Water Supply

Water Supply Vol 23 No 4, 1700 doi: 10.2166/ws.2023.093

Investigation of water utility efficiency features: production variables and the DEA model choices

Dickson K. Gidion

Ministry of Water, Water Institute, Off Sam Nujoma Road, P.O. Box 35059, Dar es Salaam, Tanzania
E-mail: dgidion@gmail.com

DKG, 0000-0003-1843-2736

ABSTRACT

Data envelopment analysis (DEA) has been used for decades to investigate utility efficiency- and management-specific features. DEA has been extended in various industries with different models that consume different types of production variables to meet industry-specific efficiencies. However, the discernment of production variables and proper DEA models is still a problem in the water industry. Ratio information has been used in DEA for absolute data to assess utility efficiency- and management-specific features; using ratios in DEA for absolute data produces wrong efficiency results because the convexity axiom of DEA for absolute data is not satisfied. This study presents a collection of information and an assessment of DEA for absolute and ratio data to identify the strengths of the production variables and the DEA model that are appropriate for assessing utility efficiency- and management-specific features. The analysis shows that the use of ratio variables in DEA for ratios improves utility economics, the working environment, and the social trust dimension.

Key words: DEA for absolute data, DEA for ratios, DEA model selection, efficiency assessment, production variables choices, water utility regulation

HIGHLIGHTS

- DEA models and production variable selection.
- Proper utility benchmarking and DEA model choices.
- Efficient utility benchmarking.
- Using performance indicators as production variables in DEA.
- Efficiency comparison.

1. INTRODUCTION

The quality of utility management is determined by the achievement of targets and service level benchmarks settings through regulated water tariffs. Improved revenue collection efficiency indicates non-revenue water (NRW) reduction and good water service delivery. Reducing NRW improves a utility's operating environment, economics, and most importantly, social trust (Gidion *et al.* 2019a; AL-Washali *et al.* 2020). To simulate the quality of services provided in a non-competitive operating environment (Pinto *et al.* 2017b), the performances of urban water utilities (UWUs) are compared using production variables referred to as key performance indicators (KPIs) (Singh *et al.* 2014; Thanassoulis & Silva 2018; Gidion *et al.* 2019b). Efficiency measurement involves the choice of a parametric or non-parametric benchmarking method (Lampe & Hilgers 2015; Gidion *et al.* 2022). Some parametric methods, such as stochastic frontier analysis (SFA) use a set of production variables (input and output), as does data envelopment analysis (DEA). It is well known that Charnes *et al.* (1978) and Banker *et al.* (1984) developed efficiency benchmarking methods with constant returns to scale (CRS) and variable returns to scale (VRS), respectively, which have been extended in various forms based on efficiency frontiers required by comparable decision-making units (DMUs). The extensions of DEA are classified mainly based on the production variables used, the technologies employed, and the number of analysis stages. The production variables can be categorised as absolute numbers or as ratios if the denominators of the variables are different (Hollingsworth & Smith 2003; Hatami-Marbini & Toloo 2019).

The majority of publications available in the water industry use input-oriented DEA for absolute data (Pinto *et al.* 2017a; Cabrera *et al.* 2018), which are available in CRS and VRS, and a few publications use DEA for ratio data (Gidion *et al.* 2019a), which are only available in VRS (Hollingsworth & Smith 2003). Recently, efficiency analysis using DEA has gained prominence in the water industry and is used to improve utilities' management capabilities (Pastor & Aparicio 2010). However, a few researchers use undesirable production variables to benchmark a DMU in the developed DEA models. For example, a recent publication by Lombardi *et al.* (2019) and lo Storto (2020) used absolute and ratio variables to generate utility efficiency using CRS and VRS DEA models developed to analyse absolute numbers, respectively.

Although a study by Romano & Guerrini (2011) employed DEA for absolute data to determine the efficiency of variables most often used to generate utility efficiency, UWU regulators use absolute data to a limited extent to rank the performance of a UWU (Berg 2010; Mugisha 2011). Regulators use KPIs to rank utility efficiency, which are ratio data in the efficiency measurement industry (Thanassoulis & Silva 2018). In general, the majority of studies extend DEA for absolute data to examine management features (Lombardi *et al.* 2019; lo Storto 2020). Using absolute numerical data to assess efficiency does not reflect the practical efficiency of a UWU (Alegre *et al.* 2016); therefore, assessing utility efficiency using ratios highlights the significant operational differences among UWUs (Gidion *et al.* 2019a, 2019b; Gidion *et al.* 2022). This study extends the use of DEA for ratio data proposed by Emrouznejad & Amin (2009) and Hatami-Marbini & Toloo (2019) and compares the utility efficiencies generated using DEA for absolute and ratio data.

DEA for absolute and ratio data are never compared to assess its suitability in water service delivery analysis, which is why most recent publications use DEA for absolute data to measure efficiency in the water industry (Lombardi *et al.* 2019; Wu *et al.* 2019; Xie *et al.* 2020). Furthermore, the types of input and output variables used to assess the efficiency of a UWU in DEA for absolute numbers do not allow assessing UWUs of different operation scales together (Romano & Guerrini 2011; Guthrie *et al.* 2017; Gidion *et al.* 2019a). It is well known that a large-scale UWU provides more information that is not likely to be readily available in a small-scale UWU (Haider *et al.* 2013) and using KPIs available in both large and small UWUs provides sound information to rank UWUs of different scales together.

