



Managerial efficiency and efficiency differentials in adult education: a conditional and bias-corrected efficiency analysis

Fritz Schiltz¹ · Kristof De Witte^{1,2}  · Deni Mazrekaj¹

Published online: 24 May 2019
© The Author(s) 2019

Abstract

In efficiency studies, inputs and outputs are often noisy measures, especially when education data is used. This article complements the conditional efficiency model by correcting for bias within conditional draws, using the m out of n bootstrap procedure. With a unique panel dataset, we estimate managerial efficiency, which accounts for nondiscretionary variables, and explain efficiency differentials of adult education programs in Flanders. Our results suggest that the characteristics of learners in a program matter for managerial efficiency, and that teacher characteristics are strongly correlated with efficiency differentials, as more homogeneity in the teacher workforce appears to result in higher program efficiency.

Keywords Adult education · Conditional efficiency · Bias-correction · Bootstrapping

JEL Classification L2 · I2

1 Introduction

A country's human capital, proxied by the educational attainment of the adult population, is an important driver of economic growth (Hanushek and Kimko 2000; Hanushek and Woessmann 2008). Yet, in 2016, about 21% of the adult population (25–64 year-olds) in the OECD countries had not completed upper secondary education (OECD 2017). Consequently, most OECD countries have devised policy objectives to enhance the educational attainment of the adult population. For instance, the European Education and Training 2020 framework aims to increase the 11.7% participation rate in adult education in 2016 to 15% by 2020. Adult

✉ Kristof De Witte
kristof.dewitte@kuleuven.be; k.dewitte@maastrichtuniversity.nl

Fritz Schiltz
fritz.schiltz@kuleuven.be

Deni Mazrekaj
deni.mazrekaj@kuleuven.be

¹ Leuven Economics of Education Research, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium

² UNU-MERIT, Maastricht University, Boschstraat 24, N-6211 AX Maastricht, The Netherlands

education is particularly important given the increased outsourcing, continuous developments in technology, and a trend towards job-polarization.

This article evaluates the efficiency of adult education programs. Since the seminal work by Charnes et al. (1978), applications of efficiency models have emerged in various sectors, ranging from health care (e.g. Cordero et al. 2015), water (Cherchye et al. 2014), banking (Puri et al. 2017), public sector (De Witte and Geys 2013), farming (Asmild et al. 2016) or energy (e.g. Chen et al. 2017; Hampf et al. 2015). In addition, a particularly rich efficiency literature focuses on education (see De Witte and López-Torres 2014; Johnes 2015; Thanassoulis et al. 2016 for a review). Although efficiency has been estimated in primary (Blackburn et al. 2014), secondary (Portela et al. 2012) and higher education (Johnes 2006), adult education has received no attention in the efficiency literature. To the best of our knowledge, no article in the efficiency literature has estimated the efficiency of adult education programs. Outside the efficiency literature, a number of articles have estimated the causal effects of such programs (Blanden et al. 2012; Martin et al. 2013; Oosterbeek et al. 2010). These studies generally conclude that adult education increases labour market outcomes. Yet, evaluating the efficiency of adult education programs is particularly important to ensure that budget resources are allocated in a way that maximizes outcomes.

This article contributes to the literature in at least three ways. First, from a methodological perspective, we introduce an innovative approach to estimate efficiency in the presence of heterogeneity and sampling noise. In particular, we extend the commonly used conditional model (Daraio and Simar 2007) by correcting for bias within conditional draws, using the m out of n bootstrap procedure (Simar and Wilson 2011). Inputs and outputs are often noisy measures, especially when education data is used, due to uncertainty when statistical procedures are not adhered to. For instance, inputs and outputs may be noisy due to imputation errors from using plausible values (PVs) in PISA (Jerrim et al. 2017),¹ due to inaccurate use of survey tests (von Davier et al. 2009), or due to measurement errors and outlying observations. As a result, efficiency estimates can be biased, despite the use of conditional models that account for observed heterogeneity.² As a second contribution, we distinguish between managerial efficiency and efficiency differentials. The former does not allow us to assume separability as nondiscretionary variables need to be accounted for in a conditional model. The latter can be regressed on a set of discretionary variables using a two-stage approach, as the attainable set is not affected by those variables (separability) (as in Cherchye et al. 2014). As a third contribution, we extend the efficiency of education literature to the yet unexplored field of adult education. We do so by exploiting a unique panel dataset from the Belgian region of Flanders. This dataset comprises detailed information on program costs and learners' outcomes, in addition to a rich set of both learners' and teachers' characteristics. Using this dataset, managerial efficiency estimates can be obtained while accounting for the observed characteristics of learners in a program, and efficiency differentials can be explained using discretionary teacher and program characteristics.

Our results indicate that a larger share of low educated learners has an unfavorable influence on the evaluation of adult education programs. By contrast, a larger share of men appears to have a favorable influence on the evaluation of adult education programs, while the age of learners does not appear decisive. Overall, we observe that characteristics of learners matter significantly for managerial efficiency, motivating the need for a conditional model. In

¹ The use of plausible values in PISA data results in imputation errors when constructing observations for students that did not complete specific subjects (for a technical description, see OECD (2009)).

² The suggested approach can also be applied in other fields where sampling noise is present. To facilitate further applications, the R code is available upon request.

addition, we find that a considerable amount (over 30%) of the bias in our unconditional estimates persists when a conditional model is used to obtain managerial efficiency estimates, despite the subsampling approach implemented to obtain the latter. This highlights the importance of a bias-correction in conditional models. With respect to the efficiency differentials, we find that teacher and program characteristics are able to explain efficiency differentials between adult education programs. In particular, female teachers and a more homogenous teacher composition, in terms of age and number of hours allocated to teachers, are likely to correspond with a higher program efficiency.

The remainder of the article is structured as follows. Section 2 proposes the conditional and bias-corrected efficiency model. Section 3 provides a brief introduction to the adult education system in Flanders and presents the data. Section 4 presents the results and Sect. 5 concludes.

2 Methodology

2.1 Efficiency in adult education

Data Envelopment Analysis (DEA) models have received considerable attention in the efficiency literature, considering their ability to aggregate information on different inputs and outputs. The DEA model emanates from work by Farrell (1957), later implemented by Charnes, et al. (1978, 1981). Our study further extends this non-parametric methodology and applies it on the yet relatively unexplored field of adult education. In particular, we estimate the efficiency of adult education program o using the following baseline output-oriented model:

$$\max_{v,u} \theta_o(y) = u_1 y_{1o} + \dots + u_q y_{qo} \quad (1)$$

$$s.t. v_1 x_{1o} + \dots + v_p x_{po} = 1 \quad (1a)$$

$$u_1 y_{1j} + \dots + u_q y_{qj} \leq v_1 x_{1j} + \dots + v_p x_{pj} \quad (j = 1, \dots, n) \quad (1b)$$

$$v_1, v_2, \dots, v_p \geq 0 \quad (1c)$$

$$u_1, u_2, \dots, u_q \geq 0$$

$$u_r \in \{0, 1\} (r = 1, \dots, q) \quad (1d)$$

Solving the above linear program essentially constructs a production frontier composed of efficient education programs, and benchmarks all programs relative to this frontier. The DEA model in (1) yields an efficiency score θ_o for program o , with q different outputs y , weighted by u , after normalizing the p different inputs x , weighted by v , in (1a). The weights u and v are determined endogenously to maximize this score. Condition (1b) requires the efficiency score of each education program o to attain at most 1, while (1c) implies that efficiency is increasing in the outputs used to construct it—i.e. weights are non-negative. The final condition (1d) imposes an integer formulation, i.e. the Free Disposal Hull (FDH) model of Deprins et al. (1984), rather than imposing a convexity assumption on the production frontier. The FDH model is a more general model than DEA, and performs better in the absence of convexity in the underlying data. An additional benefit of FDH estimates is that the omission

of convexity allows the use of ratio measures in our output-oriented model (Olesen et al. 2015). Nevertheless, condition (1d) can be relaxed such that the DEA model is formulated.³

