

Ahn, Heinz; Clermont, Marcel; Langner, Julia

Article — Published Version

Corrected normalized additive analysis as alternative method for easy measurement of efficiency

Journal of Business Economics

Provided in Cooperation with:

Springer Nature

Suggested Citation: Ahn, Heinz; Clermont, Marcel; Langner, Julia (2025) : Corrected normalized additive analysis as alternative method for easy measurement of efficiency, Journal of Business Economics, ISSN 1861-8928, Springer, Berlin, Heidelberg, Vol. 95, Iss. 6, pp. 777-808, <https://doi.org/10.1007/s11573-024-01219-2>

This Version is available at:

<https://hdl.handle.net/10419/330611>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Corrected normalized additive analysis as alternative method for easy measurement of efficiency

Heinz Ahn¹ · Marcel Clermont² · Julia Langner¹

Accepted: 4 December 2025 / Published online: 6 March 2025
© The Author(s) 2025

Abstract

Various best-practice benchmarking methods for measuring the efficiency of decision making units are intensively discussed in the literature. Compared with such methods as Data Envelopment Analysis (DEA), the comparatively straightforward Normalized Additive Analysis (NAA) has received little attention so far. Nevertheless, a recent study revealed that NAA performs similarly to established methods regarding the accurate determination of the ranking of decision making units based on their efficiency scores. However, NAA systematically underestimates the true efficiency scores with respect to the underlying production system and the chosen Farrell efficiency. To counteract this underestimation, we present two modifications of NAA in the sense of a Corrected NAA (CNAA). We conduct a Monte Carlo simulation to analyze the performance of these modifications. For our simulation design, we show that these CNAA specifications estimate the efficiency scores more accurately than NAA while determining the ranking of decision making units at an equally good level. To further analyze the estimation quality of CNAA, we use the well-known basic DEA models with constant and variable returns to scale as benchmarks. It is shown that the performance of the CNAA specifications is largely comparable to that of the DEA models. Hence, CNAA can be considered to be a suitable method.

Keywords Best-practice benchmarking methods · Corrected normalized additive analysis · Data envelopment analysis · Monte Carlo simulation · Practical applicability

✉ Heinz Ahn
hw.ahn@tu-braunschweig.de

Marcel Clermont
marcel.clermont@dhge.de

Julia Langner
j.langner@tu-braunschweig.de

¹ Department of Business Sciences, Technische Universität Braunschweig, Fallersleber-Tor-Wall 23, 38100 Brunswick, Germany

² Cooperative University Gera-Eisenach, Am Wartenberg 2, 99817 Eisenach, Germany

JEL Classification C14 · C15 · L25

1 Introduction

There are numerous methods for measuring the relative efficiency of decision making units (DMUs) using sets of multi-dimensional performance indicators. Common methods based on production theory are Data Envelopment Analysis (DEA) developed by Charnes et al. (1978), as well as Stochastic Frontier Analysis (SFA) developed by Aigner et al. (1977) and Meeusen and Van den Broeck (1977). Various Monte Carlo simulations have demonstrated that, in numerous scenarios, these methods provide a suitable estimation of the efficiency scores (see, e.g., Kuosmanen and Kortelainen 2012, Schaefer and Clermont 2018, and Andor et al. 2019).

However, even though the application literature has steadily grown over the years, there is no evidence that these methods play a noteworthy role as management control tools outside the research context (see Kerpen 2016, p. 117). An evident reason is that users must have profound knowledge of production theory and mathematics to understand the methods' procedures. Therefore, it is reasonable to consider alternative methods that exhibit a similar estimation quality and are easier to understand. Although such alternatives may come along with some limitations compared with common methods like DEA, they have the potential to make multi-dimensional efficiency measurement more accessible to practitioners.

Several approaches are discussed in the literature to enhance the user-friendliness of multi-dimensional efficiency measurement. For example, Büschken (2009) proposed the "naïve efficiency measurement", wherein the average over all input–output ratios is calculated and subsequently normalized. Bornmann et al. (2017) presented the "excellence shift" approach to measure the ability of U.S. universities to produce highly cited publications; to do so, the shares of the respective university with respect to all publications, to the 10% most-cited publications, and to the total budget are put into relation. Furthermore, Ahn et al. (2007) introduced the Normalized Additive Analysis (NAA).¹ Based on single–input–single–output productivities, this method estimates single–input–single–output efficiency scores and aggregates them into an overall efficiency score.

Besides some recent applications of NAA (see, e.g., Ahn et al. 2022), there is also a Monte Carlo simulation which examined the estimation quality of NAA (see Ahn et al. 2023). This study distinguishes between two purposes of measuring efficiency: the accurate determination of the ranking of DMUs based on their efficiency scores and the accurate estimation of these scores. Regarding the former purpose, the authors found that in numerous scenarios, NAA performs similarly to other methods, such as DEA or SFA. However, they further showed that NAA systematically underestimates the true efficiency scores with respect to the underlying production system and the chosen Farrell efficiency (see Farrell 1957).

Against this background, the question arises of how NAA's methodology can be improved to achieve not only a suitable determination of the ranking but also

¹ This designation is based on Ahn et al. (2012).

a better estimation of the efficiency scores. We propose two respective modifications of NAA that are inspired by the Corrected Ordinary Least Squares method (COLS; Winsten 1957; Greene 1980). Following the designation of COLS, we summarize these modifications under the term Corrected NAA (CNAA). We run a Monte Carlo simulation to analyze the estimation quality of these modifications with respect to the aforementioned purposes for a variety of scenarios using appropriate performance criteria. In addition to comparing the performance of CNAA and NAA, our objective is also to analyze CNAA's suitability for measuring efficiency. To this end, we evaluate CNAA's performance using the basic DEA models with constant returns to scale (CRS) and variable returns to scale (VRS) as benchmarks. We selected these benchmarks because of their popularity and their connection to NAA (see Ahn et al. 2007, pp. 635–636).

With respect to our simulation design, we show that the CNAA specifications estimate the efficiency scores clearly better than NAA; at the same time, they maintain the performance in terms of accurately determining the ranking of DMUs. Moreover, it is shown that the performance of the CNAA specifications is largely comparable to that of the DEA models with respect to both purposes. In this sense, the CNAA specifications are suitable.

Our paper is structured as follows: Sect. 2 presents the NAA methodology and our two CNAA specifications. The simulation design, data generation process, and performance criteria used are described in Sect. 3. Section 4 presents and discusses the results. First, we evaluate the improvement in performance by CNAA compared with NAA. Thereafter, we evaluate the impact of changes in parameter specifications on the performance as well as the performance of CNAA using DEA as benchmark. While Sect. 5 discusses several assumptions of our paper, Sect. 6 summarizes the paper and provides an outlook on future research.

2 Corrected normalized additive analysis

First, we introduce NAA according to Ahn et al. (2007, pp. 632–633). Subsequently, we present two modifications. Following Ahn et al.'s (2023, p. 295) notation, we assume that n DMUs with m input factors and s output factors are given. Here, $x_{ij} > 0$ denotes the i th input, and $y_{rj} > 0$ denotes the r th output of DMU j . On this basis, NAA estimates the efficiency score of DMU j in three steps:

1. For every DMU j , input factor i , and output factor r , single-input–single-output productivities (IOPs) are determined as the ratio of the respective output to the input:

$$\text{IOP}_j^{ir} = \frac{y_{rj}}{x_{ij}} \quad \forall i, j, r. \quad (1)$$

2. Based on the highest IOP per input–output combination, the IOPs are normalized to estimate single-input–single-output efficiency scores (IOEs):

$$IOP_*^{ir} = \max \left\{ IOP_j^{ir} \mid j \in \{1, \dots, n\} \right\} \forall i, r, \quad (2)$$

$$\widehat{IOE}_j^{ir} = \frac{IOP_j^{ir}}{IOP_*^{ir}} \in [0, 1] \forall i, j, r. \quad (3)$$

3. For DMU j , the IOEs are aggregated into an overall (technical) efficiency score \widehat{TE}_j :

$$\widehat{TE}_j = \left(\frac{1}{m * s} \sum_{i=1}^m \sum_{r=1}^s \left(\widehat{IOE}_j^{ir} \right)^\omega \right)^{1/\omega}. \quad (4)$$

This generalized mean can be specified depending on the choice of the parameter ω , resulting in, e.g., the arithmetic mean ($\omega = 1$), geometric mean ($\omega \rightarrow 0$), maximum ($\omega \rightarrow +\infty$) or minimum ($\omega \rightarrow -\infty$) of the IOEs (see Ahn et al. 2007, p. 633). In the following, we focus on NAA with arithmetic mean, labeled here as NAA:

$$\widehat{TE}_j = \frac{1}{m * s} \sum_{i=1}^m \sum_{r=1}^s \widehat{IOE}_j^{ir}. \quad (5)$$

As mentioned in the introduction, Ahn et al. (2023) conducted a Monte Carlo simulation and found that NAA performs well in accurately determining the ranking of DMUs based on their efficiency scores. However, the efficiency scores estimated by NAA deviate strongly from the true efficiency scores since they are systematically underestimated. This is because the independent consideration of IOPs neglects the interdependencies between inputs and outputs; thus, the DMUs are not evaluated in terms of Farrell efficiency.²

To counteract this systematic underestimation, a modification of NAA is necessary. Inspired by COLS, we shift the efficiency scores \widehat{TE}_j of NAA upward, such that at least one DMU exhibits an efficiency score of 100%. This normalization corresponds to the concept of relative efficiency, whereby DMUs are evaluated in relation to each other. The normalization is carried out in two different ways:

- $CNAA^+$ (additive CNAA): absolute increase in NAA efficiency scores for all DMUs by the same amount;
- $CNAA^x$ (multiplicative CNAA): proportional increase in NAA efficiency scores for all DMUs by the same factor.

² As widely known, radial efficiency measures such as Farrell efficiency that is used by Ahn et al. (2023) and also in the present paper measure weak efficiency. However, there are also stronger definitions of efficiency, e.g., Pareto efficiency (see Cooper et al. 2007, pp. 45–46). In these cases, NAA may be more suitable since DEA^{CCR} and DEA^{BCC} may overestimate efficiency.

Neither modification changes the ranking of DMUs based on their efficiency scores, i.e., the corresponding performance of NAA with respect to determining the ranking is retained for CNAA^+ and CNAA^x .

