

Integrating quality into the nonparametric analysis of efficiency: a simulation comparison of popular methods

Yauheniya Varabyova¹ · Jonas Schreyögg¹

Published online: 1 September 2017
© Springer Science+Business Media, LLC 2017

Abstract Measuring efficiency in service-producing industries is challenging, because output quality may have to be considered in addition to output quantity. Previous studies have relied on different nonparametric methods to model the relationship between efficiency and quality; however, limited work has been done so far to compare and integrate these methods. Our literature search identified the following popular methods: a one-stage approach, a congestion analysis, and a two-stage approach. The first two methods treat quality as an additional output (or input) of the production process, whereas the two-stage approach first estimates efficiency without considering quality and then regresses the efficiency estimates on quality in the second stage. In this study, we compare the performance of these three conventional methods to a fairly novel method, the so-called conditional approach, that can also be used to incorporate quality into the analysis of efficiency. We simulate data using eight data generating processes that reflect potential relationships between efficiency and quality in empirical settings. Our simulation exercise illustrates the dominance of the conditional approach over the conventional methods in terms of both predicting the true efficiency scores and capturing the original relationship between efficiency and quality. The conditional approach represents a powerful and flexible tool that can be used when the theoretical link between efficiency and quality is unclear.

Keywords Simulation · Data envelopment analysis (DEA) · Conditional efficiency estimator · Managerial efficiency

✉ Yauheniya Varabyova
yauheniya.varabyova@uni-hamburg.de

Jonas Schreyögg
jonas.schreyoegg@uni-hamburg.de

¹ Hamburg Center for Health Economics, Universität Hamburg, Esplanade 36, 20354 Hamburg, Germany

1 Introduction

Frontier techniques measure the efficiency of decision-making units in transforming inputs into outputs against the best-practice benchmarks. Measuring efficiency in service-producing industries is more challenging than in manufacturing, because output quality needs to be considered in addition to output quantity (Becker et al. 2011). An organization producing higher levels of output quantity but of low quality may be efficient compared to other organizations at the same resource level. However, by compromising the quality of produced outputs this organization could leave the customers dissatisfied. Quality concerns are critical to many service industries, such as banking, education, hospitality, or health care. For instance, in the health care sector, low quality of provided health care services may compromise health outcomes or even the lives of patients. Because quality plays an important role in the relation between inputs and outputs, neglecting to control for quality in determining the best-practice frontier may distort the efficiency estimates.

The theoretical relationship between efficiency and quality is ambiguous. On the one hand, producing higher levels of quality may require sacrificing input resources and thereby reduce efficiency (Newhouse 1970; Olesen and Petersen 1995). On the other hand, the proponents of total quality management (TQM) argue that an increase in quality may lead to an increase in efficiency (Prior 2006). Likewise, the bad management hypothesis suggests that poor management can lead to both low efficiency and low quality (Berger and DeYoung 1997).

Several methods have been suggested to integrate quality into the analysis of efficiency; however, these methods differ regarding their underlying assumptions. There is a lack of methodological guidance on which models are appropriate or inappropriate both to obtain efficiency scores corrected for quality and to investigate the underlying relationship between efficiency and quality. Our study contributes to the existing literature dealing with the integration of external variables into efficiency analysis and explaining their influence on the production process. While there are two main streams of thought in the efficiency literature: parametric and nonparametric, this study will pursue only the nonparametric routes of integrating quality into the analysis.

The objective of this study is, therefore, to compare the available nonparametric methods that allow integrating quality into the analysis of efficiency. To accomplish this objective, we first provide an overview of the nonparametric methods used in previous studies to integrate quality into efficiency analysis. We subsequently compare the performance of the three most commonly used methods—the one-stage approach, the congestion analysis, and the two-stage approach—to a rather novel method, namely, the conditional approach (Bádin et al. 2012). To evaluate the performance of these four methods, we simulate data using eight different data generating processes (DGPs) that capture potential relationships between efficiency and quality. Using simulated datasets is convenient because we know both the true value of the simulated inefficiencies and the underlying relationship between quality and efficiency. Thus, we are able to examine how precisely the compared methods capture the true efficiency and which conclusion they reach regarding the relationship between efficiency and quality. To enhance the understanding of the strengths and limitations of the available methods we present the results in a simple and intuitive way using graphical illustrations.

2 Previous literature

We conducted a systematic search of studies published in peer-reviewed journals through June 2016 that integrated quality into a nonparametric analysis of the efficiency of service

providers. The electronic search was performed in the following databases: Business Source Complete, EconLit, and Web of Science. We complemented the search results by going through the references listed in the relevant articles.

Table 1 summarizes 57 reviewed studies in view of the utilized quality measures and methods to integrate quality into efficiency analysis. The majority of reviewed studies (33 of 57) are in the health care sector, and the remaining 24 studies are in other sectors, including banking, education, electricity, government, and hospitality sectors. The reviewed studies used, on average, 3 quality indicators (median 1, range 1–41). In the health care sector, the measures of quality were selected along the Donabedian dimensions of structure, process, and outcome quality (Donabedian 1966) with the most frequent measure being mortality rate. Some examples of quality measures utilized in other sectors included service quality score and problem loans (banking), student grades (education), energy loss (electricity), crime rate (government), and star rating (hospitality).

The most commonly used methods to integrate quality into efficiency analysis can be categorized into the one-stage approach, the congestion analysis, and the two-stage approach. The one-stage approach includes an external variable as an additional freely disposable input or output into the production function (Banker and Morey 1986). If the hypothesized influence of an external variable on the production is positive then it is treated as another input variable, and if the influence is negative then the variable is treated as another output variable. Several studies treated quality as another output in the model on the assumption that producing quality binds additional input resources and thus has a negative effect on the production (Chang et al. 2011; Lee et al. 2009; Mark et al. 2009; Navarro-Espigares and Torres 2011; Nayar and Ozcan 2008; Varabyova and Schreyögg 2013; Yang and Zeng 2014; Du et al. 2014; Garavaglia et al. 2011; Tiemann and Schreyögg 2009; Ray and Mukherjee 1998; Guccio et al. 2016; Jablonsky 2016; McMillan and Chan 2006). Alternatively, some studies used a reverse measure of quality, the lack of quality, as an additional input in the model (Giannakis et al. 2005; Growitsch et al. 2010; Jamasb and Pollitt 2003; Miguéis et al. 2012; Prior 2006; Von Hirschhausen et al. 2006; Yu et al. 2009; Drake and Hall 2003; Drake et al. 2006). In either case, the major methodological difficulty in using the one-stage approach is the need to know a priori whether the influence of quality on the production process is positive or negative. Moreover, the one-stage approach does not allow a non-monotonic (e.g., *U*-shaped) influence of the external variable. Additionally, the assumptions of convexity and returns to scale are imposed on the production set augmented by quality measures if data envelopment analysis (DEA) is used. The fact that the one-stage approach makes a lot of restrictive assumptions about the production set augmented by the quality measures makes it unsuitable for situations when there exists ambiguity in the effect of quality on the production process (Bádmín et al. 2010).

In contrast to the one-stage approach, which adds external variables to the production model under the assumption of free (strong) disposability, the congestion analysis relaxes this assumption to allow for weak disposability. The concept of congestion has been first applied to the joint production of electric power and pollution, whereby pollution represents a “bad” output of the production process (Färe and Grosskopf 1983). Under the assumption of strong disposability, a production unit could dispose of pollution at no cost; therefore, weak disposability is assumed to impose an opportunity cost on the disposal of “bad” outputs (Färe et al. 1994). The measure of congestion is defined as the output loss due to non-disposability (Färe and Grosskopf 1983). The idea to treat the lack of quality as a “bad” output of the production process has found appeal in a range of studies (Bilsel and Davutyan 2014; Clement et al. 2008; Davutyan et al. 2014; Prior 2006; Valdmanis et al. 2008; Mutter et al. 2010; Wu et al. 2013; Olesen and Petersen 1995; Matranga and Sapienza 2015; Dismuke

Table 1 Studies that integrated quality into the nonparametric analysis of efficiency in different service sectors

Study	Measures of quality	Methods
Health care		
Anderson et al. (2003)	Multidimensional score of quality of care	Two-stage approach
Bates et al. (2006)	Mortality rate	Two-stage approach
Bilsel and Davutyan (2014)	Risk adjusted mortality	Congestion analysis
Chang et al. (2011)	Inverse mortality rate	One-stage approach
Clement et al. (2008)	15 risk-adjusted mortality inpatients quality indicators	Congestion analysis
Davutyan et al. (2014)	Mortality rate	Congestion analysis
Dismuke and Sena (2001)	Dead discharges	Congestion analysis
Du et al. (2014)	In-hospital survival rate	One-stage approach
Ferrier and Trivitt (2013)	Four outcome quality measures (e.g., AMI mortality rate), eight process quality measures	One-stage approach, output adjustment
Ferrier and Valdmanis (1996)	Ratio of predicted to actual mortality rate	Two-stage approach
Garavaglia et al. (2011)	Structure quality measures (extra nursing hours, daily case-mix), residential charges	One-stage approach
Gok and Sezen (2013)	Structural quality index, patient satisfaction index	Two-stage approach
Kooreman (1994)	Four process quality measures (e.g., the presence of a patients' council)	Two-stage approach
Laine et al. (2005)	38 outcome and process quality measures and 3 structure quality measures	Two-stage approach
Lee et al. (2009)	Seven care planning deficiencies (e.g., failures to do comprehensive assessments)	Two-stage approach
Lenard and Shimshak (2009)	Residents with a catheter, physical restraints, pressure sores	Separate DEA models
Mark et al. (2009)	Patient satisfaction, the numbers of reported medication errors and patient falls	One-stage approach
Matranga and Sapienza (2015)	Inappropriate discharges, inappropriate day hospitals	Congestion analysis
Mutter et al. (2010)	Four patient safety indicators (e.g., infection due to medical care)	Congestion analysis

Table 1 continued

Study	Measures of quality	Methods
Navarro-Espigares and Torres (2011)	Four process indicators, surgical performance, % of caesareans to assisted births, collection of information, satisfaction by the users	One-stage approach
Nayar and Ozcan (2008)	Three process indicators (e.g., % of patients given initial antibiotic timing)	One-stage approach
Nyman and Bricker (1989)	Process quality (the numbers of deficiencies respect to the federally imposed standard), structure quality	Two-stage approach
Prior (2006)	Nosocomial infections	Congestion analysis, one-stage-approach
Shimshak et al. (2009)	Three process and outcome measures (e.g., number of residents without a catheter)	Separate DEA models, weight restrictions, multiple objectives DEA
Thanassoulis et al. (1995)	Satisfaction with care, survival rates	Weight restrictions
Tiemann and Schreyögg (2009)	Inverse mortality rate	One-stage approach
Valdmanis et al. (2008)	Four patient safety events (e.g., failure to rescue, infection due to medical care)	Congestion analysis
van Ineveld et al. (2015)	Decubitus ulcers prevalence	Output adjustment
Varabyova and Schreyögg (2013)	Inverse mortality rate	One-stage approach
Varabyova et al. (2016)	Patient satisfaction, mortality rate, radiation exposure	Conditional approach
Wu et al. (2013)	Readmissions	Congestion analysis
Yang and Zeng (2014)	Mortality rate, average length of stay	One-stage approach
Zhang et al. (2008)	Summation of all deficiency citations: clinical, administrative, resident rights, physical environment	Output adjustment
Banking		
Das and Ghosh (2006)	The ratio of operating expenses to total assets	Two-stage approach, output adjustment
Drake and Hall (2003)	Problem loans	One-stage approach
Drake et al. (2006)	Loan loss provisions	One-stage approach
Isik and Hassan (2003)	Problem loans	Two-stage approach

Table 1 continued

Study	Measures of quality	Methods
Kamakura et al. (2002)	Service quality perception of customers	Separate DEA models
Kenjegalieva et al. (2009)	Loan loss provisions	Congestion analysis
Kyj and Isik (2008)	Nonperforming loans	Output adjustment
Özkan-Günay et al. (2013)	Nonperforming loans	Congestion analysis
Park and Weber (2006)	Nonperforming loans	Congestion analysis
Sherman and Zhu (2006)	Service quality score	Separate DEA models
Education		
Guccio et al. (2016)	Regular graduates, degree classification	One-stage approach, output adjustment
Jablonsky (2016)	Quality of publications	One-stage approach
McMillan and Chan (2006)	Success at obtaining research grants in two areas	One-stage approach, Two-stage approach
Olesen and Petersen (1995)	Distribution of grades	Congestion analysis
Ray and Mukherjee (1998)	Average test score in an equivalency exam, percentage of graduating students pursuing further education	One-stage approach
Electricity		
Cambini et al. (2014)	Regulatory rewards and penalties for meeting quality targets, cost of energy not served	Input adjustment
Giannakis et al. (2005)	Number of minutes lost, number of interruptions	One-stage approach
Growitsch et al. (2010)	Cost of energy not served	One-stage approach
Jamasp and Pollitt (2003)	Energy losses	One-stage approach
Miguéis et al. (2012)	Cost of energy not served	One-stage approach
Von Hirschhausen et al. (2006)	Energy losses	One-stage approach
Yu et al. (2009)	Customer minutes lost, customer's WTP for avoiding power interruption	One-stage approach
Government		
Io Storto (2016)	Six variables that measure public service effectiveness (e.g., micro-crimes, urban waste, etc.)	Two-stage approach
Hospitality		
Poldrugovac et al. (2016)	Hotel star rating	Two-stage approach

and Sena 2001; Park and Weber 2006; Kenjegalieva et al. 2009; Özkan-Günay et al. 2013). By relaxing the assumption of strong disposability, the congestion analysis allows, in contrast to the one-stage approach, for a non-monotonic relationship between efficiency and quality. However, the convexity and returns to scale assumptions of the resulting augmented set are still required to be compatible with the quality measures. Moreover, the congestion analysis

has been criticized because the efficiency of production units may improve with an increase in a “bad” output (Chen 2013).