This study presents DEA for absolute numbers and ratios to assess UWU efficiency. DEA for ratios uses absolute and ratio variables (KPIs) as inputs/outputs. Empirical studies such as Romano & Guerrini (2011) and Lombardi *et al.* (2019) examined the sustainability of UWU in efficiency analysis and concluded that the efficiency of UWU differs among firms with dissimilar scale characteristics (Worthington 2014). Based on the benchmarking technology of DEA for ratios, this study has a different perspective, as the assessment shows that the methodology ignores scale groups during the analysis and provides homogeneous efficiency results through the same input/output variables used, removing the scale size limitation of UWUs. KPIs provide an equivalent assessment of utility performance (Vilanova *et al.* 2015; Alegre *et al.* 2016) and thus, provide the opportunity for small-scale UWUs to be compared with large-scale UWUs and apply common management strategies to improve UWU efficiency.

2. MATERIALS AND METHODS

2.1. Production variables of choice

During efficiency assessment, the number of measuring units should be more than three times the number of production variables to avoid DEA evaluation failure that produces efficiency results suffering from discrimination (Eskelinen 2017; Li *et al.* 2017); this means that even an underperforming unit can become efficient (Wagner & Shimshak 2007; Toloo *et al.* 2015). Thus, reducing the number of production variables to less than one-third of the measuring units reduces the chance that many units will become efficient. The selection of production variables should be done by experts because using unreliable or omitting relevant production inputs/outputs in the DEA analysis makes a large number of units efficient (Ruggiero 2005; Lombardi *et al.* 2019). Variable selection can be done using judgement screening, statistical tests, or DEA-based analysis to determine the importance of the original and influential input/output with the greatest impact (Golany & Roll 1989). However, in a limited environment, selecting production variables based on expert judgement alone is acceptable (Wagner & Shimshak 2007; Eskelinen 2017). Utility performance improvement depends on schemes that take into account the selection of input and output variables that focus directly on operating costs and, at the same time, on the quality of service provided; the majority of available publications in the water industry rely on production variables that have been widely used in previous studies, without understanding that time and technological changes represent a new working environment that nullifies the importance of some production variables. A few studies use correlation and regression analysis to determine the potential of selected production variables before DEA assessment. A highly correlated variable is considered redundant to avoid biased measurement (Ruggiero 2005; Wagner & Shimshak 2007; Lombardi *et al.* 2019).

The current study recommends the use of correlation analysis and analyses the efficiency of UWU using DEA for ratios. Thus, the omission of highly correlated variables depends on the assessment of the significance of the variable, because not all highly correlated variables are redundant, sometimes they are an indicator to examine a variable more closely (Golany & Roll 1989; Wagner & Shimshak 2007). Some researchers in the water industry have used ratios for assessing utility efficiency in DEA for absolute data. For example, Table 1 lists some publications that use DEA for absolute numbers (Equation (1)) and analyse production sets that include ratios and absolute numbers. The results analysed were used to derive observations and suggest measures to improve performance.

The confusion is due to the nature of a utility's operating environment, which provides information in the form of ratios to rank a utility's performance (Thanassoulis & Silva 2018) and DEA models in absolute data. The use of ratio variables in DEA for absolute data technically leads to incorrect results (Hollingsworth & Smith 2003; Emrouznejad & Amin 2009). In this study, absolute variables that make ratios are used to assess the efficiency of a utility using DEA for absolute data and DEA for ratios, respectively; the absolute and ratio information generated is verified based on the results of correlation analysis. In the case of more than one redundant variable, the study uses judgement and knowledge to determine the similarity and importance of the variables in relation to utility efficiency- and management-specific features.

2.2. Specific efficiency assessment DEA models

Using standard DEA for absolute numbers, Wagner & Shimshak (2007) proposed a stepwise variable selection approach that involves efficiency analysis by variables redundant with constant units till the number of efficient units is reduced with a reduced number of production variables. This study arguably again criticises the fact that practitioners cannot rely upon the proposed approach, because a utility cannot be measured using one input and output variable over several production sets of variables used during service delivery and conclude on managerial improvement. Gidion *et al.* (2019a) developed a network-DEA model for ratios analysing utility efficiency in the same way as Wagner & Shimshak (2007); however, the recent study benchmarks utility efficiency with units and production variables that are redundant. Gidion *et al.* (2019a) study included utilities with zero values in the production variables and benchmarked them together with utilities that outperformed targets in their production variables; the analysis benchmarks a utility under a yardstick competition regime while still allowing for the examination of efficiency- and management-specific features. The two studies are presented to illustrate an example of two different DEA models that use different production variables in their analysis. DEA for absolute numbers uses a linear programme technology, while DEA for ratios uses a non-linear programme technology to benchmark utility efficiency. DEA benchmarks utility efficiency in input or output orientations; the input DEA model aims to minimise input and increase outputs, while the output orientation aims to minimise output and maximise input (Charnes *et al.* 1978; Banker *et al.* 1984).

Equations (1) and (2) present VRS DEA models for absolute numbers and ratios, respectively, employed to compare the efficiencies of a UWU. Thus, consider a set of z UWUs, each consuming i input to generate r output in a limitation that $z \geq 3(m+s)$. Let $x_j = (x_{1j} \ldots\ldots, x_{mj})^T$ and $y_j = (y_{1j} \ldots\ldots, y_{sj})^T$, respectively, represent the vectors of m consumed input ratio(s) and s output ratios. $\bar{x}_{pj}$ and $\underline{x}_{pj}$ represent a numerator and denominator of the p^{th} input (x_{pj}), and $\bar{y}_{kj}$ and $\underline{y}_{kj}$ represent a numerator and denominator of the k^{th} output (y_{kj}) for UWU_j. In the DEA model, x_{i0} and y_{r0} composed of the known input ratio and output-ratio vectors, respectively, of the target UWU_0, λ is a vector describing the percentages of other producers used for constructing the virtual producer, and θ is the producer's efficiency score.