Let us consider these two CNAA specifications in detail. First, NAA's highest efficiency score is determined in both cases by:

$$\text{TE}^* = \max_{j \in \{1, \dots, n\}} \{ \widehat{\text{TE}}_j \}. \quad (6)$$

Thereafter, the following transformation is conducted for CNAA^+ :

$$\widehat{\text{TE}}_j^+ = \widehat{\text{TE}}_j + (1 - \text{TE}^*) \quad \forall j. \quad (7)$$

This results in an efficiency score of 100% for DMU h with $\widehat{\text{TE}}_h = \text{TE}^* = \max_{j \in \{1, \dots, n\}} \{ \widehat{\text{TE}}_j \}$ because

$$\widehat{\text{TE}}_h^+ = \widehat{\text{TE}}_h + (1 - \text{TE}^*) = \text{TE}^* + (1 - \text{TE}^*) = 100\%.$$

By contrast, the transformation for CNAA^x is conducted as

$$\widehat{\text{TE}}_j^x = \widehat{\text{TE}}_j * \frac{1}{\text{TE}^*} \quad \forall j. \quad (8)$$

This again results in an efficiency score of 100% for DMU h with $\widehat{\text{TE}}_h = \text{TE}^* = \max_{j \in \{1, \dots, n\}} \{ \widehat{\text{TE}}_j \}$ because

$$\widehat{\text{TE}}_h^x = \widehat{\text{TE}}_h * 1/\text{TE}^* = \text{TE}^* * 1/\text{TE}^* = 100\%.$$

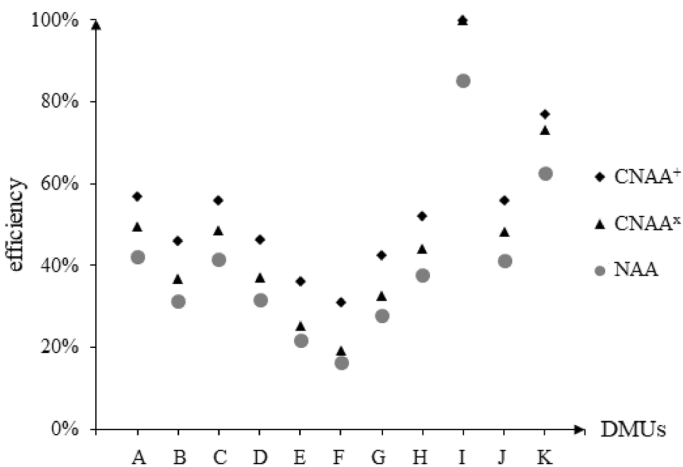
For illustration, we present the efficiency scores of NAA, CNAA^+ , and CNAA^x for an exemplary data set with eleven DMUs, two input factors, and two output factors. The underlying data and calculations are given in Table 1, whereby the respective highest IOPs and $\widehat{\text{TE}}$ are marked in bold. Figure 1 demonstrates the shift in NAA efficiency scores by CNAA^+ and CNAA^x compared with NAA. The gray circles represent the efficiency scores estimated by NAA. The black diamonds and triangles represent the efficiency scores estimated by the CNAA specifications. As can be seen, the respective efficiency scores of DMU I are set to 100% because this DMU exhibits the highest NAA efficiency score. For the other DMUs, the NAA efficiency scores are accordingly increased absolutely by the same amount (CNAA^+) or proportionally by the same factor (CNAA^x).

To conclude, let us look at the characteristics of CNAA. The CNAA specifications are deterministic because the discrepancy between the observed and the ideal productivity of a DMU is entirely interpreted as inefficiency, i.e., a stochastic noise term is not considered. Moreover, although the CNAA specifications do not require the stipulation of a form of the production function, they are less flexible

Table 1 Data and calculations for NAA, CNAA⁺, and CNAA^x for exemplary DMUs

	A	B	C	D	E	F	G	H	I	J	K
Input 1	1.00	3.00	1.25	2.50	4.00	5.00	3.50	1.25	0.75	0.50	2.50
Input 2	2.00	2.50	2.50	1.00	3.00	6.00	3.70	4.00	0.25	2.00	0.50
Output 1	1.00	4.00	3.50	3.00	5.00	6.50	6.25	5.00	3.00	1.50	6.00
Output 2	20.00	24.00	15.00	12.00	16.00	14.50	21.00	8.00	6.25	7.00	9.00
IOP ¹¹	1.00	1.33	2.80	1.20	1.25	1.30	1.79	4.00	4.00	3.00	2.40
IOP ²¹	0.50	1.60	1.40	3.00	1.67	1.08	1.69	1.25	12.00	0.75	12.00
IOP ¹²	20.00	8.00	12.00	4.80	4.00	2.90	6.00	6.40	8.33	14.00	3.60
IOP ²²	10.00	9.60	6.00	12.00	5.33	2.42	5.68	2.00	25.00	3.50	18.00
\widehat{IOE}^{11}	0.25	0.33	0.70	0.30	0.31	0.33	0.45	1.00	1.00	0.75	0.60
\widehat{IOE}^{21}	0.04	0.13	0.12	0.25	0.14	0.09	0.14	0.10	1.00	0.06	1.00
\widehat{IOE}^{12}	1.00	0.40	0.60	0.24	0.20	0.15	0.30	0.32	0.42	0.70	0.18
\widehat{IOE}^{22}	0.40	0.38	0.24	0.48	0.21	0.10	0.23	0.08	1.00	0.14	0.72
\widehat{TE}	0.42	0.31	0.41	0.32	0.22	0.16	0.28	0.38	0.85	0.41	0.63
\widehat{TE}^+	0.57	0.46	0.56	0.46	0.36	0.31	0.42	0.52	1.00	0.56	0.77
\widehat{TE}^x	0.50	0.37	0.48	0.37	0.25	0.19	0.33	0.44	1.00	0.48	0.73

The calculations were carried out using unrounded values in each step. The values in Table 1 may differ slightly, as they are rounded

**Fig. 1** Efficiency scores of CNAA⁺ and CNAA^x compared with those of NAA for exemplary DMUs

than non-parametric methods owing to their more rigid calculation method (see Ahn et al. 2023, p. 295). Furthermore, as the IOPs and IOEs are determined independently of size, CRS are implicitly assumed (see Ahn et al. 2012, p. 115). Therefore, as for DEA^{CCR}, it is not necessary to specify the (input/output) orientation.

3 Simulation design

A Monte Carlo simulation is a suitable tool for making a valid statement on the estimation quality of the CNAA specifications because it enables an analysis of the discrepancies between the simulated true efficiency scores and the efficiency scores estimated by a method (see, e.g., Oh and Shin 2015, Nieswand and Seifert 2018, and Cordero et al. 2020). As explained in the introduction, we compare the CNAA specifications with NAA and DEA. We include the basic DEA models with CRS (DEA^{CCR}; Charnes et al. 1978) and VRS (DEA^{BCC}; Banker et al. 1984) in their output-oriented version (for a comprehensive overview of basic DEA models, see Thanassoulis 2001; for a more recent overview and various further developments, see Emrouznejad et al. 2022).

Ahn et al. (2023) included NAA in a simulation study for the first time. We adopt their simulation design and argumentation structure to some extent, as far as this seems appropriate for our study. Their study aimed to investigate the suitability of different methods for measuring efficiency in numerous application-oriented scenarios. By contrast, we focus on the performance of the CNAA specifications, which were not addressed in the above-mentioned study.

To design our Monte Carlo simulation, we first specify certain parameters during the data generation process. Usually, different parameter specifications are selected and combined to analyze a large number of scenarios. In this context, the number of DMUs and the number of input and output factors must be determined. Furthermore, a distribution for generating the inputs has to be selected. Using these inputs, the corresponding efficient outputs can be generated by means of a production function. These outputs are modified by an error term consisting of an inefficiency term and a stochastic noise term. Finally, the efficiency measurement methods to be evaluated can be applied to the resulting input and output data.

To analyze the performance of CNAA, we focus on three aspects when selecting the parameter specifications. First, we vary parameters that are frequently varied in the literature and follow the corresponding parameter specifications. Second, we focus on those parameters for which a rather strong influence on the performance of methods for measuring efficiency has been found in the literature. Third, we select the parameter specifications such that they cover scenarios that are as realistic as possible.

Table 2 lists the selected parameter specifications, which are briefly explained below. The combination of these parameter specifications results in 720 scenarios. For each scenario, $R = 500$ repetitions are carried out to ensure sufficient validity. We used R for the implementation of the simulation design.

We consider 25, 50, 100, 200, and 400 DMUs to depict typical sizes (see, e.g., Banker et al. 1993, p. 337, Schaefer and Clermont 2018, p. 145, and Andor et al. 2019, p. 244).³ Regarding the number of input and output factors, the literature

³ In addition to the publications that are cited in the following, Verheyen et al. (2023) also conducted a comprehensive Monte Carlo simulation to evaluate the performance of methods for measuring efficiency. They based their simulation design largely on Schaefer and Clermont (2018) but assumed a different production system at which we look in more detail in the discussion section.

Table 2 Parametrizations for the data generation process

	Parameter	Parameter specifications
Data structure	# of DMUs	25; 50; 100; 200; 400
	# of input factors	3
	Input distribution in [5, 15]	Uniform; beta
	# of output factors	3
Production function	Functional form	Cobb–Douglas function
	Returns to scale	Decreasing returns to scale; constant returns to scale; increasing returns to scale
Error term	Consideration of the error term	Multiplicative
	Average level of efficiency	70%; 80%; 90%
	Distribution of the inefficiency term	Half-normal; gamma
	Distribution of the noise term	Normal
	Level of noise	None ($\sigma_v = 0$); Little ($\sigma_v = 0.068$); Medium ($\sigma_v = 0.136$); Strong ($\sigma_v = 0.272$)

usually considers up to three factors each, whereby variations in their numbers are rare (see, e.g., Badunenko et al. 2012, p. 869, Di Giorgio et al. 2016, pp. 5–6, and Andor et al. 2019, pp. 243–244). Even though the number of input and output factors may be relevant for the efficiency scores and their spread (see Dyson et al. 2001, p. 248), Ahn et al. (2023) found a rather small/ambiguous effect of their number on the method performance with respect to our purposes. Hence, we model three input and three output factors here.

To model the distribution of the input values, we adopt common assumptions from the literature. On the one hand, a uniform distribution $U[5, 15]$ is assumed. In this case, there is no accumulation of a certain input quantity (see, e.g., Banker et al. 1993, p. 335, Andor and Hesse 2014, p. 91, and Schaefer and Clermont 2018, p. 145). On the other hand, we include a shifted right-skewed beta distribution $10 * B(4, 16) + 5$ (for right-skewed input distributions, see, e.g., Gong and Sickles 1992, p. 271, Yu 1998, p. 573, and Andor and Hesse 2014, p. 104). Thus, scenarios with few large and many small DMUs are considered, as is often the case in reality.

