Torres, L. (2017). Efficiency in education: A review of literature and a way forward. *Journal of the Operational Research Society*, *68*(4), 339–363. https://doi.org/10.1057/jors.2015.92

De Witte, K., & Marques, R. C. (2009). Capturing the environment, a metafrontier approach to the drinking water sector. *International Transactions in Operational Research*, *16*(2), 257–271. https://doi.org/10.1111/j.1475-3995.2009.00675.x

Deprins, D., Simar, L., & Tulkens, H. (1984). *Measuring labor-efficiency in post offices*. Technical report, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

Färe, R., & Grosskopf, S. (1992). Malmquist productivity indexes and fisher ideal indexes. *The Economic Journal*, *102*(410), 158–160. https://doi.org/10.2307/2234861

Färe, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *The American Economic Review*, *84*(1), 66–83.

Grifell-Tatjé, E., & Lovell, C. A. K. (1995). A note on the malmquist productivity index. *Economics Letters*, *47*(2), 169–175. https://doi.org/10.1016/0165-1765(94)00497-P

He, D., Yang, J., Wang, Z., & Li, W. (2020). Has the manufacturing policy helped to promote the logistics industry? *PLoS One*, *15*(7), e0235292. https://doi.org/10.1371/journal.pone.0235292

Hitt, L., & Tambe, P. (2016). Health care information technology, work organization, and nursing home performance. *Ilr Review*, *69*(4), 834–859. https://doi.org/10.1177/0019793916640493

Hurtado González, J. M., & Herrero Chacón, I. (2014). The causal effects of product innovation, web technology and vertical integration on firm efficiency in the fashion industry. *Innovation*, *16*(1), 144–157. https://doi.org/10.5172/impp.2014.16.1.144

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Johnson, A. L., & Ruggiero, J. (2014). Nonparametric measurement of productivity and efficiency in education. *Annals of Operations Research*, *221*(1), 197–210. https://doi.org/10.1007/s10479-011-0880-9

Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2007). The effect of extra funding for disadvantaged pupils on achievement. *Review of Economics and Statistics*, *89*(4), 721–736. https://doi.org/10.1162/rest.89.4.721

Lin, B., & Zhu, J. (2019). Impact of energy saving and emission reduction policy on urban sustainable development: Empirical evidence from china. *Applied Energy*, *239*, 12–22. https://doi.org/10.1016/j.apenergy.2019.01.166

Luca, D., & Modrego, F. (2021). Stronger together? assessing the causal effect of inter-municipal cooperation on the efficiency of small italian municipalities. *Journal of Regional Science*, *61*(1), 261–293. https://doi.org/10.1111/jors.12509

Ma, J., Zhang, Z., Lu, C., & Xue, B. (2020). Could the construction of sustainable development pilot zones improve the urban environment efficiency in china? *Discrete Dynamics in Nature and Society*, *2020*, 1–9. https://doi.org/10.1155/2020/7678525

Mayston, D. J. (2003). Measuring and managing educational performance. *Journal of the Operational Research Society*, *54*(7), 679–691. https://doi.org/10.1057/palgrave.jors.2601576

Mergoni, A., & De Witte, K. (2021). Policy evaluation and efficiency: A systematic literature review. *International Transactions in Operational Research*, https://doi.org/10.1111/itor.13012

Mergoni, A., De Witte, K., & D'Inverno, G. Dealing with imperfect compliance in causal frontier evaluation: A probabilistic efficiency model.

Nusche, D., Miron, G., Santiago, P., & Teese, R. (2015). *OECD reviews of school resources: Flemish community of Belgium 2015*. OECD Publishing.

O'Donnell, C., Rao, D. S. P., & Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empirical Economics*, *34*(2), 231–255. https://doi.org/10.1007/s00181-007-0119-4

OECD. (2017). *Educational opportunity for all*.

Ooghe, E. (2011). The impact of'equal educational opportunity'funds: A regression discontinuity design. Available at SSRN 1833251.

Orme, C., & Smith, P. (1996). The potential for endogeneity bias in data envelopment analysis. *Journal of the Operational Research Society*, *47*(1), 73–83. https://doi.org/10.1057/jors.1996.7

Pan, X., Pan, X., Li, C., Song, J., & Zhang, J. (2019). Effects of china's environmental policy on carbon emission efficiency. *International Journal of Climate Change Strategies and Management*, *11*(3), 326–340. https://doi.org/10.1108/IJCCSM-12-2017-0206

Pastor, J. T., & Lovell, C. A. K. (2005). A global malmquist productivity index. *Economics Letters*, *88*(2), 266–271. https://doi.org/10.1016/j.econlet.2005.02.013

Pastor, J. T., Lovell, C. A. K., & Aparicio, J. (2020). Defining a new graph inefficiency measure for the proportional directional distance function and introducing a new malmquist productivity index. *European Journal of Operational Research*, *281*(1), 222–230. https://doi.org/10.1016/j.ejor.2019.08.021

