frontier economics

BENCHMARKING TECHNIQUES AND PRACTICES FOR ELECTRICITY AND NATURAL GAS NETWORK OPERATORS

Report prepared for ACM

12 OCTOBER 2023

CONTENTS

Executive summary

For the current regulatory period 2022-2026, ACM has used benchmarking techniques to determine efficient costs of the Dutch electricity and gas TSOs and DSOs

The Authority for Consumers and Markets (ACM) is responsible for the regulation of Dutch electricity and gas transmission system operators (TSOs) and distribution system operators (DSOs). For the current regulatory period 2022-2026, ACM has used the following benchmarking techniques to determine the efficient costs of the Dutch electricity and gas TSOs and DSOs:

- For TSOs, ACM identified static efficiencies using **Data Envelopment Analysis** (DEA) in a pan-European benchmarking study of TSOs;
- For DSOs, ACM identified static efficiencies based on a **unit cost analysis** of Dutch DSOs.

For future regulatory periods ACM will likely need to consider the challenges of the energy transition and implications from CBb's ruling for benchmarking analysis

The energy transition requires new investments. For electricity system operators these investments will include additional decentralised generation and demand connection for solar PV, wind, and EVs. For the gas system operators these investments will focus on potential gas phase-out and repurposing. Benchmarking techniques will likely need to be able to support these future efficient investment needs.

The Dutch Trade and Industry Appeals Tribunal (College van Beroep voor het bedrijfsleven (CBb)) has recently made its decision on the appeals lodged by the Dutch TSOs and DSOs against the Method Decisions for the period 2022-2026. The CBb has determined that:

ACM must adjust the way in which the dynamic efficiency of regional grid operators is estimated taking into account the impact from new challenges due to the energy transition. With regards to the pan-European cost benchmarking study for TSOs, the CBb criticised the lack of transparency of data. In addition, the CBb criticised that the cost-benchmark study does not allow to identify the source(s) of cost inefficiency.

Hence, ACM wants to identify benchmarking techniques that could be used in future regulatory periods. In particular, ACM asked Frontier Economics to investigate advantages and disadvantages of available benchmarking techniques and how they can be used in practice. They also asked us to identify and summarise relevant case studies. These case studies can be useful to ACM when considering some of the practical aspects of implementing a technique.

We identified a short list of benchmarking techniques that can be used to address the challenges of the energy transition and implications of CBb's ruling

Benchmarking techniques can be classified in three broad groups with different characteristics.

Descriptive techniques. Descriptive techniques are simple techniques that are used to show how inputs and outputs are related. These techniques include unit cost analysis or comparative analysis of key characteristics of businesses (e.g. density of supply volumes or connections).

- **Economic-based techniques (parametric, non-parametric, semi-parametric).** Economic-based techniques include all techniques that are grounded in microeconomic production theory, where a cost function links inputs used to outputs produced allowing for an inefficiency component and potentially some random noise. Parametric techniques require a priori assumptions around the cost function (relationship between costs and cost drivers) and the distribution of the inefficiency component. Non-parametric techniques do not require such a priori assumptions, but all residual deviation from the frontier is assumed to be inefficiency. Semi-parametric techniques usually do not require an assumption on the cost function, but require an assumption on the distribution of the error terms.
- **Engineering-based techniques**. Engineering-based models rely more heavily on engineering or expert insights to develop an assessment of efficient costs. The methods can either be applied to aggregated costs (e.g. Reference Network Analysis) or to disaggregated costs for different activities (e.g. process benchmarking) or projects (e.g. engineering assessment). Hence, these techniques range in complexity, from simple relationship between cost and cost drivers, to more complex models for calculating an 'optimal network'.

For each of these groups, first we identified a long list of available techniques from regulatory precedent and the academic literature. Then we determined a short list of techniques that can be used to address the challenges of the energy transition and the implications of CBb's ruling by applying four evaluation criteria based on ACM's regulatory context and expected future changes: promotion of efficiency, transparency, robustness, applicability. We also consider whether there is relevant regulatory precedent for the use of each technique. The table below shows the techniques that we have short-listed.