The investigation of the relationship between efficiency and quality has also been accomplished in a two-stage approach (Anderson et al. 2003; Bates et al. 2006; Gok and Sezen 2013; Kooreman 1994; Nyman and Bricker 1989; Laine et al. 2005; Ferrier and Valdmanis 1996; Io Storto 2016; Poldrugovac et al. 2016; McMillan and Chan 2006; Isik and Hassan 2003; Das and Ghosh 2006). The idea of the two-stage approach is to estimate efficiency using the information on input and output quantities in the first stage and then to regress the obtained efficiency estimates on quality indicators in the second stage. The fundamental difference between the two-stage approach and the two methods described above is the treatment of quality as an external variable, which is not directly related to either the inputs or the outputs of the production process. The reviewed studies used correlation analysis, analysis of variance, ordinary least squares (OLS), and Tobit regression to examine the relationship between the obtained efficiency scores and quality. The seminal work of Simar and Wilson (2007) compared the performance of different regression methods in the second stage of the two-stage analysis and showed that the truncated regression with double bootstrap outperformed other conventionally used methods. The work of Simar and Wilson (2007) assumed a situation, in which external factors analyzed in the second stage did not have an effect on the attainable input–output set. This situation requires the verification of a so-called separability condition between external variables (such as quality) and the set of inputs and outputs (Simar and Wilson 2011). The assumption of separability may be quite restrictive, because, in some situations, pursuing the objective of maintaining high quality standard may compromise other production objectives (Sherman and Zhu 2006) and thereby shift the production possibility frontier.

Other methods to integrate quality into efficiency analysis are less frequently used. Thus, some studies adjusted the quantity of outputs or inputs for quality. The concept of output adjustment is quite familiar in the health care sector because output quantity (e.g., number of treated patients) is commonly adjusted for case-mix on the basis of diagnosis-related groups (DRGs). The adjustment of output for case-mix is justified by the fact that the treatment of some DRGs requires much more labor and capital resources than the treatment of other DRGs. Extending this concept, some studies suggested adjusting health care output for quality indicators (Ferrier and Trivitt 2013; Zhang et al. 2008; van Ineveld et al. 2015). However, in contrast to the DRG-based adjustment weights, which are resource-based, no unambiguous relationship exists between health care quality and input resources, which may lead to a distortion in the obtained findings. Similarly, in the education sector, Guccio et al. (2016) weighted the number of graduates by degree classification. Other studies subtracted quality-related costs from inputs or the value of undesirable outputs from total outputs (Cambini et al. 2014; Park and Weber 2006; Kyj and Isik 2008). Yet other studies integrated quality by estimating separate DEA models for quality efficiency and operating efficiency, by using a DEA model with weight restrictions, or by considering a DEA model with multiple objectives (Lenard and Shimshak 2009; Shimshak et al. 2009; Thanassoulis et al. 1995; Kamakura et al. 2002; Sherman and Zhu 2006). The method involving separate DEA models generates separate scores for quality efficiency and operating efficiency, which complicates the interpretation and application of this method in managerial practice. The weight restrictions method yields a single efficiency score; however, the analyst must decide on the level of restrictions, which introduces arbitrariness into the model. The method based on the DEA model with multiple objectives uses the sum of the objectives for individual models while restricting the weights assigned to measures appearing in more than one individual model. Similar to the weight restrictions method, the multiple-objectives DEA model requires the value judgment of the

analyst to determine the weight preference for the utilized measures. Finally, one study relied on the innovative conditional approach to integrate quality into efficiency analysis without making a priori assumptions on quality measures and showed the different ways in which quality indicators may affect the production process (Varabyova et al. 2016).

In the next sections, we will focus on the one-stage approach, the congestion analysis, and the two-stage approach and compare the performance of these three popular methods to the recently introduced conditional approach.

3 Methodology of the conditional approach

The analysis of efficiency typically contains two parts: one is the estimation of a benchmark frontier and the other involves the investigation of the influence of external factors on the production process. Simar and Wilson (2015) provided an excellent guide of the definitions and assumptions that are important to understand the differences between the nonparametric methods compared in this study: the one-stage approach, the congestion analysis, the two-stage approach, and the conditional approach. Because the former three methods have been described extensively in previous literature we will focus on providing a brief description of the methodology of the conditional approach along the lines of Bădin et al. (2014) and the references therein.

3.1 Conditional estimator

Let $x \in \mathbb{R}_+^p$ characterize a vector of p inputs and $y \in \mathbb{R}_+^q$ characterize a vector of q outputs. The production set describes the set of physically attainable points (x, y) and is defined by

$$\Psi = \{(x, y) \mid x \text{ can produce } y\}. \tag{3.1}$$

Cazals et al. (2002) suggested characterizing the production process by the joint probability measure on (X, Y) . In terms of the joint probability measure, the output-oriented efficiency under free disposability is given by

$$\lambda(x, y) = \sup \{ \lambda \mid S_{Y|X}(\lambda y | x) > 0 \}, \tag{3.2}$$

where $S_{Y|X}(\lambda y | x) = \Pr(Y \geq \lambda y \mid X \leq x)$.

The probabilistic interpretation of the production process provides a convenient means to incorporate external factors $Z \in \mathbb{R}^r$ into the estimation, whereby the attainable set is adjusted for the condition $Z = z$:

$$\Psi^Z = \{(x, y) \mid Z = z, x \text{ can produce } y\}. \tag{3.3}$$

The conditional output-oriented efficiency measure is adapted to the condition $Z = z$:

$$\lambda(x, y | z) = \sup \{ \lambda \mid S_{Y|X,Z}(\lambda y | x, z) > 0 \}, \tag{3.4}$$

where $S_{Y|X,Z}(\lambda y | x, z) = \Pr(Y \geq \lambda y \mid X \leq x, Z = z)$.

The empirical estimators are obtained from a sample of n observations $S_n = \{(X_i, Y_i, Z_i)\}_{i=1}^n$. The unconditional estimator, which does not take Z into account, is given by

$$\hat{S}_{Y|X}(y|x) = \frac{\sum_{i=1}^n I(Y_i \geq y, X_i \leq x)}{\sum_{i=1}^n I(X_i \leq x)}, \tag{3.5}$$

where $I(\cdot)$ is an indicator function that takes the value of one if the argument is true and zero otherwise. The corresponding conditional output-oriented estimator, which accounts for a univariate Z , is obtained using nonparametric kernel smoothing:

$$\hat{S}_{Y|X,Z}(y|x, z) = \frac{\sum_{i=1}^n I(Y_i \geq y, X_i \leq x) K((Z_i - z)/h)}{\sum_{i=1}^n I(X_i \leq x) K((Z_i - z)/h)}, \tag{3.6}$$

where $K(\cdot)$ is some kernel with compact support (e.g., Epanechnikov) and h is the bandwidth. The value of h determines the comparison group whereby large values of h allow identifying the non-influence of Z by oversmoothing. Bădin et al. (2010) described the method of optimal bandwidth selection using a data-driven approach based on Least Squares Cross Validation (LSCV) procedure. The authors also provided a MATLAB routine to estimate the LSCV criterion.

The resulting unconditional estimator $\hat{\lambda}(x, y)$ based on probabilistic theory is obtained by plugging $\hat{S}_{Y|X}(y|x)$ into Eq. (3.2):

$$\hat{\lambda}(x, y) = \sup \left\{ \lambda | \hat{S}_{Y|X}(\lambda y|x) > 0 \right\}. \tag{3.7}$$

This estimator for the firm operating at level (x, y) coincides with the conventional Free Disposal Hull (FDH) estimator for nonconvex functions: $\hat{\lambda}_{FDH}(x, y) = \max_{i|X_i \leq x} \left\{ \min_{j=1, \dots, q} \left(\frac{Y_i^j}{y^j} \right) \right\}$. Thus, obtaining the unconditional efficiency estimator using health care data is equivalent to the estimation of technical efficiency based on a FDH model using only the information on health care inputs and outputs.

The corresponding conditional estimator is derived by using $\hat{S}_{Y|X,Z}(y|x, z)$ in formula (3.4):

$$\begin{aligned} \hat{\lambda}(x, y|z) &= \sup \left\{ \lambda | \hat{S}_{Y|X,Z}(\lambda y|x, z) > 0 \right\} = \hat{\lambda}(x, y|z) \\ &= \max_{i|X_i \leq x, \|Z_i - z\| \leq h} \left\{ \min_{j=1, \dots, q} \left(\frac{Y_i^j}{y^j} \right) \right\}. \end{aligned} \tag{3.8}$$

In contrast to the unconditional estimator, which does not account for Z , the conditional estimator is based on a benchmark frontier of production units with similar values of Z (the comparison group is limited by the size of h). It is, therefore, a localized version (by h) of the FDH estimator.

An external variable Z may represent the value of output quality. In contrast to the unconditional efficiency estimator, obtaining the conditional efficiency estimator is equivalent to analyzing technical efficiency only across service providers with similar levels of quality. In other words, high-quality providers are benchmarked against other high-quality providers whereas low-quality providers are benchmarked against other low-quality providers.

3.2 Investigating the influence of external factors

The analysis of the unconditional and conditional estimates allows disentangling the influence of the external factor Z on the production process. The influence may be twofold: first, Z may influence the shape of the attainable frontier, and, second, Z may affect the distribution of inefficiencies inside the production set. The first type of influence suggests that the support of (X, Y) depends on Z , and thus, $\Psi^Z \neq \Psi$. In the second case, the shape of the attainable frontier is not affected, $\Psi^Z = \Psi$, but the distribution of inefficiencies inside Ψ^Z is changed

because of Z . Note that if $\Psi^Z = \Psi$ for all $Z \in \mathbb{R}^r$ then the “separability” condition described in [Simar and Wilson \(2007\)](#) holds.

To investigate the effect of Z on the shape of the attainable frontier, the individual ratios of conditional to unconditional estimates, $\hat{R}(x, y|z) = \hat{\lambda}(x, y|z) / \hat{\lambda}(x, y) \leq 1$, are regressed on Z :

$$\hat{R}(x, y|z) = g(z) + \epsilon, \quad (3.9)$$

where ϵ is the error term and g is the mean regression function. [Daraio and Simar \(2005\)](#) suggested examining the scatterplot of the ratios $\hat{R}(x, y|z)$ against Z and the smoothed regression line to detect the influence of Z on the attainable frontier. In the output-oriented framework, an increasing regression line indicates a positive influence of Z on the production process, whereas a decreasing line indicates a negative influence of Z on the production process.