Peyrache, A., & Coelli, T. (2009). Testing procedures for detection of linear dependencies in efficiency models. *European Journal of Operational Research*, *198*(2), 647–654. https://doi.org/10.1016/j.ejor.2008.08.014

Piot-Lepetit, I., & Tchakoute Tchuigoua, H. (2021). Ownership and performance of microfinance institutions in Latin America: A pseudo-panel malmquist index approach. *Journal of the Operational Research Society*, 1–14. https://www.tandfonline.com/action/doSearch?AllField=Ownership+and+performance+of+microfinance+institutions+in+Latin+America%3A&SeriesKey=tjor20

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701. https://doi.org/10.1037/h0037350

Ruggiero, J. (1998). Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research*, *111*(3), 461–469. https://doi.org/10.1016/S0377-2217(97)00306-8

Samuelson, P. A., & Swamy, S. (1974). Invariant economic index numbers and canonical duality: Survey and synthesis. *The American Economic Review*, *64*(4), 566–593.

Santín, D., & Sicilia, G. (2017). Dealing with endogeneity in data envelopment analysis applications. *Expert Systems with Applications*, *68*, 173–184. https://doi.org/10.1016/j.eswa.2016.10.002

Simar, L., Vanhems, A., & Van Keilegom, I. (2016). Unobserved heterogeneity and endogeneity in nonparametric frontier estimation. *Journal of Econometrics*, *190*(2), 360–373. https://doi.org/10.1016/j.jeconom.2015.06.015

Thanassoulis, E., Kortelainen, M., Johnes, G., & Johnes, J. (2011). Costs and efficiency of higher education institutions in england: A dea analysis. *Journal of the Operational Research Society*, *62*(7), 1282–1297. https://doi.org/10.1057/jors.2010.68

The Flemish Ministry of Education and Training. (2020a). *Cijfers over schooluitval. beleidsdomein onderwijs en vorming*.

The Flemish Ministry of Education and Training. (2020b). *Program for international student assessment (PISA)*.

Tulkens, H. (1993). On fdh efficiency analysis: Some methodological issues and applications to retail banking, courts, and urban transit. *Journal of Productivity Analysis*, *4*(1–2), 183–210. https://doi.org/10.1007/BF01073473

Tulkens, H., & Eeckaut, P. V. (1995). Non-parametric efficiency, progress and regress measures for panel data: Methodological aspects. *European Journal of Operational Research*, *80*(3), 474–499. https://doi.org/10.1016/0377-2217(94)00132-V

Varghese, S. K., Veettil, P. C., Buysse, J., Speelman, S., Frija, A., & Van Huylenbroeck, G. (2011). Effect of water scarcity on groundwater use efficiency in the rice sector of karnataka: A case study. *WIT Transactions on Ecology and the Environment*, *145*, 783–791.

Wilson, P. W. (1995). Detecting influential observations in data envelopment analysis. *Journal of Productivity Analysis*, *6*(1), 27–45. https://doi.org/10.1007/BF01073493

Wooldridge, J. (2016). *Introductory econometrics: A modern approach*. Nelson Education.

Appendix A.

Descriptive statistics (Tables 5 and 7)

Table 4. Sample means in 2010 for the schools involved in question 1.

	treated	*control*	*overall*	*p – value*
INPUTS				
Operating Grant per student	936.1	877.1	889.7	0.2606
	(273.3479)	(226.5254)	(237.5539)	
Teaching Hours per student	2.3294	2.085	2.1371	0.108
	(0.7598809)	(0.7650415)	(0.7680596)	
OUTPUTS				
% students with 'A certificate'	85.76	89.10	88.39	0.1016
	(10.91393)	(6.633541)	(7.823609)	
% students with no problem of absenteeism	98.91	99.79	99.60	0.08993
	(2.878542)	(0.8356366)	(1.550007)	
% students progressing through school	96.30	97.76	97.45	0.06446
	(4.137448)	(2.873909)	3.225498	
ENVIRONMENTAL VARIABLES				
% disadvantaged students	0.3743	0.16763	0.21162	1.182e-05
	(0.225298)	(0.1354396)	(0.1793352)	
School size	386.5	429.2	420.1	0.3833
	(237.2495)	(283.8129)	(274.3959)	
Grade size	148.2	158.3	156.2	0.5707
	(84.1732)	(108.7109)	(103.8024)	
% male	0.5226	0.5170	0.5182	0.8933
	(0.2221844)	(0.1609274)	(0.1749607)	
% students that changed school	0.1431	0.20171	0.1892	0.02063
	(0.1148871)	(0.161064)	(0.1539644)	
% senior teacher	3.880	3.938	3.926	0.3681
	(0.3213181)	(0.3505061)	(0.3443208)	
% diploma teacher	0.9627	0.9763	0.9734	0.1395
	(0.04837496)	(0.03580698)	(0.03904893)	
Age teacher	4.111	4.115	4.114	0.9455
	(0.3228568)	(0.370112)	(0.3595725)	
% full time teacher	0.19890	0.191039	0.19271	0.7829
	(0.1429099)	(0.1506262)	(0.1485957)	
% female teacher	0.5865	0.5830	0.5838	0.8895
	(0.1310249)	(0.1199462)	(0.1219571)	
% senior director	5.677	5.587	5.606	0.6963
	(1.131109)	(1.262625)	(1.232801)	
% diploma director	0.9848	0.9918	0.9903	0.6881
	(0.08703883)	(0.09053575)	0.08956909	
Age director	6.035	6.037	6.037	0.9946
	(1.127126)	(1.265947)	(1.234174)	
% full time director	0.8636	0.944	0.9269	0.1072
	(0.2617202)	(0.1935729)	(0.2115734)	
%female director in the sample	0.3889	0.3566	0.3634	0.7048
	(0.4283096)	(0.4480934)	(0.4427836)	
# observations	33	122	155	

