European Journal of Operational Research 310 (2023) 432–447

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

On distinguishing the direct causal effect of an intervention from its efficiency-enhancing effects

Oleg Badunenko[a], Giovanna D'Inverno[b,c,*], Kristof De Witte[c,d]

[a] Department of Economics and Finance, College of Business, Arts and Social Sciences, Brunel University London, Uxbridge, UB8 3PH, UK
[b] Department of Economics and Management, University of Pisa, Via C. Ridolfi 10, Pisa 56124, Italy
[c] Leuven Economics of Education Research, Faculty of Economics and Business, KU Leuven, Naamsestraat 69, Leuven 3000, Belgium
[d] UNU-Merit, Maastricht University, Boschstraat 24, AX Maastricht 6211, Netherlands

A R T I C L E I N F O

Article history:
Received 17 February 2022
Accepted 24 March 2023
Available online 30 March 2023

JEL classification:
C54
H52
I22
I24
I28

Keywords:
Education
Stochastic frontier
Persistent and transient technical inefficiency
Impact evaluation
Causal inference

A B S T R A C T

This paper proposes an innovative methodology for handling endogeneity issues in the evaluation of policy performance. By estimating a regression discontinuity design with a four-component stochastic frontier panel data model, we estimate the causal impact of a policy intervention on the outcome variable, whenever the treatment status depends on an exogenous threshold. We distinguish between (i) the direct effect of the intervention, (ii) the efficiency-enhancing effect, or (iii) their combination. Moreover, we distinguish between persistent (time-invariant) and transient (time-varying) inefficiency components while accounting for unobserved heterogeneity, which is important for policy implications. We showcase the practical usefulness of the proposed approach by estimating the effect of providing additional resources on schools that exceed an exogenously set share of disadvantaged students in secondary schools in Flanders, Belgium. We also demonstrate the trade-off between balance of the covariates in the treated and control group and statistical power. Thus, despite insignificant effects in a balanced but smaller sample close to the discontinuity, the results become significant in the unbalanced sample with more statistical power. In both samples, we observe that the policy had an effect on the outcome mostly through the efficiency-enhancing channel. To this extent, we show that the model specification including both direct and indirect effects outperforms the other two specifications and it offers a more exhaustive perspective from a policy view point.

1. Introduction

Policy evaluation is receiving an increasing amount of attention in many fields (Abadie & Cattaneo, 2018). In the absence of the possibility to conduct randomized controlled trials, the regression discontinuity design (RDD) has become one of the workhouses of policy evaluation. In a nutshell, the RDD uses exogenous information in a locality defined by the policy and then generalizes the outcome to infer about the effectiveness of the policy (Lee & Lemieux, 2010; Wing & Bello-Gomez, 2018). The explanatory variables are usually indicators of performance. However, this literature ignores that these variables might not achieve their full potential, such that effectiveness and efficiency are confused. A typical hypothesis tested in an RDD study is whether a policy has had a direct effect on the outcome variable (after controlling for the environment, see for example Calonico, Cattaneo, Farrell, & Titiunik (2019)). However, if one accounts for the possibility of inefficiency, a policy may also be successful through its efficiency-enhancing effect. In the present paper, we propose an approach to analyze whether a policy had a (causal) effect on the outcome variable and whether this effect is (i) direct, (ii) efficiency-enhancing, or (iii) both. To do so, we rely on stochastic frontier (SF) analysis, which is an econometric approach to model a production process and which allows us to identify observations that operate below their potential (Aigner, Lovell, & Schmidt, 1977; Meeusen & van den Broeck, 1977). By combining recent developments of the SF modeling for panel data with policy evaluation, we consider an RDD design in the SF framework that allows for identifying two types of inefficiency: persistent (time-invariant) and transient (time-varying).

Panel-specific SF models have significantly evolved over time. The first-generation SF models for panel data assume that the inefficiency is different across observations but is time-invariant (see Battese & Coelli, 1988; Pitt & Lee, 1981). The second-generation SF models allow time-varying inefficiency by using the scale device.

* Corresponding author.
E-mail addresses: oleg.badunenko@brunel.ac.uk (O. Badunenko), giovanna.dinverno@unipi.it (G. D'Inverno), kristof.dewitte@kuleuven.be (K. De Witte).

https://doi.org/10.1016/j.ejor.2023.03.030
0377-2217/

More specifically, the time function shrinks or inflates the distribution of time-invariant inefficiency without changing its basic shape. Several functions of time have been suggested (see Battese & Coelli, 1992; Cornwell, Schmidt, & Sickles, 1990; Kumbhakar, 1990). Kumbhakar, Ghosh, & McGuckin (1991), Reifschneider & Stevenson (1991) followed by Battese & Coelli (1995) have suggested more complicated models that allow determinants of inefficiency; Caudill, Ford, & Gropper (1995) developed a model that allows non-*iid* errors and inefficiency components. These previous models, though, do not account for the potential heterogeneity of observations. The third-generation SF models for panel data include three components. The noise and time-varying inefficiency components stem from the second-generation SF models, and are augmented with a time-invariant component. Kumbhakar & Hjalmarsson (1993), Kumbhakar & Heshmati (1995), Kumbhakar & Hjalmarsson (1995) interpreted this time-invariant component as persistent inefficiency. The third-generation models by Park, Sickles, & Simar (1998, 2003, 2007) extended the first-generation SF models by assuming the third component to be a random effect and inefficiency as time-constant. The third-generation models by Greene (2005) adopt the view that the time-invariant component is either random or fixed effect and inefficiency is time-varying. Kneip, Sickles, & Song (2012) consider nonconstant individual effects and a general average function in place of the intercept. The fourth-generation SF models for panel data, also known as the four-component model, were introduced by Colombi, Kumbhakar, Martini, & Vittadini (2014) and Kumbhakar, Lien, & Hardaker (2014). This class of models encompasses all previous generations of SF models by considering both time-invariant (persistent) and time-varying (transient) inefficiencies, heterogeneity, and usual stochastic noise. Filippini & Greene (2016) and Badunenko & Kumbhakar (2016) study the model with *iid* efficiencies, while Badunenko & Kumbhakar (2017) and Lai & Kumbhakar (2018b) consider determinants of efficiency. Recently, a model that accounts for spatially autocorrelated persistent and transient inefficiency has been proposed (Skevas & Skevas, 2021), where the estimation procedure relies on Bayesian approach. Numerous latest applications of the 4-component model that breaks inefficiency down into persistent and transient parts can be found in manufacturing (Amjadi & Lundgren, 2022), energy economics (Filippini, Greene, & Masiero, 2018), agricultural production (Adom & Adams, 2020), banking (Badunenko, Kumbhakar, & Lozano-Vivas, 2021), and among others, education (Badunenko, Mazrekaj, Kumbhakar, & De Witte, 2021).

Next to a panel data structure, the SF models have been advanced to accommodate endogeneity issues. For the cross-sectional data, Kutlu (2010) deals with endogeneity where the regressors are correlated with the noise, but independent of inefficiency. Mutter, Greene, Spector, Rosko, & Mukamel (2013) consider cases where endogeneity arises in cross-sectional SF models because of the correlation between a regressor and noise, and the correlation between a regressor and inefficiency component. They find that if the endogenous regressor is included in the regression, the effect of noise-related endogeneity on efficiency estimates is small, while the bias size due to inefficiency component-related endogeneity can be large depending on the strength and direction of the correlation between a regressor and inefficiency component. Amsler, Prokhorov, & Schmidt (2016) survey various procedures for treating endogeneity of both types and adopt some standard procedures in an SF setting. Amsler, Prokhorov, & Schmidt (2017) consider a model where determinants of inefficiency modeled using the scaling device can be correlated with both the noise and inefficiency component.

Some of the models that accommodate endogeneity issues were suggested to deal with panel data. Tran & Tsionas (2013) extend the approach of Kutlu (2010) to estimate the panel data SF model which has two components, noise, and time-varying inefficiency component also known as the second-generation SF model. Karakaplan & Kutlu (2017a) employ a scaling device and propose a model that accounts for both types of endogeneity in the second-generation SF model. Griffiths & Hajargasht (2016) consider the possibility of both the noise and inefficiency component to be correlated with regressors. They consider the first-generation SF model where the inefficiency term is time-constant as well as the fourth-generation. Lai & Kumbhakar (2018a) extend the four-component model studied by Griffiths & Hajargasht (2016) by allowing determinants of inefficiency. Lien, Kumbhakar, & Alem (2018) do the same, but instead of econometric treatment, they use a behavioral assumption that firms maximize returns to the outlay to account for possible endogeneity of inputs and outputs.

However, the focus on the endogeneity of regressors might be insufficient in the evaluation of a policy, where endogeneity might come from other angles as well (e.g. self-selection, omitted variable bias, reversed causality). Therefore, SF analysis has also been employed in assessing the effectiveness of policies. Perelman & Santin (2011) investigate whether the choice of a school is important for explaining differences in students' underperformance in Spanish public and private-voucher schools. The potential for school self-selection bias to corrupt underperformance estimates is studied by Crespo-Cebada, Pedraja-Chaparro, & Santín (2014). The authors proceed in two stages. First, they use propensity score matching to identify similar-characteristic samples of students from public and private schools. They then estimate two different SF models. Karakaplan & Kutlu (2017c) investigate the effect of consolidation of public school districts on cost savings. They argue that since the decision to consolidate is not random, it introduces endogeneity, which once handled reduces the estimates of potential savings. They proceed by applying the model of Karakaplan & Kutlu (2017b) to accommodate the endogeneity of district achievement and education market concentration. The contribution closest to the standard policy evaluation is by Johnes & Tsionas (2019), where the RDD framework is introduced into the Kumbhakar et al. (1991) model. The authors include all available observations as opposed to only those that are close to the known threshold defined by the policy. In this case, some confounding elements might still be present in the estimates of the policy thus this approach departs from the threshold SF models (e.g. Lai, 2013; Wang & Huang, 2009; Yélou, Larue, & Tran, 2010).

This paper evaluates the effectiveness of an education policy that aims to improve the school outcomes of disadvantaged students. Already in their seminal paper, Charnes, Cooper, & Rhodes (1981) evaluated the benefits of Program Follow Through (PFT), a federally sponsored program aimed at providing remedial assistance to educationally disadvantaged early primary school students, while allowing managerial inefficiency. Having estimated three different technologies and disentangling managerial efficiency from program efficiency, the authors observed that the PFT program has not resulted in greater managerial efficiencies. More recently, D'Inverno, Smet, & De Witte (2021) have proposed to use a quasi-experimental setting within a metafrontier conditional order-m framework to decompose the overall efficiency into managerial and program efficiency. Unlike Charnes et al. (1981), the approach of D'Inverno et al. (2021) can be used for making causal interpretations since it accounts for potential endogeneity by exploiting an exogenous threshold set by the policy. We refer the interested reader to the recent systematic review on policy evaluation and efficiency by Mergoni & De Witte (2022).

If the policy determines outcomes, the regression discontinuity must be thought of as the data generating process rather than the estimation approach. This paper builds on D'Inverno et al. (2021) to advance Johnes & Tsionas (2019) methodology to account for school heterogeneity as well as persistent underperformance by considering the Badunenko & Kumbhakar (2017) model.

The proposed approach is applied to study the effect of the equal educational opportunities program, implemented in the Flemish Community of Belgium, where additional resources are provided to schools with a share of disadvantaged students above a fixed threshold. We find that the policy has no direct effect on student achievement, but it does expand the production possibility set and it shows a mild positive impact through the efficiency-enhancing channel.

The paper proceeds as follows. Section 2 introduces the empirical setting to showcase the use of the proposed approach in a secondary education context. Section 3 explains the approach we propose to evaluate a policy intervention allowing underperformance in the production process. Section 4 presents the results and discusses some evidence-based policy insights. Section 5 concludes the paper.

2. Empirical application in secondary education

In this section, we discuss the empirical application. Specifically, we evaluate the causal impact on school performance of a policy intervention promoted in Flanders (Belgium), while accounting for latent school heterogeneity and distinguishing time-invariant from time-varying components of inefficiency. First, we give an overview of the policy under focus and then we outline the variables included in the analysis.

2.1. The policy under evaluation

Ensuring "Equal Educational Opportunities for all" is a widely regarded issue in many countries and increasingly spread over time. Several programs and policies have been fostered to reduce the impact of socio-economically disadvantaged backgrounds on educational achievement and accordingly to help individuals reach their full potential and life satisfaction. Among other countries, Flanders is experiencing a high social inequality level (OECD, 2017), which is in turn reflected in the polarization of the educational system and the secondary school track choice. On average, low socio-economic students are more prone to choose vocational schools, while students with high socio-economic backgrounds tend to choose general education. Besides the quiescent segregation among students, this leads also to an unequal distribution of skillful teachers.

To address the depicted issue, since 2002 the Flemish Ministry of Education has promoted a program to provide additional funding to secondary schools with a significant proportion of disadvantaged students, known as the 'Equal Educational Opportunities ("*gelijkeonderwijskansenbeleid*, GOK") program'. According to this initiative, a student can be considered "disadvantaged" if he/she meets at least one of these five criteria: a low mother education level, a low family income, a non-Dutch speaking family, being part of the traveling population, and/or living alone. The school's eligibility for the additional resources depends on an exogenous cut-off set by law in relation to the share of disadvantaged students in the school (10% for the first two years and 25% for the next four years of secondary school) and it is assessed every three years. These additional resources are meant for teaching-related purposes, like teacher support, or for hiring additional teachers (for that reason, henceforth defined alternatively in terms of extra-teaching hours).[4]

2.2. The education production, the data and variables

To estimate the causal impact of additional funding on school efficiency, we consider education production as a process of converting multiple inputs (e.g., school resources, teaching hours) into multiple outputs (e.g., student achievement, educational results), following the idea of the education production function originally proposed by Levin (1974), Hanushek (1979) and Ruggiero (1996). As discussed in the efficiency in the education literature review by De Witte & López-Torres (2017), the education production function can be estimated at different levels and include different inputs and outputs. In the present application, we focus on the underperformance observed at the school level (see for example Ruggiero, 2003).