Table 1 | Publications identified to use DEA for absolute data in analysing ratio variables

S/N	Publication	Total production variables employed	Ratio variables used
1	Singh *et al.* (2014)	4	2
2	Lombardi *et al.* (2019)	10	1
3	lo Storto (2020)	12	3
4	Cheng *et al.* (2020)	4	3
5	Le *et al.* (2022)	14	8
6	Robles-Velasco *et al.* (2022)	7	1

DEA for absolute model

$$
\begin{aligned}
E_0 = \ & \min \theta_0 \\
\text{s.t:} \\
& \sum_{j=1}^{z} x_{ij}\lambda_j - x_{ij_0}\theta_0 \le 0, \quad i = 1, \ldots, m \\
& \sum_{j=1}^{z} y_{rj}\lambda_j \ge y_{rj_0}, \quad r = 1, \ldots, s \\
& \sum_{j=1}^{z} \lambda_j = 1 \\
& \lambda_j \ge 0, \quad j = 1, \ldots, z
\end{aligned}
\tag{1}
$$

DEA for ratios model

$$
\begin{aligned}
E_0 = \ & \min \theta_0 \\
\text{s.t} \\
& \sum_{j=1}^{z} \lambda_j x_{ij} \le \theta_0 x_{i0}; \quad i = 1, \ldots, m; \quad i \ne p \\
& \sum_{j=1}^{z} \lambda_j y_{rj} \ge y_{r0}; \quad r = 1, \ldots, s; \quad r \ne k \\
& \sum_{j=1}^{z} \lambda_j (\bar{x}_{pj} - x_{p0}\underline{x}_{pj}) \le 0; \quad i = p \\
& \sum_{j=1}^{z} \lambda_j (\bar{y}_{kj} - y_{k0}\underline{y}_{kj}) \ge 0; \quad r = k \\
& x_{p0} = \frac{\bar{x}_{pj}}{\underline{x}_{pj}}, \quad y_{k0} = \frac{\bar{y}_{kj}}{\underline{y}_{kj}} \\
& \sum_{j=1}^{z} \lambda_j = 1; \quad \lambda_j \ge 0; \quad j = 1, \ldots, z
\end{aligned}
\tag{2}
$$

These input-oriented DEA models have been used to investigate managerial features in drinking water companies, while the output-oriented DEA models are mostly used to measure the efficiency of wastewater companies. The VRS form of DEA is mostly used to investigate the efficiency of UWUs; this is due to the nature of UWUs operating under VRS (Berg & Lin 2008; Berg 2010). In practice, DEA for absolute numbers has been extended in many forms compared to DEA for ratios; the standard DEA models developed by Charnes *et al.* (1978) and Banker *et al.* (1984) are the original models. Studying the efficiency of a UWU requires an appropriate DEA model that examines the efficiency- and managerial-specific features, wrong models have been used to investigate managerial efficiencies (Singh *et al.* 2014; Lombardi *et al.* 2019). In this study, the standard input-oriented VRS DEA models for absolute and ratio data are used to ascertain their features and reliability in benchmarking utility efficiency; the two kinds of DEA models have been previously employed in UWU efficiency assessment. However, in a limited condition, DEA for ratio data have used. This study analyses a utility efficiency using DEA for absolute data developed by Banker *et al.* (1984) and DEA for ratio data extended by Emrouznejad & Amin (2009) and later modified by Hatami-Marbini & Toloo (2019).

2.3. Materials/data

The study focuses on Tanzanian regional utilities to produce a set of utility characteristics that significantly identify the proper DEA model suitable to assess and investigate efficiency- and management-specific features of a utility. This study employs the same data used in the Gidion *et al.* (2019b) study, which analysed 32 utilities with 10 production variables. Due to the reason that the assessment focuses on comparing DEA for absolute numbers and ratios, the analysis considers two kinds of data to investigate UWU's efficiencies- and management-specific features. Gidion *et al.* (2019b) used information published by the EWURA (2018). EWURA (Energy and Water Utilities Regulatory Authority) is an autonomous multi-sectoral regulatory authority established under Act Cap 414′ of the laws of Tanzania. Recently, the EWURA monitored the performance of 94 utilities that are geographically distributed and publicly owned, as shown in Tables 2 and 3.

The regulator uses empirical analysis to rank the performance of 94 utilities; during ranking, regional and national projects UWUs are ranked together, whereas district towns UWUs are ranked separately. District towns UWUs are considered small-scale UWUs and regional and national projects UWUs are considered large-scale UWUs. In our analysis, we are comparing the efficiencies of regional and national projects UWUs to investigate the quality of DEA for absolute and ratio variables in the UWU efficiency investigation. The KPIs used by the EWURA to rank UWUs are also published in the Alegre *et al.* (2016) study; using the input/output (I/O) coding provided in this study to simplify the analysis, Alegre *et al.* (2016) defined the KPIs employed in Gidion *et al.* (2019b) as follows:

- (I)A: NRW (%) is the ratio of the difference between the volume of water produced and the actual revenue volume of water to the volume of water produced by a utility. The difference between the volume of water produced and the actual revenue volume of water is referred to as an absolute number.
- (I)B: Working ratio (–) is the ratio of operational expenses to operational revenue. To generate a working ratio, depreciation, interests, and debit services are *excluded* from the operational expenses. Operational expenses are an absolute variable.
- (I)C: Operating ratio (–) is the ratio of operating costs to operational revenues. To generate an operating ratio, depreciation, interests, and debit services are *included* in the analysis of the operational costs. The operating revenue is an absolute variable.
- (I)D: Personnel expenditure (%) is the personnel expenditure expressed as a percentage of total revenue collection. Personnel expenditure is an absolute variable.
- (I)E: Staff/1,000 connections (FTE/1,000 Con.) is an indicator that measures the ratio of staff distribution per available connections/customers. The number of staff is an absolute variable.
- (O)A: Proportion of the population served with water (%) is the percentage of the population connected to water service compared to the total population of the area with water service extension. Population served is an absolute variable.
- (O)B: Average hours of supply (%) is the percentage of average daily hours used to pressurise the system. The actual average water supply hours are counted as an absolute variable.
- (O)C: Metering ratio (%) is the ratio of metered customers to total customers registered in a utility database expressed in percentage. The number of active metered connections is an absolute variable.