Notes: The p-values (obtained from a t-test) show that the treated and the control group do not statistically significantly differ (at 1%) for the environmental variables, except for the share of disadvantaged students.

Table 5. Sample means in 2010 for the schools involved in question 2.

	treated	*control*	*overall*	*p − value*
INPUTS				
Operating Grant per student	984.4	963.8	965.4	0.6568
	(282.6928)	(172.9375)	(183.6631)	
Teaching Hours per student	2.681	2.4763	2.4925	0.2856
	(1.164998)	(0.6537022)	(0.7079448)	
OUTPUTS				
% students with 'A certificate'	88.59	82.42	82.91	0.0000
	(7.475411)	(13.31541)	(13.0522)	
% students with no problem of absenteeism	99.65	98.75	98.82	0.0000
	(0.7579673)	(2.884149)	(2.7863)	
% students progressing through school	97.33	96.02	96.12	0.01518
	(3.043412)	(3.859265)	(3.815009)	
ENVIRONMENTAL VARIABLES				
% disadvantaged students	0.2667	0.4096	0.39833	5.504e-06
	(0.1655324)	(0.1793379)	(0.1822579)	
School size	393.3	470.7	464.6	0.06281
	(240.9295)	(268.927)	(267.4152)	
Grade size	137.6	204.9	199.6	0.0000
	(82.26107)	(94.3286)	(95.11148)	
% male	0.5292	0.51896	0.51978	0.7756
	(0.2149695)	(0.2131428)	(0.2130861)	
% students that changed school	0.136803	0.10529	0.10778	0.1768
	(0.1403386)	(0.09706514)	(0.1013249)	
% senior teacher	3.936	3.843	3.851	0.1179
	(0.3418761)	(0.4199496)	(0.4147775)	
% diploma teacher	0.9664	0.9544	0.9554	0.1175
	(0.04507506)	(0.04379324)	(0.04396812)	
Age teacher	4.152	4.083	4.089	0.2438
	(0.3484326)	(0.3367726)	(0.337851)	
% full time teacher	0.2257	0.2638	0.2608	0.09827
	(0.1359717)	(0.1249585)	(0.1261355)	
% female teacher	0.5727	0.5954	0.5936	0.3557
	(0.1467669)	(0.133465)	(0.1345423)	
% senior director	5.184	5.506	5.481	0.1607
	(1.374791)	(1.057899)	(1.088057)	
% diploma director	0.9744	0.996	0.9943	0.406
	(0.1601282)	(0.04336415)	(0.06117279)	
Age director	5.692	5.939	5.920	0.2717
	(1.346176)	(1.12278)	(1.142384)	
% full time director	0.9915	0.9393	0.9434	2.396e-05
	(0.05337605)	(0.1769421)	(0.1710249)	
%female director in the sample	0.2479	0.3281	0.3217	0.2357
	(0.3989603)	(0.4107215)	(0.4099812)	
# observations	39	455	494	

Notes: The p-values show that the treated and the control group do not statistically significantly differ (at 1%) for the environmental variables, except for the share of disadvantaged students, grade size and the share of fulltime director.

Table 6. Sample means in 2011 for the schools involved in question 1.