Group	Sub-group	Technique
Descriptive technique	Performance indicators	Partial Performance Indicators (PPIs)
Mainly based on economic theory	Parametric	Corrected OLS (COLS), Modified OLS (MOLS) Stochastic Frontier Analysis (SFA)
	Non-parametric	Data Envelopment Analysis (DEA)

		Bootstrap DEA
	Semi-parametric	Stochastic non-smooth envelopment data (StoNED)
Mainly based on engineering rationale	Engineering based	Engineering models Reference Network Analysis (RNA)

Each technique in our short list has strengths and weaknesses

Our in-depth comparative analysis of the short-listed techniques against four evaluation criteria revealed that each technique has strengths and weaknesses. Our evaluation is summarised in the table below.

Criterion	Question	Findings
Promotion of efficiency	Which technique should be used to benchmark a given cost category?	Econometric techniques likely more appropriate for high-level cost categories and business as usual activities Engineering models and RNA likely more appropriate for lower-level business as usual activities, or new activities (especially where significant new investments can be anticipated)
	Which technique can be used to identify where inefficiency is coming from?	PPIs can provide a high level indication and are relatively simple to implement Econometric techniques can be applied to disaggregated cost data, but usually the quality of disaggregated data is lower Engineering based models can be used for specific activities
Transparency	Which technique is more transparent?	More transparent: PPIs, simple engineering models, econometric models (COLS, MOLS, SFA, DEA, StoNED). We consider that all these econometric models are transparent as the implementation of the model is clear. DEA and StoNED might be considered less transparent as they do not show the relationship between costs and cost

Criterion	Question	Findings
		drivers as explicitly as the other models. There is also limited precedent in the use of StoNED. Less transparent: complex engineering models, RNA
Applicability	What is an appropriate sample size?	Econometric techniques require a larger sample size than engineering models, RNA and PPI (particularly relevant since sample sizes may be small in the Dutch context)
	How to account for heterogeneity?	All techniques can account for heterogeneity. This can be done within the technique or by adjusting cost data ex-ante or results ex-post.
	How to account for economies of scale?	Econometric techniques (e.g. DEA, SFA) and PPIs can be used to estimate economies of scale Economies of scale can be assumed for all techniques
	Which technique can be used with forecast data?	All techniques can accommodate forecast data, which can either be included directly in the estimation or used to forecast efficient allowances
	Which technique is likely to be more resource intensive?	PPI is the least resource intensive, followed by economic-based models, and complex engineering models and RNA
Robustness	How do you ensure robustness?	By ensuring good data quality and testing results of models for small variations in data and assumptions

Given that each technique has its strengths and weaknesses, there may be merit in combining different benchmarking techniques

Benchmarking is a valuable instrument of the regulatory toolbox. However, when applying this tool, it is unlikely that there exists a unique right benchmarking model, as each technique has its own strengths and weaknesses. Therefore, there may be merit in combining different benchmarking techniques that complement each other in order to offset possible shortcomings of a single benchmarking technique.

A combination of techniques can be used to:

- Improve the robustness of the assessment of efficiency of a given cost category and

- Benchmark different cost categories.

On the first point, it is possible to combine results from different techniques applied at the same level of cost aggregation to mitigate some of the weaknesses of specific techniques (e.g. benchmarking totex using both SFA and DEA). It is also possible to combine results from techniques applied at different level of cost aggregation (e.g. results from a top-down benchmarking of totex are combined with results from a bottom-up benchmarking of components of totex).

On the second point, different techniques can be applied to different cost categories, for example SFA might be better suited for benchmarking business as usual activities, while engineering models might be better for bespoke large capex investments.

Some techniques can also be used to understand the sources of inefficiency

Some of the techniques we identified can be used to understand where inefficiency is coming from, which is one of the key implications from CBb's ruling. For example, even if the overall efficiency of costs is assessed using a top-down econometric model, it would be possible to apply specific models to disaggregated costs to understand where inefficiency comes from. It is also possible to use some descriptive statistics like PPIs to understand how unit costs might differ between companies and use this information to attempt to understand the source of inefficiencies.