Table 2 | Geographic distribution of Tanzanian UWUs (Source: EWURA 2021)

Current zone	No. of utilities	Percentage of utilities
Eastern	16	17
Southern	24	26
Northern	18	19
Lake	25	26
Central	11	12

Table 3 | Regulated utilities by public ownership (Source: EWURA 2021)

UWU category	No. of utilities	Percentage of utilities
Regional	26	28
National projects[a]	7	7
District towns	61	65

[a]Delivering water service to more than one district or town.

- (O)D: Revenue collection efficiency (%) is the ratio of revenue collected to the billed amount expressed in percentage. The revenue collected is counted as an absolute variable.
- (O)E: Water quality compliance (%) is the ratio of the number of samples that have passed a water quality test to the total samples collected. The number of samples that passed the test is an absolute variable.

The numerator of the KPI is an absolute number used in DEA for absolute data analysis, and the ratio variable that results in KPI analysis is the data used in DEA for ratio analysis. Because DEA for absolute numbers uses part of the information used in DEA for ratios, the study applies correlation analysis generated from the KPIs to identify redundant variables in the analysis, as previously discussed. The information obtained from the correlation analysis of the KPI variables is presented in Table 4.

3. RESULTS AND DISCUSSIONS

3.1. Dataset

Table 4 presents the correlation matrix for 10 ratio variables; the analysis shows that the correlations between input-to-input and output-to-output correlations are positive, while the correlations between inputs and outputs are negative. The positive observations signify that the independent and dependent variables increase in value simultaneously, while the negative observation signifies that the independent variable increases while the dependent variable decreases in value (Clemen & Reilly 1999; Yang *et al.* 2019). The correlation identity guide from Yang *et al.* (2019) helps to determine the degree of correlation of the variables analysed. If the correlation degree is 0.000, it means that the correlation degree is independent, if the correlation value is between 0.000 and 0.300, the correlation degree is a micro correlation. If the correlation coefficient is between 0.300 and 0.500, the correlation degree is termed a real correlation, if the correlation coefficient is between 0.500 and 0.800, the correlation degree is a significant correlation. Note that the correlation range is between $+1$ and -1, if the coefficients range above 0.800 is termed as highly correlated. Referring to the analysis in Table 4, the majority of the correlations of the variables fall within the micro- and real-correlation degrees. None of the variables are highly correlated. However, variable (O)D, which is the revenue collection efficiency should be considered redundant because it is the opposite of NRW; reducing NRW will increase revenue collection efficiency and vice versa. Furthermore, variables (I)B and (I)C use the same numerator name differentiated with either depreciation, interests, and debit services are included or excluded thus, to avoid confusion in the analysis variable (I)C is counted as redundant. Thus, our new variables to use in DEA for absolute numbers and ratios are indicated in Table 5.

3.2. Efficiency results in comparison

3.2.1. All ratio data in DEA for absolute numbers and ratios

The production variables in Table 5 were employed to assess utility efficiency. The two forms of DEA resulted in two different charts, which are shown in Figure 1, DEA for absolute numbers and ratios produced 27 and 20 efficient UWUs, respectively, after the analysis. This is an indicator of benchmarking an underperforming UWU as efficient/inefficient when ratios are used

Table 4 | Variables correlation matrix

KPI	(I)A	(I)B	(I)C	(I)D	(I)E	(O)A	(O)B	(O)C	(O)D	(O)E
(I)A	1									
(I)B	0.474	1								
(I)C	0.290	0.755	1							
(I)D	0.409	0.192	0.374	1						
(I)E	0.491	0.529	0.441	0.138	1					
(O)A	0.140	− 0.212	− 0.203	− 0.011	− 0.295	1				
(O)B	− 0.466	− 0.188	− 0.161	− 0.436	− 0.127	0.248	1			
(O)C	− 0.381	− 0.068	− 0.173	− 0.538	− 0.219	0.210	0.417	1		
(O)D	− 0.039	− 0.091	− 0.336	− 0.156	− 0.225	0.352	0.250	0.381	1	
(O)E	− 0.686	− 0.531	− 0.471	− 0.175	− 0.68	.014	0.254	0.46	0.305	1

Table 5 | Production variables for DEA efficiency

No.	DEA for absolute: Input	DEA for absolute: Output	DEA for ratio: Input	DEA for ratio: Output
1.	NRW (m^3)	Population served with water (number.)	NRW (%)	The proportion of the population served with water (%)
2.	Average operational expenses (TZS)	Average hours of supply (h)	Working ratio (–)	Average hours of supply (%)
3.	Average personnel expenditure (TZS)	Metered customers (Number.)	Personnel expenditure (%)	Metering ratio (%)
4.	Staff (Number.)	Average of tested samples (Number.)	Staff/1,000 connections (FTE/1,000 Con.)	Water quality compliance (%)