	treated	*control*	*overall*	*p − value*
INPUTS				
Operating Grant per student	1001.7	865.6	894.6	0.0004679
	(190.5457)	(154.7815)	(171.7345)	
Teaching Hours per student	2.475	2.017	2.115	0.0006162
	(0.6391851)	(0.6340315)	(0.6603352)	
OUTPUTS				
% students with 'A certificate'	86.84	90.59	89.79	0.02331
	(8.484661)	(6.682222)	(7.240065)	
% students with no problem of absenteeism	99.10	99.78	99.63	0.1024
	(2.257923)	(1.014307)	(1.394621)	
% students progressing through school	96.98	98.32	98.04	0.01791
	(2.94588)	2.013347	(2.30055)	
ENVIRONMENTAL VARIABLES				
% disadvantaged students	0.3715	0.16054	0.20545	4.123e-06
	(0.2158472)	(0.1268982)	(0.1727391)	
School size	379.7	426.3	416.4	0.3391
	(235.4844)	(283.6676)	(274.0685)	
Grade size	144.9	157.81	155.1	0.4496
	(80.66666)	(105.8395)	(100.9055)	
% male	0.5262	0.51090	0.51416	0.7067
	(0.2157745)	(0.1650449)	(0.1763995)	
% students that changed school	0.13653	0.18700	0.17626	0.05224
	(0.1213704)	(0.1581608)	(0.1521353)	
% senior teacher	3.904	3.966	3.953	0.4099
	(0.3856587)	(0.3579052)	(0.3635934)	
% diploma teacher	0.9623	0.9753	0.9725	0.1425
	(0.04603574)	(0.03644109)	(0.03888484)	
Age teacher	4.131	4.147	4.143	0.8239
	(0.3430423)	(0.384139)	(0.3747466)	
% full time teacher	0.20729	0.197149	0.19931	0.7313
	(0.14842)	(0.1542786)	(0.1526308)	
% female teacher	0.5974	0.5855	0.5881	0.6515
	(0.1368495)	(0.1150701)	(0.1196613)	
% senior director	5.745	5.679	5.693	0.7726
	(1.134975)	(1.239125)	(1.21442)	
% diploma director	1	1	1	–
	0	0	0	
Age director	6.116	6.117	6.117	0.9957
	(1.242676)	(1.264709)	(1.256035)	
% full time director	0.8838	0.9276	0.9183	0.3582
	(0.2446825)	(0.2236968)	(0.228212)	
%female director in the sample	0.4066	0.3593	0.3694	0.584
	(0.4358887)	(0.4422688)	(0.4399367)	
# observations	33	122	155	

Notes: The p-values show that the treated and the control group do not statistically significantly differ (at 1%) for the environmental variables, except for the share of disadvantaged students.

Table 7. Sample means in 2011 for the schools involved in question 2.

	treated	*control*	*overall*	*p − value*
INPUTS				
Operating Grant per student	887.4	1002.8	993.7	0.0001065
	(156.5869)	(250.079)	(245.8712)	
Teaching Hours per student	2.226	2.508	2.486	0.01024
	(0.6151478)	(0.8374979)	(0.8251728)	
OUTPUTS				
% students with 'A certificate'	88.43	83.20	83.61	0.0003153
	(7.594137)	(13.29989)	(13.01259)	
% students with no problem of absenteeism	99.53	98.80	98.86	0.0006807
	(1.069318)	(2.527001)	(2.451149)	
% students progressing through school	97.97	96.59	96.70	0.0005955
	(2.18339)	(3.121399)	(3.07891)	
ENVIRONMENTAL VARIABLES				
% disadvantaged students	0.23531	0.4076	0.39399	4.032e-10
	(0.1294017)	(0.1810883)	(0.1834448)	
School size	387.5	466.8	460.5	0.05393
	(237.5262)	(268.8114)	(267.1129)	
Grade size	141.0 5	202.0	197.2	0.0000
	(78.84795)	(94.10131)	(94.3642)	
% male	0.5279	0.519151	0.519845	0.8042
	(0.21128)	(0.2123551)	(0.2120701)	
% students that changed school	0.14605	0.10492	0.10817	0.1095
	(0.1545449)	(0.09385799)	(0.1003827)	
% senior teacher	3.898	3.843	3.847	0.4369
	(0.4196154)	(0.442257)	(0.4403544)	
% diploma teacher	0.9694	0.9577	0.9587	0.0806
	(0.0387241)	(0.04331641)	(0.04305068)	
Age teacher	4.136	4.092	4.096	0.2438
	(0.3890938)	(0.351417)	(0.354307)	
% full time teacher	0.2273	0.2743	0.2705	0.04439
	(0.1365289)	(0.1308032)	(0.1317331)	
% female teacher	0.5788	0.5969	0.5955	0.4129
	(0.1317365)	(0.1306346)	(0.1306793)	
% senior director	5.38	5.562	5.548	0.433
	(1.399117)	(1.069024)	(1.098043)	
% diploma director	0.9615	0.9934	0.9909	0.2702
	(0.1771342)	(0.05709988)	(0.0741279)	
Age director	5.731	5.975	5.956	
	(1.281683)	(1.143207)	(1.155213)	0.255
% full time director	0.9872	0.9305	0.935	0.0005587
	(0.08006408)	(0.1976313)	(0.1915639)	
%female director in the sample	0.2436	0.3444	0.3364	0.1349
	(0.3952314)	(0.4145555)	(0.4135715)	
# observations	39	455	494	

Notes: The p-values show that the treated and the control group do not statistically significantly differ (at 1%) for the environmental variables, except for the share of disadvantaged students, grade size and the share of fulltime director.