The baseline FDH model, as presented above, can be improved along three lines. First, environmental variables which are not at the discretion of managers need to be included in the analysis to obtain an estimate of managerial efficiency. In the application at hand, the observed differences in learners' characteristics between programs call for a method that accounts for the environment when assessing performance, as the inflow of learners cannot be directly manipulated. Second, in the presence of noisy data, emanating from inaccurate use of survey tests (e.g. Jerrim et al. 2017; von Davier et al. 2009), invalid assessments, measurement errors or outliers in the inputs or outputs, conditional efficiency estimates should be corrected for bias. Parametric evidence suggests that ignoring similar uncertainty in the data results in biased estimates (e.g. Jerrim et al. 2017; von Davier et al. 2009). Third, variables at the discretion of managers can be used to explain differentials in conditional and bias-corrected efficiency differentials and draw policy conclusions. The next subsections provide more details on these extensions.

2.2 Conditional efficiency

One possible approach to account for nondiscretionary variables that are beyond the influence of the observation's managers (the 'environment') would be to obtain efficiency estimates using our baseline model (1) and regress these estimates on nondiscretionary variables (Z). However, this approach would impose a 'separability' condition, claiming that variables in Z are not affecting the attainable set, or production frontier (Simar and Wilson 2007). Clearly, this assumption is unlikely to hold when different types of learners sort into programs (see Sect. 3). Therefore, we evaluate adult education programs relative to a sample of programs with similar characteristics z . For each program o , we draw B subsamples of size m where the likelihood of being drawn depends on z and the kernel function estimated around z .⁴ Each subsample is then used to compute a 'conditional' efficiency score⁵:

$$\theta_o(y|z) \quad (2)$$

For example, education programs with mostly low educated and male learners are more likely to be drawn into the reference set of program o when this program has a similar inflow of learners.⁶ This contrasts to an unconditional model where the likelihood of being included in the reference set is independent of Z , and hence equal for all units. As a result, when applying the conditional model, the assumption that variables in Z are independent of the attainable set (i.e. separability) need not be imposed as programs are 'liked with likes'. In our application, we only include continuous (Epanechnikov kernel function) and ordered discrete variables (Li and Racine 2007; De Witte and Kortelainen 2013) and estimate bandwidths using least squares cross-validation (Li and Racine 2008). The issue of environmental variables in the estimation of efficiency has received considerable attention in the literature (see for example, Dyson et al. 2001; Estelle et al. 2010; Ruggiero 1996, 1998) and has been identified as a major concern in many efficiency studies in the field of, amongst others, education (De Witte and López-Torres 2017) and local governments (Narbón-Perpiñá and De Witte 2018a, b).

³ Note that our efficiency estimates are analogous when imposing convexity ($\rho = 0.979$) and equivalent findings are obtained when repeating our analyses with a BCC model.

⁴ To avoid super-efficient observations, the evaluated observation is always added to the reference set.

⁵ We set $m = 200$, in accordance with the procedure described by Daraio and Simar (2007), see Fig. 1.

⁶ This approach resembles to a traditional matching analysis (see De Witte and López-Torres 2014).

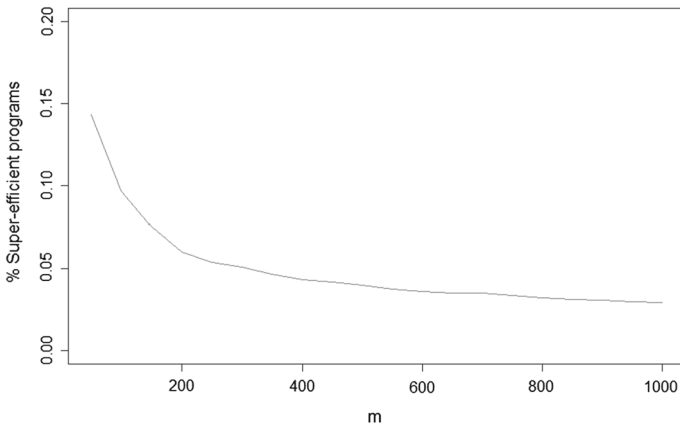


Fig. 1 Choosing m . Depending on the choice of m , the proportion of super-efficient observations varies since the size of the drawn sample (m) with respect to the total sample size n influences the probability of observation o not belonging to the production frontier. The value of m is set in a way to attain a sufficiently small decrease in the proportion of super-efficient observations (here, $m = 200$)

2.3 Conditional and bias-corrected efficiency

We correct for possible remaining sampling noise in the conditional sample. A naive bootstrap is invalid for FDH (Model 1) estimators as the sample maximum is drawn too frequently (Simar and Wilson 2011). Therefore, we follow Simar and Wilson (2011) by implementing the m out of n bootstrap procedure. As we are drawing B subsamples (of size s) from within the conditional draws (of size m), without replacement, we essentially implement an s out of m approach.⁷ Denote the bias-corrected efficiency score as follows:

$$\hat{\theta}_o(y|Z) = \theta_o(y|Z) - bias_o[\theta_o(y|Z)]$$

$$\text{with } bias_o[\theta_o(y|Z)] = \left(\frac{m}{n}\right)^{\frac{2}{(p+q+1)}} \frac{1}{B} \sum_{b=1}^B \theta_{bo}(y|Z) - \theta_o(y|Z) \tag{3}$$

Correcting for bias can prove useful in applications with noisy data in general and education data in particular. In education, inputs and outputs are often noisy measures (e.g. Jerrim et al. 2017). Bias-correction is especially important when using education data, as noise resulting from measurement error is widespread (see for example De Witte and Schiltz (2018) who illustrate issues when using survey data). Moreover, often the design of data collection methods leads to substantial bias in education data. For example, the use of plausible values in PISA data results in imputation errors when constructing observations for students that did not complete specific subjects (for a technical description, see OECD (2009)).