Based on the inputs, the outputs can be generated using a production function. In the literature, neo-classical production functions that allow for input substitution are widely used. Following this pattern and considering the rather small influence of the specific form of the neo-classical production function on the performance (see Andor and Hesse 2014, p. 99, and Ahn et al. 2023, pp. 299–300), we only use a Cobb–Douglas (CD) function $\|y_j^*\|_2 = e^{\beta_0} x_{ij}^{\beta_1} \dots x_{mj}^{\beta_m}$ (Cobb and Douglas 1928).⁴ This production function is typical for numerous production systems (see, e.g., Ruggiero 2007, pp. 260–261, Kuosmanen and Kortelainen 2012, pp. 23–24, and Oh and Shin 2015, p. 520). Thereby, the one-dimensional $\|y_j^*\|_2$ represents the optimal output

⁴ Hereby, $\|\cdot\|_2$ denotes the Euclidean norm.

Table 3 Parametrizations of the Cobb–Douglas function

Returns to scale	Parametrizations
DRS	$\beta_0 = 0, \beta_1 = \beta_2 = \beta_3 = 0.267$
CRS	$\beta_0 = 0, \beta_1 = \beta_2 = \beta_3 = 0.333$
IRS	$\beta_0 = 0, \beta_1 = \beta_2 = \beta_3 = 0.4$

capacity of DMU j , which is used to generate three outputs as described below. Regarding the CD function, we consider decreasing returns to scale (DRS), CRS, and increasing returns to scale (IRS) (see, e.g., Badunenko et al. 2012, p. 882, Krüger 2012, pp. 142–143, and Andor and Hesse 2014, p. 99). The parametrizations of the CD function are provided in Table 3.

To distribute the output capacity $\|y_j^*\|_2$ resulting from the production functions among the three output factors, for every DMU j , three truncated normally distributed variables $\alpha_{rj} \sim \mathcal{N}_{[0,1]}(1/3, (1/3)^2)$, $r \in \{1, 2, 3\}$,⁵ are simulated, following Gstach (1999, pp. 109–110). To determine the share of the r th output factor for DMU j in the output capacity, a_{rj} is set as $a_{rj} = \alpha_{rj} / (\sum_{r=1}^3 \alpha_{rj}) \forall r, j$, resulting in $y_{rj}^* = \sqrt{a_{rj}} \|y_j^*\|_2 \forall r, j$.⁶ For a specific input combination, this modeling enables a flexible allocation of the output capacity to the different output factors (see also Ahn et al. 2023, p. 297).

The error term ε_j , which consists of the inefficiency term u_j and the stochastic noise term v_j of DMU j , is included multiplicatively by $y_j = y_j^* e^{\varepsilon_j} = y_j^* e^{v_j} e^{-u_j}$. Hence, we model efficiency as well as the stochastic noise in a radial sense in accordance with Farrell efficiency (see Farrell 1957) and corresponding efficiency measurement methods, and we define ‘true’ efficiency accordingly with respect to the underlying production system and the chosen Farrell efficiency.

With respect to the inefficiency term, average levels of efficiency of 70%, 80%, and 90% are simulated. In this way, scenarios with weaker competition (lower average efficiency) and stronger competition (higher average efficiency) are modeled (see, for example, Giraleas et al. 2012, p. 677, Oh and Shin 2015, p. 520, and Andor et al. 2019, p. 244). On the one hand, we assume a half-normal distribution $\mathcal{N}^+(0, \tilde{\sigma}_u^2)$. On the other hand, we consider a right-skewed gamma distribution $\mathcal{G}(p, b)$ with the shape parameter p and the inverse scale parameter b such that the mode of efficiency is below 100% (see for both distributions, e.g., Gong and Sickles 1992, p. 263, Badunenko et al. 2012, p. 881, and Andor et al. 2019, p. 244). The parametrizations of the inefficiency term are listed in Table 4.

We model the stochastic noise term using a normal distribution, i.e., little noise is more likely than strong noise, and the noise is symmetrically distributed. With respect to the noise-to-signal ratio $\rho = \sigma_v / \tilde{\sigma}_u$, which is frequently used in the literature, we

⁵ Hereby, $\mathcal{N}_{[0,1]}$ denotes the normal distribution which is truncated to the interval $[0, 1]$.

⁶ This procedure assumes that all three outputs are measured on the same scale. However, radial efficiency measures are independent of the choice of units since they measure efficiency proportionally. Therefore, the outputs’ units do not have any impact on the following results.

simulate the stochastic noise term as $\mathcal{N}(0, (\rho * 0.136)^2)$. Here, $\tilde{\sigma}_u = 0.136$ represents the standard deviation of the normal distribution belonging to the half-normally distributed inefficiency term for an average efficiency of 90%. Thus, the stochastic noise is identically distributed for both distributions of the inefficiency term. We consider $\rho \in \{0, 0.5, 1, 2\}$ to model no noise as well as little, medium, and strong noise (see, e.g., Kuosmanen and Kortelainen 2012, p. 23, Andor and Hesse 2014, pp. 95–96, and Schaefer and Clermont 2018, p. 145).

To assess the estimation quality of the methods in the different scenarios, we need performance criteria that allow us to evaluate the discrepancies between the true and estimated efficiency scores. We select two performance criteria to quantify two main purposes of measuring efficiency: the accurate determination of the ranking of DMUs based on their efficiency scores and the accurate estimation of these scores (see, e.g., the Monte Carlo simulations by Andor and Hesse 2014, p. 92, Oh and Shin 2015, p. 520, and Ahn et al. 2023, p. 299).

To evaluate the extent to which the ranking is determined correctly, we use the Mean Rank Correlation (MRC) according to Spearman (1904) between the true and estimated efficiency scores. For the $R = 500$ repetitions of a scenario, it is calculated as

$$\text{MRC} = \frac{1}{R} \sum_{l=1}^R \frac{\text{Cov}\left(r\left(\widehat{\text{TE}}_l\right), r\left(\text{TE}_l\right)\right)}{\sqrt{\text{Var}\left(r\left(\widehat{\text{TE}}_l\right)\right) * \text{Var}\left(r\left(\text{TE}_l\right)\right)}}, \quad (9)$$

where $\widehat{\text{TE}}_{jl}$ denotes the estimated (technical) efficiency score of DMU j , TE_{jl} denotes the corresponding true (technical) efficiency score of DMU j , and $r(\text{TE}_l)$ denotes the corresponding ranking of DMUs based on their efficiency scores in the l th repetition.

To analyze the estimation quality in terms of estimating the efficiency scores as accurately as possible, we use the Mean Absolute Deviation (MAD) between the true and estimated efficiency scores. For all n DMUs and $R = 500$ repetitions, it is calculated as:

$$\text{MAD} = \frac{1}{R * n} \sum_{l=1}^R \sum_{j=1}^n \left| \widehat{\text{TE}}_{jl} - \text{TE}_{jl} \right|. \quad (10)$$

Table 4 Parametrizations of the inefficiency term

Average level of efficiency %	Half-normally distributed inefficiency term	Gamma distributed inefficiency term
70	$\mathcal{N}^+(0, 0.498^2)$	$\mathcal{G}(4, 10.730)$
80	$\mathcal{N}^+(0, 0.299^2)$	$\mathcal{G}(4, 17.440)$
90	$\mathcal{N}^+(0, 0.136^2)$	$\mathcal{G}(4, 37.439)$

4 Results and discussion

In this section, we present the simulation results. In Subsection 4.1, we assess whether CNAA is superior to NAA. In Subsections 4.2 and 4.3, we evaluate the impact of changes in parameter specifications on the performance as well as the performance of CNAA using the (basic) DEA models as benchmarks. Thereby, a distinction is made between the two considered purposes of measuring efficiency.

4.1 Validation of CNAA's improvement compared with NAA

Elucidating whether the CNAA specifications achieve a performance improvement compared with NAA, Table 5 presents the average MRCs and MADs over all scenarios and repetitions. As described in Sect. 2, the CNAA specifications do not change the ranking of DMUs, which is why both specifications as well as NAA have an identical average MRC of 69.3%. By contrast, the average MAD considerably improves for the CNAA specifications. While there is a high average MAD of 28.9 percentage points for NAA due to the systematic underestimation of the efficiency scores, the MAD reduces to 11.7 percentage points for CNAA⁺ and to 14.3 percentage points for CNAA^x.

For detailed analyses, we must consider the 720 individual scenarios instead of average values because deviations may arise for individual scenarios that are not discernible due to the aggregation. For example, compared with NAA, CNAA⁺ reduces the MAD in 706 scenarios (98.1% of the scenarios), and CNAA^x reduces the MAD in 714 scenarios (99.2% of the scenarios). Scenarios wherein CNAA^x has a higher MAD than NAA are a subset of the scenarios wherein CNAA⁺ has a higher MAD than NAA. This can be explained by the fact that the worse performance of CNAA^x occurs when the efficiency scores are shifted so strongly that the underestimation of true efficiency turns into an even greater overestimation. In these cases, CNAA⁺ overestimates the efficiency as well, because it shifts the efficiency scores more strongly than CNAA^x.

It can be stated that our objective of improving the performance of NAA through the CNAA specifications is achieved for the present simulation design. Therefore, NAA is not considered in the following analyses.

4.2 Evaluation of CNAA's performance: purpose of determining the ranking

Regarding the impact of changes in parameter specifications on the performance as well as the performance of CNAA using the DEA models as benchmarks, we

Table 5 Overview of the average results

	NAA	CNAA ⁺	CNAA ^x
MRC	0.693	0.693	0.693
MAD	0.289	0.117	0.143

first consider the purpose of accurately determining the ranking of DMUs. To capture interdependencies between the influences of the single parameters, we subdivide the 720 scenarios as follows:

- Analysis 1: variation in the number of DMUs, input distribution, and returns to scale;
- Analysis 2: variation in the number of DMUs, level and distribution of (in)efficiency, and level of noise.

Table 6 lists the MRCs for the methods considered according to Analysis 1. Here and in the following, the values of the performance criteria in each case indicate the mean value over the 500 repetitions and over those parameter specifications for which no differentiation is made in the respective analysis. Thus, the tables always refer to the 720 scenarios, whereby the focus is on different parameters. Furthermore, those CNAA cells are marked in bold for which the CNAA specifications show a better performance than both DEA models in the respective scenario. As the rankings of DMUs for CNAA⁺ and CNAA^x are identical, the corresponding results are presented jointly in this subsection.