To analyze the effect of Z on the distribution of inefficiencies, [Bădin et al. \(2012\)](#) proposed regressing the conditional efficiency estimates on Z . Note that this is similar to the conventional two-stage analysis, in which the values of unconditional efficiency estimates are regressed on external factors. However, [Bădin et al. \(2012\)](#) suggested exploring the average behavior of the conditional estimator instead of the unconditional estimator as a function of Z using a flexible regression model:

$$\hat{\lambda}(x, y|z) = \mu(z) + \sigma(z)\epsilon, \quad (3.10)$$

where $\mu(z)$ measures the average effect of Z on efficiency and $\sigma(z)$ is the variance function, which provides the information on the dispersion of the inefficiencies. The scatterplot of $\hat{\lambda}(x, y|z)$ against Z and the smoothed regression line allow detecting the influence of Z on the distribution of inefficiencies.

An additional benefit of this approach is the derivation of the values of managerial efficiency, which allows benchmarking the production units facing different levels of z . The managerial efficiencies are obtained by analyzing the residual, ϵ , from Eq. (3.10). The estimation of $\hat{\epsilon}$ is accomplished in two steps. In the first step, $\hat{\mu}(z)$ is estimated using a nonparametric regression. In the second step, the estimate of the variance function $\hat{\sigma}(z)$ is obtained by using standard procedures, for example, regressing the logarithm of the squares of the residuals obtained in the first step on Z ([Wasserman 2006](#)). The managerial efficiency of a production unit is derived by rearranging Eq. (3.10):

$$\hat{\epsilon} = \frac{\hat{\lambda}(x, y|z) - \hat{\mu}(z)}{\hat{\sigma}(z)}. \quad (3.11)$$

All analyses in this study were performed using R software. To encourage future research, we will provide the estimation code to interested researchers on request.

4 Simulation design

To illustrate how quality can be related to the production process, we introduce eight simulated examples along the lines of [Bădin et al. \(2012, 2014\)](#) and [Varabyova et al. \(2016\)](#). To keep the analysis simple and to enable the graphical presentation of results, we use two-dimensional models; however, the extension to a higher number of dimensions is possible ([Bădin et al. 2014](#)). In all eight simulation scenarios, we consider the case of a univariate frontier, in which the amount of input is set to equal one ($X \equiv 1$). The inefficiency term is

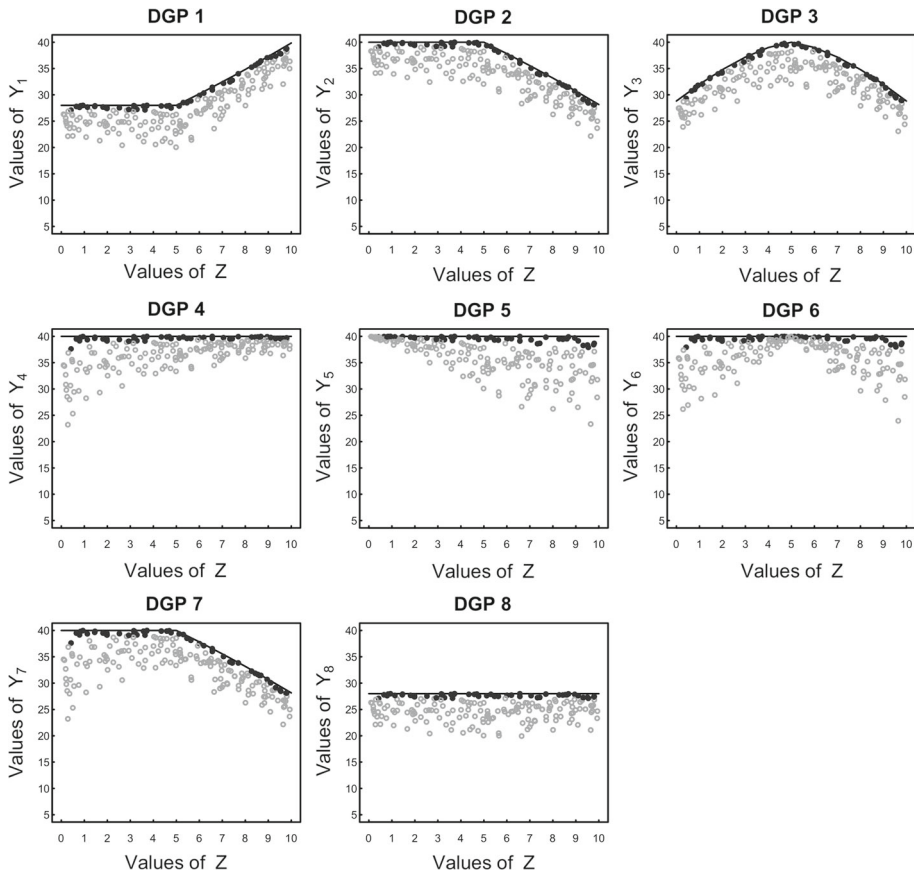


Fig. 1 The scatterplots of Z against Y using simulated data ($n = 200$). *Black dots* represent the most efficient observations $U \leq Q_1$. The *solid line* denotes the true attainable frontier. In DGP 1–DGP 3, Z shifts the attainable frontier. In DGP 4–DGP 6, Z affects the distribution of the inefficiencies. In DGP 7, Z affects both the distribution of the inefficiencies and the attainable frontier. In DGP 8, Z has no effect on the production process

half-normally distributed, $U \sim \mathcal{N}^+(0, \sigma_U^2)$ with $\sigma_U^2 = 3$, and the measure of quality is uniformly distributed, $Z \sim unif(0, 10)$. Suppose that Z represents an undesirable measure of quality, for instance, in-hospital mortality, with higher values of Z representing worse quality, because higher mortality values are undesirable. Suppose further that the amount of output Y denotes the quantity of produced output.

Figure 1 provides the scatterplots of Y against Z using observations ($n = 200$) simulated according to eight different DGPs described below. In all subsequent illustrations, the most efficient units (with the values of inefficiency, U , smaller than the first quartile value Q_1) are highlighted in black. Obviously, the most efficient observations (with small values of U) should be located at the attainable efficient frontier (solid black line).

The upper panel of Fig. 1 illustrates three DGPs, for which the separability condition is violated, resulting in a shift of the attainable frontier due to the effect of Z on Y . The observations depicted in the upper panel of Fig. 1 are simulated according to the three following DGPs:

$$Y_1 = 28I(Z \leq 5) + (28 - 5^{1.3} + Z^{1.3})I(Z > 5) - 2U, \tag{4.1}$$

$$Y_2 = 40I(Z \leq 5) + (40 + 5^{1.3} - Z^{1.3})I(Z > 5) - 2U, \tag{4.2}$$

$$Y_3 = 40 - |Z - 5|^{1.5} - 2U, \tag{4.3}$$

where $I(\cdot)$ is an indicator function that takes the value of one if the argument is true and zero otherwise.

In DGP 1 (4.1), Z has no effect on Y for $Z \leq 5$, but Z has a positive impact on Y for $Z > 5$. In DGP 2 (4.2), Z has no effect on Y for $Z \leq 5$, but Z has a negative impact on Y for $Z > 5$, shifting the boundary of the attainable frontier. In DGP 3 (4.3), there is an inverted U -shaped relationship between Z and Y , the shift of the attainable frontier is more pronounced when $|Z - 5|$ increases.

The middle panel of Fig. 1 shows three DGPs for which the boundaries of the attainable frontier are not affected by Z (i.e., the separability condition is verified). In these DGPs, Z has an effect only on the distribution of the inefficiencies inside the production set. In DGP 4–DGP 6, the output is generated by the following expressions:

$$Y_4 = 40 - 7U(Z + 1)^{-0.8}, \tag{4.4}$$

$$Y_5 = 40 - 0.5UZ, \tag{4.5}$$

$$Y_6 = 40 - U|Z - 5|. \tag{4.6}$$

DGP 4 (4.4) reflects a positive effect of Z on the production, because for higher levels of Z the output is higher. In DGP 5 (4.5), Z is negatively associated with the production of output. In DGP 6 (4.6), there is a higher probability of being inefficient when $|Z - 5|$ increases.

Finally, the lower panel of Fig. 1 displays two following situations: a situation in which Z has an effect on both the production frontier and the distribution of the inefficiencies, and a situation, in which Z has no effect on the production process:

$$Y_7 = (40 - 7U(Z + 1)^{-0.8})I(Z \leq 5) + (40 + 5^{1.3} - Z^{1.3})I(Z > 5) - 2U, \tag{4.7}$$

$$Y_8 = 28 - 2U. \tag{4.8}$$

In DGP 7 (4.7), Z positively affects the production for $Z \leq 5$ but has a negative effect on the production frontier for $Z > 5$. In DGP 8 (4.8), Z is not a part of the production process.

5 Results

We compare the performance of the four methods in terms of how precisely they capture the true efficiency estimates and which conclusion they reach regarding the relationship between efficiency and quality. The precision in capturing the true efficiencies is measured using correlation analyses and the mean squared error (MSE). Pearson linear correlation evaluates the linear relationship between the true and estimated efficiencies, whereas Spearman rank correlation evaluates the monotonic relationship between the true and estimated efficiencies. The MSE measures the average of the squares of the errors—that is, the difference between the simulated and estimated inefficiency. We use the standardized measures of both the simulated and estimated inefficiency (mean of zero and a standard deviation of one) to account for the scaling effects. The MSE is a non-negative measure, and values closer to zero are better. The underlying relationships between quality and efficiency are examined according to the typical procedure used in each method. The one-stage approach and the congestion analysis compare

the sets of efficiency estimates derived using different assumptions on quality, whereas the two-stage approach and the conditional approach rely on regression analyses.

5.1 One-stage approach

Following the majority of the reviewed studies relying on the one-stage approach, we use a DEA model with VRS (Banker et al. 1984). The estimation model includes one input (X) and two outputs (Y, \tilde{Z}). In the simulation exercise, we denoted Z an undesirable measure of quality. To redefine it in a favorable way, we apply a linear transformation $\tilde{Z} = 100 - Z$. Now, the expansion of both outputs (Y, \tilde{Z}) in the DEA model is desirable. The efficiency estimates from the one-stage approach, $\hat{\lambda}_{VRS}(x, y, \tilde{z})$, are obtained using an output-oriented model. The values of $\hat{\lambda}_{VRS}(x, y, \tilde{z})$ equal to 1 denote efficient units and the values of $\hat{\lambda}_{VRS}(x, y, \tilde{z})$ greater than 1 represent inefficient units.

Figure 2 shows eight scatterplots of \tilde{Z} against Y , which slightly differ from Fig. 1 due to the transformation of Z . The graphical illustration of the first three DGPs (the upper panel of Fig. 2) makes evident that the one-stage approach does not fully capture the true efficiency, because the most efficient black points (i.e., $\hat{\lambda}_{VRS}(x, y, \tilde{z})$ smaller than the first quartile value) are not always located close on the true efficient frontier. The Pearson correlations between the true values of inefficiency, U , and $\hat{\lambda}_{VRS}(x, y, \tilde{z})$ are rather low and equal to 0.54, 0.11, and 0.17 for DGP 1, DGP 2, and DGP 3, respectively. The corresponding Spearman correlations are 0.50, 0.13, and 0.29. The MSE are equal to 0.92, 1.77, and 1.66. In DGP 1, the resulting frontier is non-convex, and the DEA model does not capture the true inefficiency values, because it assumes convexity. In DGP 2, the external variable \tilde{Z} has a positive effect on production, and should be used not as an extra output, but rather as an extra input. In DGP 3, the relationship between quality and efficiency is non-monotonic and, therefore, the DEA model assuming free (strong) disposability of outputs fails to capture the true frontier.