Despite mitigating the influence of outlying observations by the B draws of size m in the conditional efficiency estimator, the noisy input and output variables can result in biased estimates. The proposed method, combining conditional and bias-corrected efficiency, can be seen as an alternative to partial frontier approaches such as the robust order- m approach (Cazals et al. 2002; Daraio and Simar 2007) or order- α approach (Aragon et al. 2005). Although the conditional estimates account for the operational environment an observation

⁷ We set the subsample size $s = m^k$ equal to 118, which corresponds to $k = 0.9$. In unconditional bias-corrected estimates, the subsample size is equal to 644 ($= n^k$).

is operating in, subsampling *within* conditional draws allows us to correct for remaining bias in the input and output variables of the conditional estimates. As we will illustrate in Sect. 4, bias-corrected estimates of the conditional model can still differ considerably from their conditional counterparts, despite an m out of n procedure being used to obtain the latter. In the application at hand, over 30% of the initially estimated bias remains *after* conditionally subsampling. Also, we compared the variation in bootstrapped conditional efficiency estimates to the bias estimates and conclude that bias-correction is required for almost all programs (see Sect. 4.2.). As an additional advantage, confidence intervals can be constructed from the distribution of bootstrap estimates, which facilitates the comparison and ranking of evaluated units (e.g. Johnes 2006).

2.4 Statistical inference for managerial efficiency and efficiency differentials

Once conditional and unconditional bias-corrected efficiency estimates are obtained for every program o , we can compute the ratio of both:

$$Q_o = \frac{\hat{\theta}_o(y|z)}{\hat{\theta}_o(y)} \quad (4)$$

Using this ratio, it is possible to gain insights into which nondiscretionary characteristics are favorable and unfavorable when measuring *managerial efficiency* (Daraio and Simar 2005). For example, when Q is increasing with the number of learners in program, it suggests that having more learners in a program is favorable. Nonparametric regression plots uncover such (nonlinear) relationships and corresponding significance levels can be computed (Li and Racine 2007).

In order to explain *efficiency differentials*, we regress the conditional and bias-corrected efficiency estimates on a set of explanatory variables. We do so in a nonparametric manner as a linear, a parabolic, or even a Fourier relationship (Schiltz and De Witte 2017) might be unable to capture the true nature of patterns in data. In fact, our flexible specification avoids imposing any functional form assumptions on the relationship between efficiency differentials and explanatory variables. As noted in Sect. 3, this set of variables consists of teacher and program characteristics that are discretionary, and hence subject to optimization by program providers. In contrast to the composition of learners choosing to enroll in a program, which cannot be directly manipulated, these variables are not included in the set of conditional variables. Again, nonparametric regression plots and significance levels can be obtained (Li and Racine 2007).

2.5 Summary of the methodology

Summarizing our approach, we proceed in four steps to measure and explain managerial efficiency. First, a subsample of size m is drawn based on a set of nondiscretionary characteristics (Z). Second, using this sample, the efficiency score is corrected for bias in inputs and outputs using bootstrapped (size s) estimates. This procedure is repeated many times to mitigate the influence of a single conditional draw.⁸ Third, after averaging the results for each procedure, we compute the ratio of unconditional over conditional estimates to gain insights into which characteristics are favorable and unfavorable when measuring managerial efficiency (Daraio and Simar 2005), and compute significance levels (Li and Racine 2007). Finally, we explain

⁸ The number of repetitions is set at 200, analogous to the number of conditional bootstrap draws of size m .

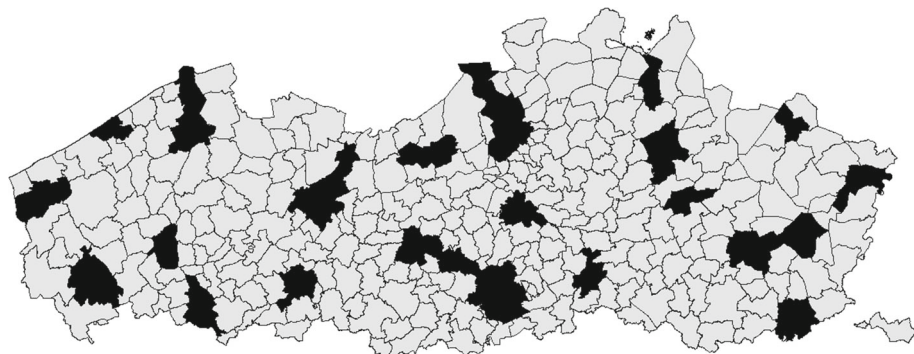


Fig. 2 Coverage of adult education programs (dark) in Flanders

efficiency differentials in conditional and bias-corrected efficiency estimates using a set of discretionary variables in a nonparametric regression.

3 Data and background

Adult education programs in Flanders aim to provide lifelong learning opportunities to adults in six sectors (i.e. technology, management, environment, food, design, and metal and wood), and are designed to foster entrepreneurship within these sectors. Following the definition of Eurydice (Eurydice 2018) which distinguishes between adult basic education, secondary adult education, and adult higher vocational schools, we focus on programs offered in adult higher vocational schools in this study. All adult education programs are privately organized and publicly funded. Programs are subsidized in an output-oriented manner by the Flemish government. Centers of adult education receive funding per learner obtaining his or her certificate, depending on the length of the program and whether or not it is an officially recognized program. In contrast to the financing of compulsory education, funding is not earmarked such that resources can be allocated freely by the centers. No prerequisites are required for enrolment, although multiple programs can be organized in an adjacent manner, increasing in the difficulty level. Completion of previous programs will then be needed to proceed to the next program. The length of programs can vary considerably, but is at the discretion of centers providing adult education. Adult education centers (with different physical locations per center) are available throughout Flanders, as displayed in Fig. 2.

Our dataset covers the period 2006–2015 and includes detailed information on learners, teachers, and programs. We analyze efficiency at the program level to explain the differences in the ability of programs to transform inputs into outputs. All models estimated here are output-oriented, as our focus is on the maximization of learners' outcomes for a given set of inputs. Accordingly, our aim is not to estimate a cost function (e.g. Costrell et al. 2008; Duncombe et al. 1995; Imazeki 2008) but rather to provide a framework which allows policy makers to benchmark adult education programs.⁹

The first panel of Table 1 reports summary statistics for the input and the outputs. As an input, we consider the cost per learner for each session. This input is measured in euros and

⁹ Although estimating a cost model would be an interesting research avenue, our model is rooted in the production literature. Naturally, cost and production models are inherently connected as weights in a DEA model can be viewed as shadow prices. Therefore, more generally, DEA models can be interesting for benchmarking decision making units when no price information is available (Cherchye et al. 2007).

Table 1 Summary statistics

Variable (N = 1220)	Mean	SD	Median	Min	Max
<i>Input</i>					
Cost per learner per session (in euro)	9.097	8.365	7.648	0.635	110.468
<i>Outputs</i>					
Average exam score	0.664	0.091	0.673	0.242	1
Total presence in class	0.761	0.099	0.777	0.405	1
Participation rate in exam	0.544	0.213	0.560	0.010	1
<i>Learners</i>					
Gender balance	0.491	0.340	0.480	0	1
Age	30.822	4.634	30.345	19.875	47.500
% low educated	0.101	0.089	0.088	0	0.667
<i>Program</i>					
Sessions per program	110.033	63.719	94	4	422
Learners per program	96.290	223.054	43	2	1581
<i>Teachers</i>					
Gender balance	0.734	0.326	0.870	0	1
Age	46.448	6.124	46.647	29.365	72.795
Age composition (sd)	8.816	2.959	8.958	0	20.506
Hours composition (sd)	59.002	33.128	52.292	0	225.931

computed by dividing the total cost of a program by the number of learners in a program and the number of sessions in a program. The total cost is obtained by multiplying the average hourly wage of teachers assigned to the program and the total number of hours taught. Hence, the cost per learner for each session is affected by the number of sessions, the number of learners, the number of hours taught, and the wage of the teachers hired for the program. Note that only the latter cost component cannot be altered by providers of entrepreneurial programs, as the teachers' wage is set by the central government. Summary statistics on the number of sessions per program and the number of learners per program are provided in Table 1.