With respect to Table 6, we first address the impact of changes in parameter specifications on the performance of CNAA⁺/CNAA^x. The corresponding MRCs range from 65.3 to 71.6%. This rather small fluctuation width indicates a limited influence of the parameters considered here on the MRC. The MRCs for the CNAA specifications increase slightly with more DMUs, though this effect weakens as the number of DMUs increases. The differences in the MRCs for scenarios with 25 and 400 DMUs range from 1.5 to 1.8 percentage points. Thus, a higher number of DMUs has a slightly positive influence on the performance of the CNAA specifications. Regarding the input distribution, the CNAA specifications perform up to 4.2 percentage points better for the beta distribution than for the uniform distribution. The returns to scale also have a rather small influence, with a maximum difference in the MRCs for the various returns to scale of 3.2 percentage points. The best performance is present for CRS, which can be attributed to CNAA's implicit assumption of CRS.

We now examine the performance of the CNAA specifications using the DEA models as benchmarks. The bold values reveal that the CNAA specifications exhibit the best performance in all scenarios considered here. The differences between the MRCs for CNAA⁺/CNAA^x and DEA^{CCR} or DEA^{BCC} range from 3.3 to 11.7 percentage points (DEA^{CCR}) and from 7.1 to 26.2 percentage points (DEA^{BCC}); thus, they can be quite large.

Table 7 presents the MRCs according to the variations in the parameter specifications in Analysis 2. Compared with Table 6, the MRCs vary much more. Hence, the considered parameters have a stronger influence on the performance than those in Analysis 1.

Regarding the impact of changes in the parameter specifications on the performance of CNAA, it also becomes evident here that an increase in the number of DMUs is accompanied by a slight improvement in performance. Additionally, the MRCs for the half-normal distribution of the inefficiency term are up to 16.6

Table 6 Mean rank correlation subdivided by number of DMUs, input distribution, and returns to scale

Input distribution	Returns to scale	CNAA ⁺ /CNAA ⁺					DEA ^{CCR}					DEA ^{BCC}				
		# of DMUs														
		25	50	100	200	400	25	50	100	200	400	25	50	100	200	400
Uniform	DRS	0.653	0.662	0.667	0.670	0.671	0.546	0.579	0.601	0.617	0.629	0.430	0.487	0.533	0.570	0.600
	CRS	0.685	0.692	0.698	0.699	0.701	0.568	0.604	0.631	0.648	0.663	0.423	0.482	0.529	0.565	0.597
	IRS	0.668	0.677	0.681	0.685	0.686	0.561	0.594	0.616	0.634	0.646	0.419	0.475	0.523	0.562	0.593
Beta	DRS	0.694	0.704	0.708	0.710	0.712	0.622	0.645	0.660	0.670	0.679	0.449	0.505	0.552	0.589	0.620
	CRS	0.701	0.709	0.713	0.715	0.716	0.626	0.648	0.663	0.673	0.683	0.447	0.506	0.550	0.587	0.617
	IRS	0.695	0.703	0.707	0.709	0.711	0.621	0.643	0.656	0.667	0.675	0.444	0.503	0.548	0.585	0.614

⁷ Average over the scenarios with: average level of efficiency: 70%/80%/90%; inefficiency term: half-normally/ gamma distributed; level of noise: none/little/medium/strong

⁷ Since we adopted the simulation design of Ahn et al. (2023) to some extent, the results for DEA as well as the MRCs for CNAA are expected to be similar. However, the simulation design underlying the individual tables differs in detail (inter alia due to the two sets of variations in Ahn et al. 2023, see the respective footers of the tables) what causes the differences in the results.

Table 7 Mean rank correlation subdivided by number of DMUs, level and distribution of (in)efficiency, and level of noise

Distribution of the inefficiency term	Average level of efficiency %	Level of noise	CNAA ⁺ /CNAA ^x					DEA ^{CCR}					DEA ^{BCC}				
			# of DMUs														
			25	50	100	200	400	25	50	100	200	400	25	50	100	200	400
Half-normal	70	None	0.983	0.987	0.989	0.990	0.990	0.910	0.937	0.954	0.965	0.973	0.670	0.755	0.819	0.869	0.908
		Little	0.941	0.949	0.953	0.956	0.957	0.882	0.903	0.918	0.927	0.934	0.650	0.722	0.781	0.826	0.863
		Medium	0.859	0.869	0.875	0.877	0.879	0.810	0.830	0.842	0.849	0.854	0.578	0.646	0.703	0.745	0.779
		Strong	0.676	0.690	0.694	0.696	0.699	0.646	0.665	0.671	0.674	0.680	0.433	0.499	0.543	0.579	0.607
	80	None	0.964	0.970	0.974	0.975	0.976	0.843	0.883	0.910	0.930	0.945	0.641	0.722	0.788	0.840	0.882
		Little	0.877	0.888	0.893	0.896	0.898	0.779	0.812	0.832	0.846	0.857	0.584	0.653	0.709	0.752	0.788
		Medium	0.730	0.739	0.746	0.749	0.751	0.662	0.683	0.699	0.708	0.714	0.483	0.532	0.578	0.617	0.646
		Strong	0.497	0.504	0.510	0.511	0.514	0.469	0.477	0.487	0.489	0.494	0.325	0.359	0.391	0.417	0.438
	90	None	0.893	0.903	0.908	0.910	0.912	0.665	0.735	0.783	0.823	0.857	0.520	0.619	0.694	0.759	0.815
		Little	0.675	0.687	0.692	0.694	0.697	0.537	0.572	0.600	0.617	0.632	0.412	0.469	0.512	0.547	0.578
		Medium	0.457	0.465	0.467	0.469	0.470	0.394	0.409	0.420	0.427	0.434	0.289	0.322	0.349	0.369	0.391
		Strong	0.258	0.260	0.262	0.265	0.265	0.238	0.244	0.247	0.251	0.253	0.160	0.183	0.199	0.214	0.224

Table 7 (continued)

Distribution of the inefficiency term	Average level of efficiency %	Level of noise	CNAA ⁺ /CNAA ^x										DEA ^{CCR}					DEA ^{BCC}				
			# of DMUs										25	50	100	200	400	25	50	100	200	400
			25	50	100	200	400	25	50	100	200	400										
Gamma	70	None	0.968	0.974	0.977	0.979	0.980	0.842	0.880	0.906	0.923	0.937	0.630	0.713	0.773	0.822	0.862					
		Little	0.892	0.900	0.905	0.909	0.910	0.790	0.819	0.840	0.855	0.865	0.586	0.652	0.709	0.753	0.790					
		Medium	0.746	0.762	0.766	0.770	0.771	0.677	0.704	0.716	0.727	0.733	0.490	0.546	0.592	0.632	0.661					
		Strong	0.517	0.524	0.531	0.533	0.536	0.486	0.499	0.507	0.511	0.516	0.331	0.375	0.407	0.437	0.458					
	80	None	0.934	0.942	0.946	0.948	0.950	0.736	0.791	0.830	0.860	0.885	0.564	0.653	0.720	0.773	0.819					
		Little	0.777	0.788	0.794	0.797	0.799	0.639	0.679	0.705	0.722	0.737	0.485	0.549	0.599	0.638	0.673					
		Medium	0.575	0.586	0.588	0.590	0.592	0.505	0.526	0.536	0.545	0.551	0.371	0.411	0.445	0.475	0.496					
		Strong	0.343	0.349	0.355	0.355	0.356	0.321	0.327	0.337	0.338	0.340	0.221	0.246	0.271	0.286	0.301					
	90	None	0.818	0.830	0.835	0.838	0.839	0.544	0.613	0.672	0.720	0.761	0.427	0.521	0.604	0.674	0.733					
		Little	0.519	0.533	0.535	0.538	0.539	0.390	0.424	0.445	0.463	0.476	0.299	0.348	0.381	0.411	0.436					
		Medium	0.312	0.317	0.328	0.329	0.330	0.260	0.279	0.290	0.297	0.301	0.194	0.219	0.241	0.258	0.270					
		Strong	0.171	0.172	0.175	0.175	0.177	0.155	0.160	0.165	0.165	0.167	0.108	0.115	0.132	0.140	0.148					

Average over the scenarios with: inputs: uniformly/beta distributed; returns to scale: DRS/CRS/IRS

percentage points higher than those for the gamma distribution. This could be because the half-normal distribution leads to a larger spread of the true efficiency scores than the gamma distribution, thus enabling better discrimination. By contrast, a higher average efficiency often precipitates a lower MRC. This can probably be explained by the fact that, in this case, there is a smaller spread of the true efficiency scores. This induces poorer discrimination, thus impeding to determine the correct ranking. Analogously, stronger noise leads to lower MRCs, because, also in this case, the ranking can be determined worse owing to the distortion. Furthermore, it can be seen that the level of noise has the strongest influence on the MRC. Overall, the best performance of the CNAA specifications is present with an MRC of 99% for 400 DMUs, an average efficiency of 70%, a half-normally distributed inefficiency term, and no noise.

The bold values show again that in all scenarios, the CNAA specifications exhibit (in some cases clearly) higher MRCs than DEA^{CCR} and DEA^{BCC} . The respective differences range from 1.0 to 27.4 percentage points ($CNAA^+/CNAA^x$ versus DEA^{CCR}) and from 2.9 to 39.1 percentage points ($CNAA^+/CNAA^x$ versus DEA^{BCC}). These differences are particularly large for scenarios with few DMUs, high average efficiency, and little noise. Looking at the 720 individual scenarios, DEA^{BCC} outperforms the CNAA specifications in only one scenario, and there is no scenario wherein DEA^{CCR} performs better.

In summary, in our simulation design, the CNAA specifications show a better performance than DEA with respect to the accurate determination of the ranking of DMUs. In this sense, the CNAA specifications seem to be fundamentally competitive here. However, it has to be noted that the MRCs are quite low in some scenarios (at times even $< 30\%$), especially in scenarios with high average efficiency and strong noise.

4.3 Evaluation of CNAA's performance: purpose of estimating efficiency scores

Now, we consider the purpose of accurately estimating the efficiency scores using the performance criterion MAD. The evaluations are done structurally in analogy with Subsection 4.2. Thus, Table 8 presents the MADs for the considered methods according to Analysis 1, which focuses on the variations in the number of DMUs, input distribution, and returns to scale.