The Flemish Ministry of Education provided us with a unique panel dataset at the school and pupil levels. The sample is a balanced panel with 2568 observations on 642 Flemish schools from 2010/2011 to 2013/2014, representing more than 90% of all the secondary schools in Flanders and spanning between two cycles of the policy program (one started in 2008/2009 and the other in 2011/2012). More specifically, we focus on the last four years of secondary education where the threshold is set at 25%, so as to have more balance among the units above and below the threshold.

2.2.1. Inputs, outputs and explanatory variables

In compliance with the related literature, we choose two inputs, namely the amount of teaching hours per student (*Teaching hours per student*) and the total budget allocated per student in each school (*Operating grants per student*). As a counterpart, we select two outputs in accordance with the existing literature and the Flemish educational context. The first output is the *Share of students without grade retention*. This variable has been included as a measure of school performance both in the impact evaluation and efficiency in education literature (see for example Grosskopf, Hayes, & Taylor, 2014; Pedraja-Chaparro, Santín, & Simancas, 2016). It seizes how successfully the schools encourage not only the interest and curiosity of their students, but also their learning and progress. The second output is the *Share of students with an "A" certificate*. As the Flemish educational system does not foresee external standardized test scores, this variable is the closest proxy to student test scores, commonly used as output in the literature (see for all De Witte & López-Torres, 2017; D'Inverno et al., 2021). It measures the student achievement a school can boost. Specifically, this variable measures the share of students that obtain an "A" certificate at the end of the school year and accordingly can proceed to the following year without program restrictions (as it would happen instead if obtaining a "B" or "C" certificate).

Besides the inputs and the outputs, the efficiency in education literature identifies another category of variables that are not under the direct control of the school management, but still influence the educational production process environment (and for this reason they are referred to as "environmental" or "explanatory" variables, see for all De Witte & López-Torres (2017); O'Donnell, Fallah-Fini, & Triantis (2017); Ruggiero (2019)). The RDD literature acknowledges the importance of these variables and foresees their inclusion in the model estimation as well (see for example the discussion in Calonico et al. (2019) and Frölich & Huber (2019)). Exploring the influence of these environmental variables on efficiency can be of interest to both the government that promotes the policy and the school management. Specifically, we define three groups of explanatory variables based on school, teacher, and student characteristics, that can have either a time-varying (transient) or time-invariant (persistent) influence on the school underperformance.

As school characteristics, first, we observe the share of students enrolled in the *vocational education* track, capturing the distribution of students across the different secondary education tracks and reflecting in turn the socio-economic background composition of the school itself as explained above. This variable addresses the open debate about the role of educational tracking and its in-

[4] More details on the program and the Flemish education system can be found in D'Inverno et al. (2021).

fluence on student performance and inequality (see for example Hanushek & Wößmann, 2006; Van de Werfhorst, 2019). Second, we consider the *school size* measured as the number of students enrolled in the school. In the education economics literature there is still a lack of consensus on the direction of its influence on school performance, if any, and accordingly, it becomes interesting to account for it (Leithwood & Jantzi, 2009). Third, there is a *share of students changing school* in the next school year. This variable seizes the share of students that are pushed away from the school they are currently enrolled in and, for this reason, it can be seen as a good proxy of the harshness that the educational production has to face. Fourth, we include a dummy variable for the school management that is equal to one in case the school is *public*. Finally, there is a dummy *previously treated school* equal to one if the school received additional funding due to the policy program also in the year before. This variable is useful to investigate the existence of a learning effect in handling the extra resource over time.

The second group of characteristics deals with the teacher characteristics measured at the school level. Among scholars, there is quite solid agreement about the importance of the teacher's role in the educational process and accordingly it is necessary to control for it (Hanushek, Piopiunik, & Wiederhold, 2019; Hanushek & Wößmann, 2015). The variable *teacher full-time* measures the share of teachers that have a full-time contract. Exploring its influence on school performance can shed light on the debated issue concerning the choice of the contract type, whether full or part-time at the institutional level. The third and last group deals with the student characteristics observed at the school level, so to capture the heterogeneity of the environment each school has to operate in while providing education. We consider the *share of students with grade retention in primary school* and the *share of special needs students in primary school*, as proxies of the student composition and, more specifically, of the difficulties that can be encountered in the educational process. Finally, we account for the student gender distribution by considering the *share of male students*, as it has been observed in the literature as an element of heterogeneity in the education achievement process (Cipollone & Rosolia, 2007).

Table 1 shows the descriptive statistics of the variables introduced above together with the variable *share of disadvantaged students,* the forcing variable used in this analysis. To capture the technological change over time, we include also a *time trend* variable, where $t = 1, \ldots, 4$ respectively for 2010/2011,..., 2013/2014.

3. Methodology

To explain our approach, we first briefly describe the implementation of the RD design in practice as well as the SF approach to multi-output multi-input production function modeling. We then amalgamate both methods to estimate treatment effects in SF models. It should be noted that the methodology section is described with an application to the educational sector in mind, but it is straightforward to apply the suggested approach to any field.

The set of models that we propose are important for practitioners in many areas, and in education in particular. The growing RD literature examines whether a policy was successful. The outlined methodology goes a step beyond as we propose a methodology which first assesses whether a policy was successful at all, and second, reveals whether the channel of success was improvement of efficiency (i.e., we account for potential underperformance). The framework that we introduce involves a policy intervention with an exogenously set threshold.

3.1. Regression discontinuity design

Consider an outcome variable Y_i and the running (or forcing) variable c_i that represents the proportion of disadvantaged students in school i. Since we know that the school is eligible for additional funding if the proportion exceeds the exogenously set level c_0, we generate a dummy variable that separates schools according to the policy,

$$D_i = \begin{cases} 1 & c_i \geq c_0 \\ 0 & c_i < c_0 \end{cases} \tag{1}$$

One way to estimate the treatment effect is first estimate two separate equations of y on c using data on schools that are above and below the threshold c_0 yielding slope coefficients β_r (above) and β_l, respectively. Then look at the difference $\beta_r - \beta_l$. The approach that is adopted by researchers is estimating a comprehensive equa-

Table 1
Descriptive statistic of the variables included in the empirical analysis.

Variable	Label	Mean	Std. dev.
Education production function			
Inputs			
x1	Teaching hours per student	2.48	0.72
x2	Operating grants per student	1014	188
Outputs			
y1	Share of students without grade retention	93.8	3.6
y2	Share of students with “A” certificate	61.7	8.7
Inefficiency determinant variables			
z1	Public School	0.087	0.282
z2	School size (log)	6.16	0.50
z3	Share of students changing school	0.098	0.052
z4	Previously treated school	0.571	0.495
z5	Teacher full-time	0.29	0.11
z6	Share of problematic students in primary school	0.174	0.123
z7	Share of special needs students in primary school	0.036	0.041
z8	Share of male students	0.51	0.23
z9	Vocational education	0.274	0.262
Assignment variable (Threshold $c_0 = 0.25$)			
c	Share of disadvantaged students	0.33	0.18
Sample size			
$\sum_{i=1}^{N} T_i$		2568	

Notes: Balanced panel for 2568 observations on 642 Flemish schools from 2010/2011 to 2013/2014.

tion using all the data in a single equation (see e.g., Lee & Lemieux, 2010)

$$Y_i = \alpha_l + \tau D_i + \beta_l (c_i - c_0) + \beta_{rl} D_i (c_i - c_0) + \epsilon_i \tag{2}$$

where D_i is defined in (1). The convenience of the approach in (2) is not only that the β_{rl} coefficient at the interaction term $D_i(c_i - c_0)$ shows the treatment effect, since $\beta_{rl} = \beta_r - \beta_l$ or the difference of the effect to the right from the threshold c_0 and that to the left from the threshold, $\beta_r - \beta_l$, but that the standard error of the treatment effect can be obtained. Note that it is not possible to obtain the standard error of the treatment effect when the two equations are estimated separately. The local linear regression in (1) posits that the effect of the running variable c is the same above threshold c_0 irrespective of whether the school is just about the threshold or substantially above it. The effect of c is assumed to be the same and measured by β_r. By the same token, the effect of c below the threshold does not depend on whether the school is just below the threshold or substantially below it and is measured by β_l. There is a compelling argument to consider the sample close to the threshold. There is probably not enough data about the counterfactual outcome y for the running variable c which is considerably away from the threshold. But even if there are data, the aim of the policy evaluation is gauging the effect of providing the additional funds. The schools that are away from the threshold to a great extent are either not counting on receiving additional funds (to the left from the threshold) or are certain to receive it (to the right from the threshold) are not comparable in the $y - c$ relationship to schools close to the threshold. The RD methods consider the sample that is within the vicinity of the threshold since due to the similarity of schools in terms of c, the difference $\beta_r - \beta_l$ represents the local treatment effect, which measures the effectiveness of the policy. Because the data that we analyze follows the data generating process (DGP) which is best described by RD, we follow the methodology that estimates the equation similar to (1) for the sample close enough to the threshold, which is determined by the bandwidth and discussed below.

To properly use the RDD technique, the check for the internal validity of the RDD design must be performed (see for all Lee & Lemieux, 2010; Wing & Bello-Gomez, 2018). Although we discuss the checks in the application section, we mention here that it is essential to ensure control over the assignment variable and the variation in long-term incentives in manipulating the running variable for the benefit of policy intervention is quasi-random. In practice, an analyst should check that a discontinuity jump is present, usually by inspecting a graph, explanatory variables are similarly distributed, and a test of continuity of the assignment variable is performed (McCrary, 2008).

In the education literature, multiple outcome variables are considered (De Witte & López-Torres, 2017). One way to investigate the policy on all student outcomes is to conduct separate education production analyses for all outcomes. In this case, the explicit assumption is made that the outcomes are generated separately, the connection between outcomes is lost, and some production laws could be violated. A more comprehensive approach is to consider a single analysis whereby the education production process is thought of as encompassing process of using inputs to obtain outcomes (see for example Ruggiero, 2006). We describe such an approach next.

3.2. A multi-input multi-output technology using stochastic frontier model for panel data

To streamline the exposition, let's consider an education production function, where 2 outputs $\boldsymbol{y} = [y_1\ y_2]$, i.e., (i) the proportion of students progressing through school, (ii) the proportion of students that obtained "A certificate", are a result of utilizing inputs $\boldsymbol{x} = [x_1\ x_2]$, i.e., total hours and budget in per capita terms (see Section 2). One way to study the production technology using multiple outputs and multiple inputs is to consider the transformation function

$$\mathcal{AF}(\boldsymbol{y}, \boldsymbol{x}) = 1 \tag{3}$$

and to proceed as follows. First, by setting $\mathcal{A} = \exp(\nu)$, where ν can be both positive and negative, the transformation function in (3) is assumed to be stochastic as both favorable and unfavorable shocks influence the performance. Second, since underperformance is allowed, which we denote by $\theta < 1$, the transformation function can be re-written as

$$\mathcal{AF}(\theta^{-1}\boldsymbol{y}, \boldsymbol{x}; \boldsymbol{\beta}) = 1, \tag{4}$$

where θ is the radial measure of technical efficiency, or output distance function (ODF), which measures the ratio of the actually observed to the maximum possible amounts of outputs. $\boldsymbol{\beta}$ is the vector of the assumed parametric form of the technology. Third, since the transformation function is homogeneous of degree 1 in outputs, (4) can be further rewritten as

$$\mathcal{AF}\left(\lambda\theta^{-1}\boldsymbol{y}, \boldsymbol{x}; \boldsymbol{\beta}\right) = \lambda, \quad \lambda > 0. \tag{5}$$

λ can be set to θy_1^{-1} in (5), where y_1 is the first output (it does not matter which output is considered for scaling). Then (5) becomes

$$\theta y_1^{-1} = f\left(\tilde{\boldsymbol{y}}_{-1}, \boldsymbol{x}; \boldsymbol{\beta}\right) \exp \nu, \tag{6}$$

where $\tilde{\boldsymbol{y}}_{-1} = y_2/y_1$.[5] Finally, taking logs of both sides of (6) and denoting $u = -\ln\theta \geq 0$, we obtain

$$-\ln y_1 = \ln f\left(\tilde{\boldsymbol{y}}_{-1}, \boldsymbol{x}; \boldsymbol{\beta}\right) + \nu + u. \tag{7}$$

which is the familiar stochastic frontier *cost* function formulation, where u is the output distance function. Equation (7) can be thought of as a parametric distance function approach (Coelli & Perelman, 1999; 2000).

3.3. Panel data model and the determinants of inefficiency

Equation (7) is generic and omits any subscripts. For panel data, we wish to make use of the state-of-the-art fourth-generation SF model discussed in the introduction, to wit,

$$-\ln y_{1,it} = \ln f(\tilde{\boldsymbol{y}}_{-1,it}, \boldsymbol{x}_{it}; \boldsymbol{\beta}) + \nu_{0i} + u_{0i} + \nu_{it} + u_{it}, \tag{8}$$

where ν_{0i} represents heterogeneity, u_{0i} and u_{it} are a time-invariant (persistent) and time-varying (transient) inefficiency, and ν_{it} is the usual noise for observation $i = 1, \ldots, N$ in time period $t = 1, \ldots, T_i$, where unbalanced panel data is explicitly allowed. Since the time-invariant and time-varying efficiencies are $\exp(-u_{0i})$ and $\exp(-u_{it})$, the overall efficiency is

$$\theta_{it}^{overall} = \theta_{0i}^{persistent} \times \theta_{it}^{transient} \tag{9}$$

and the sum of u_{0i} and u_{it} is the overall inefficiency.