The selection of the outputs follows the insights obtained by De Leebeeck et al. (2018), who analyse adult education by a qualitative research design. They report the outcomes of 5 focus groups and interviews with 13 program directors of adult education programs, and several interviews of representatives of the national association of adult education (SYN-TRA). De Leebeeck et al. (2018) indicate that adult education programs aim to maximize the participation of learning both in classes and in evaluations in order to make all learners succeed. Therefore, subject to data limitations, we use three outputs: the reported presence during classes (reported by teachers and very accurate as it is used for funding purposes), the share of learners participating in exams and the average score on the final exam. It should be noted that these outputs are also in accordance with the criteria for evaluation put forward by the Flemish government. All outputs are measured as ratios, ranging from 0 to 1.

The second panel of Table 1 presents summary statistics for learners enrolled in an adult education program between 2006 and 2015. Our dataset includes information on the share of boys relative to girls ('gender balance'), the average age of learners, the share of low educated (at most high school) learners, the number of sessions per program, and the size of the program—measured as the number of learners. Apart from the two latter variables, learner characteristics cannot be directly manipulated by providers of adult education programs.

Therefore, in order to benchmark programs in a more equitable manner, we compare programs that started with a similar composition of learners using a conditional efficiency model. It can be seen intuitively that non-discretionary variables (e.g. prior education of learners choosing a program) should be taken into account when benchmarking programs, while discretionary variables (e.g. the size of the groups created per program) should not. Along the same lines, the panel data structure requires programs to be evaluated relative to programs instructed in the same reference period to account for time-varying efficiency.

In addition to characteristics of learners following adult education programs, our unique dataset allows the inclusion of the gender balance of teachers, their average age, but also the variation in age between teachers and the way teaching hours are allocated between teachers. In the third panel, we present this set of variables related to programs' teacher workforce and the corresponding summary statistics. It is debatable whether teacher characteristics can be altered at will. Evaluating adult education programs in Flanders, two arguments can be raised to not include them as conditions in our model. First, the employment restrictions in adult education are less strict than in the compulsory education system. Teachers in adult education can be hired and fired more easily compared to those employed in the compulsory education system, making them more prone to managerial forces—and hence can be considered discretionary. Second, using our dataset, it is unfeasible to include all learners' and teachers' characteristics as conditions, as the curse of dimensionality would render all programs efficient. Therefore, once managerial efficiency estimates are obtained for all programs accounting for the variation in the inflow of learners, the teacher variables will be used in a second stage to explain efficiency differentials across programs. Our final sample includes 1200 observations, which corresponds to 120 programs observed over a 10 years' time span.

4 Results

4.1 Measuring efficiency

Table 2 presents the efficiency scores of adult education programs under different model assumptions, introduced in the previous section. We start from our baseline FDH model without bias-correction (Model 1) and gradually extend it to measure managerial efficiency with bias-correction. On average, the FDH unconditional efficiency score amounts to 0.87, implying that programs could improve their performance by 13% if they were as efficient as those programs in their reference set. This rate of improvement amounts to over 50% for the worst performing programs. We enrich this unconditional FDH model to account for sampling noise in the data by adding bias-correction (Model 2), using the s out of m bootstrap, and identify an upward bias of around 1 percentage points.

However, as performance is expressed in terms of learners' presence, participation in exams, and exam results, the inflow of learners in a program will be decisive for its evaluation. Failing to account for differences in inflow will complicate the comparison of programs. For example, two programs with the same budget (input) but with a different composition in learners might perform differently, which does not necessarily lead to the conclusion that one program is more efficient than the other. It might be that the program deemed efficient is attracting more motivated or mature learners who always attend classes and study well for their exams. In order to account for programs' environment, not at the discretion of providers of adult education, we employ conditional models to obtain managerial efficiency scores. Efficiency differences are then due to managerial decisions as similar programs are used to benchmark each other.

Table 2 Efficiency estimates

Model		Mean	SD	Median	Min	Max
1	Unconditional	0.869	0.084	0.875	0.500	1
2	Unconditional and bias-corrected	0.860	0.085	0.866	0.499	1
	Lower bound	0.805	0.088	0.810	0.432	1
	Upper bound	0.870	0.084	0.875	0.500	1
3	Conditional	0.963	0.058	0.991	0.568	1
4	Conditional and bias-corrected	0.961	0.059	0.990	0.567	1
	Lower bound	0.949	0.074	0.984	0.560	1
	Upper bound	0.963	0.058	0.991	0.568	1
5	Unconditional BC/conditional BC	0.894	0.062	0.898	0.619	1

Note Conditional models are obtained by including the following set of variables as conditions: the year of assessment, the gender balance, the average age, and the share of low educated learners. Bounds are constructed at the 95% level

4.2 Measuring managerial efficiency

Model 3 in Table 2 presents the results from the conditional model, accounting for the age, gender, and education background of learners in a program. In addition, we include the year of assessment as an ordered conditional variable. On average, conditional efficiency estimates are higher and less dispersed. This follows directly from programs being benchmarked to more similar programs in smaller subsamples. Nonetheless, efficiency estimates still vary considerably, ranging from full efficiency (efficiency score equal to 1) to a projected improvement of 43% for the worst performing programs. In Model 4, we further extend the conditional model to its bias-corrected form by accounting for sampling noise. Bias-corrected estimates again suggest an upward bias of conditional estimates, around 0.2 percentage points. Comparing the ratio of bias in the unconditional model to bias in the conditional model reveals that over 30% of the initial bias persists in the conditional model, despite the subsampling approach implemented to obtain the conditional estimates.¹⁰ Accordingly, 95% confidence intervals of conditional estimates are markedly narrower on average. The final row of Table 2 presents the ratio of unconditional over conditional bias-corrected efficiency estimates. Figure 3 presents the plots of nonparametric regressions, relating this ratio to the set of conditional variables.¹¹ By considering the slope graphically, an interpretation can be inferred for the marginal effect of a given variable on the attainable set. The share of low educated learners appears to have an unfavorable influence on the evaluation of adult education programs, although its influence only appears once a certain threshold is reached (around 40% low educated learners in a program). This negative correlation implies that the frontier used to evaluate programs with many low educated learners is positioned below the frontier used in the unconditional model. The timing of data collection and the age of learners does not appear to be of major

¹⁰ For each adult education program, we calculated the ratio $\frac{\frac{1}{3}(\text{bias}_o[\theta_o(y|Z)])}{\hat{\sigma}^2}$, based on Simar and Wilson (2000, p. 790) and conclude that it is well above unity for over 95% of conditional efficiency estimates. This suggests that (estimated) bias is substantially larger than the variation between bootstrapped estimates, providing additional justification for the bias-correction in this particular application.