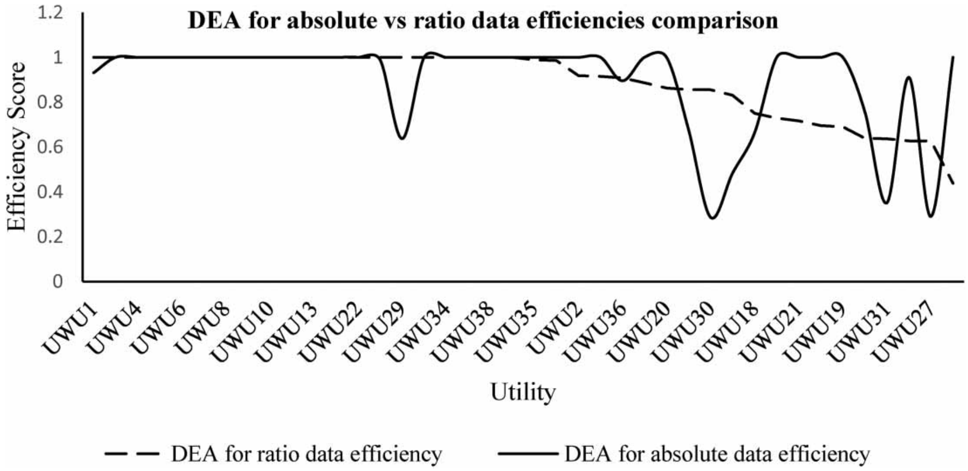

Figure 1 | Comparison of efficiencies of DEA for ratio and absolute data.

in DEA for absolute numbers; this means that the convexity problem can increase or decrease the number of efficient UWUs and lead to wrong recommendations on efficiency- and management-specific features. In the analysis in Figure 1, absolute numbers and their ratio were used to generate the two different efficiencies of the same UWU; some UWUs scored the same in both models, while others did not. For example, UWU1 was found to be efficient when using absolute and ratio data when analysed with DEA for ratios and inefficient when the ratio variables were analysed with DEA for absolute data. The same efficient results in DEA for ratios were observed in UWU_7, UWU_{26}, and UWU_{29}. Conversely, the seven UWUs were benchmarked as inefficient in using DEA for ratios while efficient in using DEA for absolute numbers, this is due to the DEA convexity assumptions being satisfied or not satisfied by the two DEA forms in the analysis (Hatami-Marbini & Toloo 2019). Therefore, when ratio variables are used in DEA for absolute numbers as in the studies by Singh *et al.* (2014) and Lombardi *et al.* (2019), the DEA convex axiom is not fully satisfied, and the efficiency results are not correct and cannot be used to improve efficiency- and management-specific features (Banker *et al.* 1984; Hollingsworth & Smith 2003; Emrouznejad & Amin 2009; Hatami-Marbini & Toloo 2019).

3.2.2. One ratio data in DEA for an absolute number

Figure 2 shows a comparison of two analyses in DEA for absolute data using production variables without ratio information and a ratio variable. When NRW was used as a ratio variable, the efficiency assessment showed that the number of efficient UWUs increased from 27 to 28 and there were efficiency shifts to many of the benchmarked units (see the comparison in Figure 2). This is a poor indication of studies concluding efficiency- and managerial-specific features observed after analysis using one or more ratio variables in DEA for absolute data, DEA for ratios is a proper model to overcome an analysis problem and DEA for absolute numbers should not be used when a ratio variable is included in a production set. A suitable model for a

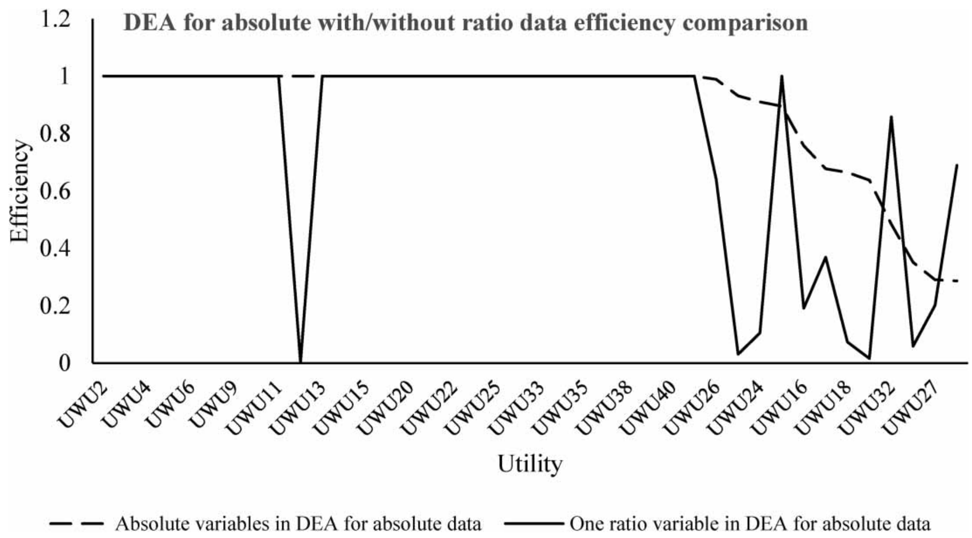

Figure 2 | Efficiency score from DEA for absolute analysed with absolute and one ratio variable.

ratio variable in an input/output production set can be obtained from the Hatami-Marbini & Toloo (2019) study. However, using absolute numbers to conclude on the improvement of management-specific features after a specific efficiency analysis does not reflect a utility economic comparison (Vilanova *et al.* 2015), so it is not possible to retrieve feasible best practice measures from the best performer to improve the weak-performing utility (Marques & Monteiro 2001). The DEA for ratios proposed in this study is used when absolute and ratio data of a KPI are available.