Evidently, for the CNAA specifications, the parameters varied in this subsection have a rather small influence on the MAD values. In Table 8, the MADs range from 10.4 to 14.0 percentage points for $CNAA^+$ and from 11.3–18.3 percentage points for $CNAA^x$, thus exhibiting a relatively small spread. The strongest effect occurs for the number of DMUs. For example, its increase from 25 to 400 DMUs increases the MADs for $CNAA^+$ by 1.2–3.1 percentage points and those for $CNAA^x$ by 5.3–6.4 percentage points. The influence of the input distribution clearly differs between $CNAA^+$ and $CNAA^x$. As with the MRC, the estimation quality of $CNAA^x$ is slightly better for the beta distribution (by up to 1.5 percentage points), whereas, for $CNAA^+$, slightly lower MADs result for the uniform distribution (by up to 2.1 percentage points). This could be because, for the beta distribution, there are fewer

Table 8 Mean absolute deviation subdivided by number of DMUs, input distribution, and returns to scale

Input distrib- ution	Returns to scale	CNAA ⁺					DEA ^{CCR}					DEA ^{BCC}				
		# of DMUs														
		25	50	100	200	400	25	50	100	200	400	25	50	100	200	400
uniform	DRS	0.104	0.107	0.110	0.113	0.118	0.119	0.119	0.134	0.150	0.167	0.183	0.133	0.123	0.117	0.118
	CRS	00.105	0.107	0.110	0.114	0.117	0.114	0.126	0.142	0.158	0.173	0.173	0.134	0.124	0.118	0.118
	IRS	00.104	0.107	0.111	0.115	0.118	0.116	0.129	0.145	0.161	0.176	0.176	0.133	0.123	0.118	0.121
beta	DRS	0.110	0.116	0.124	0.132	0.139	0.115	0.126	0.140	0.154	0.168	0.168	0.123	0.117	0.116	0.120
	CRS	0.109	0.116	0.124	0.132	0.140	0.113	0.125	0.139	0.152	0.167	0.167	0.122	0.117	0.116	0.120
	IRS	0.108	0.115	0.123	0.132	0.139	0.114	0.126	0.140	0.155	0.169	0.169	0.122	0.117	0.116	0.120

Average over the scenarios with: average level of efficiency: 70%/80%/90%; inefficiency term: half-normally/gamma distributed; level of noise: none/little/medium/strong

high inputs, which tend to result in lower IOPs and IOEs. The induced upward shift in efficiency scores is method-immanently stronger for $CNAA^+$ than for $CNAA^x$ and may result in a shift that is too strong. Looking at the returns to scale, almost no differences exist in the MADs for $CNAA^+$, with a maximum difference of 0.2 percentage points. For $CNAA^x$, the differences are also small, amounting to a maximum of 1 percentage point for the uniform distribution and only up to 0.3 percentage points for the beta distribution. Here, the best performance is present for CRS, which is again consistent with the implicit assumption of CRS by $CNAA$.

In the following, the performances of the $CNAA$ specifications are compared with each other and with those of the DEA models as benchmarks. Once more, those cells are marked in bold for which the $CNAA$ specifications show a better performance than both DEA models in the respective scenario, i.e., a lower MAD. For $CNAA^+$, this is the case in most scenarios, namely for the beta distribution combined with 25 or 50 DMUs and for the uniform distribution. The difference to the MADs for DEA is large, up to 5.5 percentage points, especially for scenarios with few DMUs. Thus, in these scenarios, $CNAA^+$ can be clearly considered as suitable method for measuring efficiency. $CNAA^x$ shows a better performance than the DEA models in scenarios with 25 DMUs. However, in these cases, $CNAA^+$ performs even better.

Table 9 presents the MADs according to Analysis 2, i.e., the number of DMUs, level and distribution of (in)efficiency, and level of noise are addressed. As for the MRC, these parameters have a considerably stronger influence on the estimation quality than those previously considered in Analysis 1. Table 8 Mean Absolute Deviation subdivided by number of DMUs, input distribution, and returns to scale.

Regarding the impact of changes in parameter specifications on the performance of $CNAA^+/CNAA^x$, in contrast to the MRC, interdependencies between the influences of the four considered parameters can be observed. For example, lower MADs are often observed with higher average efficiency. This could be because the $CNAA$ specifications tend to estimate high efficiency scores owing to the shift. However, this effect only occurs for little noise, as strong noise can occasionally lead to very high observed outputs due to distortions. In the latter case, the IOPs and IOEs of single DMUs are estimated so high that the other efficiency scores are not sufficiently shifted upward, resulting in efficiency scores that are too low. For the same reason, when the average efficiency is low, little noise often exhibits a positive effect compared with no noise.

By contrast, for the gamma distribution, the probability of higher IOPs and IOEs for single DMUs decreases because of the lower mode compared with that of the half-normal distribution. The resulting stronger shift usually leads to a slightly better estimation in scenarios with high average efficiency and stronger noise. Increasing the number of DMUs increases the occurrence of single high IOPs and IOEs. In the case of rather low average efficiency, rather little noise, and the gamma distribution, this tends to lead to lower MADs owing to the influences of the other parameters and to higher MADs otherwise.

It can be observed that the relations between the influences of the four parameters considered here are complex. The evaluations show that the performance first improves and then deteriorates again with increases in the number of DMUs,

Table 9 Mean absolute deviation subdivided by number of DMUs, level and distribution of (in)efficiency, and level of noise

Distri- bution of the ineff- iciency term	Aver- age level of effi- ciency %	Level of noise	CNAA ⁺					DEA ^{CCR}					DEA ^{BCC}									
			# of DMUs																			
			25	50	100	200	400	25	50	100	200	400	25	50	100	200	400					
Half- nor- mal	70	None	0.075	0.077	0.080	0.085	0.090	0.027	0.021	0.021	0.023	0.027	0.134	0.105	0.081	0.062	0.047	0.194	0.157	0.126	0.099	0.076
		Little	0.070	0.067	0.068	0.070	0.072	0.050	0.055	0.065	0.076	0.088	0.124	0.095	0.073	0.058	0.051	0.187	0.149	0.116	0.090	0.071
		Medium	0.085	0.087	0.091	0.095	0.101	0.096	0.110	0.130	0.150	0.168	0.114	0.094	0.084	0.085	0.092	0.177	0.143	0.116	0.101	0.096
		Strong	0.156	0.172	0.186	0.200	0.213	0.189	0.222	0.253	0.282	0.308	0.143	0.149	0.165	0.188	0.214	0.186	0.168	0.166	0.174	0.190
	80	None	0.054	0.053	0.054	0.056	0.058	0.025	0.023	0.026	0.029	0.034	0.112	0.089	0.070	0.054	0.041	0.146	0.123	0.101	0.081	0.063
		Little	0.055	0.054	0.056	0.058	0.059	0.060	0.070	0.083	0.097	0.109	0.101	0.080	0.065	0.057	0.055	0.138	0.114	0.092	0.075	0.064
		Medium	0.090	0.097	0.104	0.112	0.121	0.120	0.142	0.164	0.187	0.208	0.099	0.091	0.091	0.100	0.113	0.132	0.114	0.102	0.100	0.104
		Strong	0.199	0.218	0.236	0.257	0.274	0.242	0.276	0.310	0.344	0.373	0.155	0.175	0.201	0.233	0.264	0.157	0.163	0.177	0.200	0.227
	90	None	0.029	0.027	0.027	0.027	0.028	0.026	0.029	0.034	0.039	0.045	0.072	0.059	0.048	0.039	0.031	0.085	0.075	0.064	0.053	0.043
		Little	0.053	0.057	0.061	0.066	0.070	0.080	0.094	0.110	0.125	0.139	0.065	0.058	0.055	0.057	0.063	0.079	0.069	0.061	0.057	0.057
		Medium	0.116	0.130	0.143	0.154	0.164	0.157	0.185	0.210	0.233	0.254	0.081	0.090	0.104	0.123	0.144	0.084	0.085	0.092	0.104	0.121
		Strong	0.254	0.279	0.307	0.325	0.343	0.302	0.342	0.382	0.414	0.444	0.175	0.211	0.251	0.289	0.325	0.137	0.165	0.200	0.237	0.275

Table 9 (continued)

Distribution of the inefficiency term	Average level of age level of efficiency %	Level of noise	CNAA ⁺					DEA ^{CCR}					DEA ^{BCC}				
			# of DMUs														
			25	50	100	200	400	25	50	100	200	400	25	50	100	200	400
Gamma 70	None	Little	0.127	0.119	0.115	0.114	0.113	0.080	0.060	0.047	0.037	0.030	0.207	0.180	0.155	0.132	0.111
			0.106	0.095	0.086	0.080	0.076	0.069	0.055	0.048	0.048	0.052	0.193	0.162	0.133	0.107	0.086
			0.093	0.083	0.078	0.076	0.076	0.087	0.088	0.097	0.110	0.125	0.166	0.133	0.109	0.094	0.087
			0.141	0.148	0.159	0.169	0.181	0.171	0.193	0.221	0.247	0.273	0.156	0.150	0.155	0.167	0.187
80	None	Little	0.084	0.079	0.076	0.073	0.072	0.051	0.040	0.032	0.027	0.025	0.152	0.134	0.115	0.098	0.083
			0.064	0.057	0.052	0.049	0.047	0.055	0.055	0.059	0.068	0.078	0.136	0.113	0.093	0.076	0.064
			0.081	0.083	0.086	0.091	0.097	0.105	0.121	0.139	0.158	0.178	0.117	0.102	0.095	0.094	0.101
			0.183	0.203	0.222	0.240	0.254	0.226	0.262	0.295	0.327	0.354	0.155	0.170	0.193	0.220	0.248
90	None	Little	0.039	0.036	0.033	0.031	0.030	0.029	0.027	0.027	0.028	0.031	0.085	0.076	0.066	0.057	0.049
			0.047	0.048	0.051	0.054	0.057	0.070	0.081	0.095	0.108	0.121	0.074	0.065	0.059	0.057	0.059
			0.110	0.121	0.134	0.144	0.153	0.150	0.175	0.200	0.222	0.242	0.084	0.090	0.102	0.118	0.136
			0.249	0.277	0.299	0.319	0.339	0.298	0.339	0.375	0.408	0.438	0.173	0.209	0.246	0.283	0.320

Average over the scenarios with: inputs: uniformly/beta distributed; returns to scale: DRS/CRS/IRS

average level of efficiency, and level of noise. An additional evaluation of the mean deviation between the true and estimated efficiency scores reveals that in these scenarios, the decreasing overestimation of the efficiency scores turns into an underestimation. As $CNAA^x$ shifts the efficiency scores less strongly than $CNAA^+$, the inversion of the overestimation into the underestimation tends to occur already for lower values of the mentioned parameters.