¹¹ We obtained p -values using nonparametric bootstraps (Li and Racine 2007), and conclude that all conditional variables are significantly related to the ratio of unconditional over conditional estimates ($p < 0.05$), except the share of low educated learners in a program ($p = 0.234$).

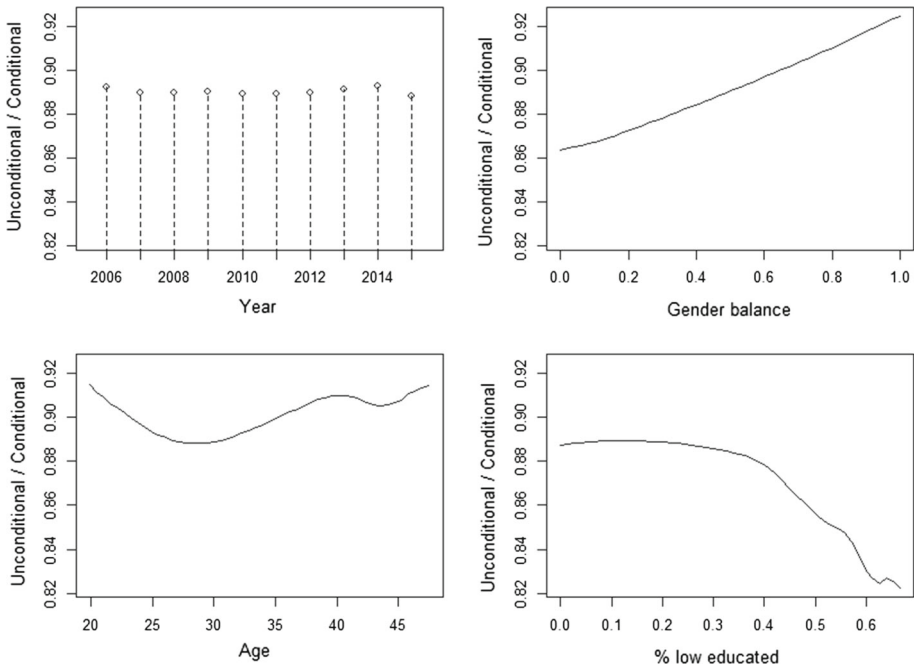


Fig. 3 Conditional efficiency. Plots of nonparametric regressions are presented linking the ratio of unconditional to conditional estimates and the conditional variables included in the analysis. A negative slope suggests an unfavorable influence on efficiency, while a positive slope suggests a favorable influence on efficiency. We obtained p values using nonparametric bootstraps (Li and Racine 2007), and conclude that all conditional variables are significantly related to the ratio of unconditional over conditional estimates ($p < 0.05$), except the share of low educated learners in a program ($p = 0.234$)

importance, whereas the gender balance of learners appears important for the assessment of programs. Overall, as the share of men increases, the evaluation of programs appears to be more favorable.

4.3 Explaining efficiency differentials

Once we obtained efficiency estimates that measure the managerial efficiency—i.e. accounting for non-discretionary variables—of adult education programs, we are interested in which discretionary variables are able to explain efficiency differentials. In particular, we focus on teacher and program characteristics (see Sect. 2). Figure 4 presents the plots of nonparametric regressions, relating the conditional and bias-corrected efficiency estimates to the size of the program (in number of learners and number of sessions), and the composition of the teacher workforce (age, gender and hours).¹² To reduce the influence of outliers, the number of learners per program and the number of sessions per program are expressed in logs.

Our results suggest that efficiency is increasing in the number of learners in a program and decreasing in the number of sessions per program, although the longest programs appear to be efficient again. In terms of demographics, it can be seen that, in contrast to learners'

¹² Again, we obtained p -values using nonparametric bootstraps (Li and Racine 2007), and conclude that all program and teacher characteristics are significantly related to the managerial efficiency estimates ($p < 0.05$), except the way teaching hours are allocated between teachers ($p = 0.132$).

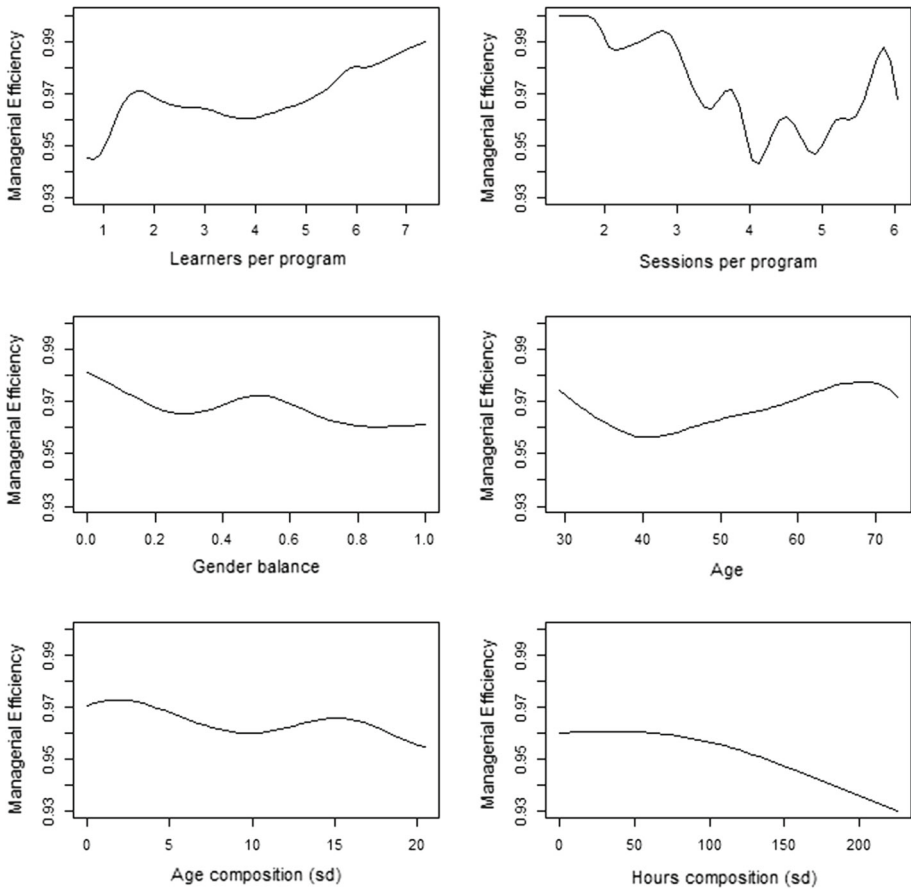


Fig. 4 Explaining efficiency differentials. Plots of nonparametric regressions are presented, relating the conditional and bias-corrected efficiency estimates to teacher and program characteristics. A negative slope suggests an unfavorable influence on efficiency, while a positive slope suggests a favorable influence on efficiency. We obtained p values using nonparametric bootstraps (Li and Racine 2007), and conclude that all program and teacher characteristics are significantly related to the managerial efficiency estimates ($p < 0.05$), except the way teaching hours are allocated between teachers ($p = 0.132$)

gender, the most efficient programs are characterized by mostly female teachers. With respect to age, a U-shaped curve is found, where the least efficient programs are taught by teachers averaging age 40. Finally, the composition of teachers seems to matter for efficiency. In general, more homogeneous groups (little variation in age, and number of hours allocated to teachers) result in more efficient adult education programs.