Appendix B

parallel trends assumption

The first step to investigate the parallel trends assumption is through a graphical representation of the trends of the outcome variable for treated and control, as reported in Figures 1–6. These graphs provide a first indication on the validity of the parallel trends assumption. Besides, this analysis is informative on at least two different levels. First, it can be noticed that, on average, the schools which receive the extra funds in 2011 perform worse than the school which not receive the extra funds (see Figures 1–3 relative to sample 1); while, the schools which stop receiving funds in 2011, on average, perform better than the school which have received the founds before and after 2011 (see Figures 4–6 relative to sample 2). This phenomenon suggest that the policy is targeting the schools which are performing worse. Second, the difference among the trend of the share of A certificate and the trend of the share of students with no problem of absenteeism highlights the importance of a multidimensional analysis (Tables 8 and 9).

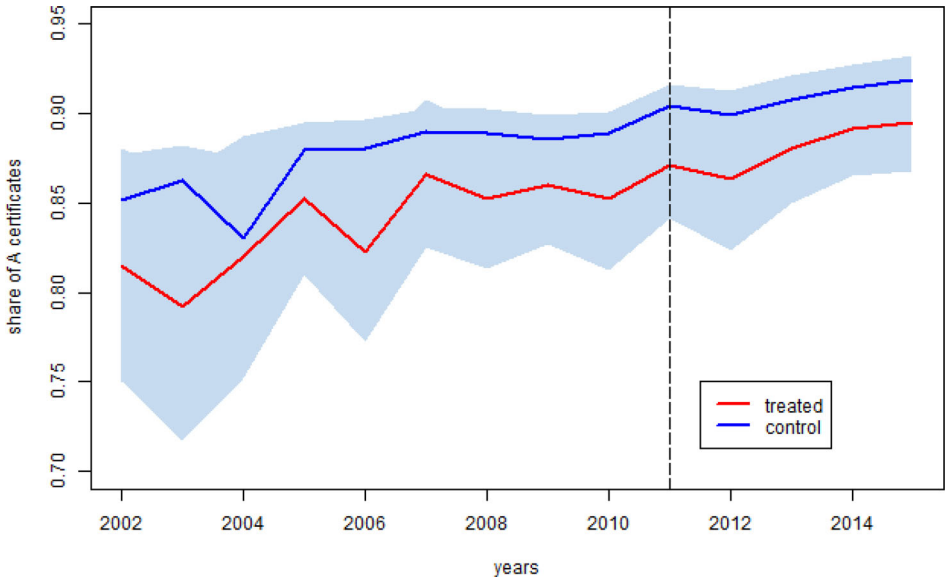

Figure 1. Trend for the variable "A certificate" for the schools involved in question 1.
The blue line represents the control group, the red line the treated group. The blue line represents the control group, i.e., the schools which did not receive funds neither in the cycle 2008-2010, neither in the cycle 2011-2013. The red line represents the treated group, i.e., the schools which started receiving the funds in 2011. This graphical representation supports the parallel trends assumption and shows that the schools in the control group, on average, perform better than the schools which received the extra funds. This can be explained by considering that the schools in the treated group are characterized by higher share of disadvantaged students, a characteristics that results in lower share of A certificates at school's level.

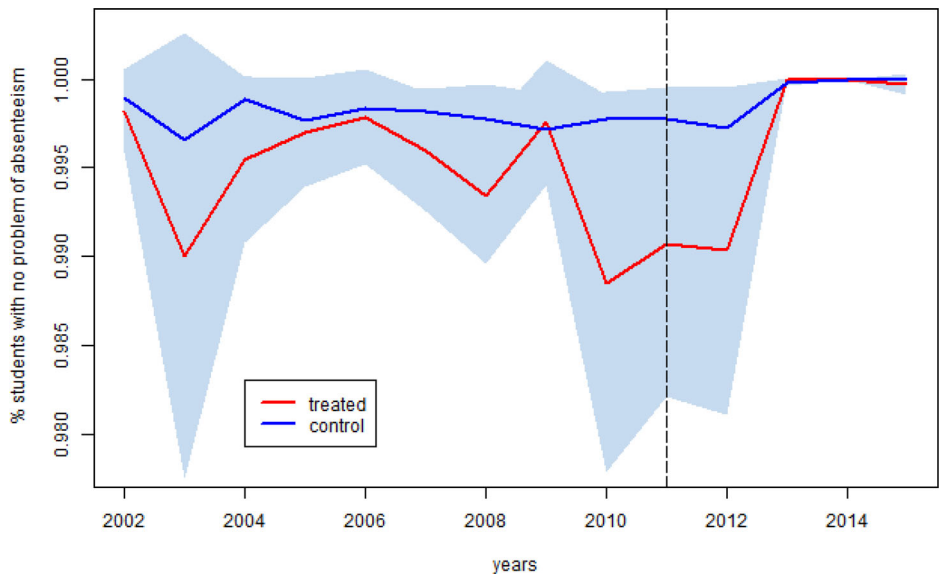

Figure 2. Trend for the variable "no problem of absenteeism" the schools involved in question 1.
The blue line represents the control group, i.e., the schools which did not receive funds neither in the cycle 2008-2010, neither in the cycle 2011-2013. The red line represents the treated group, i.e., the schools which started receiving the funds in 2011. This graphical representation supports the parallel trends assumption and shows that the schools in the control group, on average, perform better than the schools which received the extra funds. This can be explained by considering that the schools in the treated group are characterized by higher share of disadvantaged students, a characteristics that results in higher share of students with absenteeism problem.