Since our application focuses on the impact of funding as well as other school-specific characteristics, we argue that both the policy makers and schools are interested in knowing what factors can influence persistent and transient cost efficiency, and magnitude of their marginal effects. Although there are many ways to introduce determinants of inefficiency such as for example varying truncation point of the truncated normal distribution (e.g., Kumbhakar et al., 1991), for simplicity and especially for estimation purposes, we follow Caudill et al. (1995); Reifschneider &

[5] For more outputs each output is scaled, e.g., for 4 outputs $\tilde{\boldsymbol{y}}_{-1} = [y_2/y_1\ y_3/y_1\ y_4/y_1]$.

Stevenson (1991) and Badunenko & Kumbhakar (2017) and introduce them via the variance. That is, we specify the (pre-truncated) variance of u_{0i} which is time-invariant,

$$u_{0i} \sim N^{+}(0, \sigma^2_{u_{0i}}) \text{ where } \sigma^2_{u_{0i}} = \exp\left(\boldsymbol{z}_{u_{0i}}\boldsymbol{\gamma}_{u_0}\right), \quad i = 1, \cdots, n, \quad (10)$$

and $\boldsymbol{z}_{u_{0i}}$ is the vector of covariates that are determinants of persistent inefficiency. By definition the variables in $\boldsymbol{z}_{u_{0i}}$ are time-invariant. Since $E(u_{0i}) = \sqrt{(2/\pi)}\sigma_{u_{0i}} = \sqrt{(2/\pi)} \times \exp(\frac{1}{2}\boldsymbol{z}_{u_{0i}}\boldsymbol{\gamma}_{u_0})$, the $\boldsymbol{z}_{u_{0i}}$ variables can be viewed as determinants of persistent inefficiency. Consider a marginal effect of a variable z_{0j} on persistent inefficiency, which is the inefficiency change (IC) due to change in z_{0j} holding everything else fixed. Since the persistent cost efficiency is $\exp(-u_{0i})$, the rate of change in it due to a change in z_{0j} (labeled as IC) is given by

$$\text{IC} :\equiv \Delta \text{E}_i = -\frac{\partial u_{0i}}{\partial z_{0j,i}} \approx -\frac{\partial E(u_{0i})}{\partial z_{0j,i}} = -\sqrt{\frac{2}{\pi}}\frac{\partial \sigma_{u_{0i}}}{\partial z_{0j,i}}. \quad (11)$$

Under the assumption $\sigma^2_{u_{0i}} = \exp(\boldsymbol{z}_{u_{0i}}\boldsymbol{\gamma}_{u_0})$, Eq. (11) becomes

$$-\sqrt{\frac{2}{\pi}}\frac{1}{2}\frac{\partial\left(\boldsymbol{z}_{u_{0i}}\boldsymbol{\gamma}_{u_0}\right)}{\partial z_{0j,i}}\exp\left(\frac{1}{2}\boldsymbol{z}_{u_{0i}}\boldsymbol{\gamma}_{u_0}\right). \quad (12)$$

Variables in $\boldsymbol{z}_{u_{0i}}$ vary across firms, but are time-invariant. This means that $\sigma^2_{u_{0i}}$ is explained only by time-invariant covariates.

Similarly, we introduce determinants of time-varying inefficiency via the pre-truncated variance of u_{it}. That is, we assume

$$u_{it} \sim N^{+}(0, \sigma^2_{u_{it}}) \text{ where } \sigma^2_{u_{it}} = \sigma^2_u(\boldsymbol{z}_{u_{it}}\boldsymbol{\gamma}_u), \\ i = 1, \cdots, n, \quad t = 1, \cdots, T_i, \quad (13)$$

where $\boldsymbol{z}_{u_{it}}$ denotes the vector of covariates that explains time-varying inefficiency. u_{it} is again half-normal, thus $E(u_{it}) = \sqrt{(2/\pi)}\sigma_{u_{it}} = \sqrt{(2/\pi)}\exp(\frac{1}{2}\boldsymbol{z}_{uit}\boldsymbol{\gamma}_u)$, and therefore, determinants of $\sigma_{u_{it}}$ also determines time-varying inefficiency. We calculate the marginal effects of time-varying determinants as (11) and (12) by replacing $\boldsymbol{z}_{u_{0i}}$ and $\boldsymbol{\gamma}_{u_0}$ with $\boldsymbol{z}_{u_{it}}$ and $\boldsymbol{\gamma}_u$, respectively. Appendix C presents details of estimation as well provides references for sources of the estimation procedures.

3.4. Endogeneity and regression discontinuity design in the stochastic frontier model for panel data

Practitioners will face potentially two sources of endogeneity in combining SF and policy analysis. We consider them one by one. The first endogeneity issue arises from a production viewpoint. The ODF is dual to the revenue function (Färe & Primont, 1995), in which the inputs are explicitly assumed exogenous and outputs are endogenous to the production process. Recall that our inputs are total hours and budget in per capita terms, which are set exogenously. It can be shown that if inputs are exogenous and individuals maximize revenue, which is reasonable to assume in our case, the output ratios are exogenous (see e.g., Kumbhakar, 2011; Lien et al., 2018). What this implies for the estimation of the ODF in (8) is that the estimator is not corrupted by the first endogeneity issue.

The second source of endogeneity comes from potential self-selection and reverse causality of schools and funding in the policy evaluation as discussed in Section 3.1. To examine the causal effect of policy on the outcome variable(s), we need to identify a group of treated and not treated observations that are otherwise similar but for the policy intervention. RDD is the technique that can ensure this holds. Hence RDD is used here to determine the sample that contains observations with similar characteristics of both regressors and efficiency determinants to ensure causal interpretation.

In the light of this discussion, it remains to account for the DGP, which is marked by the regression discontinuity. The interest lies in analyzing policy effect on multiple outcomes. In the previous section we described how multiple-input multiple-output production process that allows for underperformance can be modeled. We also established that the DGP that involves RD must be analyzed using the equivalent of (2). In Eq. (2), that is in a standard RDD approach, the policy is studied under the assumption that underperformance in terms of outcome variable Y is precluded. If we allow underperformance by splitting the error term into the four components, v_{it}, random effects v_{0i}, and transient (u_{it}) and persistent (u_i) underperformances, the Eq. (2) can be rewritten in the light of the previous section as:

$$-\ln y_{1,it} = \ln f\left(\tilde{\boldsymbol{y}}_{-1,it}, \boldsymbol{x}_{it}; \boldsymbol{\beta}\right) + \tau D + \beta_l(c - c_0) \\ +(\beta_r - \beta_l)D(c - c_0) + v_{0i} + u_{0i} + v_{it} + u_{it}. \quad (14)$$

If u is *i.i.d.* in (14) or does not depend on $c - c_0$, the regression discontinuity is only in the main equation and thus the policy evaluation is in the sense about the policy impact on the outcome variable *directly* while accounting for potential underperformance. We will call this **policy evaluation Method I**.

One of the important channels how policy can influence the outcome variable is through efficiency with which the outcome variable is achieved. Thus, one can convincingly argue that the discontinuity occurs in the inefficiency term u, more specifically, $c - c_0$ is the determinant of u. More formally, the Eq. (14) can be generically written as:

$$-\ln y_{1,it} = \ln f(\tilde{\boldsymbol{y}}_{-1,it}, \boldsymbol{x}_{it}; \boldsymbol{\beta}) + v_{0i} + u_{0i} + v_{it} + u^*_{it}, \quad (15)$$

in which the regression discontinuity is built into the underachievement by making u^*_{it} a function of three variables D, $(c - c_0)$, and $D(c - c_0)$, that are a part of the $\boldsymbol{z}_{u_{it}}$ vector as introduced in (13) in previous section. Equation (15) represents what we call the **policy evaluation Method II**, whereby the only channel through which a policy affects the outcome variable is under-achievement with which the outcome is attained.

Finally, the effect of the policy can be twofold. A policy can have both a direct effect and effect through underperformance with which the outcome is attained. The combination of Method I and Method II yields what we refer to as the **policy evaluation Method III**, which can be expressed as

$$-\ln y_{1,it} = \ln f\left(\tilde{\boldsymbol{y}}_{-1,it}, \boldsymbol{x}_{it}; \boldsymbol{\beta}\right) + \tau D + \beta_l(c - c_0) \\ + (\beta_r - \beta_l)D(c - c_0) + v_{0i} + u_{0i} + v_{it} + u^*_{it}, \quad (16)$$

There are several advantages of Method III over either Method I or Method II. First, it is a comprehensive analysis since conceptually none of the effects is missed. Second, empirically the channels are allowed to have either (i) confounding, i.e., the two effects of the same direction, or (ii) confronting, that is the two effects of different directions. The effects are also allowed to cancel each other out.

The approach that we suggest ultimately combines the above three methods with a multiple-input multiple-output production process via the output distance function method, which is estimated for the sample in the vicinity of the threshold. In our empirical application, we will evaluate the policy using Methods **I**, **II**, and **III** and show comparatively the differences they entail.

3.5. Size of the treatment effect in models with determinants of inefficiency

We conclude this section by discussing how to compute the size of the treatment effect. Consider (16), where u^*_{it} is a function of a generic variable $z_{k,it}$ that belongs to the vector $\boldsymbol{z}_{u_{it}}$. The partial derivative of $\ln y_{1,it}$ with respect to z_k (we omit subscript *it* to streamline the exposition) can be obtained analogously to (12).

More specifically,

$$\frac{\partial \ln y_{1,it}}{\partial z_k} = -\sqrt{\frac{2}{\pi}}\frac{1}{2}\frac{\partial(\boldsymbol{z}_{u_{it}}\boldsymbol{\gamma}_u)}{\partial z_k}\exp\left(\frac{1}{2}\boldsymbol{z}_{u_{it}}\boldsymbol{\gamma}_u\right), \tag{17}$$

which is straightforward to compute using estimates $\hat{\boldsymbol{\gamma}}_u$. Noting that $\frac{\partial \ln y_{1,it}}{\partial z_k} = \frac{\frac{1}{y_{1,it}}\partial y_{1,it}}{\partial z_k}$, we can write

$$\frac{\partial y_{1,it}}{\partial z_k} = y_{1,it}\frac{\partial \ln y_{1,it}}{\partial z_k} = -y_{1,it}\sqrt{\frac{2}{\pi}}\frac{1}{2}\frac{\partial(\boldsymbol{z}_{u_{it}}\boldsymbol{\gamma}_u)}{\partial z_k}\exp\left(\frac{1}{2}\boldsymbol{z}_{u_{it}}\boldsymbol{\gamma}_u\right). \tag{18}$$

Due to the discussion in Section 3.2, the impact of z_k on $y_{2,it}$ is calculated analogously. If we choose the interaction term $c * D$ to be the variable z_k, Eq. (18) quantifies the treatment effect of the policy at the running variable. The treatment effect is time- and school-specific due to the highly nonlinear nature of the (18).

4. Results

In the following, we implement the procedure described in Section 3 and applied in the Flemish education context depicted in Section 2. Specifically, we evaluate the impact of additional funding on outputs of schools in which the share of disadvantaged students exceeds the 25% threshold. We exploit a regression discontinuity setting while decomposing the error term into four components, namely time-varying and time-invariant inefficiency, school latent heterogeneity, and random shocks. As we have argued earlier, applying three different models allows us to see whether the effect is direct, via efficiency, or both.

4.1. General setting

To defend the causal interpretation of the findings, we have to show that the control group is a valid counterfactual for the treated schools. To do so, we restrict the attention (and consequently the sample) around the exogenous threshold and we focus on the so-called "discontinuity sample". The discontinuity sample has been obtained by restricting the analysis around the threshold relying on the optimal bandwidth procedure developed by Calonico, Cattaneo, & Titiunik (2014), for more details, see also Calonico et al. (2019) and updated by Cattaneo, Jansson, & Ma (2020). The optimal bandwidth is computed using 2011 as a reference year for two reasons. First, the yearly determined forcing variable and the three-year based policy are by definition not synchronized, so we have to choose one reference year. Second, 2011 represents the starting year of a new cycle that covers three out of four school years under analysis. Moreover, in the discontinuity sample, we keep the observations that are observed for at least two years.

The optimal bandwidth procedure requires the specification of an output variable along with the running variable. Then, the RDD estimation takes one output at a time as dependent variable. Differently from the traditional RDD methods, the approach we propose can handle more than one output at a time in the estimation and yet the optimal bandwidths need to be computed separately. Specifically, in the present application we have two outputs. As we have pointed out in Section 3.2, it does not matter which output is used for scaling. This implies that equation

$$\theta y_1^{-1} = f(\tilde{\boldsymbol{y}}_{-1}, \boldsymbol{x}; \boldsymbol{\beta})\exp v, \tag{19}$$

where $\tilde{\boldsymbol{y}}_{-1} = y_2/y_1$ is the equivalent of the Eq. (6). Accordingly, we get two optimal bandwidths (one for each output) that can be identified as optimal lower and upper bounds. The size of the bandwidth governs the trade-off between statistical power and covariates balance. To ensure that underlying assumptions of the RD approach are credible and the covariates are balanced in the control and treated groups, we give preference to a smaller bandwidth at the expense of statistical power in our empirical application. The bandwidth that is chosen in the illustration of the proposed methods is the average of bandwidths for two outputs selected by applying the procedure that accounts for included regressors (Calonico, Cattaneo, & Farrell, 2020). More precisely, the results are discussed considering the bandwidth of 4%, which is the rounded average of bandwidths of 4.4 and 3.5% for the outputs y_1 and y_2, respectively. We refer the reader to Appendix A for a more extensive discussion of the optimal bandwidth choice and usual checks for the internal validity. Even though the analysis is performed using the 4% discontinuity sample, we also discuss the results using the 6 and 7% discontinuity samples, the results of which are presented in Appendix B. These extended analyses are performed to showcase the trade-off between statistical power and covariates balance which are not atypical in empirical work (Lee & Lemieux, 2010).