In DGP 4–6 (the middle panel of Fig. 2), Z does not affect Y directly (the separability condition holds), but it affects the distribution of the inefficiencies U . In this situation, including \tilde{Z} as another output is inappropriate, because it means adding a variable, which is not part of the production process. The introduction of an additional variable increases the number of constraints, and therefore increases the average measured efficiency. However, the new constraint is most likely to be binding only for the production units with relatively high value of \tilde{Z} in relation to input (Smith 1997). The Pearson correlations between U and $\hat{\lambda}_{VRS}(x, y, \tilde{z})$ are equal to 0.62, 0.38, and 0.35 for DGP 4, DGP 5, and DGP 6, respectively. The corresponding Spearman correlations are 0.69, 0.45, and 0.43. The MSE are equal to 0.66, 1.24, and 1.30. The low correlation values and high MSE values are due to the fact that the effect of Z on the distribution of efficiencies is not accounted for.

The Pearson correlations for DGP 7 and DGP 8 are equal to 0.11 and 0.46, whereas the corresponding Spearman correlations are equal to 0.14 and 0.51. The MSE are equal to 1.78 and 1.08 for DGP 7 and DGP 8, respectively. In DGP 7, the low correlation values and the MSE quite different from zero are due to the fact that the assumptions of the one-stage approach are violated. In DGP 8, the inclusion of a variable, which is not part of the production process, is inappropriate.

To examine the underlying relationship efficiency and quality, some studies relying on the one-stage approach (Navarro-Espigares and Torres 2011; Nayar and Ozcan 2008; Tiemann and Schreyögg 2009) compared the DEA estimates obtained using only input and output quantities with the efficiency estimates obtained after a quality measure was added. Adding

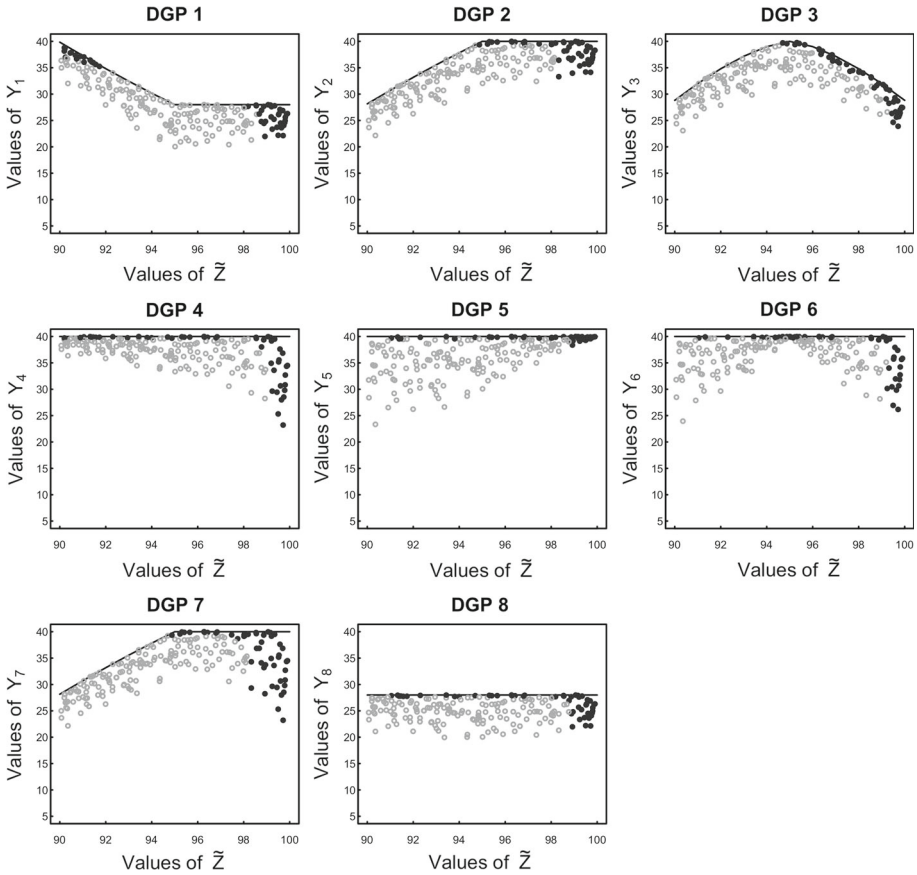


Fig. 2 The results of the one-stage approach using the transformed quality variable, \tilde{Z} . *Black dots* represent the most efficient observations, $\hat{\lambda}_{VRS}(x, y, \tilde{z}) \leq Q_1$. The *solid line* denotes the true attainable frontier. In DGP 1–DGP 3, \tilde{Z} shifts the attainable frontier. In DGP 4–DGP 6, \tilde{Z} affects the distribution of the inefficiencies. In DGP 7, \tilde{Z} affects both the distribution of the inefficiencies and the attainable frontier. In DGP 8, \tilde{Z} has no effect on the production process

an additional variable improves the average efficiency through improving the efficiency estimates of the observations with high values of quality. The more dimensions of quality are added to the model the higher is the increase of the average efficiency. However, the interpretation of results may be difficult. For instance, consider a comparison of DGP 1 (a negative relationship between efficiency and quality) and DGP 2 (a positive relationship between efficiency and quality). The average efficiency of 200 production units in the model without quality, $\hat{\lambda}_{VRS}(x, y)$, equals 1.40 and 1.19 for DGP 1 and DGP 2, respectively. When quality \tilde{Z} is added to the model, $\hat{\lambda}_{VRS}(x, y, \tilde{z})$, the average efficiency among 200 units equals 1.03 and 1.05 for DGP 1 and DGP 2, respectively. Thus, in both cases there has been a substantial increase in the average efficiency; however, only in DGP 1 there is a negative relationship between efficiency and quality. Without a clear picture of the underlying DGP, making conclusions about the relationship between quality and efficiency on the basis of the average efficiency score is problematic.

5.2 Congestion analysis

To conduct the congestion analysis, we relax the strong disposability assumption in the DEA model with VRS to allow for the weak disposability of the “bad” output using the model of Fare et al. (1994). The estimation model includes one input (X) and two outputs (Y, Z), where Z represents the “bad” output (the lack of quality). We denote the efficiency estimates obtained under the weak disposability assumption as $\hat{\lambda}_{WEAK}(x, y, z)$ and the efficiency estimates obtained under the strong disposability assumption as $\hat{\lambda}_{VRS}(x, y, z)$.

The scatterplots illustrated in the upper panel of Fig. 3 suggest that the models including Z as a bad output under weak disposability provide a better fit than the one-stage approach in DGP 2 and DGP 3, because the most efficient observations, $\hat{\lambda}_{WEAK}(x, y, z)$, are located closer to the true efficient frontier. The improved estimates are due to the relaxation of the assumption of strong disposability. The Pearson correlations between the true values of inefficiency, U , and $\hat{\lambda}_{WEAK}(x, y, z)$ are equal to 0.70, 0.93, and 0.87 for DGP 1, DGP 2,

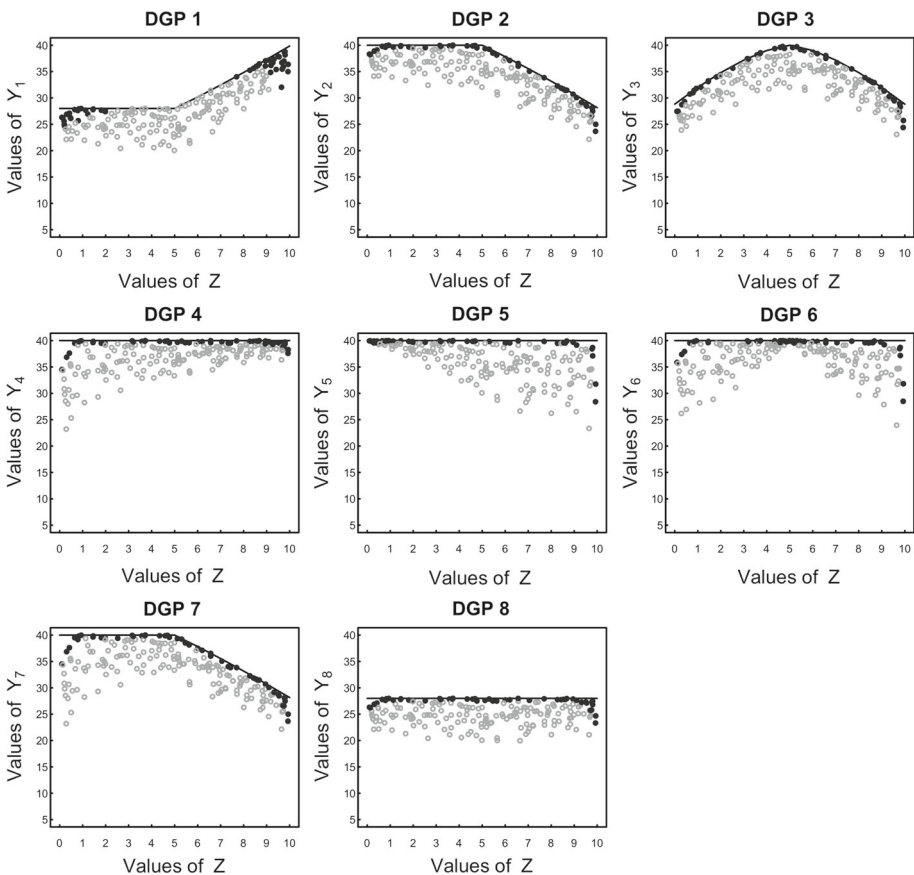


Fig. 3 The results of congestion analysis under weak disposability of outputs. *Black dots* represent the most efficient observations, $\hat{\lambda}_{WEAK}(x, y, z) \leq Q_1$. The *solid line* denotes the true attainable frontier. In DGP 1–DGP 3, Z shifts the attainable frontier. In DGP 4–DGP 6, Z affects the distribution of the inefficiencies. In DGP 7, Z affects both the distribution of the inefficiencies and the attainable frontier. In DGP 8, Z has no effect on the production process

and DGP 3, respectively. The corresponding Spearman rank correlations are 0.62, 0.92, and 0.90. The MSE are equal to 0.59, 0.15, and 0.26. In DGP 1, the estimates are rather poor because the assumption of convexity is violated.

The middle panel of Fig. 3 presents the DGPs where Z only affects the distribution of U . As discussed above when addressing the results of the one-stage approach, Z should not be included as an output. Although the inclusion of Z does not affect the efficiency of production units with low values of Z , it leads to the improvement of efficiency for the production units with relatively high values of Z . This may seem particularly problematic from the managerial perspective, because Z represents an undesirable measure of quality. The Pearson correlations between U and $\hat{\lambda}_{WEAK}(x, y, z)$ equal to 0.67, 0.75, and 0.60 for DGP 4, DGP 5, and DGP 6, respectively. The corresponding Spearman correlations are 0.85, 0.74, and 0.68. The MSE are equal to 0.65, 0.49, and 0.80 for DGP 4, DGP 5, and DGP 6, respectively.

The lower panel of Fig. 3 contains results for DGP 7 and DGP 8. The Pearson correlations are equal to 0.68 and 0.92 for DGP 7 and DGP 8, respectively. The corresponding Spearman correlations are equal to 0.87 and 0.91. The MSE are equal to 0.63 and 0.15. In DGP 7, the effect of Z on the distribution of the inefficiencies is not captured. In DGP 8, Z wrongly affects the efficiency estimates with high and low values of Z .

The measure of output congestion, C , is obtained by taking the ratio $\hat{\lambda}_{VRS}(x, y, z) / \hat{\lambda}_{WEAK}(x, y, z)$. C represents the output loss due to non-disposability. Note that in the output-oriented framework, $\hat{\lambda}_{VRS}(x, y, z) \geq \hat{\lambda}_{WEAK}(x, y, z)$ and thus $C \geq 1$. Looking at the first three DGPs, the mean value of C equals 1.14, 1.00, and 1.06 for DGP 1, DGP 2, and DGP 3, respectively. Thus, there is no congestion in case of DGP 2, but there is some congestion between efficiency and quality outputs in DGP 1 and DGP 3. The results are as we expect from the simulation exercise. On the other hand, the mean value of C equals 1.00 for all three DGPs in which Z has an effect on the distribution of inefficiencies (i.e., DGP 4–6). Thus, in this situation the congestion approach fails to identify the relationship between quality and efficiency for DGP 4 and DGP 6. Moreover, the analysis of congestion ratio doesn't allow identifying the inverted U -shaped relationship between efficiency and quality in DGP 3 or DGP 6. The values of C for DGP 7 and DGP 8 are equal to 1.01 and 1.00. Thus, DGP 7 identifies some congestion and DGP 8 correctly identifies no underlying relationship between efficiency and quality.