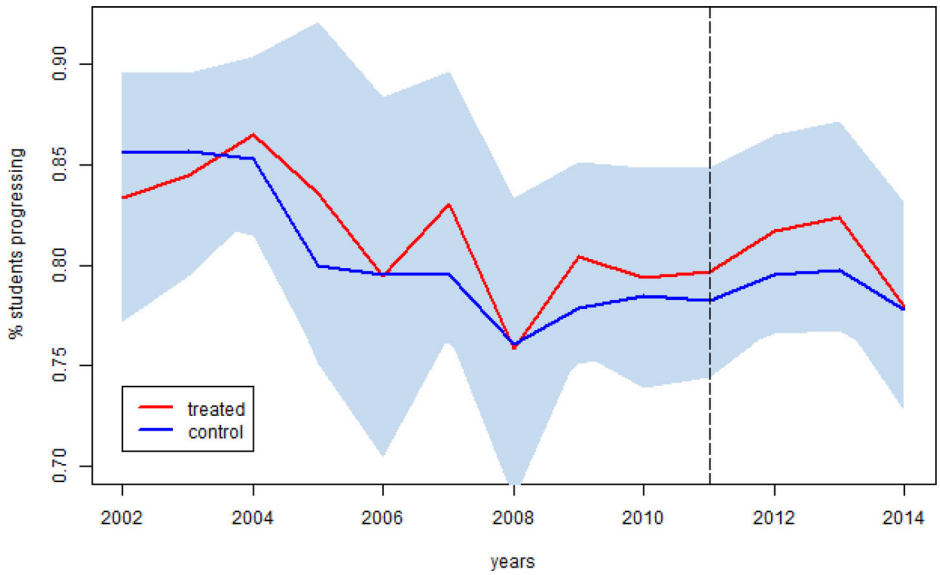

Figure 3. Trend for the variable "Progress" for the schools involved in question 1.
The blue line represents the control group, i.e., the schools which did not receive funds neither in the cycle 2008-2010, neither in the cycle 2011-2013, the red line represents the treated group, i.e., the schools which started receiving the funds in 2011. This graphical representation supports the parallel trends assumption and shows that the schools in the treated group, on average, perform better than the schools which did not receive extra funds. This suggest that the presence of higher share of disadvantaged students results, on average, in higher share of students able to progress in schools for the schools that did not receive the funds in the first cycle.

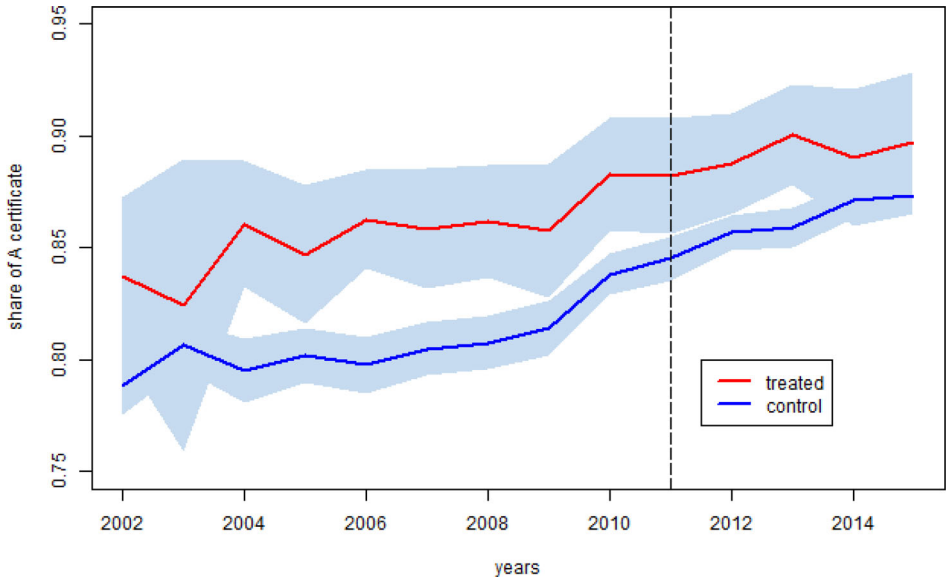

Figure 4. Trend for the variable "A certificate" for the schools involved in question 2.
The blue line represents the control group, i.e., the schools which received funds in the cycle 2008-2010 and in the cycle 2011-2013, the red line represents the treated group, i.e., the schools which stopped receiving the funds in 2011. This graphical representation supports the parallel trend assumption and shows that the schools in the treated group, on average, perform better than the schools which stop receiving the extra funds. This can be explained by considering that the schools in the treated group are characterized by lower share of disadvantaged students, a characteristics that results in better share of A certificates at school's level.