The schools between (25%-4%) and 25% form the control group, while the schools between 25% and (25%+4%) form the treated group. We identify 29 control schools and 36 treated schools in year 2011. Table 2 presents the sample means for the control and the treated group, together with the p-values obtained from the t-test to examine whether the variables are statistically different in means across the two groups, so to inspect the balance of the covariates.[6] The table provides a suitable evidence that the covariates are balanced at the cut-off. The differences in covariates' means are not statistically different in most of the cases, but with a few exceptions that mainly coincide with pre-treatment characteristics and are so reasonably more different. In this case, the RDD literature suggests to tackle the left heterogeneity by including the imbalanced variables in the estimation to provide consistent estimates (we refer to Frölich & Huber, 2019, for further discussion). More in general, also the balanced variables that are not pre-determinants of the treatment might be included in the estimation to improve the precision and accuracy of the estimation (Calonico et al., 2019; Lee & Lemieux, 2010). For this reason both imbalanced and balanced variables have been included in the efficiency estimation.

4.2. Educational production and policy evaluation

For the estimation we choose a translog specification of the $f(\tilde{\boldsymbol{y}}_{-1}, \boldsymbol{x}; \boldsymbol{\beta})$ function introduced in Eq. (8).[7] Table 3 presents results for the "discontinuity sample". The parameter of interest for the policy evaluation framework is D, together with the centered forcing variable $(c - c_0)$ and the interaction term $(c - c_0) * D$, that lets the regression function differ on both sides of the cut-off c_0 (Lee & Lemieux, 2010). Specifically, D is a dummy equal to 1 if the units are above the 25% threshold or 0 otherwise, c is the share of disadvantaged students and c_0 the cut-off at 25%. The three models are meant to disentangle the different ways in which the policy intervention might cause changes in schools' performance.

- Specification M.I, where the variables D, c, and their interaction enter the model in the technology part only, asks the question "Did the policy have a direct effect on schools' outcomes?"
- Specification M.II, in which the variables D, c, and their interaction enter the model in the inefficiency part seeks to answer the question of whether the policy had an effect

[6] For the same motivations provided above, we restrict the focus on 2011.

[7] The translog specification is widely used in the parametric distance function estimations (Coelli & Perelman, 1999; 2000).

Table 2
Sample means for control/treated group and population. The 4% discontinuity sample.

Variable	Label	Control		Treated		Combined		t-test
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	p-value
Education production function								
Inputs								
x1	Teaching hours per student	2.330	0.464	2.396	0.420	2.367	0.438	0.548
x2	Operating grants per student	942.161	94.178	966.464	122.167	955.621	110.411	0.382
Outputs								
y1	Share of students progressing through school	93.637	2.672	93.086	3.842	93.331	3.357	0.515
y2	Share of students that obtained "A certificate"	63.728	5.455	60.504	6.904	61.942	6.457	0.044
Inefficiency determinant variables								
z1	Public School	0.103	0.310	0.111	0.319	0.108	0.312	0.923
z2	School size (log)	6.257	0.346	6.153	0.483	6.200	0.427	0.333
z3	Share of students changing school	0.096	0.029	0.099	0.036	0.098	0.033	0.717
z4	Previously treated school	0.379	0.494	0.639	0.487	0.523	0.503	0.038
z5	Teacher full-time	0.322	0.088	0.303	0.098	0.311	0.093	0.438
z6	Share of problematic students in primary school	0.133	0.046	0.153	0.053	0.144	0.051	0.107
z7	Share of special needs students in primary school	0.026	0.029	0.033	0.031	0.030	0.030	0.332
z8	Share of male students	0.501	0.215	0.510	0.212	0.506	0.212	0.865
z9	Vocational education	0.173	0.172	0.229	0.180	0.204	0.178	0.211
Assignment variable (Threshold $c_0 = 0.25$)								
c	Share of disadvantaged students	0.235	0.012	0.267	0.012	0.253	0.020	<0.0001
Sample size								
N		29		36		65		

Notes: p-values obtained from a t-test to examine whether the control and the treated group variables are statistically different in means (computed for 2011, the first year of the new cycle - for the other years, similar results are obtained and available upon request).

Table 3
Education production function. Dependent variable: $-\log(y_1)$. p-values in parentheses. The 4% Discontinuity sample.

Parameter	M.I		M.II		M.III	
Education production frontier						
Intercept	−0.055	(0.093)	−0.047	(<1e-9)	−0.049	(0.271)
log(x1)	0.013	(0.389)	0.011	(0.480)	0.016	(0.311)
log(x2)	−0.016	(0.442)	−0.013	(0.483)	−0.017	(0.431)
log(y2/y1)	−0.084	(<1e-9)	−0.080	(<1e-9)	−0.081	(<1e-9)
0.5*log(x1)^2	0.138	(0.336)	0.155	(0.282)	0.144	(0.349)
0.5*log(x2)^2	0.337	(0.201)	0.267	(0.297)	0.372	(0.161)
0.5*log(y2/y1)^2	−0.265	(0.095)	−0.280	(0.072)	−0.255	(0.120)
t	−0.002	(0.869)	−4.6e-4	(0.953)	−0.002	(0.822)
t^2	1.8e-4	(0.936)	−5.5e-5	(0.971)	2.5e-4	(0.897)
log(x1)∗log(x2)	−0.296	(0.069)	−0.268	(0.094)	−0.311	(0.062)
log(x1)∗log(y2/y1)	0.170	(0.201)	0.193	(0.133)	0.164	(0.210)
log(x2)∗log(y2/y1)	0.060	(0.763)	0.035	(0.847)	0.069	(0.713)
c	0.057	(0.587)			0.042	(0.770)
D	0.047	(0.123)			0.041	(0.285)
$c * D$	−0.193	(0.124)			−0.178	(0.249)
Random effects component: $\log\sigma^2_{v_{0i}}$						
Intercept	−34.334	(0.996)	−20.941	(0.846)	−19.790	(0.874)
Persistent underperformance component: $\log\sigma^2_{u_{0i}}$						
Intercept	−6.824	(<1e-9)	−6.820	(<1e-9)	−6.824	(<1e-9)
z1: Public School	0.462	(0.310)	0.485	(0.293)	0.466	(0.307)
Random noise component: $\log\sigma^2_{v_{it}}$						
Intercept	−10.420	(<1e-9)	−10.532	(<1e-9)	−10.233	(<1e-9)
Transient underperformance component: $\log\sigma^2_{u_{it}}$						
Intercept	−2.409	(0.090)	−4.009	(0.158)	−3.918	(0.260)
z2: School size (log)	−1.000	(1e-4)	−0.927	(1e-4)	−0.935	(0.004)
z3: Share of students changing school	12.288	(<1e-9)	12.077	(<1e-9)	12.244	(0.002)
z4: Previously treated school	−0.053	(0.784)	−0.117	(0.586)	−0.206	(0.362)
z5: Teacher full-time	0.124	(0.924)	−0.061	(0.963)	0.155	(0.914)
z6: Share of problematic students in primary school	3.846	(0.074)	3.418	(0.139)	3.542	(0.210)
z7: Share of special needs students in primary school	2.744	(0.567)	2.680	(0.593)	3.116	(0.592)
z8: Share of male students	0.093	(0.881)	0.105	(0.863)	0.170	(0.799)
z9: Vocational education	−1.802	(0.056)	−1.818	(0.055)	−2.104	(0.032)
D			3.141	(0.244)	1.280	(0.738)
c			5.503	(0.582)	4.129	(0.772)
c∗D			−11.666	(0.294)	−3.914	(0.801)
Sample Characteristics						
N	98		98		98	
$\sum_{i=1}^{N} T_i$	380		380		380	
Sim. logL	910.93		909.33		912.07	

Note: A parameter with positive sign in the technology component suggests a negative effect on school outcomes. A positive sign in the underperformance component implies that the inefficiency is larger, thereby reducing school performance.

on schools' outcomes through changes in educational attainment.

- Finally, specification M.III, in which the variables D, c, and their interaction enter the model both in the technology and inefficiency parts seeks to establish if the policy had an effect and whether the effect was direct on educational attainment enhancing.

For the chosen output distance function discussed in Section 3, the dependent variable of the educational output distance function is $-\log(y_1)$ (where y_1 is the share of students progressing through school). A parameter with positive sign in the technology function $f(\tilde{\boldsymbol{y}}_{-1}, \boldsymbol{x}; \boldsymbol{\beta})$ has a negative effect on school outcomes. Overall, we observe that the forcing variable c, namely the share of disadvantaged students, has a negative effect on school outcomes and this is the reason why equal educational opportunity programs are promoted in the first place. A positive sign in the inefficiency function as in Eq. (13) implies that the inefficiency is larger thereby reducing school performance.

The channel through which policy intervention had an effect on schools' performance will be determined by the significance and sign of the interaction term $c * D$. In model M.I this coefficient is negative (-0.185) suggesting that the effect of the policy is positive and direct, i.e., not attainment enhancing. The coefficient is, however, not statistically significant at the 10% level (p-value is 0.15). The interaction term $c * D$ in model M.II is also negative (-11.633) implying positive effect of the policy through attainment-enhancing channel. As in the first model, the coefficient is not statistically significant (p-value is 0.29).

A word of caution needs to be said about the magnitudes of coefficients and marginal effects. Since the left-hand variable is logged, the coefficients in the technology part of the model can be interpreted directly as the marginal effect of a variable on the outcome variable with a minimal additional computation. More specifically, if the right-hand variable is logged, the coefficient can be interpreted as elasticity. If a right hand variable w is in level, the usual log-level interpretation applies. More specifically, the marginal effect of this variable on y_1 is approximately obtained as $\beta_w \times 100\%$, where β_w is the coefficient of w. This approximation works well if β_w is small. In the example, the change of w by 1 implies change of y_1 by $(\exp(\beta_w) - 1) \times 100\%$. To obtain the marginal effects of a variable in the inefficiency part, we need to perform even more additional calculations discussed right after Eq. (13). In model M.III, we have two interaction terms $c * D$, so we can determine through which channel policy had an effect. Even though the signs of both model M.I and M.II are preserved, the interaction terms are not statistically significant (p-values are 0.32 and 0.73, respectively). Hence for the 4% discontinuity sample we do not observe any statistically significant impact of the policy and the channel through which it works. As we mentioned previously, this may be due to loss of statistical power due to using a very narrow bandwidth of 4%.

To showcase what effect the trade-off between statistical power and covariates balance can have in practice, we perform the same regression analysis for wider bandwidths, namely 6 and 7%. The likes of the Table 2 for these bandwidths appear in Appendix A and those of Table 3 in Appendix B. The tables in Appendix A reveal that as the bandwidth becomes wider, the number of observations increases. However, the balance of covariates in the treated and control groups dwindles, as the differences in means of few more variables become statistically significant. The increased number of observations releases statistical power. The results shown in Appendix B suggest that with the bandwidth of 6% we already observe one statistically significant interaction term $c * D$ in model M.II. The even larger bandwidth of 7% enlarges the sample size, which manifests itself in statistical significance of the interaction term $c * D$ in all three models. One important finding is that the signs of the interaction terms $c * D$ in all three models are the same for all considered bandwidths. This allows us to make a suggestion that the results in Table 3 are insignificant due to a very small sample, which goes in accord with argumentation and evidence in previous empirical studies (Lee & Lemieux, 2010). Applying the LR test (whereby the statistic follows a mixed χ^2 distribution), we establish that model M.III, which allows two different channels of the policy effect, is econometrically superior to either model M.I or M.II. The combined results using the three bandwidths give us some confidence that the effect of the policy exists and it is realized through the efficiency-enhancing mechanism.

The proposed approach distinguishes from the previous ones considerably. First, it allows to simultaneously check whether the policy had an impact on the outcome variable directly (while accounting for potential underperformance) and whether the policy had an effect on schools' outcomes through changes in educational attainment. Every proposed model specification takes into account the possibility of underperformance, that is, schools might not operate at their most efficient level, while this is not the case in a standard RDD where underperformance is ignored. It follows that the conventional RDD and the proposed SF models have different assumptions, different error terms and might lead to different estimates. In addition, conventional RDD analysis consider one output at a time, while our model deals with multiple outputs, capturing variable interactions. For the specific program under assessment, a study has been conducted using a conventional RDD analysis (please see De Witte, Smet, & Van Assche, 2017) and another one using a conditional metafrontier approach (D'Inverno et al., 2021). The most comprehensive model estimation (M. III) presented in this paper synthesizes the partial evidence obtained from the previous two studies, where the policy has been found ineffective (De Witte et al., 2017), but some schools managed to expand their production possibility set (D'Inverno et al., 2021). Also, additional information is provided, as for example on the size of the treatment effect through the efficiency channel (see Section 4.3).