5.3 Two-stage approach

To obtain the efficiency estimates in the first stage of the two-stage approach, we estimate a DEA model with VRS. The output-oriented estimation model includes one input, X , and one output, Y . We denote the obtained efficiency estimates as $\hat{\lambda}_{VRS}(x, y)$. As before, the values of efficiency estimates equal to 1 represent efficient units, and the values greater than 1 represent inefficient units.

The upper panel of Fig. 4 shows the obtained DEA estimates in the DGPs, where the effect of Z causes a shift in the attainable efficient frontier. Figure 4 confirms that the fit of the efficient frontier is very poor in these situations, because only observations with large Y are deemed efficient. The Pearson correlations between the true values of inefficiency, U , and $\hat{\lambda}_{VRS}(x, y)$ are equal to 0.58, 0.39, and 0.42 for DGP 1, DGP 2, and DGP 3, respectively. The corresponding Spearman correlations are 0.55, 0.48, and 0.43 and the MSE are 0.83, 1.21, and 1.16.

The middle panel of Fig. 4 shows the obtained DEA estimates in the DGPs, where the influence of Z leads to a change in the distribution of inefficiencies. The most efficient

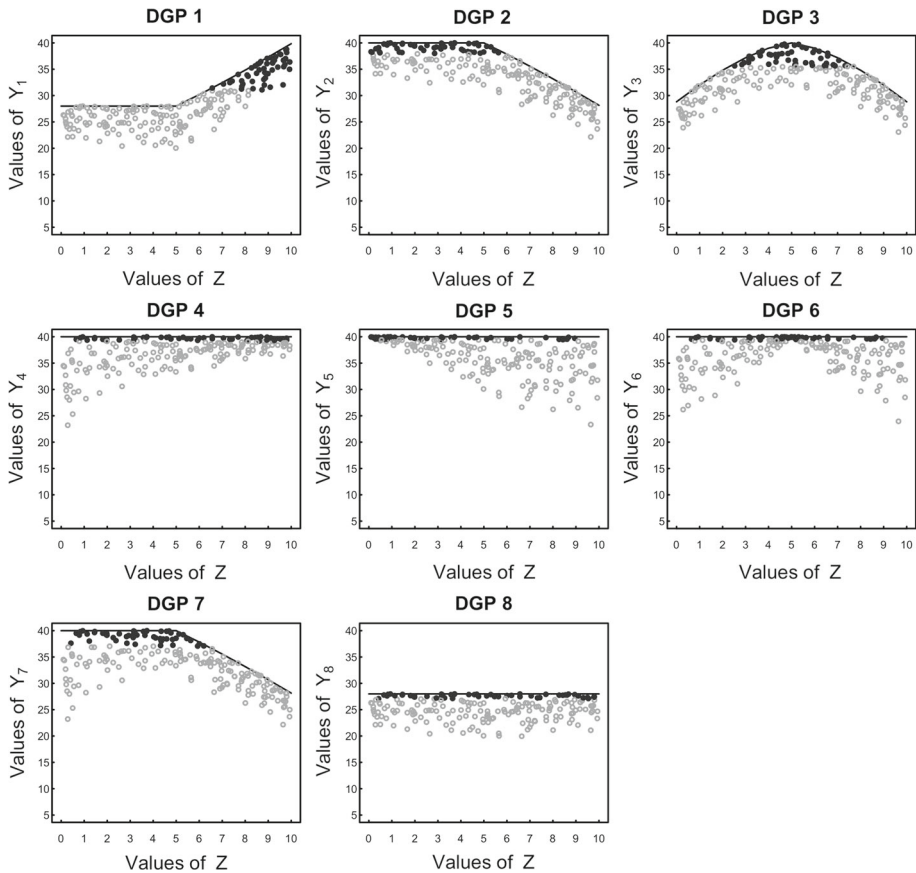


Fig. 4 The efficiency estimates from the first stage of the two-stage analysis. *Black dots* represent the most efficient observations, $\hat{\lambda}_{VRS}(x, y) \leq Q_1$. The *solid line* denotes the true attainable frontier. In DGP 1–DGP 3, Z shifts the attainable frontier. In DGP 4–DGP 6, Z affects the distribution of the inefficiencies. In DGP 7, Z affects both the distribution of the inefficiencies and the attainable frontier. In DGP 8, Z has no effect on the production process

observations are indeed located at the true efficient frontier. However, the Pearson correlations between the true values of inefficiency, U , and $\hat{\lambda}_{VRS}(x, y)$ are rather low and equal to 0.63, 0.73, and 0.59 for DGP 4, DGP 5, and DGP 6, respectively. The corresponding Spearman correlations are 0.88, 0.76, and 0.67. These weak correlations are due to the fact that the effect of Z on the distribution of inefficiencies is not taken into account. The MSE are equal to 0.73, 0.53, and 0.81 for DGP 4, DGP 5, and DGP 6, respectively.

The lower panel of Fig. 4 shows the results for DGP 7 and DGP 8. The Pearson correlations equal 0.47 and 1.00 for DGP 7 and DGP 8, respectively. The corresponding Spearman correlations equal 0.52 and 1.00, and the MSE equal 1.06 and 0.01. The results for DGP 7 are rather poor, because the effects of Z both on the distribution of the inefficiencies and on the shift of the frontier are not considered. The results for DGP 8 are very good because Z is not included into the estimation.

In the second stage, the efficiencies estimates obtained in the first stage, $\hat{\lambda}_{VRS}(x, y)$, are regressed on Z using a Tobit regression censored at 1. Simar and Wilson (2011) do not

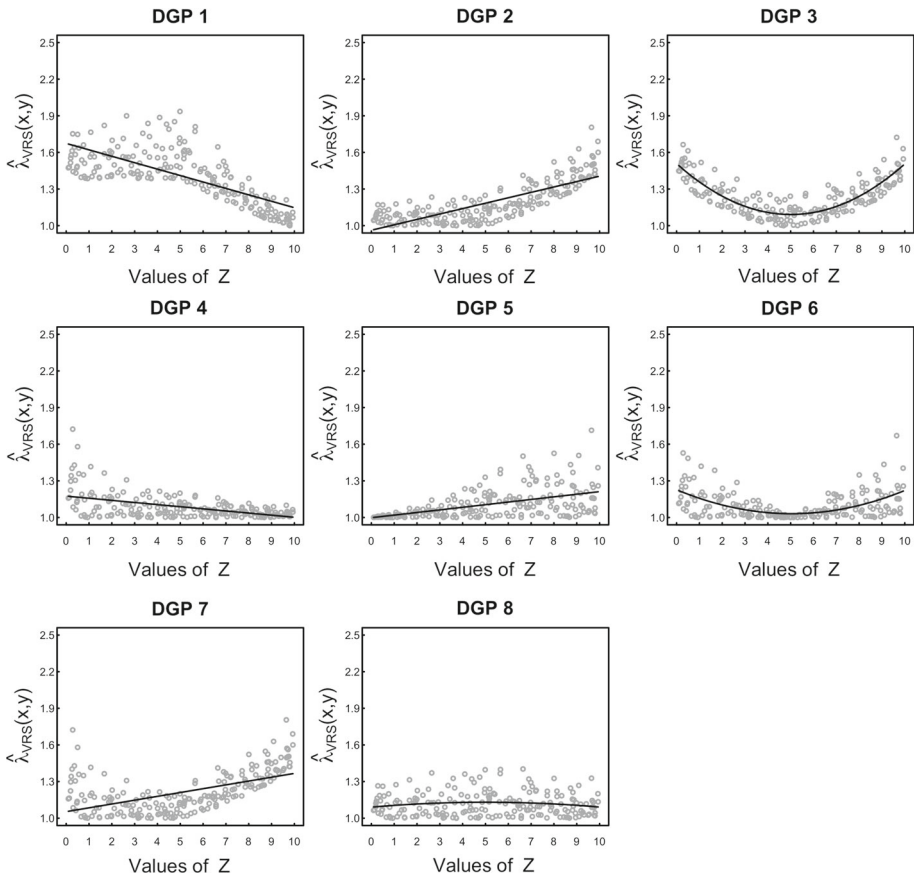


Fig. 5 The scatterplots and the Tobit regression of the efficiency scores $\hat{\lambda}_{VRS}(x, y)$ on Z

recommend using a Tobit regression in the second stage analysis; however, we apply this model for illustration purposes following the majority of the reviewed studies (e.g., [Bates et al. 2006](#); [Kooreman 1994](#)). In DGP 3 and DGP 6, we add a quadratic term to account for the non-linear influence of Z . The scatterplot of Z against $\hat{\lambda}_{VRS}(x, y)$ and the fitted regression lines are shown in Fig. 5. The fitted lines roughly correspond to the simulated direction of the relationship between quality and efficiency for all DGPs except DGP 7. However, in DGP 1 and DGP2, the parametrically fitted line does not identify the area of non-influence of Z on Y for $Z \leq 5$. In DGP 7, the complex underlying effect of Z both on the shift of the frontier and the distribution of inefficiencies is not captured correctly. Note that [Simar and Wilson \(2011\)](#) recommend using a two-stage approach only in case when the separability condition is verified (DGP 4–DGP 6). In order to make valid inference, they recommend using a truncated regression of the bootstrapped efficiency estimates.

5.4 Conditional approach

To obtain the values of managerial efficiency, $\hat{\varepsilon}$, whitened from the effect of Z , we use a two-step procedure described in Sect. 3 and derive $\hat{\varepsilon}$ from Eq. (3.11). Higher values of the

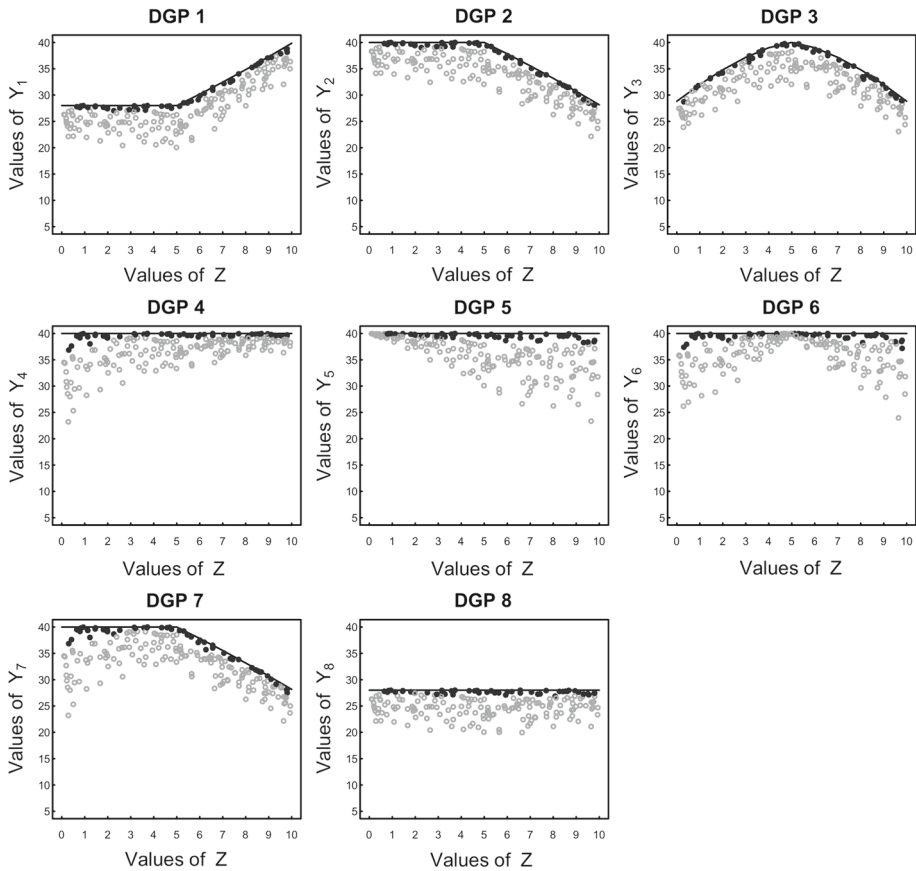


Fig. 6 Managerial efficiencies obtained using the conditional approach. *Black dots* represent the most efficient observations, $\hat{\varepsilon} \leq Q_1$. The *solid line* denotes the true attainable frontier. In DGP 1–DGP 3, Z shifts the attainable frontier. In DGP 4–DGP 6, Z affects the distribution of the inefficiencies. In DGP 7, Z affects both the distribution of the inefficiencies and the attainable frontier. In DGP 8, Z has no effect on the production process

residual $\hat{\varepsilon}$ represent higher inefficiency of the production units. Figure 6 displays the estimates of managerial efficiencies for the eight DGPs. The estimates of most efficient observations correspond closely to the true most efficient units in Fig. 1. The Pearson correlations between U and $\hat{\varepsilon}$ are equal to 0.98, 0.98, and 0.97 for DGP 1, DGP 2, and DGP 3, respectively. The corresponding Spearman correlations are 0.98, 0.98, and 0.97. The corresponding MSE are equal to 0.03, 0.03, and 0.07. High correlations with true values and low values of MSE confirm that the effect of Z on the attainable frontier has been effectively whitened from the estimates of managerial efficiency.