Case studies form other jurisdictions show how techniques can be applied and combined

Country	Sector	Regulator	Reason for selecting case study
Great Britain	Gas DSOs	Ofgem	Use of different techniques to benchmark different cost categories (economic and engineering-based techniques) Use of forward looking data Estimation of econometric models with a small sample
Great Britain	Electricity DSOs	Ofgem	Use of totex regression models and disaggregated benchmarking to benchmark a given cost category (totex) Use of forward looking data Estimation of econometric models with a small sample
Finland	Electricity DSOs	Energy Authority	Development of benchmarking (from DEA, to DEA and SFA, to StoNED)

Country	Sector	Regulator	Reason for selecting case study
			Application of StoNED in a regulatory context
Australia	Electricity DSOs	AER	Combination of different benchmarking techniques (SFA, OLS with fixed effects), and PPIs as cross-checks Use of international sample to address challenges with small samples
Germany	Electricity TSOs	BNetzA	Application of RNA
Germany	Electricity and gas DSOs	BNetzA	Different benchmarking techniques applied (DEA, SFA) Outlier analysis and cost driver analysis

When undertaking a benchmarking study, other aspects of the broader benchmarking framework should be considered together with the choice of techniques

Benchmarking techniques are only one component of the broader benchmarking framework. The benchmarking framework to determine static efficiency usually consists of two steps:

- Undertaking a benchmarking analysis to estimate relative static efficiencies. There are four components of a benchmarking analysis: technique, cost, cost drivers, and sample.
- Applying the results of the benchmarking analysis to set revenue allowances.

Therefore, when undertaking a benchmarking study other aspects of the broader benchmarking framework are as important as the choice of technique. For example, when undertaking a benchmarking analysis it is important to also consider the set of comparators, how the cost drivers are defined, whether the data is consistent across operators and over time, how the results of benchmarking are used (e.g. mechanistically or not), and which incentives are in place (e.g. whether the operators are incentivised to provide accurate forecasts; what the implications of benchmarking opex and capex separately are).

1 Introduction

1.1 Context

The Authority for Consumers and Markets (ACM) is responsible for the regulation of Dutch electricity and gas transmission system operators (TSOs) and distribution system operators (DSOs). For the current regulatory period 2022-2026, ACM has used benchmarking techniques to determine the efficient costs of the Dutch electricity and gas TSOs and DSOs. For TSOs, ACM identified static efficiencies using Data Envelopment Analysis (DEA) in a pan-European benchmarking study of TSOs; for DSOs, ACM identified static efficiencies based on a unit cost analysis of Dutch DSOs.

For future regulatory periods, the benchmarking techniques will likely need to take into account two important factors:

- The **energy transition**. The energy transition requires new investments. For electricity system operators these investments will include additional decentralised generation and demand connection for solar PV, wind, and EVs; voltage control; for the gas system operators these investments will focus on potential gas phase-out and repurposing. Benchmarking techniques will likely need to be able to support these future efficient investment needs.
- The **recent decision by the Dutch Trade and Industry Appeals Tribunal (CBb)**[1] on the appeals lodged by the grid operators (DSOs and TSOs) against the Method Decisions for the period 2022 – 2026. The CBb has determined that ACM must adjust the way in which the dynamic efficiency of regional grid operators is estimated taking into account the impact from new challenges due to the energy transition. With regards to the pan-European cost benchmarking study for TSOs, the CBb criticised the lack of transparency of data. In addition, the CBb criticised that the cost-benchmark study does not allow to identify the source(s) of cost inefficiency.

1.2 Scope of this study

In this context, ACM has commissioned Frontier Economics to identify benchmarking techniques that could be used to determine the efficient costs of gas and electricity TSOs and DSOs in future regulatory periods. ACM asked us to investigate advantages and disadvantages of available benchmarking techniques and how they can be used in practice for benchmarking. They also asked us to identify and summarise relevant case studies. The objective of this study is to support ACM's choice of benchmarking techniques for future regulatory periods.