In order to test the robustness of our results, we repeated second stage regressions for several alternative specifications of the conditional efficiency estimates. First, we added an additional specification by also including the share of non-Belgians as a conditional variable. When non-natives sort into specific programs, this is expected to affect efficiency scores—if the share of non-natives affects the production function. Second, we take care of heterogeneity among programs by specifying an additional model with sector fixed effects instead of year fixed effects. This allows us to take out any sector-level variations, essentially comparing programs within the same sector, and hence similar learners and teachers. Third, we include center fixed effects. This specification captures all variation at

the center level (taking out urban–rural variation, for example), while exploiting variation across programs within centers. For all specifications, the explained variation (R^2) is rather high (between 0.417 and 0.471, and 0.454 in the original model), suggesting the set of six explanatory variables is able to explain the efficiency differentials relatively well. Empirically, nonparametric plots are equivalent to Fig. 4 (see Figs. 5, 6, 7 in the “Appendix”) when running the second stage regressions using these alternative efficiency estimation procedures. These findings imply that the set of conditional variables in the original model captures the heterogeneity in programs rather well.

5 Discussion and conclusion

This article introduced an innovative approach to estimate managerial efficiency in the presence of sampling noise. Using the m out of n bootstrap procedure (Simar and Wilson 2011), we extended the conditional efficiency model (Daraio and Simar 2007) by correcting for bias within conditional draws. Inputs and outputs are often noisy measures, which is especially the case when education data is used due to uncertainty emanating from survey sample designs, measurement errors or outlying observations. As a result, efficiency estimates can be biased, also when conditional models are used. We further distinguished between managerial efficiency and efficiency differentials. The former does not allow separability such that nondiscretionary variables need to be accounted for in a conditional model. The latter can be regressed on a set of discretionary variables using a two-stage approach to provide an insight into factors that explain why some units are efficient and others are not.

Using a rich panel dataset from Flanders, we illustrate the usefulness of this framework to measure and explain managerial efficiency of adult education programs. Our results suggest that a higher share of low educated learners has an unfavorable influence on the evaluation of adult education programs. By contrast, a higher share of men appears to have a favorable influence on the evaluation of adult education programs. Furthermore, we find that teacher characteristics influence the efficiency of adult education programs. In particular, more female teachers and a more homogenous teacher composition in terms of age and number of hours allocated to teachers are likely to result in a higher program efficiency. Finally, the estimated bias in our conditional model amounts to 30% of the initial bias estimate in the unconditional model, despite the subsampling approach used when estimating the conditional model.

Nonetheless, several limitations of our approach are worth noting. First, we do not claim to present causal evidence, but we contribute by using a novel empirical framework to benchmark adult education programs. Future research might expand the range of environmental variables or exploit exogenous shocks that may address causality. For example, including socio-economic characteristics (e.g. ethnicity or income) that influence both the costs of adult education (inputs) and the in-class performance (outputs) might prove an interesting robustness check beyond the ones presented here—considering our data limitations.

Second, although the education system in Flanders is similar to the education system in many OECD countries, results cannot be simply extrapolated to other countries. Nonetheless, the framework proposed here can serve as a starting point for further research. As the main use of conditional models is to account for the exogenous environment, the methodology can be applied to other educational systems as well. Finally, a possible avenue to explore is to extend the proposed model to the linearized version of FDH (‘LFDH’), introduced by Jeong and Simar (2006). This extension provides a means to overcome limitations of the subsampling approach when constructing confidence intervals for the FDH estimator in

finite sample applications, as suggested by Jeong and Simar (2006). Further research should address these topics.

Acknowledgements The authors are grateful to participants of DEA40 in Birmingham for many useful suggestions, in particular Léopold Simar and Paul Wilson, and two anonymous referees. The authors acknowledge financial support from SYNTRA Flanders. Corresponding author: Fritz Schiltz.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

See Figs. 5, 6 and 7.

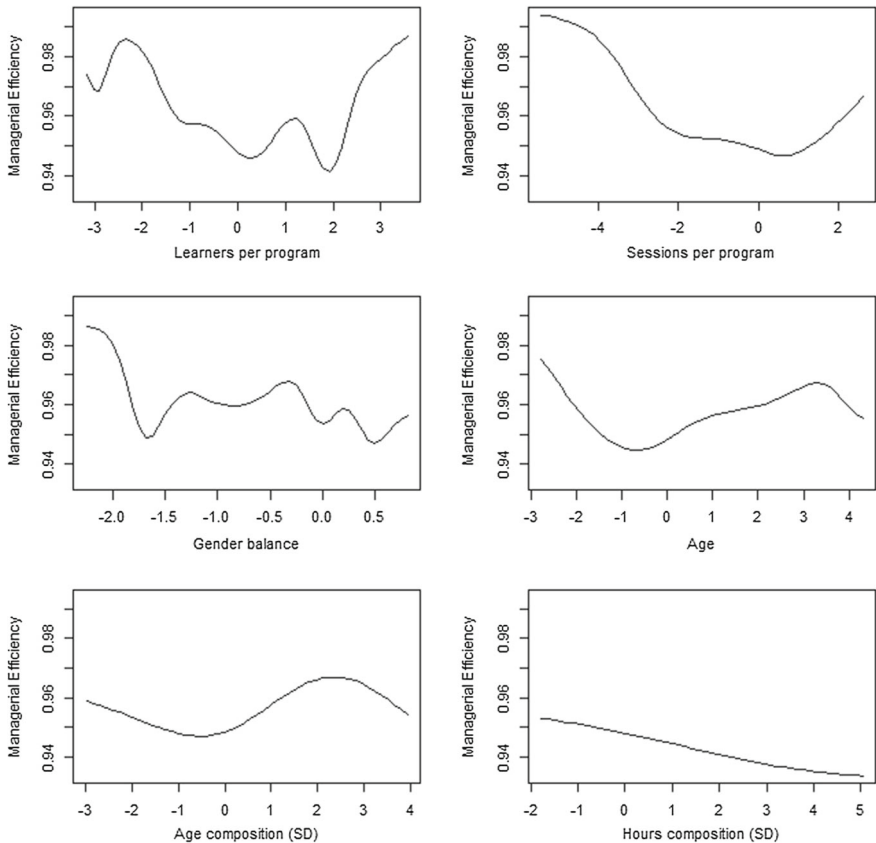


Fig. 5 Explaining efficiency differentials (including % non-Belgians as additional conditional variable. Plots of nonparametric regressions are presented, relating the conditional and bias-corrected efficiency estimates to teacher and program characteristics. A negative slope suggests an unfavorable influence on efficiency, while a positive slope suggests a favorable influence on efficiency. We obtained p values using nonparametric bootstraps (Li and Racine 2007), and conclude that all program and teacher characteristics are significantly related to the managerial efficiency estimates ($p < 0.1$), except the way teaching hours are allocated between teachers ($p = 0.547$)