3.3. DEA for ratios: significance in utility management

3.3.1. Sustainability of water service delivery

Investigating utilities management-specific features starts after comparing the degree of achievement of established performance indicators (PIs) and best practices. The International Water Association (IWA) and other water industry stakeholders have established several PIs, and regulators summarise some of them and rank utilities' performance. Regulators use alternative techniques that involve empirical analysis by using the PIs as input variables and identifying a utility with excellent, medium, and low performance on selected PIs in a competitive scenario (Davis 2011; Chen & Chen 2014). However, the approaches are prone to human error and favouritism (Gidion *et al.* 2019b). DEA for absolute data cannot use PIs as input variables, and when they are used in the analysis, the results are not reliable because the convexity axiom is not satisfied; DEA for ratio data overcomes the convexity problem when analysing UWU with ratio data (Hatami-Marbini & Toloo 2019). Aggregating UWUs and assessing their efficiency using PIs opens up the possibility for a researcher to identify a utility that outperforms the level of several PIs compared to other utilities, where a weak-performing utility will have a chance to learn how to improve the underlying poorly rated PIs. Regulators use the same approach to examine through empirical analysis. and propose measures for performance improvement. The efficiency assessment proposed in this study does not rely on assumptions, thus avoiding human errors. Figure 1 shows the trends of efficiency assessment using DEA for absolute and ratio data; DEA for ratios significantly outperformed DEA for absolute data in efficiency benchmarking and has added value to improve the economic dimension of a utility. Examining the level of PIs achieved through efficiency benchmarking and proposing measures to improve performance improves not only economic factors but also environmental and social trust factors, which are indicators of utility sustainability (Molinos-Senante *et al.* 2016).

3.3.2. Performance indicator significance in efficiency assessment

Multidimensional efficiency assessment can be achieved by aggregating a set of selected PIs, referred to as production variables; representativeness, relevance, reliability, sensitivity, understandability, comparability, and transparency are the criteria recommended in the PIs selection (Vilanova *et al.* 2015; Molinos-Senante *et al.* 2016). Due to the nature of the utility operating environment, which includes the provision of water services in an area without other competitors, the use of PIs in efficiency assessments provides holistic utility management in different environments (Marques & Monteiro 2001; Haider *et al.* 2016). The role of PIs in utility efficiency assessment remains important (Alegre *et al.* 2016; Molinos-Senante *et al.* 2016); management complexity is determined upfront after efficiency assessment by combining efficiency score and production variables (Xie *et al.* 2020). Based on the researcher's preferences, PIs are classified into different groups (Marques & Monteiro 2001; Molinos-Senante *et al.* 2016); however, based on the importance of sustainability of services, PIs are classified into economic, environmental, and social dimensions (Vilanova *et al.* 2015). Global technological progress is on the rise and the importance of economic management has improved the utility's operating environment and social trust. Advances in the design of pumps that operate optimally and cost-effectively are improving the operating environment. However, the increase of water losses in the system slows the efforts. Efficiency assessment has the general objective of improving performance in utility management, and efficiency scores and production variables quantify specific efficiency measures. While efficiency scores rank a utility's position within a group of DMUs, PIs are used to make decisions about the efficiency achieved and to verify the effectiveness of optimisation measures that an efficient utility uses to improve efficiency that are appropriate for an inefficient utility. In this way, the adaptation of management sharing techniques between utilities to improve efficiency is encouraged; the measure leads utilities to work together as a group without regard to the environmental barrier (Korhonen & Syrjänen 2004; Lejarraga & Müller-Trede 2017).

4. CONCLUSIONS

According to Berg (2010), regulators have used empirical methods to rank utility performance and identify specific management feature improvements; empirical methods are prone to human error. DEA is a non-parametric efficiency assessment method; human error does not exist during the analysis and the selection of the DEA model for UWU efficiency assessment depends on available production data. The assessment shows that the use of inappropriate variable data in the analysis leads to incorrect efficiency results that cannot be relied upon in decision-making. Production variables and choosing the right DEA model in efficiency assessment is one of the best approaches during efficiency assessment. Using ratio data in DEA for absolute numbers increases the number of efficient UWUs; thus, underperforming UWUs identified efficiently due to incorrect data selection can also be included in efficiency- and management-specific feature investigations.

The majority of available publications investigate UWU management features using DEA for absolute data. However, DEA for absolute data has been extended to analyse ratios, which overcomes the convexity problem in DEA for absolute data when ratio data are used. The assessment shows that DEA for ratios achieves efficiency gains that can be relied upon. The use of DEA for ratios helps researchers and regulators identify efficiency- and management-specific features based on efficiency outcomes and production sets (KPIs) that are used. DEA for ratio data has been extended in the water industry and compares the efficiency of UWUs in a competitive scenario, as the empirical analysis does (Gidion *et al.* 2019a, 2019b, 2022).

The assessment of this study is limited to variables and DEA model choices; the remark helps researchers and practitioners to conclude the best management features for efficient utility management. Future researchers should explore efficiency- and management-specific features using proper production variables and the DEA model.

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories. https://data.mendeley.com/datasets/2mz55y3b4w.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Alegre, H., Baptista, J. M., Cabrera, E., Cubillo, F., Duarte, P., Hirner, W., Merkel, W. & Parena, R. 2016 *Performance Indicators for Water Supply Services*. IWA Publishing, London. https://doi.org/10.2166/9781780406336.