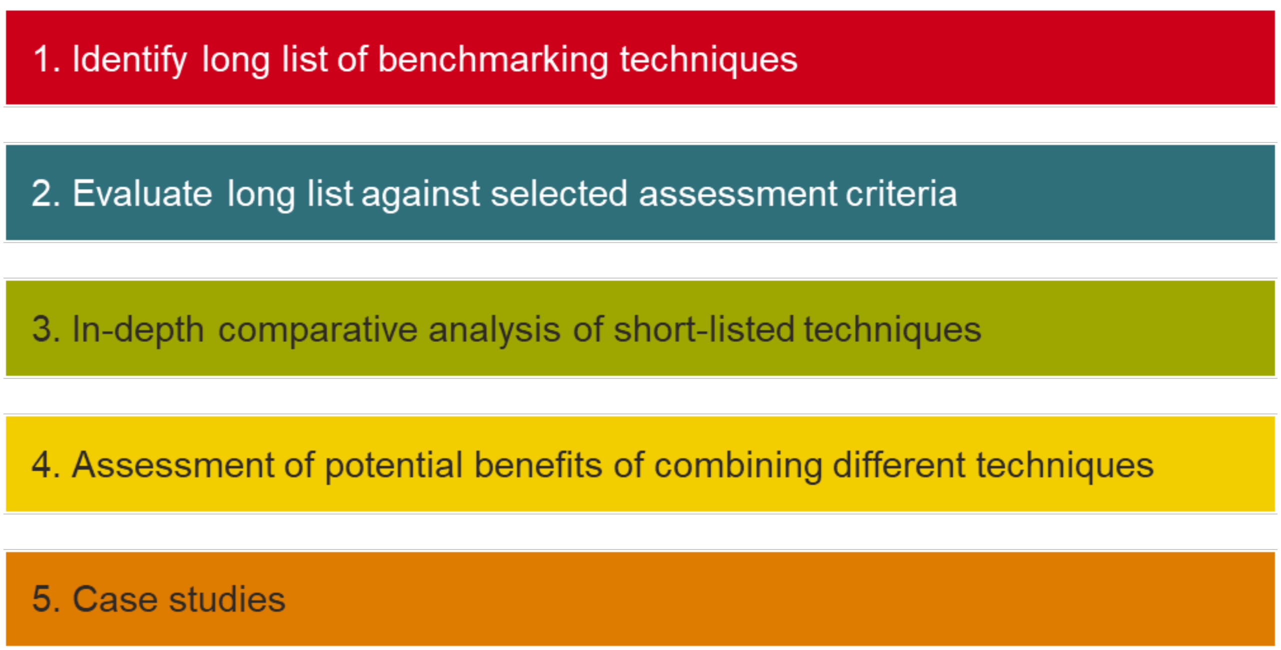
[1] CBb = College van Beroep voor het bedrijfsleven.

1.3 Our approach

The figure below outlines the approach that we have adopted to undertake this study, based on the above diagnosis:

- First, we identified a long list of benchmarking techniques through a combination of regulatory precedent and literature review.
- Second, we developed a set of evaluation criteria and in applying those to the long list we identified a short list of techniques that might be suitable for application by the ACM. The evaluation criteria reflect ACM's objectives for benchmarking and benchmarking best practice.
- Third, we undertook a more comprehensive comparative assessment of the short-listed techniques against the evaluation criteria, focussing on how the techniques could be used in practice for benchmarking.
- Fourth, we investigated how different techniques (or elements thereof) can be combined when undertaking a benchmarking analysis and the benefits that this might bring.
- Finally, we identified a number of case studies that show how some of the short-listed techniques have been used by regulators in other jurisdictions. These case studies can be useful to ACM when considering some of the practical aspects of implementing a technique.