In the middle panel, the effect of Z is on the distribution of the inefficiencies. The Pearson correlations between U and managerial efficiencies are equal to 0.97, 0.97, and 0.86 for DGP 4, DGP 5, and DGP 6, respectively. The corresponding Spearman correlations are 0.98, 0.97, and 0.87. The MSE are equal to 0.05, 0.05, and 0.31 for DGP 4, DGP 5, and DGP 6, respectively. Therefore, the effect of Z on the distribution of inefficiencies has also been whitened from the estimates of managerial efficiency.

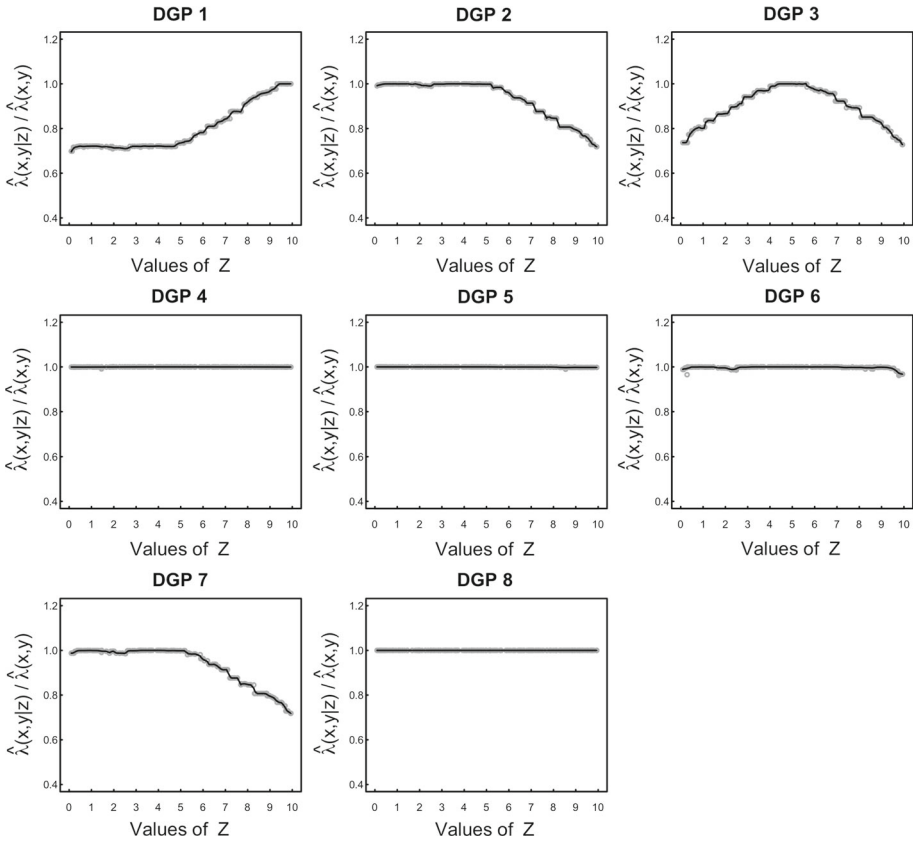


Fig. 7 The scatterplots and smoothed nonparametric regression of the ratios $\hat{R}(x, y|z) = \hat{\lambda}(x, y|z) / \hat{\lambda}(x, y)$ on Z

In the lower panel, the results for DPG 7 and DPG 8 are displayed. The Pearson correlations for these DGPs equal 0.97 and 0.98, whereas the Spearman correlations equal 0.98 and 0.99. The MSE are equal to 0.07 and 0.03 for DPG 7 and DPG 8, respectively. Thus, in DPG 7, the effects of Z on the shift of the frontier and the distribution of the inefficiencies have been correctly accounted for. In DPG 8, the fact that Z is unrelated to the production process has also been correctly identified.

The conditional analysis allows differentiating between two possible effects of Z on the production process. First, we uncover the effect of Z on the shift in the attainable frontier by nonparametrically regressing the ratios of conditional to unconditional estimates $R(x, y|z) = \hat{\lambda}(x, y|z) / \hat{\lambda}(x, y)$ on Z (Fig. 7). Second, the effect of Z on the distribution of inefficiencies is analyzed by nonparametrically regressing the conditional estimates $\hat{\lambda}(x, y|z)$ on Z (Fig. 8).

The upper panel of Fig. 7 presents the scatterplot of the ratios $\hat{R}(x, y|z)$ against Z and the nonparametrically fitted regression line. Remember that in the output-oriented framework, an increasing regression line corresponds to a positive influence of Z on the production process, whereas a decreasing regression line corresponds to a negative influence of Z on the production process. The displayed results suggest that in DGP 1 there is a positive effect of Z on the attainable frontier for $Z > 5$, which exactly corresponds to the simulated relationship. In DGP 2, there is a negative effect of Z on the attainable frontier for $Z > 5$, which again

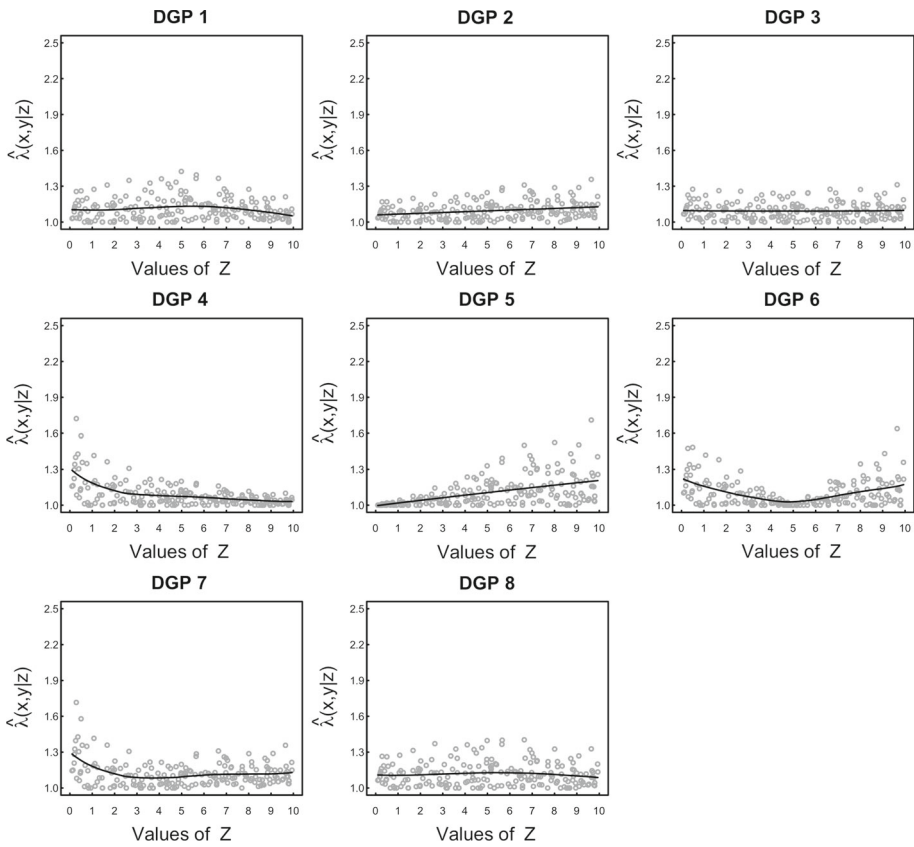


Fig. 8 The scatterplots and smoothed nonparametric regression of the conditional efficiencies $\hat{\lambda}(x, y|z)$ on Z

confirms the simulated relationship. The area of non-influence of Z on Y for $Z \leq 5$ is also clearly discernible for both DGP1 and DGP 2. In DGP 3, an inverted U -shaped effect of Z on the attainable frontier is exactly as we expect from the simulated data. The fitted regression lines of $\hat{R}(x, y|z)$ on Z for DGP 4–DGP 6 presented in the middle panel are flat, because Z does not affect the attainable frontier in these DGPs. The results for DGP 7 and DGP 8 are presented in the lower panel. In DGP 7, there is a negative effect of Z on the frontier for $Z > 5$. In DGP 8, the fitted regression line is flat, because Z has no effect on the production process.

Figure 8 presents the scatterplot of $\hat{\lambda}(x, y|z)$ against Z and the nonparametrically fitted regression line. Recall that higher values of output-oriented efficiency estimates, $\hat{\lambda}(x, y|z)$, represent higher inefficiency. Therefore, an increasing regression line in this case corresponds to a negative influence of Z on efficiency, whereas a decreasing regression line corresponds to a positive influence of Z on efficiency. The fitted lines in DGP 1–DGP 3 are flat because Z does not influence the distribution of inefficiencies in the first three DGPs. The fitted lines in DGP 4–DGP 6 correspond to the simulated relationship between quality and efficiency. The fitted line for DGP 7 correctly displays a positive effect of Z on the distribution of the inefficiencies for $Z \leq 5$. In DGP 8, the scatterplot confirms the fact that Z has no effect on the production process.

5.5 Comparison of the four methods

To compare the four methods, we first examine how precisely they capture the true efficiency estimates. Table 2 provides the summary of the mean Pearson and Spearman correlations between the inefficiency estimates and the true values of inefficiency, U , and the mean of the MSE based on 100 Monte Carlo simulations of the corresponding DGPs. The managerial efficiency estimates derived using the conditional approach provide the strongest prediction of the true values of efficiency in all eight DGPs (mean Pearson correlations are in the range 0.91–0.98, mean Spearman correlations are in the range 0.92–0.98, and mean MSE are in the range 0.04–0.18). The performance of the other three methods, namely, the one-stage approach, the congestion analysis, and the two-stage approach, depends on the compatibility of the underlying assumptions with the generated production sets and on the mechanism of how quality influences efficiency. In situations, when quality influences the attainable frontier (DGP 1–DGP 3), the performance of the three popular methods is highly dependent on the assumptions on the augmented production set, such as strong disposability, monotonicity, and convexity. For instance, the non-convexity of the augmented production set in DGP 1 leads to poor predictions of the true inefficiency when the one-stage approach or the congestion analysis is used. In the situations, when quality influences distribution of inefficiencies inside the production set (DGP 4 – DGP 6) the three approaches perform rather poorly, because none of the three approaches takes the effect of Z on the distribution of inefficiencies into account. For DGP 7, the performance of the one-stage approach and the two-stage approach is quite poor, because these approaches do not accommodate for the effects both on the distribution of the inefficiencies and the shift of the frontier. The congestion analysis performs quite well in this situation, because it captures the effect on the shift of the frontier; however, it does not account for the effect on the distribution of the inefficiencies. For DGP 8, the results of the two-stage approach are the best, because it correctly does not include the external variable in the estimation.