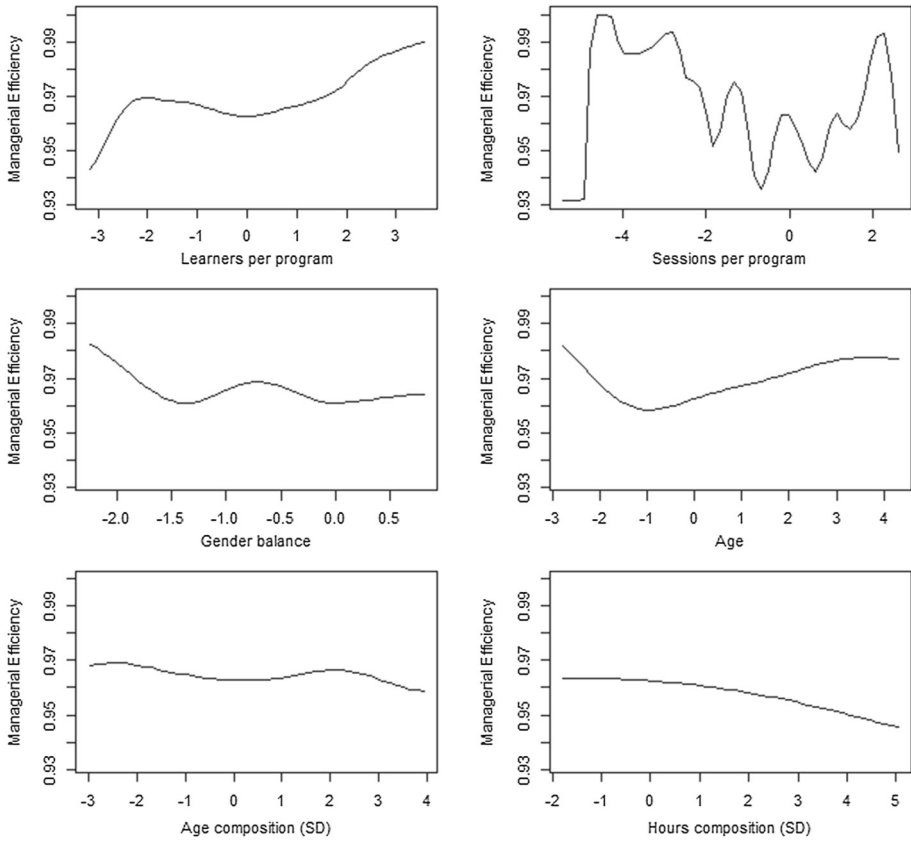


Fig. 6 Explaining efficiency differentials (including sector fixed effect as additional conditional variable. Plots of nonparametric regressions are presented, relating the conditional and bias-corrected efficiency estimates to teacher and program characteristics. A negative slope suggests an unfavorable influence on efficiency, while a positive slope suggests a favorable influence on efficiency. We obtained p values using nonparametric bootstraps (Li and Racine 2007), and conclude that all program and teacher characteristics are significantly related to the managerial efficiency estimates ($p < 0.1$), except the way teaching hours are allocated between teachers ($p = 0.394$)

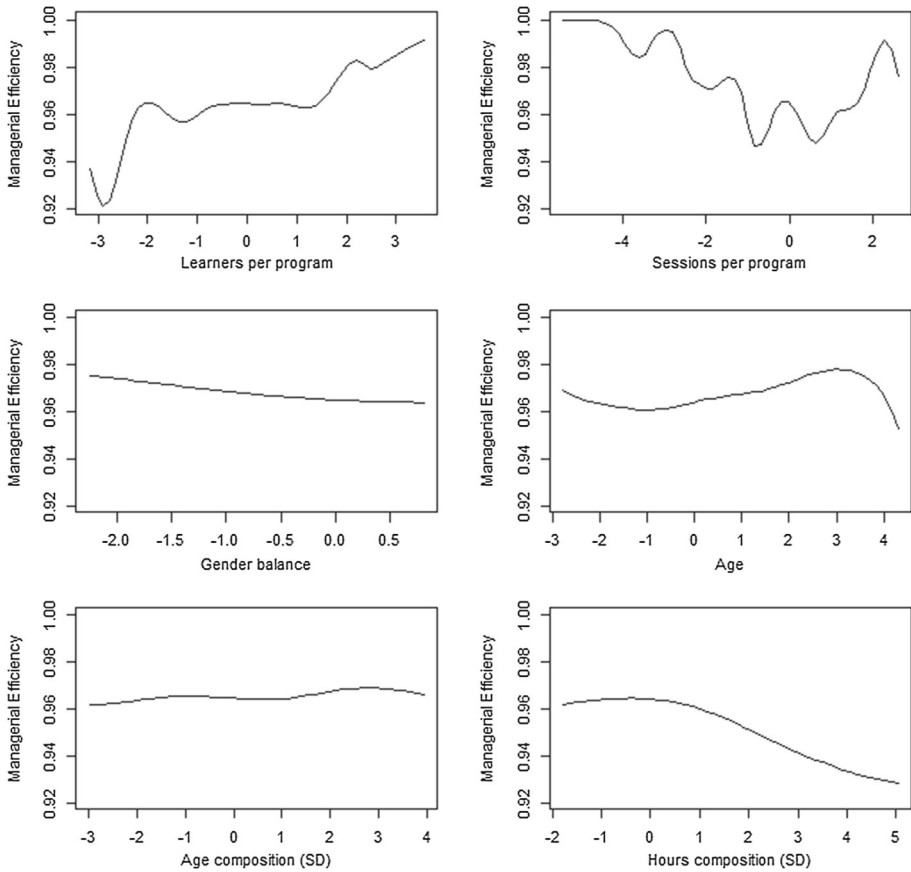


Fig. 7 Explaining efficiency differentials (including center fixed effect as additional conditional variable). Plots of nonparametric regressions are presented, relating the conditional and bias-corrected efficiency estimates to teacher and program characteristics. A negative slope suggests an unfavorable influence on efficiency, while a positive slope suggests a favorable influence on efficiency. We obtained p values using nonparametric bootstraps (Li and Racine 2007), and conclude that all program and teacher characteristics are significantly related to the managerial efficiency estimates ($p < 0.1$)

References

- Aragon, Y., Daouia, A., & Thomas-Agnan, C. (2005). Nonparametric frontier estimation: A conditional quantile-based approach. *Econometric Theory*, 21(2), 358–389.
- Asmild, M., Balezentis, T., & Hougaard, J. L. (2016). Multi-directional program efficiency: The case of Lithuanian family farms. *Journal of Productivity Analysis*, 45(1), 23–33.
- Blackburn, V., Brennan, S., & Ruggiero, J. (2014). Measuring efficiency in Australian Schools: A preliminary analysis. *Socio-Economic Planning Sciences*, 48(1), 4–9.
- Blanden, J., Buscha, F., Sturgis, P., & Urwin, P. (2012). Measuring the earnings returns to lifelong learning in the UK. *Economics of Education Review*, 31, 501–514.
- Cazals, C., Florens, J.-P., & Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics*, 106(1), 1–25.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.
- 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.