Figure 1 Overview of our approach



Source: Frontier Economics

This report is the output of the work we have undertaken for ACM. Our work has benefitted from expert input from Professor Dr Christopher Parmeter and Dr Mark Andreas Andor, as well as feedback from ACM. The overall findings are solely the responsibility of Frontier Economics.

1.4 Structure of this report

The remainder of this report is structured as follows:

- In Section 2 we provide more details on the context of this work against which our findings should be interpreted.
- In Section 3 we present the long list of techniques we have identified.
- In Section 4 we explain how we have identified the short list of techniques.
- In Section 5 we present our comparative analysis of the short-listed techniques.
- In Section 6 we discuss how different techniques can be combined.
- In Section 7 we present a range of case studies.
- In Section 8 we summarise our findings.

2 Use of benchmarking as a regulatory toolbox in the Netherlands

A benchmarking study is part of the toolbox that regulators can use to set efficient revenues or tariff allowances of electricity and gas TSOs and DSOs.

Typically, benchmarking involves identifying a range of 'inputs' to the business process (e.g. physical inputs such as labour and materials, or financial resources) and a range of 'outputs' (e.g. services delivered and the quality of those services). A business is regarded as performing more efficiently if it is able to deliver more outputs while using the same or less inputs. Differences in the operating environment that affect the possible level of performance across different operators will need to be taken into account when assessing relative efficiency.

A number of different techniques can be deployed to conduct a benchmarking study. The choice of specific benchmarking techniques used in benchmarking studies depend on a number of factors, including the objectives that the regulator wants to achieve as well as which data is available.

In this section, we provide more details on the context in which ACM is operating which is relevant for choosing a benchmarking technique. This context informed our selection of the evaluation criteria and case studies. The findings of our report should be considered within this context.

This section is organised as follows:

- First, we summarise how ACM used benchmarking to set allowances for electricity and gas TSOs and DSOs in the current regulatory period;
- Second, we explain some of the potential implications for future use of benchmarking due to the energy transition and the CBb's recent ruling; and
- Third, we explain the different components of a benchmarking analysis: benchmarking techniques, costs, cost drivers and the data sample. We also discuss briefly the choices to be made when implementing the results of a benchmarking analysis within the overall regulatory framework.

2.1 Overview of the role of benchmarking in the current Dutch regulatory framework

ACM sets the revenue allowances for the Dutch electricity and gas TSOs and DSOs. There are 1 TSO each for electricity and gas, and 6 DSOs for both electricity and gas. In the current regulatory period (2022-26), ACM assessed the static efficiency of the total costs of each individual operator to inform its revenue allowances. The static efficiency represents the relative cost efficiency of an operator compared to its peers. Static efficiency is only one component of the revenue allowance calculations, other components include dynamic efficiency, regulatory WACC, and for electricity DSOs, a quality indicator called 'q-factor'. The total costs include operating expenditures, depreciation, and financing costs.

ACM has used different benchmarking techniques for TSOs and DSOs as illustrated in the table below. We describe how the ACM uses these techniques in turn.

Table 1 Comparison of ACM's current approach to determine static efficiency for electricity and gas networks

	TSOs	DSOs
Costs	Totex (based mainly on historical costs)	Totex (based mainly on historical costs)
Static efficiency	Derived from efficient peer from pan-European TSO study with physical assets as key output parameter	Based on yardstick approach, where physical assets are not taken into account
Benchmarking technique	DEA	Unit cost approach

Source: Frontier Economics

2.1.1 Electricity and gas TSOs

The Netherlands has one electricity TSO, TenneT, and one gas TSO, GTS. ACM used a sample of European TSOs to assess the static efficiency of TenneT and GTS.

ACM sets the static efficiency of TenneT based on the efficiency estimated from a pan-European TSO study (TCB18-electricity).[2] The study estimates the efficiency of TenneT's costs compared to those of other European TSOs using DEA, a non-parametric benchmarking technique. DEA estimates the efficiency frontier using total costs as inputs, and three physical outputs: weighted physical assets (NormGrid), total installed transformer power, and total line length weighted by share of angular towers.