Next, we examine the ability of the four methods to identify the underlying relationship between efficiency and quality. The four methods differ considerably in their approach to make inferences about this relationship. In the one-stage approach, comparing the efficiency scores obtained without considering quality, $\hat{\lambda}_{VRS}(x, y)$, and the efficiency scores including quality measures, $\hat{\lambda}_{VRS}(x, y, \tilde{z})$, is problematic because adding an additional dimension

Table 2 Comparison of the four methods

Method	One-stage approach			Congestion analysis			Two-stage approach			Conditional approach		
	P	S	MSE	P	S	MSE	P	S	MSE	P	S	MSE
DGP 1	0.47	0.44	1.06	0.65	0.59	0.70	0.54	0.49	0.92	0.97	0.96	0.06
DGP 2	0.07	0.09	1.85	0.91	0.92	0.17	0.47	0.50	1.05	0.97	0.96	0.06
DGP 3	0.19	0.28	1.61	0.85	0.90	0.29	0.57	0.53	0.86	0.96	0.95	0.07
DGP 4	0.63	0.66	0.74	0.78	0.87	0.45	0.79	0.92	0.41	0.98	0.98	0.05
DGP 5	0.31	0.38	1.38	0.75	0.78	2.02	0.74	0.83	0.51	0.98	0.98	0.05
DGP 6	0.36	0.46	1.26	0.67	0.72	0.65	0.69	0.73	0.62	0.91	0.92	0.18
DGP 7	0.08	0.10	1.84	0.80	0.89	0.40	0.52	0.53	0.95	0.96	0.96	0.08
DGP 8	0.39	0.46	1.21	0.90	0.90	0.21	0.99	1.00	0.02	0.98	0.98	0.04

Average results using 100 Monte Carlo simulations ($n = 200$)

DGP data generating process, MSE mean standard error, P Pearson correlation, S Spearman correlation

always results in an increase in average efficiency. The congestion analysis compares the efficiency estimates, which include the lack of quality as a “bad” output, obtained under the assumption of strong disposability, $\hat{\lambda}_{VRS}(x, y, z)$, with the estimates obtained under weak disposability, $\hat{\lambda}_{WEAK}(x, y, z)$. The ratio of the two estimates provides the congestion score, which signals the negative relationship between quality and efficiency. In our simulation exercise, the congestion approach correctly identifies the negative relationship between efficiency and quality when quality causes the shift of the attainable frontier, but it does not identify this relationship in the DGPs, where the influence of quality is on the distribution of inefficiencies. Moreover, the congestion analysis does not identify the inverted *U*-shaped relationship between quality and efficiency. Both the two-stage approach and the conditional approach rely on regressions to make an inference about the relationship between efficiency and quality. In our simulation exercise, both approaches correctly identify the direction of the association between efficiency and quality. However, in contrast to the two-stage approach, the conditional approach allows differentiating between the influence of quality on the attainable efficient frontier from its influence on the distribution of inefficiencies inside the production set. Additionally, because the conditional approach makes use of a nonparametric regression, it provides a better fit in case of non-linear relationships than parametric regressions without the need to test additional model specifications (such as a quadratic function).

6 Discussion

While the earliest analyses of efficiency in the service sector have been mostly concerned with comparing input to output quantities, subsequent studies have tried to integrate output quality using various methods. Our objective was to compare the available nonparametric methods that allow integrating quality into the analysis of efficiency and to provide guidance on the strengths and limitations on the existing approaches. The four compared methods include three commonly used methods, namely, the one-stage approach, the congestion analysis, and the two-stage approach, and one innovative method based on new developments in the nonparametric methodology—the conditional approach. To compare these methods, we simulated data in eight scenarios that reflect potential relationships between efficiency and quality and evaluated the performance of the compared methods in terms of their ability to estimate the true values of efficiency and to identify the underlying relationship between efficiency and quality. The simulation exercise revealed that the conditional approach outperformed the three popular methods in terms of both predicting the true efficiency scores and capturing the original relationship between efficiency and quality.

The theoretical relationship between efficiency and quality is ambiguous in two aspects: the direction of the relationship between efficiency and quality and the mechanism by which quality enters the production process. Identifying the direction of the relationship between efficiency and quality is often pursued in the empirical applications; however, some methods implicitly require that this relationship is known a priori; for instance, in the one-stage approach, quality should be added to the inputs if the relationship is positive but it should be added to the outputs if the relationship is negative. Failing to identify this relationship and the inappropriate use of the quality measure in the production model leads to a major bias in the estimated efficiencies. The aspect of recognizing the mechanism by which quality enters the production process—that is, by shifting the attainable frontier or by changing the distribution of inefficiencies within the production set—is imperative to identify the appropriateness of using a particular method. Thus, the two-stage approach should not be used if quality shifts the attainable frontier, whereas the one-stage approach and the congestion analysis should

not be used if quality only changes the distribution of inefficiencies within the production set. Moreover, the analyst needs to keep in mind that the methods that augment the production set by quality measures, such as the one-stage approach and the congestion analysis, require imposing additional assumptions on the measures of quality, such as free disposability (or monotonicity), convexity, and returns to scale, that may be quite restrictive in real-life applications. Furthermore, in situations when a quality measure is included inappropriately, its inclusion will affect efficiency estimates for units with extreme values of this measure. As a consequence, in the congestion analysis, production units with extreme values of the “bad” output may be identified as fully efficient.

The conditional approach overcomes the limitations of the conventional methods, because it does not require making a priori assumptions about the relationship between quality and efficiency and it accounts for the potential influence of quality both on the attainable frontier and on the distribution of inefficiencies. The conditional efficiency estimates are obtained by benchmarking operational units with similar quality levels and thereby account for the fact that quality may shift the attainable frontier. The residuals from the regression of conditional efficiency measures on quality account for the influence of quality on the distribution of inefficiency and represent the value of the managerial efficiency, which can be used to compare and rank organizations. A recent empirical application of the conditional approach has shown that depending on the utilized quality measure different effects on the production function can be observed (Varabyova et al. 2016). In the study by Varabyova et al. (2016) analyzing the efficiency of cardiology departments, patient satisfaction had no effect on the attainable frontier but affected the distribution of the inefficiencies, whereas mortality rate did not affect the distribution of the inefficiencies but was instead responsible for the shift in the attainable frontier. Further empirical applications of the conditional approach have confirmed the possibility that an external variable can have an effect on either the shift of the frontier or the distribution of the inefficiencies (Bădin et al. 2014).

Identifying the underlying mechanism whereby quality enters the production process amounts to the need of testing the separability condition between the measure of quality and the set of input and output factors. The formal test of the separability condition was described by Daraio et al. (2015). Moreover, Bădin et al. (2012) showed how to use partial frontier analysis to disentangle the influence of the external factor on the attainable frontier from its influence on the distribution of the inefficiencies. The use partial frontier measures of order- m or order- α provides an additional benefit of being robust to extreme values, which are not uncommon to empirical data. Partial frontier measures also benefit from better convergence rates of the estimators than full frontier measures. In our analysis, we used full frontier measures to keep the analysis simple. Because we do not allow for random variation in the simulated data, the current results would be very similar if we used partial frontier measures. However, empirical applications should rely on partial frontier measures to derive efficiency measures robust to outliers and random variation.

Future empirical applications should provide graphical illustrations of their results, which enhances transparency and provides an indication to the reader that the results are robust and not an artifact of chance. In the two-stage analysis, the relationship between efficiency and quality should be plotted in order to verify whether this relationship is linear and whether the utilized model specification is appropriate. In the methods relying on augmented production set, the graphical illustration needs to encompass inputs, outputs, and quality measures. Bădin et al. (2014) demonstrated a fully graphical solution to three-dimensional and four-dimensional problems. However, because many empirical studies rely on more than four dimensions, the graphical examination of the results may become complicated. To reduce the number of dimensions, inputs and outputs could be aggregated without much loss of

information using the procedure described in [Daraio and Simar \(2007\)](#). The aggregation of inputs, outputs, and quality measures is also useful for avoiding the convergence problems of most nonparametric methods, known as the curse of dimensionality. In fact, a number of reviewed studies considered multiple measures of quality (up to 41). Because the studies relying on the one-stage approach or the congestion analysis integrate quality measures directly into the production model, an increase in the number of model dimensions may lead to low rates of convergence of the estimator (i.e., curse of dimensionality). Moreover, the assumptions made on the production set augmented by multiple quality measures apply for each additional quality measure and may thus affect the results in an unpredictable manner. The development of appropriate quality measures should be targeted in future research in regard to different service sectors.

7 Conclusions

External factors may affect an organizational production process by either shifting the attainable frontier or changing the distribution of inefficiencies within the attainable set, or both. In particular, the need to reach multiple and at times conflicting objectives, such as providing high quantity but also high quality of services, necessitates an extension of classical efficiency models. There has been a rapid expansion in empirical studies that attempted to integrate quality into the nonparametric analysis of efficiency. However, little work has been done so far to evaluate and compare the nonparametric methods that allow the integration of quality. Hence, the user faces the challenge of selecting the most appropriate method from among several feasible alternatives.

The present study provides an overview of existing methods to integrate quality into the analysis of efficiency in various service sectors. We show an application of three popular methods, namely, the one-stage approach, the congestion analysis, and the two-stage approach. Due to the fact that these methods require making very restrictive assumptions about the production process, their usefulness is quite limited when there is uncertainty about the underlying relationship between efficiency and quality. However, the use of conventional methods is justified in situations, when the theoretical relationships are straightforward and the underlying assumptions are verified.

We further demonstrate the benefits of using the conditional approach to integrate quality in the analysis of health care efficiency in a flexible and non-restrictive way. The comparison of the four methods in a simple two-dimensional framework shows the clear dominance of the conditional approach over the three conventional methods in terms of both predicting the true values of efficiency and examining the influence of quality on the production process. The conditional approach allows differentiating between the influence of quality on the attainable frontier and the influence of quality on the distribution of inefficiencies within the production set.

We believe that considerable progress in our understanding of the relationship between efficiency and quality in service sectors can occur if scholars apply methods that allow testing the underlying assumptions and adjust their estimation strategy accordingly. While the nonparametric literature witnesses a growth in the computationally complex methodological advances, their added value may be diminished to the extent that the basic assumptions are violated. The conditional approach represents a powerful and simple tool that can be used when the theoretical link between efficiency and quality is unclear. Future research should elaborate on the theory of the relationship between efficiency and quality in the service-producing sectors. A wide variety of different measures to account for quality need to be

classified and linked to a sound theory. Finally, empirical applications based on nonrestrictive models should deliver further evidence about the relationship between the different measures of quality and efficiency.

References

- Anderson, R., Weeks, H., Hobbs, B., & Webb, J. (2003). Nursing home quality, chain affiliation, profit status and performance. *Journal of Real Estate Research*, 25(1), 43–60.
- Bădin, L., Daraio, C., & Simar, L. (2010). Optimal bandwidth selection for conditional efficiency measures: A data-driven approach. *European Journal of Operational Research*, 201(2), 633–640.
- Bădin, L., Daraio, C., & Simar, L. (2012). How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research*, 223(3), 818–833.
- Bădin, L., Daraio, C., & Simar, L. (2014). Explaining inefficiency in nonparametric production models: The state of the art. *Annals of Operations Research*, 214(1), 5–30.
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30, 1078–1092.
- Banker, R. D., & Morey, R. C. (1986). Efficiency analysis for exogenously fixed inputs and outputs. *Operations Research*, 34, 513–521.
- Bates, L. J., Mukherjee, K., & Santerre, R. E. (2006). Market structure and technical efficiency in the hospital services industry: A DEA approach. *Medical Care Research and Review*, 63(4), 499–524.
- Becker, J., Beverungen, D., Breuker, D., Dietrich, H.-A., Knackstedt, R., & Rauer, H. P. (2011). How to model service productivity for data envelopment analysis? A meta-design approach. In *European conference on information systems (ECIS) 2011 proceedings*.
- Berger, A. N., & DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks. *Journal of Banking & Finance*, 21(6), 849–870.
- Bilsel, M., & Davutyan, N. (2014). Hospital efficiency with risk adjusted mortality as undesirable output: The Turkish case. *Annals of Operations Research*, 221(1), 73–88.
- Cambini, C., Croce, A., & Fumagalli, E. (2014). Output-based incentive regulation in electricity distribution: Evidence from Italy. *Energy Economics*, 45, 205–216.
- Cazals, C., Florens, J.-P., & Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of econometrics*, 106(1), 1–25.
- Chang, S.-J., Hsiao, H.-C., Huang, L.-H., & Chang, H. (2011). Taiwan quality indicator project and hospital productivity growth. *Omega*, 39(1), 14–22.
- Chen, C.-M. (2013). A critique of non-parametric efficiency analysis in energy economics studies. *Energy Economics*, 38, 146–152.
- Clement, J. P., Valdmanis, V. G., Bazzoli, G. J., Zhao, M., & Chukmaitov, A. (2008). Is more better? An analysis of hospital outcomes and efficiency with a DEA model of output congestion. *Health Care Management Science*, 11(1), 67–77.
- 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). *Advanced robust and nonparametric methods in efficiency analysis [electronic resource]: Methodology and applications* (Vol. 4). Berlin: Springer.
- Daraio, C., Simar, L., & Wilson, P. W. (2015). Testing the “separability” condition in two-stage nonparametric models of production. LEM working paper series.
- Das, A., & Ghosh, S. (2006). Financial deregulation and efficiency: An empirical analysis of Indian banks during the post reform period. *Review of Financial Economics*, 15(3), 193–221.
- Davutyan, N., Bilsel, M., & Tarcan, M. (2014). Migration, risk-adjusted mortality, varieties of congestion and patient satisfaction in Turkish provincial general hospitals. *The Data Envelopment Analysis Journal*, 1(2), 135–169.
- Dismuke, C. E., & Sena, V. (2001). Is there a trade-off between quality and productivity? The case of diagnostic technologies in Portugal. *Annals of Operations Research*, 107(1–4), 101–116.
- Donabedian, A. (1966). Evaluating the quality of medical care. *The Milbank Memorial Fund Quarterly*, 44, 66–206.
- Drake, L., & Hall, M. J. (2003). Efficiency in Japanese banking: An empirical analysis. *Journal of Banking & Finance*, 27(5), 891–917.
- Drake, L., Hall, M. J., & Simper, R. (2006). The impact of macroeconomic and regulatory factors on bank efficiency: A non-parametric analysis of Hong Kong’s banking system. *Journal of Banking & Finance*, 30(5), 1443–1466.