Let us finally look again at the comparison of the CNAA specifications with DEA. Here, the results are more differentiated than in the previous steps of the investigation. Considering the bold values in Table 9, it can be seen that both CNAA specifications show a better performance than the DEA models in numerous scenarios with a combination of rather few DMUs, low average efficiency, and little noise. The aforementioned scenarios tend to be found at the top left of the 4×5 fields in Table 9. Moreover, an inefficiency term that is gamma distributed seems to be more advantageous for $CNAA^+$ and $CNAA^x$. The CNAA specifications sometimes exhibit clearly lower MADs than the DEA models (up to 16.9 percentage points). This is particularly the case in scenarios with few DMUs, low average efficiency, and little noise.

$CNAA^+$ is marked in bold slightly more often than $CNAA^x$. With respect to the 720 individual scenarios, $CNAA^+/CNAA^x$ shows a better performance than both DEA models in 53.1%/40% of the scenarios. These scenarios overlap in 35.8% of the 720 scenarios. As $CNAA^+$ has the lowest MAD in 27.8%, $CNAA^x$ in 28.6%, DEA^{CCR} in 15.6% and DEA^{BCC} in 27.2% of the scenarios, the best performance is relatively evenly distributed across the methods.

Comparing the two CNAA specifications, $CNAA^x$ has a lower MAD than $CNAA^+$ in 29.9% of all 720 scenarios, namely for the rather low parameter values (few DMUs, low average efficiency, and little noise). This can be explained by the fact that $CNAA^x$ conducts a smaller shift in efficiency scores than $CNAA^+$ and, thus, tends to estimate lower true efficiency scores better. By contrast, $CNAA^+$ is the better specification in the remaining 70.1% of the scenarios with somewhat higher parameter values. This is especially the case in scenarios wherein neither of the two CNAA specifications performs better than the DEA models.

In summary, our results reveal that the performance of the CNAA specifications is largely comparable to that of the DEA models. Hence, CNAA can be considered as suitable method in terms of accurately estimating the efficiency scores. This is especially true for scenarios with a combination of rather few DMUs, low average efficiency, and little noise. Moreover, $CNAA^+$ outperforms $CNAA^x$ in about 70% of the scenarios. It has to be noted, however, that some scenarios exhibit very high MADs (in some cases even > 30 percentage points). This is predominantly the case in scenarios with many DMUs, high average efficiency, and strong noise.

5 Discussion

As with any simulation study, our Monte Carlo simulation is subject to the chosen simulation design. Since our paper aims to look at the application context and we want to ensure comparability of our results with the literature, we have based our

Table 10 Overview of the average results, now including NAA^{Max}

	NAA	CNAA ⁺	CNAA ^x	NAA ^{Max}
MRC	0.693	0.693	0.693	0.551
MAD	0.289	0.117	0.143	0.148

simulation design in particular on common assumptions. In the following, we discuss some corresponding aspects of our paper.

First, particularly regarding analyses that compare DEA with other methods, it is sometimes criticized that such a simulation design is unfavorable for DEA models and therefore potentially gives an advantage to other methods. We want to take up this aspect and investigate what happens to the performance if we make appropriate modifications. For example, DEA ordinarily estimates a large share of DMUs (about 10–30%) to be 100% efficient. With our distributions of the inefficiency term, it is only rarely the case that a DMU is simulated as 100% efficient. As robustness check and to further verify CNAA's suitability, we simulated one more parametrization of (in)efficiency. In accordance with simulations in the literature that assume a decent share of DMUs to be efficient, we assume a half-normally distributed inefficiency term with an average efficiency of 80% (i.e., $u_i \sim \mathcal{N}^+(0, 0.299^2)$) as basis. Then, we set 20% of DMUs to be 100% efficient to also depict scenarios with a substantial number of efficient DMUs (see, e.g., Ruggiero 1998, p. 465, Holland and Lee 2002, p. 13, and Muñoz et al. 2006, p. 1179). The parametrizations of the other parameters remain as for the initial simulation.

Regarding the MRC, the CNAA specifications again exhibit the best performance in all considered scenarios, and the specific values are similar to those in Subsection 4.2. Regarding the MAD, the results are more differentiated. There are still numerous scenarios in which CNAA shows a better performance than DEA but not as many as in the initial simulation, apparently due to the better estimation of 100% efficient DMUs by DEA. However, the specific CNAA results again do not differ much from the results in Subsection 4.3. There are therefore no major differences in the results we obtained from this modified data generation process.

Second, in our study, we decided to base the two CNAA specifications on NAA with arithmetic mean. Nevertheless, other choices of the parameter ω for the generalized mean in Eq. (5) are possible. Since Ahn et al. (2007, pp. 635–636) found a certain connection between DEA^{CCR} and NAA with $\omega \rightarrow +\infty$, i.e., the maximum of the IOEs (in the labeled as NAA^{Max}), it is conceivable to investigate the performance of NAA^{Max} .⁸ The following Table 10 presents the average results of a respective additional simulation.⁹ It is an extension of Table 5 in Subsection 4.1 that now also includes the average MRC and MAD for NAA^{Max} .

It can be seen that, overall, NAA^{Max} exhibits a clearly lower average MRC and a slightly higher average MAD than the CNAA specifications, even though the

⁸ Since NAA^{Max} always estimates at least one DMU to be 100% efficient, it is not necessary to modify NAA^{Max} in our sense of a Corrected NAA^{Max} .

⁹ We present the detailed results in the Appendix.

average MAD is lower than that for NAA. When looking at the 720 individual scenarios instead of average values, there is only one scenario wherein NAA^{Max} exhibits a higher MRC than CNAA; regarding the MAD, NAA^{Max} exhibits a lower value than $CNAA^+/CNAA^x$ in 25.0%/46.0% of the scenarios. To summarize, CNAA performs better than NAA^{Max} , especially regarding the MRC but also in the majority of scenarios regarding the MAD. It seems therefore more appropriate to carry out efficiency measurements using our proposed CNAA instead of using NAA^{Max} .

A third important aspect to discuss is the underlying production system. We modeled it based on the well-known output-oriented neo-classical CD function which allows for input substitution. However, such production functions are sometimes criticized since they are representative of production systems at an aggregate level in contexts of, e.g., national or regional economics. They do not describe industrial and other business contexts with a lack of substitutability between production factors or co-products well (see Verheyen et al. 2023, pp. 680–681). Against this background, other production systems could be modeled, e.g., systems that allow for only very limited input substitution and/or undesirable output factors. Verheyen et al. (2023) considered a Leontief production as well as a rigidly coupled production with no possibility of substitution between the inputs.¹⁰ They conducted an extensive Monte Carlo simulation to compare the performance of DEA, SFA and Stochastic Non-Smooth-Envelopment of Data (StoNED; see Kuosmanen and Kortelainen 2012). The authors found a decisive influence of the underlying production system on the performance of efficiency measurement methods. Therefore, extending our investigations to the modeling of such alternative production systems seems to be an interesting approach for evaluating CNAA's performance. In particular, it would then have to be analyzed whether CNAA's performance also changes if the simulation design is chosen according to the considerations of Verheyen et al. (2023). Are the differences in the performance similar to those found in Verheyen et al. (2023) or is CNAA's performance more robust? However, a corresponding analysis would go beyond the scope of this paper and is therefore left to future research.

As already stated by Ahn et al. (2023, p. 310), it should be considered that the choice between methods may not only depend on their estimation quality but also on aspects that cannot be modeled in simulation studies. For example, the simplicity and ease of use of a method can be of particular relevance for applications by practitioners. In this respect, the CNAA specifications have a central advantage: their methodology is easy to understand, even with little knowledge of production theory and mathematics. Additionally, no special computing proficiency or specialized software is required for implementation because the calculations can be done in spreadsheet programs, such as MS Excel. This can be decisive for practical use in real-world applications without expert support, especially in comparison with more complex methods, such as DEA or SFA, where, e.g., linear or stochastic optimization problems have to be solved.

¹⁰ The working paper by Schaefer and Dyckhoff (2018) contains further results not included in the cited publication.

Moreover, CNAA generates less information than other methods. For example, unlike CNAA, DEA generates Pareto-optimal benchmarks as well as input and output weights in the form of multipliers. When selecting a method, it must be determined whether such information is relevant in the respective context. Furthermore, CNAA is primarily an option if CRS are assumed. Besides respective cases with empirical evidence about CRS (see, e.g., Johnes 2006, pp. 279–280, for universities), CRS are a common assumption in many applications. For example, popular cost accounting systems assume linear relations between costs and their influencing factors. This is, e.g., the case for the German Grenzplankostenrechnung (Marginal Planned Cost Accounting), which is implemented in the SAP software (see, e.g., Friedl et al. 2009, p. 39).

A last aspect, we want to address is that the choice of a method can also depend on the number of input and output factors. If this number is relatively high, DEA's discriminating power may be limited. This is due to the method's endogenous factor weighting that puts each DMU in the best light and, thereby, may *ceteris paribus* lead to higher estimated efficiency scores for more input and output factors and a relatively high number of DMUs with an efficiency score of 100% (see Dyson et al. 2001, p. 248). For CNAA, there is no similar issue.

Table 11 provides a brief comparison of general characteristics of CNAA and basic DEA models (DEA^{CCR} and DEA^{BCC}), i.e., of their assumptions and generated information.

As noted in Subsections 4.2 and 4.3, for some scenarios, the methods estimate extremely low MRCs or extremely high MADs, especially for a combination of high average efficiency and strong noise. However, this is not a specific characteristic of CNAA, since it is also present for DEA (and for other efficiency measurement methods, see, e.g., Ahn et al. 2023). In such situations, using these methods generally does not seem appropriate.