AL-Washali, T. M., Elkhider, M. E., Sharma, S. K. & Kennedy, M. D. 2020 A review of nonrevenue water assessment software tools. *WIRES Water* **7** (2), e1413. https://doi.org/10.1002/wat2.1413.

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

Berg, S. 2010 *Water Utility Benchmarking Measurement, Methodologies, and Performance Incentives*. IWA Publishing, London. https://doi.org/10.2166/9781780401690.

Berg, S. & Lin, C. 2008 Consistency in performance rankings: the Peru water sector. *Applied Economics* **40** (6), 793–805. http://dx.doi.org/10.1080/00036840600749409.

Cabrera, E., Estruch-Juan, E. & Molinos-Senante, M. 2018 Adequacy of DEA as a regulatory tool in the water sector. The impact of data uncertainty. *Environmental Science & Policy* **85**, 155–162. https://doi.org/10.1016/j.envsci.2018.03.028.

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

Chen, Y.-T. & Chen, C.-C. 2014 An analysis of domestic water management performance across regions in Taiwan. *Water Policy* **16** (4), 704–719. https://doi.org/10.2166/wp.2014.052.

Cheng, P., Jin, Q., Jiang, H., Hua, M. & Ye, Z. 2020 Efficiency assessment of rural domestic sewage treatment facilities by a slacked-based DEA model. *Journal of Cleaner Production* **267**, 122111. https://doi.org/10.1016/j.jclepro.2020.122111.

Clemen, R. T. & Reilly, T. 1999 Correlations and copulas for decision and risk analysis. *Management Science* **45** (2), 208–224. https://doi.org/10.1287/mnsc.45.2.208.

Davis, C. 2011 Water utility benchmarking: measurement, methodologies, and performance incentives. *Water Policy* **13** (6), 895–896. https://doi.org/10.2166/wp.2011.071.

Emrouznejad, A. & Amin, G. R. 2009 DEA models for ratio data: convexity consideration. *Applied Mathematical Modelling* **33** (1), 486–498. https://doi.org/10.1016/j.apm.2007.11.018.

Eskelinen, J. 2017 Comparison of variable selection techniques for data envelopment analysis in a retail bank. *European Journal of Operational Research* **259** (2), 778–788. https://doi.org/10.1016/j.ejor.2016.11.009.

EWURA 2018 *Annual Performance Reports*. Available from: http://www.ewura.go.tz/?page_id=1109 (accessed 30 March 2018).

EWURA 2021 *General Information*. EWURA. Available from: https://www.ewura.go.tz/water-and-sanitation-2/ (accessed 28 December 2021).

Gidion, D. K., Hong, J., Adams, M. Z. A. & Khoveyni, M. 2019a Network DEA models for assessing urban water utility efficiency. *Utilities Policy* **57**, 48–58. https://doi.org/10.1016/j.jup.2019.02.001.

Gidion, D. K., Hong, J., Hou, B., Shillamkwese, S. S., Adams, M. Z. A. & Khoveyni, M. 2019b Consistency in efficiency benchmarking: urban water utility regulation with performance improvement. *Water Supply* **19** (7), 1955–1964. https://doi.org/10.2166/ws.2019.072.

Gidion, D. K., Karia, A. O., Nyangi, M. J., Mwamila, T. B., Senkondo, W., Mmasi, D. B., Alphayo, S. & Khoveyni, M. 2022 Stepwise efficiency assessment in a competitive scenario using DEA for ratios: service level benchmark. *Water Supply*. https://doi.org/10.2166/ws.2022.351.

Golany, B. & Roll, Y. 1989 An application procedure for DEA. *Omega* **17** (3), 237–250. https://doi.org/10.1016/0305-0483(89)90029-7.

Guthrie, L., De Silva, S. & Furlong, C. 2017 A categorisation system for Australia's integrated urban water management plans. *Utilities Policy* **48**, 92–102. https://doi.org/10.1016/j.jup.2017.08.007.

Haider, H., Sadiq, R. & Tesfamariam, S. 2013 Performance indicators for small- and medium-sized water supply systems: a review. *Environmental Reviews* **22** (1), 1–40. https://doi.org/10.1139/er-2013-0013.

Haider, H., Sadiq, R. & Tesfamariam, S. 2016 Inter-utility performance benchmarking model for small-to-medium-sized water utilities: aggregated performance indices. *Journal of Water Resources Planning and Management* **142** (1), 04015039. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000552.

Hatami-Marbini, A. & Toloo, M. 2019 Data envelopment analysis models with ratio data: a revisit. *Computers & Industrial Engineering* **133**, 331–338. https://doi.org/10.1016/j.cie.2019.04.041.

Hollingsworth, B. & Smith, P. 2003 Use of ratios in data envelopment analysis. *Applied Economics Letters* **10** (11), 733–735. https://doi.org/10.1080/1350485032000133381.

Korhonen, P. & Syrjänen, M. 2004 Resource allocation based on efficiency analysis. *Management Science* **50** (8), 1134–1144.

Lampe, H. W. & Hilgers, D. 2015 Trajectories of efficiency measurement: a bibliometric analysis of DEA and SFA. *European Journal of Operational Research* **240** (1), 1–21. https://doi.org/10.1016/j.ejor.2014.04.041.

Le, N. T., Thinh, N. A., Ha, N. T. V., Tien, N. D., Lam, N. D., Hong, N. V., Tuan, N. T., Hanh, T. V., Khanh, N. N., Thanh N. N. & Hens L. 2022 Measuring water resource use efficiency of the Dong Nai River Basin (Vietnam): an application of the two-stage data envelopment analysis (DEA). *Environment, Development and Sustainability* **24** (10), 12427–12445. https://doi.org/10.1007/s10668-021-01940-w.