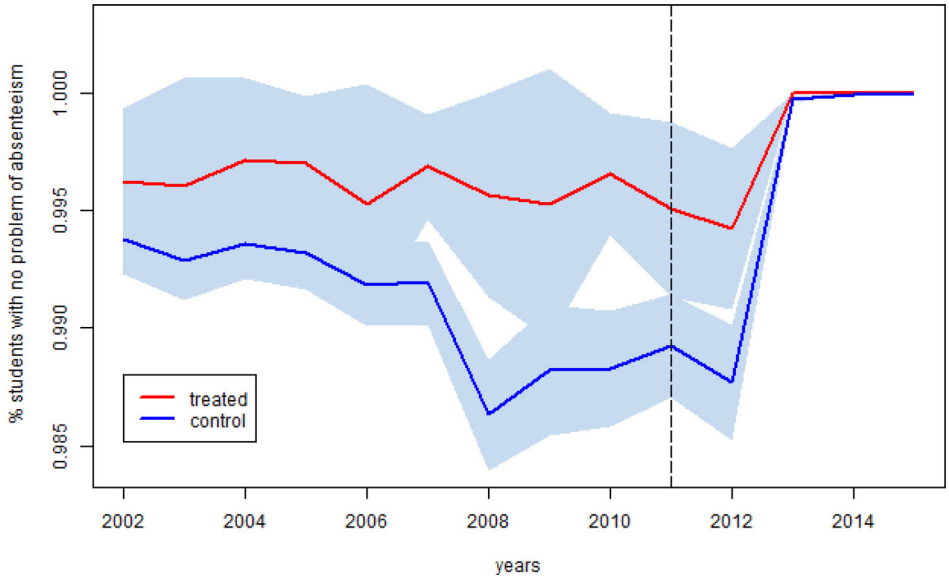

Figure 5. Representation of the trend for the variable "no problem of absenteeism" for the schools involved in question 2.
The blue line represents the control group, i.e., the schools which received funds in the cycle 2008-2010 and in the cycle 2011-2013, the red line represents the treated group, i.e., the schools which stopped receiving the funds in 2011. This graphical representation supports the parallel trend assumption and shows that the schools in the treated group, on average, perform better than the schools which stop receiving the extra funds. This can be explained by considering that the schools in the treated group are characterized by lower share of disadvantaged students, a characteristics that results in higher share of students with no problem of absenteeism at school's level.

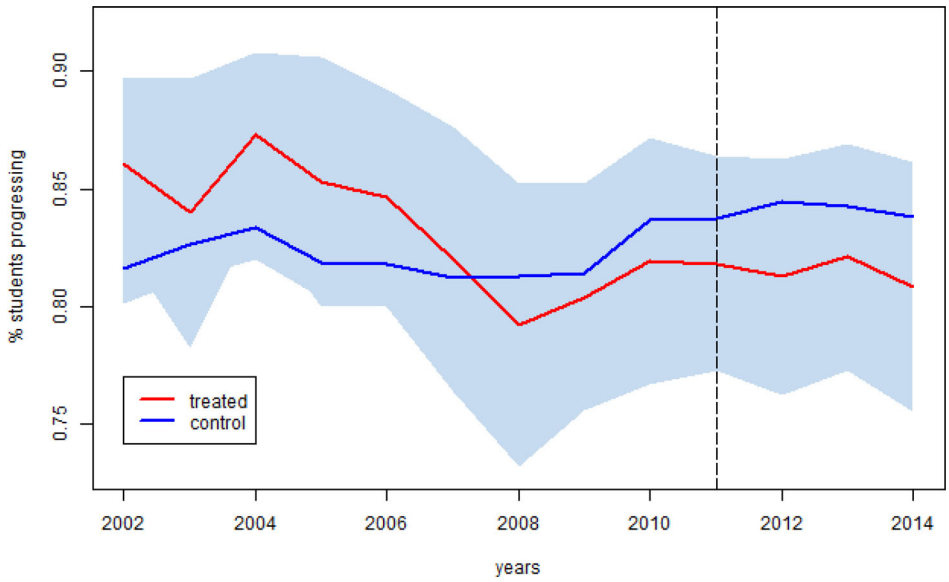

Figure 6. Representation of the trend for the variable "Progress" for the schools involved in question 2.
The blue line represents the control group, i.e., the schools which received funds in the cycle 2008-2010 and in the cycle 2011-2013, the red line represents the treated group, i.e., the schools which stopped receiving the funds in 2011. This graphical representation supports the parallel trend assumption and shows that the schools in the treated group, on average, performed better than the schools which stop receiving the extra funds only until 2007, while after that year we observe an inversion in the ranking. This suggest that there is only a weak correlation between the share of disadvantaged students in a school and the share of students in that schools who are able to progress.