2 https://www.ceer.eu/documents/104400/6742745/TCB18_final_report_elec_190717.pdf/559c7df0-9cf3-2153-07bd-855bdf9a6a13

ACM set the static efficiency of GTS using a benchmarking approach similar to the approach used for TenneT (TCB18-gas)[3]. The study estimates the efficiency of GTS's costs compared to those of other European TSOs using DEA. The study uses total costs as inputs and four outputs: weighted physical assets (NormGrid), number of connection points, installed compressor capacity, and pipeline length weighted by a wetness factor.

2.1.2 Electricity and gas DSOs

The Netherlands has 6 electricity DSOs and 6 gas DSOs. ACM uses 'yardstick competition' to derive static efficiency, i.e. by comparing unit costs of each DSOs with an estimate of efficient unit costs.

In particular, for electricity DSOs, ACM sets static efficiency by comparing the unit costs of each DSO with an estimate of the efficient unit costs. Unit costs are derived by dividing total costs by a composite output. The composite output is designed to condense the multiple dimensions of an operator's sales into a single statistic.[4] The efficient unit costs is the industry average unit cost, derived as the ratio of total costs in the industry and the total composite output in the industry. When calculating the total efficient costs ACM applies a parameter for dynamic efficiency based on historic outturn productivity growth averaged over all six DSOs. If an operator provides its services at a unit cost below this benchmark it realises some extra return. For gas DSOs, ACM used the same approach adopted for electricity DSOs.

2.2 Challenges for future regulation and implications for benchmarking

Two factors are likely to affect the future use of benchmarking: challenges related to the energy transition and the CBb's ruling.

Implications of the energy transition

The energy transition is expected to cause changes to the infrastructure and activities undertaken by electricity and gas networks. These changes affect electricity and gas networks in different ways:

- **Electricity networks** will likely require large investments to support increased electrification (e.g. in the mobility and heating sectors) and the integration of renewable energy sources. For example, transmission and distribution networks will need to prepare their networks to deal with additional or enhanced decentralised generation and demand (e.g. from solar or wind infeed and the increased utilisation of electric vehicles or heat pumps). There is also a need for new

3 Sumicsid. (2019). Pan-European Cost-Efficiency Benchmark for Gas Transmission System Operators. Report for CEER.

4 Outputs having a larger impact on operators' costs are assigned a larger weight in the calculation of the composite output than outputs with a smaller impact on costs.

capabilities for voltage (and frequency) control and congestion management on the distribution networks level.

- **Gas networks** will instead likely focus on natural gas phase-out and repurposing of gas assets for use with sustainable gases, sometimes referred to as 'green gases', such as biogas, biomethane, hydrogen from renewable electricity ('green hydrogen'), and hydrogen produced using carbon capture technology ('blue hydrogen'). Investment needs are likely to focus on the integration of green gases on a more regional level. In the medium to long run repurposing of the existing gas grid infrastructure to get it 'H2 ready' will trigger further investments. The decarbonisation impacts overall gas demand and part of the existing network (in particular distribution networks) will not be utilised anymore.

Future regulation (and therefore benchmarking analysis) will likely need to be adapted to take these changes into account and **avoid jeopardising future efficient investment needs** (as also indicated in the recent CBb's decision – see next section), while there is some degree of uncertainty over the timing of the described trends and disruptions. For example, for electricity networks, setting incentives for efficient future investments might become more important vis-à-vis identifying cost inefficiencies of historical investments; for gas networks, supporting efficient investments for the integration of green gases will likely become more important.

Implications of CBb's ruling

On Tuesday, July 4, 2023, the CBb ruled on appeals lodged by the grid operators (regional grid operators as well as TenneT and GTS) against ACM's Method Decisions for the period 2022-2026. The CBb confirmed many points of the ACM decision. However, it also ruled that ACM shall adjust the Method Decisions in some areas. Concerning the TSO benchmarks, the CBb ruled that there was a lack of data transparency. ACM was therefore obligated to amend the Method Decisions for TenneT and GTS and set their static efficiency at 100%.