4.3. Educational underperformance and its determinants

Although the focus of the paper is policy evaluation, it is useful to discuss inefficiency, which signals underperformance in an educational framework. From an efficiency perspective, we distinguish time-invariant and time-varying characteristics affecting respectively the persistent and the transient inefficiency components, while tackling latent school heterogeneity and statistical noise. The distinction between persistent and transient inefficiency components is relevant for the policy implications this entails: to reduce the first one structural changes should be made, while to address the second one less invasive measures should be sufficient. Persistent inefficiency is a 'structural' inefficiency that persists over time. It is also known as a long-term inefficiency, which in our case is a 4-year period. When we say that it is more difficult to address persistent inefficiency because it is structural, we mean that it is like a 'negative' (or adverse) individual effect, which is a characteristic of this school. It may be a factor or a combination of the following factors that define the school in our 'long-term' period: the overall size of the school, geographical region, 'cultural' characteristics such as diversity of students, diversity of staff members, other goals of school such as sporting performance, greater emphasis on best performers than on average performers, a school with stem or humanities emphasis, etc. To properly study persistent inefficiency much more granular data are required. The transient inefficiency is 'easier' to address as it depends on the current school management. Among the covariates listed in the previous section, only the dummy variable equal to 1 for *public schools* can

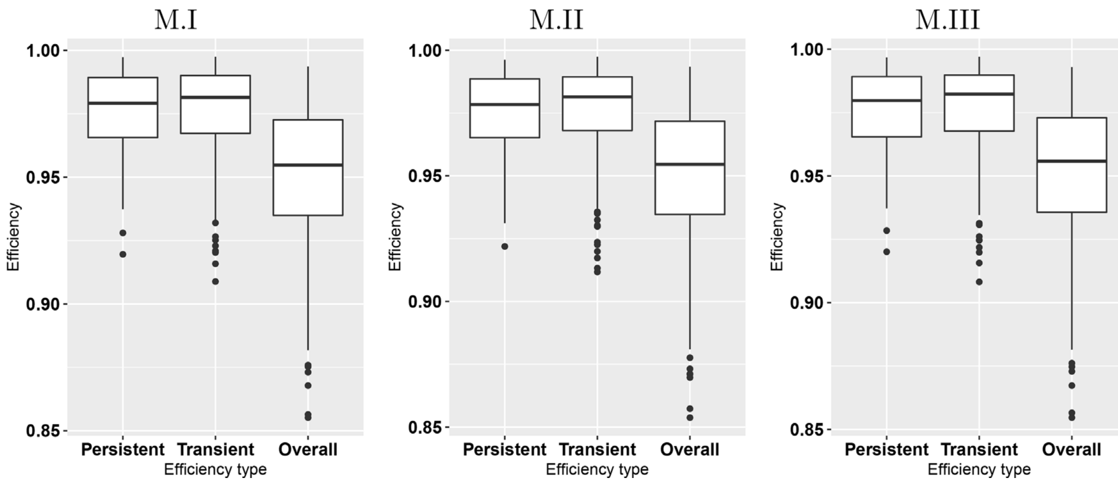

Fig. 1. Efficiencies.

be related to a longer-term feature that might require a structural change in order to affect the level of efficiency and for this reason linked to the persistent component. Figure 1 shows the distributions of transient and persistent efficiency for the discontinuity sample for all three considered models. Transient inefficiency is quite small compared to a rather sizable persistent inefficiency. Taking into account that persistent inefficiency is in every year, it is a great concern. It can be seen as a structural inefficiency that is not easy to improve.

In addition to knowing the efficiency level, stakeholders might be also interested in investigating the influence of the direction of the explanatory variables. In Table 3, the panel with the persistent and transient underperformance components shows the estimated coefficients. We briefly recall that a negative coefficient denotes reducing school underperformance, while a positive sign implies that the inefficiency is increasing with the covariate. To focus the discussion, we only comment on the statistically significant estimates for model M.III, being the most complete and econometrically superior specification, as argued above.

In Table 3, as well as in Tables B.1 and B.2, the parameter estimates outline a scenario consistent with the consensus and the main findings of the education economics literature (Bénabou, Kramarz, & Prost, 2009; Leuven, Lindahl, Oosterbeek, & Webbink, 2007; van der Klaauw, 2008). For example, the regression results show that public schools are less efficient in the long run. They persistently achieve lower educational attainment with available resources. Considering the time-varying variables, the variable that measures the share of students changing schools in the next year negatively influences school performance. These students are often more problematic and accordingly boosting the same level of educational outcome becomes more difficult. Following the same reasoning, also the share of students that experienced grade retention in primary school has a negative influence on school efficiency (D'Inverno et al., 2021). On the contrary, a larger share of students in vocational education helps improve school performance, or stated differently, focusing on a specific track helps improve learning efficiency (Traini, Kleinert, & Bittmann, 2021). With respect to the role of school size, the literature is mixed. For the sample under analysis, we observe a positive influence on school performance, pointing at the existence of scale economies in the resources and teaching activities (for an extensive review, please see Leithwood & Jantzi, 2009).

Finally, we discuss the size of the treatment effect as introduced in Section 3.5. The boxplot for treatment effects on both output variables is shown in Fig. 2. The application of the policy has a sizeable impact on the schools in need. More precisely, the first and third quartiles of the treatment effect on the percentage of students progressing through school are 3.3 to 5.7% (this percentage ranges from 80 to 100 in our sample). The treatment effect on the share of students that obtained 'A certificate' is smaller: first and third quartiles are 2.2 and 3.7%, respectively. The mean of this share is 63% and the share ranges from 5 to 80% in our sample (Table 2).

5. Conclusions and policy discussion

This paper suggested an innovative tool to assess the causal impact of a policy intervention whenever the treatment status depends on an exogenous threshold, reconciling different existing approaches in the literature. We use a state-of-the-art panel data

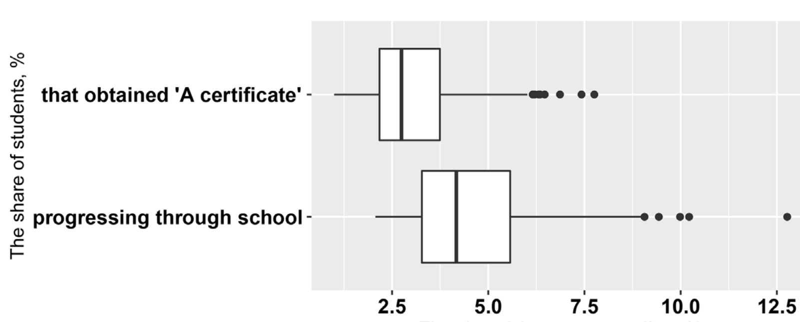

Fig. 2. The size of the treatment effect.

stochastic frontier model combined with a Regression Discontinuity Design framework to evaluate the effectiveness and efficiency of a policy intervention in a comprehensive way. First, we can check whether the policy had an effect on the outcome variables and if the effect is (i) direct, (ii) indirect, i.e. boosted through the efficiency channel, or (iii) both. Second, this approach can tackle two sources of endogeneity that might distort the estimates and prevent from a causal interpretation of the results: one from the production perspective, the other from self-selection and reverse causality in the policy intervention assignment. Third, we can disentangle the inefficiency components between persistent and transient determinants, while accounting for unobserved heterogeneity and random noise.

The method can be applied in various settings to analyze the policy, which is characterized by a known exogenously set threshold. To facilitate the application to other settings, the code will be available upon request. In this paper, the practical usefulness of the proposed approach is shown by evaluating the "Equal Educational Opportunity" program promoted in the Flemish region of Belgium. Schools with a share of disadvantaged students above 25% are the target of the policy intervention, consisting of additional resources to further support teaching. The model tackles multiple inputs and multiple outputs chosen in compliance with the institutional setting, the data availability, and the education economics literature. Even though the results are insignificant in a sample with balanced covariates, they become significant in the unbalanced sample through gained statistical power. Coherently with previous evidence, we found that the policy exerted a rather small but positive effect mostly through the efficiency-enhancing channel. From an econometric perspective, we showed that the most exhaustive model performs better than the other ones accounting only for partial effects. More importantly, the method that we propose in this paper allows us to identify the channel of the effect as well as, in contrast to existing literature, to account for schools' heterogeneity and persistent or structural underachievement. More specifically, public schools are found to be more underachieving. Although persistent underachievement does not seem large in comparison to transient underachievement, it is more difficult to eradicate. Further, the study shows that the policy helps reduce transient underachievement, thus serving as a catching-up mechanism for eliminating the differences in achievement between best-performing schools and relatively poorly performing schools.

The model suggested in this paper builds on a policy evaluation framework where the threshold is exogenous and cannot be chosen arbitrarily. Future research might extend the advocated method to settings where the threshold can be endogenized, following insights from other regression discontinuity methods (see for example the search for tipping points as proposed by Card, Mas, & Rothstein, 2008).

Acknowledgments

We owe gratitude to three anonymous referees and to participants of 'GPAC2 Workshop' 2019 at UNU-MERIT, Virtual NAPW XI 2020 in Miami, KU Leuven LEER seminar, K Hervé Dakpo, Raffaele Miniaci, Maria Laura Parisi, Chris Parmenter, and Robin Sickles for their useful comments and insights on a previous version of this paper. Giovanna D'Inverno also gratefully acknowledges financial support from Research Foundation – Flanders, FWO (Postdoctoral Fellowship 12U0219N).

Appendix A. Optimal bandwidth and internal validity checks

As introduced in the main text (see Section 4), we consider 2011 a reference year to compute the optimal bandwidth and we rely on the procedure proposed by Calonico et al. (2020), which can be viewed as an improved version of the earlier procedures suggested by Calonico et al. (2014) and Calonico et al. (2019).

The optimal bandwidth procedure requires the specification of an output variable along with the running variable. Then, the RDD estimation takes one output at a time as dependent variable. Differently from the traditional RDD methods, the approach we propose can handle more than one output at a time in the estimation and yet the optimal bandwidths need to be computed separately. Specifically, in the present application we have two outputs. As we have pointed out in Section 3.2, it does not matter which output is used for scaling. This implies that equation

$$\theta y_1^{-1} = f(\tilde{\boldsymbol{y}}_{-1}, \boldsymbol{x}; \boldsymbol{\beta}) \exp \upsilon, \tag{A.1}$$

where $\tilde{\boldsymbol{y}}_{-1} = y_2/y_1$ is the equivalent of the Eq. (6) in the paper. Accordingly, two optimal bandwidths (one for each output, see Table A.1) are computed and identify lower and upper bounds.

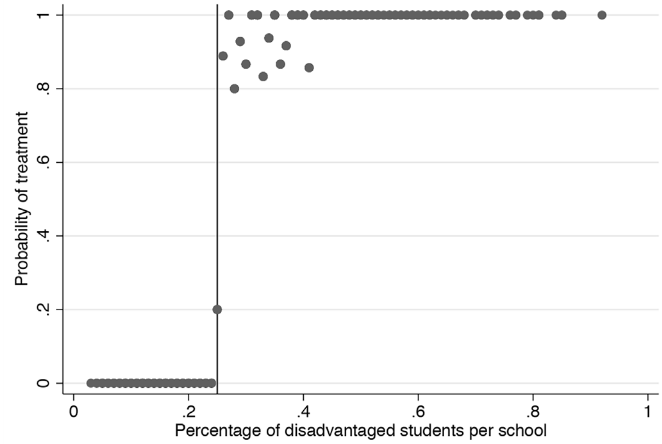

Fig. A.1. Discontinuity in the probability of treatment.

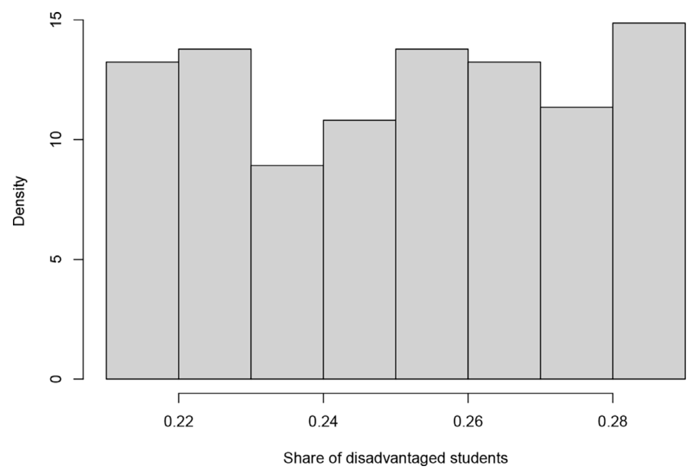

Fig. A.2. Distribution of the schools with respect to the share of disadvantaged students.

Table A.1
Optimal bandwidths. Threshold at 25% of disadvantaged students.

Outputs	Bandwidth
Share of students progressing to next year without restrictions	0.0441
Share of students without grade retention	0.0350

Note: Bandwidths computed using the 'rdrobust package' in R for year 2011 (Cattaneo et al., 2020).

Table A.2
Sample means for control/treated group and population. The 6% Discontinuity sample.

Variable	Label	Control		Treated		Combined		t-test
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	p-value
Education production function								
Inputs								
x1	Teaching hours per student	2.086	0.443	2.479	0.449	2.297	0.486	<0.0001
x2	Operating grants per student	910.458	80.932	1002.834	151.244	960.014	131.783	<0.0001
Outputs								
y1	Share of students progressing through school	94.600	2.762	93.588	3.302	94.057	3.097	0.024
y2	Share of students that obtained “A certificate”	66.033	5.592	61.289	8.054	63.488	7.394	<0.0001
Inefficiency determinant variables								
z1	Public School	0.056	0.232	0.136	0.344	0.099	0.299	0.066
z2	School size (log)	6.181	0.455	6.166	0.486	6.173	0.471	0.822
z3	Share of students changing school	0.094	0.035	0.095	0.038	0.095	0.036	0.866
z4	Previously treated school	0.191	0.395	0.728	0.447	0.479	0.501	<0.0001
z5	Teacher full-time	0.294	0.113	0.307	0.100	0.301	0.106	0.400
z6	Share of problematic students in primary school	0.091	0.060	0.158	0.067	0.127	0.072	<0.0001
z7	Share of special needs students in primary school	0.012	0.022	0.037	0.034	0.025	0.031	<0.0001
z8	Share of male students	0.461	0.148	0.539	0.248	0.503	0.211	0.009
z9	Vocational education	0.081	0.135	0.256	0.197	0.175	0.192	<0.0001
Assignment variable (Threshold $c_0 = 0.25$)								
c	Share of disadvantaged students	0.210	0.025	0.296	0.029	0.256	0.051	<0.0001
N		89		103		192		

Notes: p-values obtained from a t-test to examine whether the control and the treated group variables are statistically different in means (computed for 2011, the first year of the new cycle - for the other years, similar results are obtained and available upon request).