- Chen, Y., Cook, W. D., Du, J., Hu, H., & Zhu, J. (2017). Bounded and discrete data and Likert scales in data envelopment analysis: Application to regional energy efficiency in China. *Annals of Operations Research*, 255, 347–366.
- Cherchye, L., Demuyne, T., De Rock, B., & De Witte, K. (2014). Non-parametric analysis of multi-output production with joint inputs. *Economic Journal*, 124(577), 735–775.
- Cherchye, L., Lovell, C. A. K., Moesen, W., & Van Puyenbroeck, T. (2007). One market, one number? A composite indicator assessment of EU internal market dynamics. *European Economic Review*, 51(3), 749–779.
- Cordero, J. M., Alonso-Morán, E., Nuño-Solinis, R., Orueta, J. F., & Arce, R. S. (2015). Efficiency assessment of primary care providers: A conditional nonparametric approach. *European Journal of Operational Research*, 240(1), 235–244.
- Costrell, R., Hanushek, E., & Loeb, S. (2008). What do cost functions tell us about the cost of an adequate education? *Peabody Journal of Education*, 83(2), 198–223.
- Daraio, C., & Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: A probabilistic approach. *Journal of Productivity Analysis*, 24(1), 93–121.
- Daraio, C., & Simar, L. (2007). Conditional nonparametric frontier models for convex and nonconvex technologies: A unifying approach. *Journal of Productivity Analysis*, 28(1), 13–32.
- De Leebeek, K., Struyven, L., De Witte, K., Schiltz, F., & Mazrekaj, D. (2018). *Ondernemerschapstrajecten in het SYNTRA netwerk. Deelrapport 2: Een kwalitatieve analyse en benchmark van efficiëntie en effectiviteit*. Leuven.
- De Witte, K., & Geys, B. (2013). Citizen coproduction and efficient public good provision: Theory and evidence from local public libraries. *European Journal of Operational Research*, 224(3), 592–602.
- De Witte, K., & Kortelainen, M. (2013). What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics*, 45(17), 2401–2412.
- De Witte, K., & López-Torres, L. (2014). Efficiency in education. A review of literature and a way forward. *Journal of the Operational Research Society*, 68, 1–25.
- 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.
- De Witte, K., & Schiltz, F. (2018). Measuring and explaining organizational effectiveness of school districts. Evidence from a robust and conditional benefit-of-the-doubt approach. *European Journal of Operational Research*, 267(3), 1172–1181.
- Deprins, D., Simar, L., & Tulkens, H. (1984). Measuring labor inefficiency in post offices. In M. Marchand, P. Pestieau, & H. Tulkens (Eds.), *The performance of public enterprises: Concepts and measurements*. Amsterdam: North-Holland.
- Duncombe, W., Miner, J., & Ruggiero, J. (1995). Potential cost savings from school district consolidation: A case study of New York. *Economics of Education Review*, 14(3), 265–284.
- Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132(2), 245–259.
- Estelle, S. M., Johnson, A. L., & Ruggiero, J. (2010). Three-stage DEA models for incorporating exogenous inputs. *Computers & Operations Research*, 37(6), 1087–1090.
- Eurydice. (2018). *Adult education and training*. Retrieved December 12, 2018, from https://eacea.ec.europa.eu/national-policies/eurydice/content/adult-education-and-training-3_en.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3), 253–290.
- Hampf, B., Rødseth, K. L., & Løvold, K. (2015). Carbon dioxide emission standards for U.S. power plants: An efficiency analysis perspective. *Energy Economics*, 50, 140–153.
- Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor force quality, and the growth of nations. *American Economic Review*, 90(5), 1184–1208.
- Hanushek, E. A. E., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3), 607–668.
- Imazeki, J. (2008). Assessing the costs of adequacy in California public schools: A cost function approach. *Education Finance and Policy*, 3(1), 90–118.
- Jeong, S. O., & Simar, L. (2006). Linearly interpolated FDH efficiency score for nonconvex frontiers. *Journal of Multivariate Analysis*, 97(10), 2141–2161.
- Jerrim, J., Lopez-Agudo, L. A., Marcenaro-Gutierrez, O. D., & Shure, N. (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. *Economics of Education Review*, 61(August), 51–58.
- Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, 25(3), 273–288.

- Johnes, J. (2015). Operational research in education. *European Journal of Operational Research*, 243(3), 683–696.
- Li, Q., & Racine, J. (2007). *Nonparametric econometrics: Theory and practice*. Princeton: Princeton University Press.
- Li, Q., & Racine, J. S. (2008). Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics*, 26(4), 423–434.
- Martin, B. C., McNally, J. J., & Kay, M. J. (2013). Examining the formation of human capital in entrepreneurship: A meta-analysis of entrepreneurship education outcomes. *Journal of Business Venturing*, 28, 211–224.
- Narbón-Perpiñá, I., & De Witte, K. (2018a). Local governments' efficiency: A systematic literature review— Part I. *International Transactions in Operational Research*, 25(2), 431–468.
- Narbón-Perpiñá, I., & De Witte, K. (2018b). Local governments' efficiency: A systematic literature review— Part II. *International Transactions in Operational Research*, 25(4), 1107–1136.
- OECD. (2009). *PISA 2006 technical report*. Paris.
- OECD. (2017). *Education at a glance 2017: OECD indicators*. OECD.
- Olesen, O. B., Petersen, N. C., & Podinovski, V. V. (2015). Efficiency analysis with ratio measures. *European Journal of Operational Research*, 245(2), 446–462.
- Oosterbeek, H., van Praag, M., & Ijsselstein, A. (2010). The impact of entrepreneurship education on entrepreneurship skills and motivation. *European Economic Review*, 54(3), 442–454.
- Portela, M. C. S., Camanho, A. S., & Borges, D. (2012). Performance assessment of secondary schools: The snapshot of a country taken by DEA. *Journal of the Operational Research Society*, 63(8), 1098–1115.
- Puri, J., Yadav, S. P., Garg, H., Prasad, S., & Harish, Y. (2017). A new multi-component DEA approach using common set of weights methodology and imprecise data: an application to public sector banks in India with undesirable and shared resources. *Annals of Operations Research*, 259(1–2), 351–388.
- Ruggiero, J. (1996). On the measurement of technical efficiency in the public sector. *European Journal of Operational Research*, 90(3), 553–565.
- Ruggiero, J. (1998). Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research*, 111(3), 461–469.
- Schiltz, F., & De Witte, K. (2017). Estimating scale economies and the optimal size of school districts: A flexible form approach. *British Educational Research Journal*, 43(6), 1048–1067.
- Simar, L., & Wilson, P. W. (2000). A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics*, 27(6), 779–802.
- Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136(1), 31–64.
- Simar, L., & Wilson, P. W. (2011). Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis*, 36(1), 33–53.
- Thanassoulis, E., De Witte, K., Johnes, J., Johnes, G., Karagiannis, G., & Portela, C. S. (2016). Applications of data envelopment analysis in education. In J. Zhu (Ed.), *Data envelopment analysis. International series in operations research & management science* (Vol. 238). Boston: Springer.
- von Davier, M., Gonzalez, E. J., & Mislevy, R. J. (2009). What are plausible values and why are they useful? *IERI monograph series: Issues and Methodologies in Large-Scale Assessments*, 2, 9–36. Retrieved from http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02_Chapter_Introduction.pdf.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.