Table 8. Regression to check the parallel trends assumption.

	Schools involved in question 1		
	(1)	(2)	(3)
	A certificate	no PA	Progress
treatment	−0.040*	−0.014	0.011
	(0.024)	(0.024)	(0.024)
year_2003 × treatment	−0.030	0.007	−0.023
	(0.053)	(0.053)	(0.053)
year_2004 × treatment	0.030	0.010	0.001
	(0.052)	(0.052)	(0.051)
year_2005 × treatment	0.013	0.013	0.025
	(0.051)	(0.051)	(0.051)
year_2006 × treatment	−0.017	0.013	−0.012
	(0.049)	(0.049)	(0.049)
year_2007 × treatment	0.017	0.011	0.025
	(0.049)	(0.049)	(0.049)
year_2008 × treatment	0.003	0.009	−0.013
	(0.043)	(0.043)	(0.043)
year_2009 × treatment	0.014	0.014	0.014
	(0.043)	(0.043)	(0.043)
year_2010 × treatment	0.004	0.004	−0.002
	(0.043)	(0.043)	(0.043)
year_2011 × treatment	0.007	0.007	0.003
	(0.043)	(0.043)	(0.043)
year_2012 × treatment	0.004	0.007	0.011
	(0.043)	(0.043)	(0.043)
year_2013 × treatment	0.013	0.014	0.015
	(0.043)	(0.043)	(0.043)
year_2014 × treatment	0.017	0.014	−0.010
	(0.043)	(0.043)	(0.043)
Constant	0.671***	0.742***	0.527***
	(0.011)	(0.011)	(0.011)
Observations	1,930	1,930	1,930
R^2	0.184	0.225	0.246
Adjusted R^2	0.174	0.214	0.236
Residual Std. Error (d*f* = 1904)	0.181	0.180	0.179
F Statistic (d*f* = 25; 1904)	17.202***	22.063***	24.882***
Time Fixed Effect	Yes	Yes	Yes

In this table we report results from the estimation of equation 9, considering three outcomes: the share of A certificate, the share of student without Problem of Absenteeism and the share of students able to Progress in school. The regression is implemented for the sample of schools that did not receive the extra funds before 2011. Since the coefficient of the interaction terms relative to the period before the treatment are not significant, the parallel trends assumption is fulfilled.
Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***} < 0.01$.

Table 9. Regression to check the parallel trends assumption.

	Schools involved in question 2		
	A certificate	no PA	progress
	(1)	(2)	(3)
treatment	0.037*	0.016	−0.004
	(0.021)	(0.021)	(0.017)
year_2003 × treatment	−0.019	−0.013	0.017
	(0.044)	(0.043)	(0.034)
year_2004 × treatment	0.029	−0.012	0.044
	(0.043)	(0.043)	(0.033)
year_2005 × treatment	0.008	−0.012	0.038
	(0.042)	(0.042)	(0.033)
year_2006 × treatment	0.028	−0.012	0.032
	(0.042)	(0.041)	(0.032)
year_2007 × treatment	0.017	−0.011	0.012
	(0.041)	(0.041)	(0.032)
year_2008 × treatment	0.017	−0.007	−0.017
	(0.040)	(0.040)	(0.031)
year_2009 × treatment	0.007	−0.009	−0.006
	(0.040)	(0.040)	(0.031)
year_2010 × treatment	0.008	−0.008	−0.013
	(0.040)	(0.040)	(0.031)
year_2011 × treatment	0.0001	−0.010	−0.016
	(0.040)	(0.040)	(0.031)
year_2012 × treatment	−0.006	−0.009	−0.028
	(0.040)	(0.040)	(0.031)
year_2013 × treatment	0.005	−0.016	−0.017
	(0.040)	(0.040)	(0.031)
year_2014 × treatment	−0.018	−0.016	−0.026
	(0.040)	(0.040)	(0.031)
Constant	0.578***	0.688***	0.548***
	(0.006)	(0.006)	(0.004)
Observations	7,038	7,038	7,038
R^2	0.199	0.259	0.331
Adjusted R^2	0.196	0.256	0.329
Residual Std. Error (d*f* = 7012)	0.201	0.200	0.156
F Statistic (d*f* = 25; 7012)	69.747***	97.991***	139.051***
Time Fixed Effect	Yes	Yes	Yes

In this table we report results from the estimation of equation 9, considering three outcomes: the share of A certificate, the share of student without Problem of Absenteeism and the share of students able to Progress in school. The regression is implemented for the sample of schools that received the extra funds before 2011. Since the coefficient of the interaction terms relative to the period before the treatment are not significant, the parallel trends assumption is fulfilled.
Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.