6 Conclusions and outlook

To counteract the systematic underestimation of the Farrell efficiency scores by NAA, we presented two modifications in this study: additive CNAA⁺ and multiplicative CNAA[×]. While the first modification shifts the NAA efficiency scores for all DMUs upward by the same amount, the second modification increases the NAA efficiency scores proportionally by the same factor. We ran a Monte Carlo simulation with 720 scenarios to analyze the estimation quality of the CNAA specifications. We first evaluated their performance compared with that of NAA and investigated the impact of changes in the parameter specifications on the performance. Furthermore, we analyzed their performance using the (basic) DEA models as benchmarks. In our analyses, we considered two main purposes of measuring efficiency: the accurate determination of the ranking of DMUs based on their efficiency scores (measured by the MRC) and the accurate estimation of these scores (measured by the MAD). The findings obtained from our investigation can be summarized as follows:

Table 11 Comparison of general characteristics of CNAA and basic DEA models

CNAA		DEA
Assumptions	Orientation	Input or output
	Production function	Monotonically increasing, continuous, concave, and section-wise linear
	Returns to scale	CRS or VRS
	Necessary mathematical background of users	Relatively high
	Efficiency scores	Are estimated
	Benchmarks	Are generated
	Ranking of DMUs	Deducible
	Relative contribution of the inputs and outputs to the efficiency scores	Deducible via multipliers
	Discriminating power	Depends on the number of input and output factors
		Relatively high
Information scope		None
		Implicitly linear for every input-output combination
		CRS
		Relatively low
		Are estimated
		None
		Deducible
		Not deducible

1. Regarding the improvement in performance by the CNAA specifications compared with NAA, the core findings for the two purposes are as follows:

- a) Accurate determination of the ranking of DMUs: The performance of the CNAA specifications is method-inherently identical to that of NAA.
- b) Accurate estimation of the efficiency scores of DMUs: In the simulated scenarios, the CNAA specifications achieve a considerable improvement in the estimation quality. The average MAD decreases from 28.9 percentage points for NAA to 11.7 percentage points for CNAA^+ and 14.3 percentage points for CNAA^x . With respect to the 720 individual scenarios, CNAA^+ and CNAA^x reduce the MAD in 98.1% and 99.2% of the scenarios, respectively.

2. The performance of the CNAA specifications depends strongly on the parameter specifications in a scenario. Especially for the level and distribution of (in)efficiency and level of noise, a rather strong influence on the values of the performance criteria is found for both CNAA specifications. The number of DMUs has a rather small influence on the MRC and a stronger influence on the MAD. The influence of the input distribution and the returns to scale is small.

3. Regarding the performance of the CNAA specifications using the DEA models as benchmarks, the following differentiated findings emerge:

- a) Accurate determination of the ranking of DMUs: The CNAA specifications exhibit a better estimation quality than both DEA models in 719 of the 720 scenarios. Thus, there is an extensive superiority of the CNAA specifications. Thereby, the difference in the MRCs and, therefore, the advantage over the DEA models is particularly large in scenarios with few DMUs, high average efficiency, and little noise.
- b) Accurate estimation of the efficiency scores of DMUs: $\text{CNAA}^+/\text{CNAA}^x$ exhibits a (sometimes clearly) better performance than both DEA models in 53.1%/40% of the scenarios. This is especially the case in scenarios with a combination of rather few DMUs, low average efficiency, and little noise. The CNAA specifications exhibit the best estimation quality even slightly more often than the DEA models: The lowest MAD results for CNAA^+ in 27.8%, for CNAA^x in 28.6%, for DEA^{CCR} in 15.6%, and for DEA^{BCC} in 27.2% of the scenarios.

In conclusion, for our simulation design and for both purposes, the CNAA specifications can be considered as reasonable addition to the pool of methods for measuring (relative) efficiency. Their application may especially be considered in the absence of profound knowledge of production theory and mathematics, since CNAA is relatively easy to understand.

4. The comparison of $CNAA^+$ and $CNAA^x$ shows that $CNAA^x$ exhibits a lower MAD than $CNAA^+$ in approximately 30% of the scenarios considered. This is predominantly the case for scenarios with a combination of rather few DMUs, low average efficiency, and little noise. However, in the other 70% of the scenarios, $CNAA^+$ exhibits a better estimation quality. Thus, $CNAA^+$ frequently outperforms $CNAA^x$, but there is no general superiority.

Our findings are based on the presented CNAA methodology and limited to the simulation design. Therefore, our analyses could be extended by further studies. First, regarding the simulation design, the chosen parametrizations cover only a specific part of all conceivable scenarios. In addition to further specifications for the varied parameters, variations in other parameters are possible. For example, the number of input and output factors, the distribution of the stochastic noise and the way multi-dimensional outputs are generated could be modified.

Second, our results depend on the specific definition of true efficiency. In accordance with the majority of publications in the literature, we modeled the true efficiency in a radial sense, which corresponds to Farrell efficiency and methods for measuring efficiency, e.g., basic DEA models. Nonetheless, the effect of other definitions of true efficiency could be considered (see Verheyen et al. 2023, pp. 683–684 and 694).

Third, the CNAA methodology could be modified and further developed, respectively. Referring to Eqs. (2 and 3), e.g., other normalization methodologies as listed by Freudenberg (2003, p. 10) could be applied, especially for the purpose of sensitivity analyses. Additionally, it would be beneficial to develop CNAA further to extend the information scope of the generated results.

Finally, we decided to use DEA as benchmark for CNAA because both are deterministic methods. The basic DEA models were chosen since they are most used by practitioners. Nevertheless, a comparison with other DEA models could provide further insights, e.g., with additive DEA models, such as the slacks-based measure by Tone (2001) that is non-radial, non-oriented and takes account of all slacks. Moreover, future research could include other methods, such as SFA or StoNED.

Appendix (see Tables 12, 13, 14, 15)

Table 12 Mean rank correlation for the additional simulation subdivided by number of DMUs, input distribution, and returns to scale

Input distribution	Returns to scale	NAA ^{Max}				
		# of DMUs				
		25	50	100	200	400
Uniform	DRS	0.472	0.486	0.490	0.495	0.497
	CRS	0.484	0.496	0.502	0.508	0.510
	IRS	0.485	0.499	0.505	0.511	0.513
Beta	DRS	0.592	0.604	0.611	0.614	0.617
	CRS	0.590	0.602	0.610	0.614	0.616
	IRS	0.587	0.599	0.605	0.610	0.611

Average over the scenarios with: average level of efficiency: 70%/80%/90%; inefficiency term: half-normally/gamma distributed; level of noise: none/little/medium/strong

Table 13 Mean Rank Correlation for the additional simulation subdivided by number of DMUs, level and distribution of (in)efficiency, and level of noise

Distribution of the inefficiency term	Average level of efficiency %	Level of noise	NAA ^{Max}				
			# of DMUs				
			25	50	100	200	400
Half-normal	70	None	0.868	0.880	0.890	0.892	0.895
		Little	0.841	0.857	0.865	0.869	0.872
		Medium	0.782	0.794	0.805	0.809	0.812
		Strong	0.640	0.651	0.655	0.662	0.662
	80	None	0.756	0.771	0.778	0.784	0.787
		Little	0.711	0.727	0.736	0.741	0.745
		Medium	0.628	0.640	0.646	0.651	0.654
		Strong	0.454	0.465	0.472	0.474	0.477
	90	None	0.501	0.515	0.527	0.529	0.531
		Little	0.447	0.459	0.465	0.472	0.473
		Medium	0.357	0.363	0.370	0.374	0.376
		Strong	0.231	0.239	0.234	0.241	0.242
Gamma	70	None	0.766	0.784	0.795	0.800	0.803
		Little	0.731	0.747	0.755	0.761	0.765
		Medium	0.642	0.662	0.667	0.672	0.675
		Strong	0.475	0.486	0.494	0.496	0.497
	80	None	0.611	0.627	0.637	0.644	0.646
		Little	0.560	0.575	0.582	0.587	0.591
		Medium	0.458	0.477	0.479	0.485	0.486
		Strong	0.314	0.318	0.321	0.325	0.328
	90	None	0.361	0.376	0.382	0.386	0.389
		Little	0.320	0.324	0.332	0.333	0.334
		Medium	0.242	0.249	0.254	0.257	0.256
		Strong	0.149	0.155	0.159	0.159	0.160

Average over the scenarios with: inputs: uniformly/beta distributed; returns to scale: DRS/CRS/IRS

Table 14 Mean absolute deviation for the additional simulation subdivided by number of DMUs, input distribution, and returns to scale

Input distribution	Returns to scale	NAA ^{Max}				
		# of DMUs				
		25	50	100	200	400
Uniform	DRS	0.129	0.141	0.163	0.189	0.216
	CRS	0.126	0.138	0.160	0.187	0.215
	IRS	0.126	0.136	0.159	0.187	0.216
Beta	DRS	0.112	0.114	0.124	0.141	0.161
	CRS	0.112	0.114	0.124	0.141	0.161
	IRS	0.113	0.114	0.126	0.142	0.163

Average over the scenarios with: average level of efficiency: 70%/80%/90%; inefficiency term: half-normally/gamma distributed; level of noise: none/little/medium/strong

Table 15 Mean absolute deviation for the additional simulation subdivided by number of DMUs, level and distribution of (in)efficiency, and level of noise

Distribution of the inefficiency term	Average level of efficiency %	Level of noise	NAA ^{Max}				
			# of DMUs				
			25	50	100	200	400
Half-normal	70	None	0.093	0.073	0.070	0.080	0.097
		Little	0.094	0.078	0.081	0.094	0.112
		Medium	0.103	0.098	0.109	0.128	0.152
		Strong	0.148	0.167	0.193	0.224	0.255
	80	None	0.079	0.075	0.086	0.103	0.125
		Little	0.085	0.085	0.099	0.120	0.144
		Medium	0.101	0.113	0.137	0.164	0.192
		Strong	0.171	0.203	0.241	0.277	0.315
	90	None	0.072	0.089	0.113	0.139	0.165
		Little	0.082	0.104	0.132	0.160	0.189
		Medium	0.113	0.143	0.178	0.212	0.245
		Strong	0.205	0.252	0.301	0.344	0.383
Gamma	70	None	0.147	0.110	0.086	0.076	0.076
		Little	0.140	0.106	0.088	0.083	0.087
		Medium	0.132	0.109	0.103	0.108	0.121
		Strong	0.154	0.158	0.174	0.199	0.226
	80	None	0.100	0.084	0.082	0.090	0.104
		Little	0.099	0.090	0.093	0.106	0.123
		Medium	0.109	0.111	0.126	0.147	0.172
		Strong	0.169	0.196	0.229	0.265	0.299
	90	None	0.074	0.087	0.107	0.131	0.154
		Little	0.083	0.102	0.126	0.152	0.179
		Medium	0.112	0.140	0.173	0.205	0.236
		Strong	0.205	0.251	0.296	0.340	0.379

Average over the scenarios with: inputs: uniformly/beta distributed; returns to scale: DRS/CRS/IRS

Acknowledgements We would like to thank three anonymous reviewers for their helpful comments.

Funding Open Access funding enabled and organized by Projekt DEAL. This paper is based on a project funded by the German Federal Ministry of Education and Research under grant number 16NISTA03. The responsibility for the content lies with the authors.