Table A.3
Sample means for control/treated group and population. The 7% Discontinuity sample.

Variable	Label	Control		Treated		Combined		t-test
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	p-value
Education production function								
Inputs								
x1	Teaching hours per student	2.096	0.491	2.528	0.438	2.336	0.509	<0.0001
x2	Operating grants per student	908.671	83.371	1016.591	149.452	968.574	135.346	<0.0001
Outputs								
y1	Share of students progressing through school	94.698	2.989	93.472	3.188	94.018	3.154	0.003
y2	Share of students that obtained “A certificate”	66.185	5.847	60.951	7.449	63.280	7.253	<0.0001
Inefficiency determinant variables								
z1	Public School	0.050	0.218	0.127	0.334	0.093	0.290	0.045
z2	School size (log)	6.143	0.484	6.201	0.479	6.175	0.481	0.362
z3	Share of students changing school	0.096	0.047	0.094	0.037	0.095	0.041	0.706
z4	Previously treated school	0.168	0.376	0.754	0.432	0.493	0.501	<0.0001
z5	Teacher full-time	0.282	0.121	0.310	0.098	0.298	0.109	0.050
z6	Share of problematic students in primary school	0.091	0.063	0.167	0.065	0.133	0.074	<0.0001
z7	Share of special needs students in primary school	0.011	0.020	0.039	0.033	0.027	0.031	<0.0001
z8	Share of male students	0.455	0.142	0.558	0.255	0.512	0.218	<0.0001
z9	Vocational education	0.075	0.129	0.287	0.198	0.193	0.201	<0.0001
Assignment variable (Threshold $c_0 = 0.25$)								
c	Share of disadvantaged students	0.219	0.023	0.288	0.022	0.261	0.040	<0.0001
N		101		126		227		

Notes: p-values obtained from a t-test to examine whether the control and the treated group variables are statistically different in means (computed for 2011, the first year of the new cycle - for the other years, similar results are obtained and available upon request).

The bandwidths are very close to one another. To avoid arbitrary choice, we use the average of the two bandwidths to illustrate the empirical application. We also round it to 4%, to make it intuitive and clear (Leuven et al., 2007). Additionally, due to the trade-off between statistical power and covariates balance along with the computational complexity of the parametric stochastic frontier, we expand the sub-sample by using bandwidths 6 and 7%. The RDD setting internal validity about the obtained restricted samples is checked to rule out precise control over the running variable. The RDD literature suggests mainly two ways: 1) the balance check of covariates and 2) the McCrary test, with a null hypothesis of no manipulation (Cattaneo et al., 2020; McCrary, 2008). Tables A.2 and A.3 below show the sample means, standard deviations and the t-tests for the 6% and the 7% restricted samples. We observe that although there are some differences in RHS variables for different samples, we also observe similarities and both tables offer ample evidence that the covariates are balanced. As expected, when the sample shrinks, the similarities rise, even in the dependent variable, hence we use the 4% bandwidth for illustration. In Appendix B, we present regression results using bandwidths 6 and 7%.

In the following, we report the check for the discontinuity in the probability of treatment (Fig. A.1), we graphically check the continuity of the assignment variable for the 4% discontinuity sample (Fig. A.2) and we formally test it by the McCrary manipulation test that rules out manipulation at the threshold, as displayed in Table A.4. Manipulation tests are performed for equal bandwidths 4, 6, and 7% as well not specified ones.

Table A.4
Manipulation tests for secondary education. Threshold at 25% share of disadvantaged students.

	Bandwidth		Test	
	Below	*Above*	T	*p*-value
$h_- = h_+$	0.04	0.04	−0.280	0.780
$h_- = h_+$	0.06	0.06	−0.125	0.901
$h_- = h_+$	0.07	0.07	−0.313	0.755
$h_- = h_+$	0.046	0.052	−0.269	0.788

Note: Results obtained using the 'rddensity package' in R (Cattaneo et al. 2018). The first three tests have been obtained by specifying the bandwidth at both sides of the cutoff to construct the density estimators on the two sides of the cutoff. The forth one has been obtained without specifying the bandwidth.

Appendix B. Extended analysis

Tables B.1 and B.2 below provide the estimates for wider discontinuity samples based on bandwidths 6 and 7%. Tables show that as the sample becomes less restricted but also less balanced, the effects become statistically significant. More specifically with reference to model III, the interaction term is still not statistically significant for the bandwidth 6%, but it becomes significant for the bandwidth 7%. Together with the results presented in Tables A.2 and A.3, this showcases the known trade-off between statistical power and covariates balance. We note that as the sample size shrinks (as in Table 2 in the main text), the *p*-value of the interaction terms increase, with the caveats related to a pretty small sample size mentioned by Lee & Lemieux (2010).

Appendix C. A maximum simulated likelihood estimator

Here we describe how to estimate the model in (8). To facilitate the discussion and without loss of generality, we rewrite

$$-\ln y_{1,it} = \ln f(\tilde{\boldsymbol{y}}_{-1,it}, \boldsymbol{x}_{it}; \boldsymbol{\beta}) + v_{0i} + u_{0i} + v_{it} + u_{it} \tag{21}$$

as

$$-\ln y_{1,it} = \ln f(\tilde{\boldsymbol{y}}_{-1,it}, \boldsymbol{x}_{it}; \boldsymbol{\beta}) + \epsilon_{0i} + \epsilon_{it},$$

where $\epsilon_{it} = v_{it} + u_{it}$ and $\epsilon_{0i} = v_{0i} + u_{0i}$ decompose the error term into two 'composed error' terms (both of which contain a two-sided and a one-sided error terms). Assume the most general case where all four components are heteroskedastic

$$\sigma^2_{u_{it}} = \exp(\boldsymbol{z}_{u_{it}} \boldsymbol{\varphi}_u), \quad i = 1, \cdots, n, \quad t = 1, \cdots, T_i, \tag{22}$$

Table B.1
±6% Education production function. Dependent variable: $-\log(y_1)$. *p*-values in parentheses.

Parameter	M.I		M.II		M.III	
Intercept	−0.047	(4e-4)	−0.042	(<1e-9)	−0.052	(9e-4)
log(x1)	0.004	(0.704)	0.005	(0.645)	0.006	(0.581)
log(x2)	0.001	(0.948)	0.001	(0.945)	−0.001	(0.942)
log(y2/y1)	−0.061	(<1e-9)	−0.053	(<1e-9)	−0.057	(<1e-9)
0.5*log(x1)^2	0.037	(0.681)	0.038	(0.685)	0.038	(0.675)
0.5*log(x2)^2	0.057	(0.732)	0.018	(0.914)	0.040	(0.813)
0.5*log(y2/y1)^2	−0.055	(2e-4)	−0.055	(<1e-9)	−0.057	(<1e-9)
t	−0.004	(0.357)	−0.004	(0.455)	−0.004	(0.334)
t^2	2.9e-4	(0.694)	4.1e-4	(0.704)	4.8e-4	(0.586)
log(x1)∗log(x2)	−0.100	(0.336)	−0.085	(0.434)	−0.092	(0.387)
log(x1)∗log(y2/y1)	−0.021	(0.297)	−0.021	(0.281)	−0.018	(0.360)
log(x2)∗log(y2/y1)	−0.019	(0.588)	0.013	(0.675)	−0.032	(0.327)
c	0.025	(0.624)			0.052	(0.388)
D	0.024	(0.166)			0.003	(0.890)
$c * D$	−0.092	(0.178)			−0.031	(0.691)
Random effects component: $\log\sigma^2_{v_{0i}}$						
Intercept	−9.986	(<1e-9)	−9.988	(<1e-9)	−9.928	(<1e-9)
Persistent underperformance component: $\log\sigma^2_{u_{0i}}$						
Intercept	−6.858	(<1e-9)	−6.862	(<1e-9)	−6.856	(<1e-9)
z1: Public School	1.013	(0.001)	0.989	(8e-4)	0.979	(0.001)
Random noise component: $\log\sigma^2_{v_{it}}$						
Intercept	−10.042	(<1e-9)	−10.290	(<1e-9)	−10.237	(<1e-9)
Transient underperformance component: $\log\sigma^2_{u_{it}}$						
Intercept	−4.075	(1e-4)	−4.275	(0.001)	−3.521	(0.012)
z2: School size (log)	−0.811	(<1e-9)	−0.761	(5e-4)	−0.794	(1e-4)
z3: Share of students changing school	13.282	(<1e-9)	13.985	(<1e-9)	13.749	(<1e-9)
z4: Previously treated school	−0.046	(0.757)	−0.035	(0.836)	−0.043	(0.806)
z5: Teacher full-time	0.739	(0.393)	0.529	(0.551)	0.721	(0.408)
z6: Share of problematic students in primary school	4.440	(0.013)	4.518	(0.018)	4.256	(0.023)
z7: Share of special needs students in primary school	3.234	(0.461)	2.994	(0.506)	2.271	(0.612)
z8: Share of male students	0.361	(0.423)	0.588	(0.183)	0.447	(0.313)
z9: Vocational education	−2.135	(0.003)	−2.171	(0.004)	−1.872	(0.016)
D			2.851	(0.023)	2.743	(0.059)
c			−1.098	(0.758)	−3.791	(0.369)
c∗D			−9.259	(0.059)	−7.844	(0.166)
Sample Characteristics						
N	193		193		193	
$\sum_{i=1}^{N} T_i$	772		772		772	
Sim. logL	1838.22		1842.43		1843.57	

Note: A parameter with positive sign in the technology component suggests a negative effect on school outcomes. A positive sign in the underperformance component implies that the inefficiency is larger, thereby reducing school performance.

Table B.2
±7% Education production function. Dependent variable: $-\log(y_1)$. p-values in parentheses.

Parameter	M.I		M.II		M.III	
Intercept	−0.061	(<1e-9)	−0.044	(<1e-9)	−0.059	(<1e-9)
log(x1)	0.013	(0.193)	0.018	(0.061)	0.017	(0.085)
log(x2)	−0.012	(0.389)	−0.015	(0.260)	−0.016	(0.229)
log(y2/y1)	−0.066	(<1e-9)	−0.063	(<1e-9)	−0.063	(<1e-9)
0.5*log(x1)^2	0.030	(0.681)	0.023	(0.745)	0.030	(0.672)
0.5*log(x2)^2	−0.011	(0.933)	−0.042	(0.739)	−0.039	(0.758)
0.5*log(y2/y1)^2	−0.053	(6e-4)	−0.054	(2e-4)	−0.054	(2e-4)
t	−0.002	(0.578)	−0.002	(0.499)	−0.002	(0.478)
t^2	3.8e-5	(0.952)	1.1e-4	(0.860)	1.6e-4	(0.804)
log(x1)*log(x2)	−0.067	(0.404)	−0.049	(0.516)	−0.053	(0.485)
log(x1)*log(y2/y1)	−0.007	(0.714)	−0.004	(0.831)	−0.003	(0.852)
log(x2)*log(y2/y1)	−0.011	(0.730)	0.003	(0.934)	2.8e-4	(0.993)
c	0.076	(0.103)			0.076	(0.156)
D	0.027	(0.075)			0.007	(0.665)
$c*D$	−0.110	(0.071)			−0.050	(0.473)
Random effects component: $\log\sigma^2_{v_{0i}}$						
Intercept	−10.078	(<1e-9)	−10.135	(<1e-9)	−10.184	(<1e-9)
Persistent underperformance component: $\log\sigma^2_{u_{0i}}$						
Intercept	−6.787	(<1e-9)	−6.762	(<1e-9)	−6.762	(<1e-9)
z1: Public School	0.833	(0.021)	0.808	(0.022)	0.800	(0.022)
Random noise component: $\log\sigma^2_{v_{it}}$						
Intercept	−9.973	(<1e-9)	−9.909	(<1e-9)	−9.936	(<1e-9)
Transient underperformance component: $\log\sigma^2_{u_{it}}$						
Intercept	−4.171	(<1e-9)	−5.347	(<1e-9)	−4.706	(3e-4)
z2: School size (log)	−0.771	(<1e-9)	−0.718	(<1e-9)	−0.713	(1e-4)
z3: Share of students changing school	12.101	(<1e-9)	12.530	(<1e-9)	12.494	(<1e-9)
z4: Previously treated school	−0.048	(0.738)	−0.130	(0.439)	−0.131	(0.447)
z5: Teacher full-time	1.014	(0.208)	0.856	(0.297)	0.853	(0.299)
z6: Share of problematic students in primary school	4.001	(0.012)	4.172	(0.011)	3.952	(0.016)
z7: Share of special needs students in primary school	2.743	(0.502)	1.692	(0.692)	1.880	(0.660)
z8: Share of male students	0.096	(0.813)	0.213	(0.605)	0.232	(0.574)
z9: Vocational education	−1.865	(0.005)	−1.975	(0.006)	−1.933	(0.008)
D			3.080	(0.006)	2.833	(0.031)
c			3.459	(0.315)	0.149	(0.970)
c*D			−10.663	(0.017)	−8.664	(0.096)
Sample Characteristics						
N	227		227		227	
$\sum_{i=1}^{N} T_i$	908		908		908	
Sim. logL	2167.91		2171.03		2172.39	

Note: A parameter with positive sign in the technology component suggests a negative effect on school outcomes. A positive sign in the underperformance component implies that the inefficiency is larger, thereby reducing school performance.

$$\sigma^2_{u_{0i}} = \exp(\boldsymbol{z}_{u_{0i}}\boldsymbol{\varphi}_{u0}), \quad i = 1, \cdots, n, \tag{23}$$

$$\sigma^2_{v_{it}} = \exp(\boldsymbol{z}_{v_{it}}\boldsymbol{\varphi}_{v}), \quad i = 1, \cdots, n, \quad t = 1, \cdots, T_i, \tag{24}$$

$$\sigma^2_{v_{0i}} = \exp(\boldsymbol{z}_{v_{0i}}\boldsymbol{\varphi}_{v0}), \quad i = 1, \cdots, n, \tag{25}$$

where $\boldsymbol{z}_{u_{it}}$ are the determinants of transient inefficiency, $\boldsymbol{z}_{u_{0i}}$ are the determinants of persistent inefficiency, and $\boldsymbol{z}_{v_{it}}$ and $\boldsymbol{z}_{v_{0i}}$ define the heteroskedasticity functions of the noise and random effects. The homoskedastic error component is easily derived from ((22)–(25)) by setting the vector of determinants to a constant. For example if v_{it} is homoskedastic, $\boldsymbol{z}_{v_{it}}$ is a vector of ones of length $\sum_{i=1}^{n} T_i$. Due to the skew-normal distribution properties, the conditional density of $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iT_i})$ is given by

$$f(\boldsymbol{\epsilon}_i|\epsilon_{0i}) = \prod_{t=1}^{T_i} \frac{2}{\sigma_{it}}\phi\left(\frac{\epsilon_{it}}{\sigma_{it}}\right)\Phi\left(\frac{\epsilon_{it}\lambda_{it}}{\sigma_{it}}\right),$$

where $\sigma_{it} = [\exp(\boldsymbol{z}_{u_{it}}\boldsymbol{\varphi}_u) + \exp(\boldsymbol{z}_{v_{it}}\boldsymbol{\varphi}_v)]^{1/2}$ and $\lambda_{it} = [\exp(\boldsymbol{z}_{u_{it}}\boldsymbol{\varphi}_u)/\exp(\boldsymbol{z}_{v_{it}}\boldsymbol{\varphi}_v)]^{1/2}$.