Lejarraga, T. & Müller-Trede, J. 2017 When experience meets description: how dyads integrate experiential and descriptive information in risky decisions. *Management Science* **63** (6), 1953–1971. https://doi.org/10.1287/mnsc.2016.2428.

Li, Y., Shi, X., Yang, M. & Liang, L. 2017 Variable selection in data envelopment analysis via Akaike's information criteria. *Annals of Operations Research* **253** (1), 453–476. https://doi.org/10.1007/s10479-016-2382-2.

Lombardi, G. V., Stefani, G., Paci, A., Becagli, C., Miliacca, M., Gastaldi, M., Giannetti, B. F. & Almeida, C. M. V. B. 2019 The sustainability of the Italian water sector: an empirical analysis by DEA. *Journal of Cleaner Production* **227**, 1035–1043. https://doi.org/10.1016/j.jclepro.2019.04.283.

lo Storto, C. 2020 Measuring the efficiency of the urban integrated water service by parallel network DEA: the case of Italy. *Journal of Cleaner Production* **276**, 123170. https://doi.org/10.1016/j.jclepro.2020.123170.

Marques, R. C. & Monteiro, A. J. 2001 Application of performance indicators in water utilities management – a case-study in Portugal. *Water Science and Technology* **44** (2–3), 95–102.

Molinos-Senante, M., Marques, R. C., Pérez, F., Gómez, T., Sala-Garrido, R. & Caballero, R. 2016 Assessing the sustainability of water companies: a synthetic indicator approach. *Ecological Indicators* **61**, 577–587. https://doi.org/10.1016/j.ecolind.2015.10.009.

Mugisha, S. 2011 *Utility Benchmarking and Regulation in Developing Countries*. IWA Publishing, London. https://doi.org/10.2166/9781780400938.

Pastor, J. T. & Aparicio, J. 2010 The relevance of DEA benchmarking information and the least-distance measure: comment. *Mathematical and Computer Modelling* **52** (1), 397–399. https://doi.org/10.1016/j.mcm.2010.03.010.

Pinto, F. S., Costa, A. S., Figueira, J. R. & Marques, R. C. 2017a The quality of service: an overall performance assessment for water utilities. *Omega the International Journal of Management Science* **69**, 115–125. https://doi.org/10.1016/j.omega.2016.08.006.

Pinto, F. S., Simões, P. & Marques, R. C. 2017b Water services performance: do operational environment and quality factors count? *Urban Water Journal* **14** (8), 773–781. https://doi.org/10.1080/1573062X.2016.1254254.

Robles-Velasco, A., Rodríguez-Palero, M., Muñuzuri, J. & Onieva, L. 2022 Sustainable development and efficiency analysis of the major urban water utilities in Spain. *Water* **14** (9), 1519.

Romano, G. & Guerrini, A. 2011 Measuring and comparing the efficiency of water utility companies: a data envelopment analysis approach. *Utilities Policy* **19** (3), 202–209. http://doi.org/10.1016/j.jup.2011.05.005.

Ruggiero, J. 2005 Impact assessment of input omission on DEA. *International Journal of Information Technology & Decision Making* **04** (03), 359–368. https://doi.org/10.1142/S021962200500160X.

Singh, M., Mittal, A. K. & Upadhyay, V. 2014 Efficient water utilities: use of performance indicator system and data envelopment analysis. *Water Supply* **14** (5), 787–794. https://doi.org/10.2166/ws.2014.036.

Thanassoulis, E. & Silva, M. C. A. 2018 Measuring efficiency through data envelopment analysis. *Impact* **2018** (1), 37–41. https://doi.org/10.1080/2058802X.2018.1440814.

Toloo, M., Barat, M. & Masoumzadeh, A. 2015 Selective measures in data envelopment analysis. *Annals of Operations Research* **226** (1), 623–642. https://doi.org/10.1007/s10479-014-1714-3.

Vilanova, M. R. N., Magalhães Filho, P. & Balestieri, J. A. P. 2015 Performance measurement and indicators for water supply management: review and international cases. *Renewable and Sustainable Energy Reviews* **43**, 1–12. https://doi.org/10.1016/j.rser.2014.11.043.

Wagner, J. M. & Shimshak, D. G. 2007 Stepwise selection of variables in data envelopment analysis: procedures and managerial perspectives. *European Journal of Operational Research* **180** (1), 57–67. https://doi.org/10.1016/j.ejor.2006.02.048.

Worthington, A. C. 2014 A review of frontier approaches to efficiency and productivity measurement in urban water utilities. *Urban Water Journal* **11** (1), 55–73. https://doi.org/10.1080/1573062X.2013.765488.

Wu, G., Hong, J., Li, D. & Wu, Z. 2019 Efficiency assessment of pollutants discharged in urban wastewater treatment: evidence from 68 key cities in China. *Journal of Cleaner Production* **233**, 1437–1450. https://doi.org/10.1016/j.jclepro.2019.06.012.

Xie, Q., Xu, Q., Zhu, D., Rao, K. & Dai, Q. 2020 Fair allocation of wastewater discharge permits based on satisfaction criteria using data envelopment analysis. *Utilities Policy* **66**, 101078. https://doi.org/10.1016/j.jup.2020.101078.

Yang, J., Xie, G., Yang, Y., Zhang, Y. & Liu, W. 2019 Deep model integrated with data correlation analysis for multiple intermittent faults diagnosis. *ISA Transactions* **95**, 306–319. https://doi.org/10.1016/j.isatra.2019.05.021.

First received 9 January 2023; accepted in revised form 29 March 2023. Available online 10 April 2023