Data availability We used R for the implementation of the simulation design. The code is available upon request from the corresponding author.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahn H, Dyckhoff H, Gilles R (2007) Datenaggregation zur Leistungsbeurteilung durch ranking: Vergleich der CHE- und DEA-Methodik sowie Ableitung eines Kompromissansatzes. *Z Betriebswirtsch* 77:615–643. <https://doi.org/10.1007/s11573-007-0046-z>
- Ahn H, Höfer-Diehl Y, Neumann L, Vazquez Novoa N (2012) Fakultätsübergreifende Beurteilung der Lehrleistung: Eine Methodik zum Vergleich der Lehreffektivität und Lehreffizienz von Studiengängen. *Hochschulmanagement* 7:110–115
- Ahn H, Clermont M, Langner J (2022) The impact of selected input and output factors on measuring research efficiency of university research fields: insights from a purpose-, field-, and method-specific perspective. *J Bus Econ* 92:1303–1335. <https://doi.org/10.1007/s11573-022-01099-4>
- Ahn H, Clermont M, Langner J (2023) Comparative performance analysis of frontier-based efficiency measurement methods: a Monte Carlo simulation. *Eur J Oper Res* 307:294–312. <https://doi.org/10.1016/j.ejor.2022.09.039>
- Aigner D, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. *J Econ* 6:21–37. [https://doi.org/10.1016/0304-4076\(77\)90052-5](https://doi.org/10.1016/0304-4076(77)90052-5)
- Andor M, Hesse F (2014) The StoNED age: The departure into a new era of efficiency analysis? A Monte Carlo comparison of StoNED and the “oldies” (SFA and DEA). *J Prod Anal* 41:85–109. <https://doi.org/10.1007/s11123-013-0354-y>
- Andor MA, Parmeter CF, Sommer S (2019) Combining uncertainty with uncertainty to get certainty? Efficiency analysis for regulation purposes. *Eur J Oper Res* 274:240–252. <https://doi.org/10.1016/j.ejor.2018.10.007>
- Badunenko O, Henderson DJ, Kumbhakar SC (2012) When, where and how to perform efficiency estimation. *J R Stat Soc A Stat Soc* 175:863–892. <https://doi.org/10.1111/j.1467-985X.2011.01023.x>
- Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manage Sci* 30:1078–1092. <https://doi.org/10.1287/mnsc.30.9.1078>
- Banker RD, Gadh VM, Gorr WL (1993) A Monte Carlo comparison of two production frontier estimation methods: corrected ordinary least squares and data envelopment analysis. *Eur J Oper Res* 67:332–343. [https://doi.org/10.1016/0377-2217\(93\)90289-Y](https://doi.org/10.1016/0377-2217(93)90289-Y)

- Bornmann L, Wohlrabe K, De Moya AF (2017) Calculating the excellence shift: how efficiently do institutions produce highly cited papers? *Scientometrics* 112:1859–1864. <https://doi.org/10.1007/s11192-017-2446-3>
- Büschken J (2009) When does data envelopment analysis outperform a naïve efficiency measurement model? *Eur J Oper Res* 192:647–657. <https://doi.org/10.1016/j.ejor.2007.09.026>
- Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. *Eur J Oper Res* 2:429–444. [https://doi.org/10.1016/0377-2217\(78\)90138-8](https://doi.org/10.1016/0377-2217(78)90138-8)
- Cobb CW, Douglas PH (1928) A theory of production. *Am Econ Rev* 18:139–165
- Cooper WW, Seiford LM, Tone K (2007) Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software, 2nd edn. Springer, New York
- Cordero JM, Polo C, Santín D (2020) Assessment of new methods for incorporating contextual variables into efficiency measures: a Monte Carlo simulation. *Oper Res Int J* 20:2245–2265. <https://doi.org/10.1007/s12351-018-0413-2>
- Di Giorgio L, Flaxman AD, Moses MW, Fullman N, Hanlon M, Conner RO, Wollum A, Murray CJL (2016) Efficiency of health care production in low-resource settings: a Monte-Carlo simulation to compare the performance of data envelopment analysis, stochastic distance functions, and an ensemble model. *PLoS ONE* 11:e0147261. <https://doi.org/10.1371/journal.pone.0147261>
- Dyson RG, Allen R, Camanho AS, Podinovski VV, Sarrico CS, Shale EA (2001) Pitfalls and protocols in DEA. *Eur J Oper Res* 132:245–259. [https://doi.org/10.1016/S0377-2217\(00\)00149-1](https://doi.org/10.1016/S0377-2217(00)00149-1)
- Emrouznejad A, Yang G, Khoveyni M, Michali M (2022) Data envelopment analysis: recent developments and challenges. In: Salhi S, Boylan J (eds) *The palgrave handbook of operations research*. Springer International Publishing, Cham, pp 307–350. https://doi.org/10.1007/978-3-030-96935-6_10
- Farrell MJ (1957) The measurement of productive efficiency. *J Roy Stat Soc Ser A (Gen)* 120:253–290. <https://doi.org/10.2307/2343100>
- Freudenberg M (2003) Composite indicators of country performance: a critical assessment. OECD Science, technology and industry working papers 2003/16, OECD, Paris. <https://doi.org/10.1787/405566708255>
- Friedl G, Hammer C, Pedell B, Küpper H-U (2009) How do German companies run their cost accounting systems? *Manag Account Q* 10:38–52
- Giraleas D, Emrouznejad A, Thanassoulis E (2012) Productivity change using growth accounting and frontier-based approaches: evidence from a Monte Carlo analysis. *Eur J Oper Res* 222:673–683. <https://doi.org/10.1016/j.ejor.2012.05.015>
- Gong B-H, Sickles RC (1992) Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. *J Econom* 51:259–284. [https://doi.org/10.1016/0304-4076\(92\)90038-S](https://doi.org/10.1016/0304-4076(92)90038-S)
- Greene WH (1980) Maximum likelihood estimation of econometric frontier functions. *J Econom* 13:27–56. [https://doi.org/10.1016/0304-4076\(80\)90041-X](https://doi.org/10.1016/0304-4076(80)90041-X)
- Gstach D (1999) Bounded vs. unbounded noise in efficiency estimation: performance of alternative estimators. In: Westermann G (ed) *Data envelopment analysis in the service sector*. Deutscher Universitäts-Verlag, Wiesbaden, pp 103–119. https://doi.org/10.1007/978-3-663-08343-6_6
- Holland D, Lee S (2002) Impacts of random noise and specification on estimates of capacity derived from data envelopment analysis. *Eur J Oper Res* 137:10–21. [https://doi.org/10.1016/S0377-2217\(01\)00087-X](https://doi.org/10.1016/S0377-2217(01)00087-X)
- Johnes J (2006) Data envelopment analysis and its application to the measurement of efficiency in higher education. *Econ Educ Rev* 25:273–288. <https://doi.org/10.1016/j.econedurev.2005.02.005>
- Kerpen P (2016) *Praxisorientierte data envelopment analysis*. Springer Gabler, Wiesbaden. <https://doi.org/10.1007/978-3-658-15460-8>
- Krüger JJ (2012) A Monte Carlo study of old and new frontier methods for efficiency measurement. *Eur J Oper Res* 222:137–148. <https://doi.org/10.1016/j.ejor.2012.04.026>
- Kuosmanen T, Kortelainen M (2012) Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *J Prod Anal* 38:11–28. <https://doi.org/10.1007/s1123-010-0201-3>
- Meeusen W, Van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. *Int Econ Rev* 18:435–444. <https://doi.org/10.2307/2525757>
- Muñiz M, Paradi J, Ruggiero J, Yang Z (2006) Evaluating alternative DEA models used to control for non-discretionary inputs. *Comput Oper Res* 33:1173–1183. <https://doi.org/10.1016/j.cor.2004.09.007>

- Nieswand M, Seifert S (2018) Environmental factors in frontier estimation: a Monte Carlo analysis. *Eur J Oper Res* 265:133–148. <https://doi.org/10.1016/j.ejor.2017.07.047>
- Oh S-C, Shin J (2015) The impact of mismeasurement in performance benchmarking: a Monte Carlo comparison of SFA and DEA with different multi-period budgeting strategies. *Eur J Oper Res* 240:518–527. <https://doi.org/10.1016/j.ejor.2014.07.026>
- Ruggiero J (1998) Non-discretionary inputs in data envelopment analysis. *Eur J Oper Res* 111:461–469. [https://doi.org/10.1016/S0377-2217\(97\)00306-8](https://doi.org/10.1016/S0377-2217(97)00306-8)
- Ruggiero J (2007) A comparison of DEA and the stochastic frontier model using panel data. *Int Trans Oper Res* 14:259–266. <https://doi.org/10.1111/j.1475-3995.2007.00585.x>
- Schaefer J, Dyckhoff H (2018) Performance of efficiency measurement by StoNED compared to DEA and SFA in cases of Leontief and by-production technologies. Unpublished Working Paper, RWTH Aachen University, School of Business and Economics. <https://doi.org/10.13140/RG.2.2.12511.20644>
- Schaefer J, Clermont M (2018) Stochastic non-smooth envelopment of data for multi-dimensional output. *J Prod Anal* 50:139–154. <https://doi.org/10.1007/s1123-018-0539-5>
- Spearman C (1904) The proof and measurement of association between two things. *Am J Psychol* 15:72–101. <https://doi.org/10.2307/1412159>
- Thanassoulis E (2001) Introduction to the theory and application of data envelopment analysis: a foundation text with integrated software. Kluwer Academic Publishers, Norwell. <https://doi.org/10.1007/978-1-4615-1407-7>
- Tone K (2001) A slacks-based measure of efficiency in data envelopment analysis. *Eur J Oper Res* 130:498–509. [https://doi.org/10.1016/S0377-2217\(99\)00407-5](https://doi.org/10.1016/S0377-2217(99)00407-5)
- Verheyen J, Dyckhoff H, Clermont M (2023) Performance comparison of efficiency measurement methods: an analysis for Leontief technologies and coupled production. *Betriebswirtschaftliche Forschung und Praxis* 75:679–695
- Winsten CB (1957) Discussion on Mr. Farrell's paper. *J Roy Statist Soc Ser A (Gen)* 120:282–284
- Yu C (1998) The effects of exogenous variables in efficiency measurement: a Monte Carlo study. *Eur J Oper Res* 105:569–580. [https://doi.org/10.1016/S0377-2217\(97\)00076-3](https://doi.org/10.1016/S0377-2217(97)00076-3)

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