Integrate ϵ_{0i} (the distribution of which we know) out to get the unconditional density of $\boldsymbol{\epsilon}_i$

$$f(\boldsymbol{\epsilon}_i) = \int_{-\infty}^{\infty}\left[\prod_{t=1}^{T_i} \frac{2}{\sigma_{it}}\phi\left(\frac{\epsilon_{it}}{\sigma_{it}}\right)\Phi\left(\frac{\epsilon_{it}\lambda_{it}}{\sigma_{it}}\right)\right] \times \frac{2}{\sigma_{0i}}\phi\left(\frac{\epsilon_{0i}}{\sigma_{0i}}\right)\Phi\left(\frac{\epsilon_{0i}\lambda_{0i}}{\sigma_{0i}}\right)d\epsilon_{0i},$$

where $\sigma_{0i} = [\exp(\boldsymbol{z}_{u_{0i}}\boldsymbol{\varphi}_{u0}) + \exp(\boldsymbol{z}_{v_{0i}}\boldsymbol{\varphi}_{v0})]^{1/2}$ and $\lambda_{0i} = [\exp(\boldsymbol{z}_{u_{0i}}\boldsymbol{\varphi}_{u0})/\exp(\boldsymbol{z}_{v_{0i}}\boldsymbol{\varphi}_{v0})]^{1/2}$. The log-likelihood function for the ith observation of model (21) is therefore given by

$$\begin{aligned}
&\log L_i(\boldsymbol{\beta}, \boldsymbol{\varphi}_{u0}, \boldsymbol{\varphi}_{v0}, \boldsymbol{\varphi}_{u}, \boldsymbol{\varphi}_{v}) \\
&= \log\left[\int_{-\infty}^{+\infty}\left(\prod_{t=1}^{T_i}\left\{\begin{array}{l}\frac{2}{\sigma_{it}}\phi\left(\frac{r_{it}-\epsilon_{0i}}{\sigma_{it}}\right)\\ \times\Phi\left(\frac{(r_{it}-\epsilon_{0i})\lambda_{it}}{\sigma_{it}}\right)\end{array}\right\}\right)\frac{2}{\sigma_{0i}}\phi\left(\frac{\epsilon_{0i}}{\sigma_{0i}}\right)\Phi\left(\frac{\epsilon_{0i}\lambda_{0i}}{\sigma_{0i}}\right)\mathrm{d}\epsilon_{0i}\right] \\
&= \log\left[\int_{-\infty}^{+\infty}\left(\prod_{t=1}^{T_i}\left\{\frac{2}{\sigma_{it}}\phi\left(\frac{\epsilon_{it}}{\sigma_{it}}\right)\Phi\left(\frac{\epsilon_{it}\lambda_{it}}{\sigma_{it}}\right)\right\}\right)\times\frac{2}{\sigma_{0i}}\phi\left(\frac{\epsilon_{0i}}{\sigma_{0i}}\right)\Phi\left(\frac{\epsilon_{0i}\lambda_{0i}}{\sigma_{0i}}\right)\mathrm{d}\epsilon_{0i}\right],
\end{aligned} \tag{26}$$

where $\epsilon_{it} = r_{it} - (v_{0i} + u_{0i})$ and $r_{it} = -\ln y_{1,it} - \ln f(\tilde{\boldsymbol{y}}_{-1,it}, \boldsymbol{x}_{it}; \boldsymbol{\beta})$. We make use of the Monte-Carlo integration to approximate the integral in (26). For estimation purposes, we write $\epsilon_{0i} = [\exp(\boldsymbol{z}_{u_{0i}}\boldsymbol{\varphi}_{u0})]^{1/2}V_i + [\exp(\boldsymbol{z}_{v_{0i}}\boldsymbol{\varphi}_{v0})]^{1/2}|U_i|$, where both V_i and U_i are standard normal random variables. The simulated log-likelihood function for the ith panel is

$$\log L_i^S(\boldsymbol{\beta}, \boldsymbol{\varphi}_{u0}, \boldsymbol{\varphi}_{v0}, \boldsymbol{\varphi}_u, \boldsymbol{\varphi}_v)$$
$$= \log\left[\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T_i}\left\{\begin{array}{l}\frac{2}{\sigma_{it}}\phi\left(\frac{r_{it} - ([\exp(\boldsymbol{z}_{u_{0i}}\boldsymbol{\varphi}_{u0})]^{1/2}V_{ir} + [\exp(\boldsymbol{z}_{v_{0i}}\boldsymbol{\varphi}_{v0})]^{1/2}|U_{ir}|)}{\sigma_{it}}\right)\\ \times\Phi\left(\frac{[r_{it} - ([\exp(\boldsymbol{z}_{u_{0i}}\boldsymbol{\varphi}_{u0})]^{1/2}V_{ir} + [\exp(\boldsymbol{z}_{v_{0i}}\boldsymbol{\varphi}_{v0})]^{1/2}|U_{ir}|)]\lambda}{\sigma_{it}}\right)\end{array}\right\}\right)\right]$$
$$= \log\left[\frac{1}{R}\sum_{r=1}^{R}\left(\prod_{t=1}^{T_i}\left\{\frac{2}{\sigma}\phi\left(\frac{\epsilon_{itr}}{\sigma}\right)\Phi\left(\frac{\epsilon_{itr}\lambda}{\sigma}\right)\right\}\right)\right], \tag{27}$$

where V_{ir} and U_{ir} are R random deviates from the standard normal distribution, and $\epsilon_{itr} = r_{it} - ([\exp(\boldsymbol{z}_{u_{0i}}\boldsymbol{\varphi}_{u0})]^{1/2}V_{ir} + [\exp(\boldsymbol{z}_{v_{0i}}\boldsymbol{\varphi}_{v0})]^{1/2}|U_{ir}|)$. R is the number of draws for approximating the log-likelihood function. The full log-likelihood is the sum of panel-i specific log-likelihoods given in (27).

The results of Colombi et al. (2014) can be used to calculate persistent and time-varying cost efficiencies. More precisely, the conditional means of $u_{0i}, u_{i1}, \cdots, u_{iT_i}$ are given by:

$$E(\exp\{\boldsymbol{t}'\boldsymbol{u}_i\}|\boldsymbol{r}_i) = \frac{\overline{\Phi}_{T_i+1}(\boldsymbol{R}_i\boldsymbol{r}_i + \boldsymbol{\Lambda}_i\boldsymbol{t}, \boldsymbol{\Lambda}_i)}{\overline{\Phi}_{T_i+1}(\boldsymbol{R}_i\boldsymbol{r}_i, \boldsymbol{\Lambda}_i)} \times \exp\left(\boldsymbol{t}'\boldsymbol{R}_i\boldsymbol{r}_i + 0.5\boldsymbol{t}'\boldsymbol{\Lambda}_i\boldsymbol{t}\right), \tag{28}$$

where $\boldsymbol{r}_i = (r_{i1}, \ldots, r_{iT_i})'$, $\boldsymbol{A} = -[\boldsymbol{1}_{T_i}\boldsymbol{I}_{T_i}]$, $\boldsymbol{1}_{T_i}$ is the column vector of length T_i and $\boldsymbol{I}_{T_i}$ is the identity matrix of dimension T_i, the diagonal elements of $\boldsymbol{V}_i$ are $[\exp(\boldsymbol{z}_{u_{0i}}\boldsymbol{\varphi}_{u0})\ \exp(\boldsymbol{z}_{u_{it}}\boldsymbol{\varphi}_u)]$, $\boldsymbol{\Sigma}_i = \exp(\boldsymbol{z}_{v_{it}}\boldsymbol{\varphi}_v)\boldsymbol{I}_{T_i} + \exp(\boldsymbol{z}_{v_{0i}}\boldsymbol{\varphi}_{v0})\boldsymbol{1}_{T_i}\boldsymbol{1}_{T_i}'$, $\boldsymbol{\Lambda}_i = \boldsymbol{V}_i - \boldsymbol{V}_i\boldsymbol{A}'(\boldsymbol{\Sigma}_i + \boldsymbol{A}\boldsymbol{V}_i\boldsymbol{A}')^{-1}\boldsymbol{A}\boldsymbol{V}_i = (\boldsymbol{V}_i^{-1} + \boldsymbol{A}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{A})^{-1}$, $\boldsymbol{R}_i = \boldsymbol{V}_i\boldsymbol{A}'(\boldsymbol{\Sigma}_i + \boldsymbol{A}\boldsymbol{V}_i\boldsymbol{A}')^{-1} = \boldsymbol{\Lambda}_i\boldsymbol{A}'\boldsymbol{\Sigma}_i^{-1}$, $\phi_q(x, \boldsymbol{\mu}, \boldsymbol{\Omega})$ is the density function of a q-dimensional normal variable with expected value $\boldsymbol{\mu}$ and variance $\boldsymbol{\Omega}$ and $\overline{\Phi}_q(\boldsymbol{\mu}, \boldsymbol{\Omega})$ is the probability that a q-variate normal variable of expected value $\boldsymbol{\mu}$ and variance $\boldsymbol{\Omega}$ belongs to the positive orthant, $\boldsymbol{u}_i = (u_{0i}, u_{i1}, \ldots, u_{iT_i})'$, and $-\boldsymbol{t}$ is a row of the identity matrix of dimension $(T_i + 1)$. If $-\boldsymbol{t}$ is the τth row, Eq. (28) provides the conditional expected value of the τth component of the cost efficiency vector $\exp(-\boldsymbol{u}_i)$. In particular, for $\tau = 1$, we get the conditional expected value of the persistent technical efficiency. The efficiencies in (28) are of the Battese & Coelli (1988) type.

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.

Adom, P. K., & Adams, S. (2020). Decomposition of technical efficiency in agricultural production in Africa into transient and persistent technical efficiency under heterogeneous technologies. *World Development, 129*, 104907. https://doi.org/10.1016/j.worlddev.2020.104907.

Aigner, D., Lovell, C., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics, 6*(1), 21–37. https://doi.org/10.1016/0304-4076(77)90052-5.

Amjadi, G., & Lundgren, T. (2022). Is industrial energy inefficiency transient or persistent? Evidence from Swedish manufacturing. *Applied Energy, 309*, 118324. https://doi.org/10.1016/j.apenergy.2021.118324.

Amsler, C., Prokhorov, A., & Schmidt, P. (2016). Endogeneity in stochastic frontier models. *Journal of Econometrics, 190*(2), 280–288. https://doi.org/10.1016/j.jeconom.2015.06.013.

Amsler, C., Prokhorov, A., & Schmidt, P. (2017). Endogenous environmental variables in stochastic frontier models. *Journal of Econometrics, 199*(2), 131–140. https://doi.org/10.1016/j.jeconom.2017.05.005.

Badunenko, O., & Kumbhakar, S. C. (2016). When, where and how to estimate persistent and transient efficiency in stochastic frontier panel data models. *European Journal of Operational Research, 255*(1), 272–287. https://doi.org/10.1016/j.ejor.2016.04.049.

Badunenko, O., & Kumbhakar, S. C. (2017). Economies of scale, technical change and persistent and time-varying cost efficiency in Indian banking: Do ownership, regulation and heterogeneity matter? *European Journal of Operational Research, 260*(2), 789–803. https://doi.org/10.1016/j.ejor.2017.01.025.

Badunenko, O., Kumbhakar, S. C., & Lozano-Vivas, A. (2021). Achieving a sustainable cost-efficient business model in banking: The case of European commercial banks. *European Journal of Operational Research, 293*(2), 773–785. https://doi.org/10.1016/j.ejor.2020.12.039.

Badunenko, O., Mazrekaj, D., Kumbhakar, S. C., & De Witte, K. (2021). Persistent and transient inefficiency in adult education. *Empirical Economics, 60*(6), 2925–2942. https://doi.org/10.1007/s00181-020-01966-5.

Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics, 38*(3), 387–399. https://doi.org/10.1016/0304-4076(88)90053-X.

Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis, 3*(1), 153–169. https://doi.org/10.1007/BF00158774.

Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics, 20*(2), 325–332. https://doi.org/10.1007/BF01205442.

Bénabou, R., Kramarz, F., & Prost, C. (2009). The french zones d'éducation prioritaire: Much ado about nothing? *Economics of Education Review, 28*(3), 345–356. https://doi.org/10.1016/j.econedurev.2008.04.005.

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *Econometrics Journal, 23*(2), 192–210. https://doi.org/10.1093/ectj/utz022.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics, 101*(3), 442–451.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica: Journal of the Econometric Society, 82*(6), 2295–2326. https://doi.org/10.3982/ECTA11757.

Card, D., Mas, A., & Rothstein, J. (2008). Tipping and the dynamics of segregation. *The Quarterly Journal of Economics, 123*(1), 177–218.

Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association, 115*(531), 1449–1455. https://doi.org/10.1080/01621459.2019.1635480.

Caudill, S. B., Ford, J. M., & Gropper, D. M. (1995). Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. *Journal of Business & Economic Statistics, 13*(1), 105–111. https://doi.org/10.2307/1392525.

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.

Cipollone, P., & Rosolia, A. (2007). Social interactions in high school: Lessons from an earthquake. *American Economic Review, 97*(3), 948–965. https://doi.org/10.1257/aer.97.3.948.

Coelli, T., & Perelman, S. (1999). A comparison of parametric and non-parametric distance functions: With application to European railways. *European Journal of Operational Research, 117*(2), 326–339. https://doi.org/10.1016/S0377-2217(98)00271-9.

Coelli, T., & Perelman, S. (2000). Technical efficiency of European railways: A distance function approach. *Applied Economics, 32*(15), 1967–1976. https://doi.org/10.1080/000368400050155896.

Colombi, R., Kumbhakar, S. C., Martini, G., & Vittadini, G. (2014). Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *Journal of Productivity Analysis, 42*(2), 123–136. https://doi.org/10.1007/s11123-014-0386-y.

Cornwell, C., Schmidt, P., & Sickles, R. C. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics, 46*(1), 185–200. https://doi.org/10.1016/0304-4076(90)90054-W.

Crespo-Cebada, E., Pedraja-Chaparro, F., & Santín, D. (2014). Does school ownership matter? An unbiased efficiency comparison for regions of Spain. *Journal of Productivity Analysis, 41*(1), 153–172. https://doi.org/10.1007/s11123-013-0338-y.

De Witte, K., & López-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., Smet, M., & Van Assche, R. (2017). The impact of additional funds for schools with disadvantaged students. *Steunpunt Onderwijsonderzoek, Gent.*

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*, 1111–1124.

Filippini, M., Greene, W., & Masiero, G. (2018). Persistent and transient productive inefficiency in a regulated industry: Electricity distribution. *Energy Economics, 69*, 325–334. https://doi.org/10.1016/j.eneco.2017.11.016.

Filippini, M., & Greene, W. H. (2016). Persistent and transient productive inefficiency: A maximum simulated likelihood approach. *Journal of Productivity Analysis*. https://doi.org/10.1007/s11123-015-0446-y.
Frölich, M., & Huber, M. (2019). Including covariates in the regression discontinuity design. *Journal of Business & Economic Statistics, 37*(4), 736–748.
Färe, R., & Primont, D. (1995). *Multi-output production and duality: Theory and applications*. Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-0651-1.
Greene, H. W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics, 126*(2), 269–303. https://doi.org/10.1016/j.jeconom.2004.05.003.
Griffiths, W. E., & Hajargasht, G. (2016). Some models for stochastic frontiers with endogeneity. *Journal of Econometrics, 190*(2), 341–348. https://doi.org/10.1016/j.jeconom.2015.06.012.
Grosskopf, S., Hayes, K. J., & Taylor, L. L. (2014). Efficiency in education: Research and implications. *Applied Economic Perspectives and Policy, 36*(2), 175–210.
Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources, 14*(3), 351–388.
Hanushek, E. A., Piopiunik, M., & Wiederhold, S. (2019). The value of smarter teachers international evidence on teacher cognitive skills and student performance. *Journal of Human Resources, 54*(4), 857–899.
Hanushek, E. A., & Wößmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal, 116*(510), C63–C76.
Hanushek, E. A., & Wößmann, L. (2015). *Universal basic skills, what countries stand to gain*. https://doi.org/10.1787/9789264234833-en.
Johnes, G., & Tsionas, M. G. (2019). A regression discontinuity stochastic frontier model with an application to educational attainment. *Stat, 8*(1), e242.
Karakaplan, M. U., & Kutlu, L. (2017a). Endogeneity in panel stochastic frontier models: An application to the Japanese cotton spinning industry. *Applied Economics, 49*(59), 5935–5939. https://doi.org/10.1080/00036846.2017.1363861.
Karakaplan, M. U., & Kutlu, L. (2017b). Handling endogeneity in stochastic frontier analysis. *Economics Bulletin, 37*(2), 889–901. http://www.accessecon.com/Pubs/EB/2017/Volume37/EB-17-V37-I2-P79.pdf
Karakaplan, M. U., & Kutlu, L. (2017c). School district consolidation policies: Endogenous cost inefficiency and saving reversals. *Empirical Economics*, 1–40. https://doi.org/10.1007/s00181-017-1398-z.
Kneip, A., Sickles, R. C., & Song, W. (2012). A new panel data treatment for heterogeneity in time trends. *Econometric Theory, 28*(3), 590–628. https://doi.org/10.1017/S026646661100034X.
Kumbhakar, S. C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics, 46*(1), 201–211. https://doi.org/10.1016/0304-4076(90)90055-X.
Kumbhakar, S. C. (2011). Estimation of production technology when the objective is to maximize return to the outlay. *European Journal of Operational Research, 208*(2), 170–176. https://doi.org/10.1016/j.ejor.2010.09.015.
Kumbhakar, S. C., Ghosh, S., & McGuckin, J. T. (1991). A generalized production frontier approach for estimating determinants of inefficiency in U.S. dairy farms. *Journal of Business & Economic Statistics, 9*(3), 279–286. https://doi.org/10.2307/1391292.
Kumbhakar, S. C., & Heshmati, A. (1995). Efficiency measurement in Swedish dariy farms: An application of rotating panel data. *American Journal of Agricultural Economics, 77*(3), 660–674. https://doi.org/10.2307/1243233.
Kumbhakar, S. C., & Hjalmarsson, L. (1993). Technical efficiency and technical progress in swedish dairy farms. In H. O. Fried, C. A. K. Lovell, & S. S. Schmidt (Eds.), *The measurement of productive efficiency. Techniques and applications* (pp. 256–270). Oxford: Oxford University Press.
Kumbhakar, S. C., & Hjalmarsson, L. (1995). Labour-use efficiency in swedish social insurance offices. *Journal of Applied Econometrics, 10*(1), 33–47. https://doi.org/10.1002/jae.3950100104.
Kumbhakar, S. C., Lien, G., & Hardaker, J. B. (2014). Technical efficiency in competing panel data models: A study of norwegian grain farming. *Journal of Productivity Analysis, 41*(2), 321–337. https://doi.org/10.1007/s11123-012-0303-1.
Kutlu, L. (2010). Battese-Coelli estimator with endogenous regressors. *Economics Letters, 109*(2), 79–81. https://doi.org/10.1016/j.econlet.2010.08.008.
Lai, H.-P. (2013). Estimation of the threshold stochastic frontier model in the presence of an endogenous sample split variable. *Journal of Productivity Analysis, 40*(2), 227–237. https://doi.org/10.1007/s11123-012-0319-6.
Lai, H.-p., & Kumbhakar, S. C. (2018a). Endogeneity in panel data stochastic frontier model with determinants of persistent and transient inefficiency. *Economics Letters, 162*, 5–9. https://doi.org/10.1016/j.econlet.2017.10.003.
Lai, H.-P., & Kumbhakar, S. C. (2018b). Panel data stochastic frontier model with determinants of persistent and transient inefficiency. *European Journal of Operational Research, 271*(2), 746–755. https://doi.org/10.1016/j.ejor.2018.04.043.
Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature, 48*(2), 281–355.
Leithwood, K., & Jantzi, D. (2009). A review of empirical evidence about school size effects: A policy perspective. *Review of Educational Research, 79*(1), 464–490. https://doi.org/10.3102/0034654308326158.
Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2007). The effect of extra funding for disadvantaged pupils on achievement. *The Review of Economics and Statistics, 89*(4), 721–736. https://doi.org/10.1162/rest.89.4.721. November
Levin, H. M. (1974). Measuring efficiency in educational production. *Public Finance Quarterly, 2*(1), 3–24.
Lien, G., Kumbhakar, S. C., & Alem, H. (2018). Endogeneity, heterogeneity, and determinants of inefficiency in Norwegian crop-producing farms. *International Journal of Production Economics, 201*(January), 53–61. https://doi.org/10.1016/j.ijpe.2018.04.023.
McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design : A density test. *Journal of Econometrics, 142*, 698–714. https://doi.org/10.1016/j.jeconom.2007.05.005.
Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International Economic Review, 18*(2), 435–444.
Mergoni, A., & De Witte, K. (2022). Policy evaluation and efficiency: A systematic literature review. *International Transactions in Operational Research, 29*(3), 1337–1359.
Mutter, R. L., Greene, W. H., Spector, W., Rosko, M. D., & Mukamel, D. B. (2013). Investigating the impact of endogeneity on inefficiency estimates in the application of stochastic frontier analysis to nursing homes. *Journal of Productivity Analysis, 39*(2), 101–110. https://doi.org/10.1007/s11123-012-0277-z.
O'Donnell, C. J., Fallah-Fini, S., & Triantis, K. (2017). Measuring and analysing productivity change in a metafrontier framework. *Journal of Productivity Analysis, 47*(2), 117–128. https://doi.org/10.1007/s11123-017-0494-6.
OECD (2017). *Educational opportunity for all: Overcoming inequality throughout the life course*.
Park, B. U., Sickles, R. C., & Simar, L. (1998). Stochastic panel frontiers: A semiparametric approach. *Journal of Econometrics, 84*(2), 273–301. https://doi.org/10.1016/S0304-4076(97)00087-0.
Park, B. U., Sickles, R. C., & Simar, L. (2003). Semiparametric-efficient estimation of AR(1) panel data models. *Journal of Econometrics, 117*(2), 279–309. https://doi.org/10.1016/S0304-4076(03)00149-0.
Park, B. U., Sickles, R. C., & Simar, L. (2007). Semiparametric efficient estimation of dynamic panel data models. *Journal of Econometrics, 136*(1), 281–301. https://doi.org/10.1016/j.jeconom.2006.03.004.
Pedraja-Chaparro, F., Santín, D., & Simancas, R. (2016). The impact of immigrant concentration in schools on grade retention in spain: A difference-in-differences approach. *Applied Economics, 48*(21), 1978–1990.
Perelman, S., & Santin, D. (2011). Measuring educational efficiency at student level with parametric stochastic distance functions: An application to Spanish PISA results. *Education Economics, 19*(1), 29–49. https://doi.org/10.1080/09645290802470475.
Pitt, M. M., & Lee, L.-F. (1981). The measurement and sources of technical inefficiency in the Indonesian weaving industry. *Journal of Development Economics, 9*(1), 43–64. https://doi.org/10.1016/0304-3878(81)90004-3.
Reifschneider, D., & Stevenson, R. (1991). Systematic departures from the frontier: A framework for the analysis of firm inefficiency on JSTOR. *International Economic Review, 32*(3), 715–723. https://www.jstor.org/stable/2527115
Ruggiero, J. (1996). Efficiency of educational production: An analysis of New York school districts. *The Review of Economics and Statistics, 78*(3), 499–509.
Ruggiero, J. (2003). Comment on estimating school efficiency. *Economics of Education Review, 22*(6), 631–634.
Ruggiero, J. (2006). Measurement error, education production and data envelopment analysis. *Economics of Education Review, 25*(3), 327–333.
Ruggiero, J. (2019). The choice of comparable DMUs and environmental variables. In *The palgrave handbook of economic performance analysis* (pp. 123–144). Springer.
Skevas, I., & Skevas, T. (2021). A generalized true random-effects model with spatially autocorrelated persistent and transient inefficiency. *European Journal of Operational Research, 293*(3), 1131–1142. https://doi.org/10.1016/j.ejor.2021.01.004.
Traini, C., Kleinert, C., & Bittmann, F. (2021). How does exposure to a different school track influence learning progress? Explaining scissor effects by track in Germany. *Research in Social Stratification and Mobility, 76*, 100625.
Tran, K. C., & Tsionas, E. G. (2013). GMM estimation of stochastic frontier model with endogenous regressors. *Economics Letters, 118*(1), 233–236. https://doi.org/10.1016/j.econlet.2012.10.028.
Wang, M. H., & Huang, T. H. (2009). Threshold effects of financial status on the cost frontiers of financial institutions in nondynamic panels. *Applied Economics, 41*(26), 3389–3401. https://doi.org/10.1080/00036840802600079.
Van de Werfhorst, H. G. (2019). Early tracking and social inequality in educational attainment: Educational reforms in 21 european countries. *American Journal of Education, 126*(1), 65–99.
van der Klaauw, W. (2008). Breaking the link between poverty and low student achievement: An evaluation of title I. *Journal of Econometrics, 142*(2), 731–756. https://doi.org/10.1016/j.jeconom.2007.05.007.
Wing, C., & Bello-Gomez, R. A. (2018). Regression discontinuity and beyond: Options for studying external validity in an internally valid design. *American Journal of Evaluation, 39*(1), 91–108.
Yélou, C., Larue, B., & Tran, K. C. (2010). Threshold effects in panel data stochastic frontier models of dairy production in Canada. *Economic Modelling, 27*(3), 641–647. https://doi.org/10.1016/j.econmod.2010.01.003.