- Du, J., Wang, J., Chen, Y., Chou, S.-Y., & Zhu, J. (2014). Incorporating health outcomes in Pennsylvania hospital efficiency: An additive super-efficiency DEA approach. *Annals of Operations Research*, 221(1), 161–172.
- Färe, R., & Grosskopf, S. (1983). Measuring output efficiency. *European Journal of Operational Research*, 13(2), 173–179.
- Färe, R., Grosskopf, S., & Lovell, C. K. (1994). *Production frontiers*. Cambridge: Cambridge University Press.
- Ferrier, G. D., & Trivitt, J. S. (2013). Incorporating quality into the measurement of hospital efficiency: A double DEA approach. *Journal of Productivity Analysis*, 40(3), 337–355.
- Ferrier, G. D., & Valdmanis, V. (1996). Rural hospital performance and its correlates. *Journal of Productivity Analysis*, 7(1), 63–80.
- Garavaglia, G., Lettieri, E., Agasisti, T., & Lopez, S. (2011). Efficiency and quality of care in nursing homes: An Italian case study. *Health Care Management Science*, 14(1), 22–35.
- Giannakis, D., Jamasb, T., & Pollitt, M. (2005). Benchmarking and incentive regulation of quality of service: An application to the UK electricity distribution networks. *Energy Policy*, 33(17), 2256–2271.
- Gok, M. S., & Sezen, B. (2013). Analyzing the ambiguous relationship between efficiency, quality and patient satisfaction in healthcare services: The case of public hospitals in Turkey. *Health Policy*, 111(3), 290–300.
- Growitsch, C., Jamasb, T., Müller, C., & Wissner, M. (2010). Social cost-efficient service quality—integrating customer valuation in incentive regulation: Evidence from the case of Norway. *Energy Policy*, 38(5), 2536–2544.
- Guccio, C., Martorana, M. F., & Monaco, L. (2016). Evaluating the impact of the Bologna process on the efficiency convergence of Italian universities: A non-parametric frontier approach. *Journal of Productivity Analysis*, 45(3), 275–298.
- Isik, I., & Hassan, M. K. (2003). Efficiency, ownership and market structure, corporate control and governance in the Turkish banking industry. *Journal of Business Finance & Accounting*, 30(9–10), 1363–1421.
- Jablonsky, J. (2016). Efficiency analysis in multi-period systems: An application to performance evaluation in Czech higher education. *Central European Journal of Operations Research*, 24(2), 283–296.
- Jamasb, T., & Pollitt, M. (2003). International benchmarking and regulation: An application to European electricity distribution utilities. *Energy Policy*, 31(15), 1609–1622.
- Kamakura, W. A., Mittal, V., De Rosa, F., & Mazzon, J. A. (2002). Assessing the service-profit chain. *Marketing Science*, 21(3), 294–317.
- Kenjegalieva, K., Simper, R., Weyman-Jones, T., & Zelenyuk, V. (2009). Comparative analysis of banking production frameworks in Eastern European financial markets. *European Journal of Operational Research*, 198(1), 326–340.
- Kooreman, P. (1994). Nursing home care in The Netherlands: A nonparametric efficiency analysis. *Journal of Health Economics*, 13(3), 301–316.
- Kyj, L., & Isik, I. (2008). Bank α -efficiency in Ukraine: An analysis of service characteristics and ownership. *Journal of Economics and Business*, 60(4), 369–393.
- Laine, J., Finne-Soveri, U. H., Björkgren, M., Linna, M., Noro, A., & Häkkinen, U. (2005). The association between quality of care and technical efficiency in long-term care. *International Journal for Quality in Health Care*, 17(3), 259–267.
- Lee, R. H., Bott, M. J., Gajewski, B., & Taunton, R. L. (2009). Modeling efficiency at the process level: An examination of the care planning process in nursing homes. *Health Services Research*, 44(1), 15–32.
- Lenard, M. L., & Shimshak, D. G. (2009). Benchmarking nursing home performance at the state level. *Health Services Management Research*, 22(2), 51–61.
- lo Storto, C. (2016). The trade-off between cost efficiency and public service quality: A non-parametric frontier analysis of Italian major municipalities. *Cities*, 51, 52–63.
- Mark, B. A., Jones, C. B., Lindley, L., & Ozcan, Y. A. (2009). An examination of technical efficiency, quality, and patient safety in acute care nursing units. *Policy, Politics, & Nursing Practice*, 10(3), 180–186.
- Matranga, D., & Sapienza, F. (2015). Congestion analysis to evaluate the efficiency and appropriateness of hospitals in Sicily. *Health Policy*, 119(3), 324–332.
- McMillan, M. L., & Chan, W. H. (2006). University efficiency: A comparison and consolidation of results from stochastic and non-stochastic methods. *Education Economics*, 14(1), 1–30.
- Miguéis, V. L., Camanho, A. S., Bjørndal, E., & Bjørndal, M. (2012). Productivity change and innovation in Norwegian electricity distribution companies. *Journal of the Operational Research Society*, 63(7), 982–990.
- Mutter, R., Valdmanis, V., & Rosko, M. (2010). High versus lower quality hospitals: A comparison of environmental characteristics and technical efficiency. *Health Services and Outcomes Research Methodology*, 10(3–4), 134–153.
- Navarro-Espigares, J. L., & Torres, E. H. (2011). Efficiency and quality in health services: A crucial link. *The Service Industries Journal*, 31(3), 385–403.

- Nayar, P., & Ozcan, Y. A. (2008). Data envelopment analysis comparison of hospital efficiency and quality. *Journal of Medical Systems*, 32(3), 193–199.
- Newhouse, J. P. (1970). Toward a theory of nonprofit institutions: An economic model of a hospital. *The American Economic Review*, 60(1), 64–74.
- Nyman, J. A., & Bricker, D. L. (1989). Profit incentives and technical efficiency in the production of nursing home care. *The Review of Economics and Statistics*, 71(4), 586–594.
- Olesen, O., & Petersen, N. (1995). Incorporating quality into data envelopment analysis: A stochastic dominance approach. *International Journal of Production Economics*, 39(1), 117–135.
- Özkan-Günay, E. N., Günay, Z. N., & Günay, G. (2013). The impact of regulatory policies on risk taking and scale efficiency of commercial banks in an emerging banking sector. *Emerging Markets Finance and Trade*, 49(sup5), 80–98.
- Park, K. H., & Weber, W. L. (2006). A note on efficiency and productivity growth in the Korean banking industry, 1992–2002. *Journal of Banking & Finance*, 30(8), 2371–2386.
- Poldrugovac, K., Tekavcic, M., & Jankovic, S. (2016). Efficiency in the hotel industry: An empirical examination of the most influential factors. *Economic Research-Ekonomska Istraživanja*, 29(1), 583–597.
- Prior, D. (2006). Efficiency and total quality management in health care organizations: A dynamic frontier approach. *Annals of Operations Research*, 145(1), 281–299.
- Ray, S. C., & Mukherjee, K. (1998). Quantity, quality, and efficiency for a partially super-additive cost function: Connecticut public schools revisited. *Journal of Productivity Analysis*, 10(1), 47–62.
- Sherman, H. D., & Zhu, J. (2006). Benchmarking with quality-adjusted DEA (Q-DEA) to seek lower-cost high-quality service: Evidence from a US bank application. *Annals of Operations Research*, 145(1), 301–319.
- Shimshak, D. G., Lenard, M. L., & Klimberg, R. K. (2009). Incorporating quality into data envelopment analysis of nursing home performance: A case study. *Omega*, 37(3), 672–685.
- 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). Two-stage DEA: Caveat emptor. *Journal of Productivity Analysis*, 36(2), 205–218.
- Simar, L., & Wilson, P. W. (2015). Statistical approaches for non-parametric frontier models: A guided tour. *International Statistical Review*, 83(1), 77–110.
- Thanassoulis, E., Boussofiene, A., & Dyson, R. (1995). Exploring output quality targets in the provision of perinatal care in England using data envelopment analysis. *European Journal of Operational Research*, 80(3), 588–607.
- Tiemann, O., & Schreyögg, J. (2009). Effects of ownership on hospital efficiency in Germany. *Business Research*, 2(2), 115–145.
- Valdmanis, V. G., Rosko, M. D., & Mutter, R. L. (2008). Hospital quality, efficiency, and input slack differentials. *Health Services Research*, 43(5p2), 1830–1848. doi:10.1111/j.1475-6773.2008.00893.x.
- van Ineveld, M., van Oostrum, J., Vermeulen, R., Steenhoek, A., & van de Klundert, J. (2015). Productivity and quality of Dutch hospitals during system reform. *Health Care Management Science*, 19, 1–12.
- Varabyova, Y., Blankart, C. R., & Schreyögg, J. (2016). Using nonparametric conditional approach to integrate quality into efficiency analysis: Empirical evidence from cardiology departments. *Health Care Management Science*, doi:10.1007/s10729-016-9372-4.
- Varabyova, Y., & Schreyögg, J. (2013). International comparisons of the technical efficiency of the hospital sector: Panel data analysis of OECD countries using parametric and non-parametric approaches. *Health Policy*, 112(1), 70–79.
- Von Hirschhausen, C., Cullmann, A., & Kappeler, A. (2006). Efficiency analysis of German electricity distribution utilities-non-parametric and parametric tests. *Applied Economics*, 38(21), 2553–2566.
- Wasserman, L. (2006). *All of nonparametric statistics*. Berlin: Springer.
- Wu, C.-H., Chang, C.-C., Chen, P.-C., & Kuo, K.-N. (2013). Efficiency and productivity change in Taiwan's hospitals: A non-radial quality-adjusted measurement. *Central European Journal of Operations Research*, 21(2), 431–453.
- Yang, J., & Zeng, W. (2014). The trade-offs between efficiency and quality in the hospital production: Some evidence from Shenzhen, China. *China Economic Review*, 31, 166–184.
- Yu, W., Jamasb, T., & Pollitt, M. (2009). Willingness-to-pay for quality of service: an application to efficiency analysis of the UK electricity distribution utilities. *The Energy Journal*, 30(4), 1–48.
- Zhang, N. J., Unruh, L., & Wan, T. T. (2008). Has the Medicare prospective payment system led to increased nursing home efficiency? *Health Services Research*, 43(3), 1043–1061.