We identified three areas which may be relevant for future benchmarking studies:

- **Challenges due to the energy transition**. The efficiency of regional grid operators should be estimated by accounting for the impact of new technology and investment.
- **Transparency of approach.** The CBb criticised the lack of transparency of the pan-European benchmarking study of TSOs, given that the data was not available to the key stakeholders. Therefore, future benchmarking studies will likely need to be more transparent, with data available to key stakeholders.
- **Sources of inefficiency**. The CBb criticised that the TSO cost benchmarking study does not allow identification of the sources of cost inefficiency.

2.3 Benchmarking techniques and the broader benchmarking framework

The benchmarking framework to determine static efficiency usually consists of two steps:

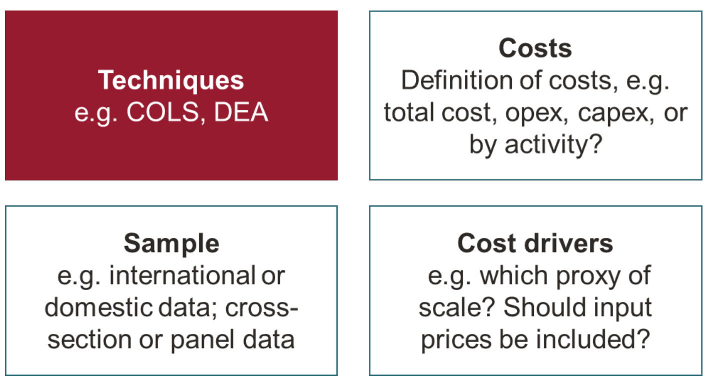
- Undertaking a benchmarking analysis to estimate relative static efficiencies.
- Applying the results of the benchmarking analysis to set revenue allowances.

The scope of this study is on benchmarking techniques, which is one of the four components of a benchmarking analysis. However, the other components of the benchmarking analysis as well as how the results are used is equally important for any benchmarking study. In the remainder of this report we comment on these other aspects where relevant.

Undertaking a benchmarking analysis

Figure 2 shows four key components of a benchmarking analysis: technique, costs, cost drivers, and sample. These components interact with each other. For example, the technique used might depend on the data available.

Figure 2 Four components of a benchmarking analysis



Source: Frontier Economics

Note: The box highlighted in red is the focus of this study.

For each of these components, the regulator is likely required to make some material choices around how to benchmark operators.

- **Benchmarking techniques.** These are the techniques (or mathematical algorithms) that can be used to bring together data on costs and cost drivers to estimate efficiency.
- **Costs**. These consist of the different ways in which costs could be defined, structured, aggregated, and treated for inclusion in a benchmarking study (e.g. benchmarking total costs vs. disaggregated costs).

- **Cost drivers**. These are the key cost drivers of the business. Cost drivers include measures of the scale of the network task that each operator undertakes; additional variables that capture values outputs, such as quality of supply; and other factors that can affect costs, such as proxies for topography or input prices.
- **Sample.** This is the group of comparators considered when undertaking a benchmarking analysis. Usually, for most techniques, the greater the sample size the more variables it is possible to include within a model and the more robust the estimation of the model.

We discussed these components in more details in our 2010 study for the UK energy regulator Ofgem on the future role of benchmarking in regulatory reviews in light of the proposals emerging from the RPI-X@20 review.[5]

Applying the results of the benchmarking analysis

The choice of how to apply the results of a benchmarking analysis in regulatory decisions is often as important as the benchmarking analysis in itself. For example it can:

- Impact upside chances and downside risks from the benchmarking analysis. For example, if the benchmark is determined by the frontier company (i.e. most efficient company) then the application mainly results in cost reduction targets for companies with efficiency scores below the frontier company, but no upside chances for the frontier company.
- Mitigate some of the weaknesses of the benchmarking analysis. For example, if there is uncertainty around data quality or modelling results, typically regulators may decide not to use the results of the benchmarking analysis mechanistically (e.g. by requiring all companies to be 100% efficient). Instead, they could apply some regulatory judgment to set the catch-up efficiency target (e.g. by using a glide-path or setting the catch-up efficiency below the estimated efficiency frontier).
- Incentivise operators to provide some expected (and unexpected) responses. It is therefore important that the regulator is clear and mindful of the incentives caused by its approach to benchmarking. For example, a regulator that benchmarks opex and capex separately might inadvertently cause a capex bias. Another example is when forecast data is used without placing incentives on companies to produce accurate forecasts (and hence companies may decide to over/under forecast in order to get a higher allowance or a better efficiency score).

---

[5] See Frontier Economics. (2010). RPI-X@20: The Future Role of Benchmarking in Regulatory Reviews. https://www.ofgem.gov.uk/sites/default/files/docs/2010/05/rpt-benchmarking.pdf

3 Long list of benchmarking techniques

In this section we present an overview of the different benchmarking techniques that can be used for a regulatory benchmarking study. We identified these techniques by surveying the academic literature and regulatory precedents. Our long list of techniques includes techniques that have been used extensively, such as DEA, as well as less commonly used techniques (e.g. StoNED and four component SFA), some of which have been developed fairly recently in the academic literature.

The remainder of this section is structured as follows:

- First, we present a classification of the range of possible techniques in three high-level groups: descriptive techniques, economic-based techniques, and engineering-based techniques. This classification is helpful when comparing the techniques.
- Then, we summarise and explain each of the techniques we have identified.

3.1 Our classification of benchmarking techniques

Benchmarking techniques can be classified in three broad groups with different characteristics.

- Descriptive techniques;
- Economic-based techniques (parametric, non-parametric, semi-parametric); and
- Engineering-based techniques.

Table 2 illustrates the long list of techniques that we have identified for each group.

Table 2 Our long list of benchmarking techniques

Group	Sub-group	Technique
Descriptive technique	Performance indicators	Partial Performance Indicators (PPIs)
Mainly based on economic theory	Parametric	OLS, Corrected OLS (COLS), Modified OLS (MOLS) Quantile regression Stochastic Frontier Analysis (SFA) Variations of SPA (e.g. four random components SFA)
	Non-parametric	Data Envelopment Analysis (DEA, bootstrap DEA) Free Disposable Hull (FDH)

		Multilateral Total Factor Productivity (MTFP) Multilateral Partial Factor Productivity (MPFP)
	Semi-parametric	Stochastic non-smooth envelopment data (StoNED) Stochastic FDH/DEA Variations of StoNED (Maindiratta; Fan, Li & Weersink; Parmeter and Racine)
Mainly based on engineering rationale	Engineering based	Process benchmarking Reference Network Analysis (RNA) Engineering models

Source: Frontier Economics

Descriptive techniques

Descriptive techniques are simple techniques that are used to show how inputs and outputs are related. These techniques include unit cost analysis or comparative analysis of key characteristics of businesses (e.g. density of supply volumes or connections).

Inputs and outputs can be measured both in monetary terms and physical terms, and usually no adjustment for exogenous factors affecting performance is done. These techniques do not usually allow identification of an efficiency frontier. However, they are sometimes used to determine benchmarks for unit costs (e.g. by taking the average of the unit costs across a range of companies). Such techniques may help assess the plausibility of benchmarking results derived from the other, more sophisticated techniques. For example, a company that performs well in all dimensions of partial benchmarks should be expected to also perform well when other techniques are applied.

Economic-based techniques

Economic-based techniques include all techniques that are grounded in microeconomic production theory, where a cost function links inputs used to outputs produced allowing for an inefficiency component and potentially some random noise. In formula,

$$c_i = m(y_i, z_i) + u_i + v_i,$$

where c_i is the cost incurred by firm i; $m(.)$ is the cost function that describes how a set of outputs, y_i, and exogenous environmental variables, z_i, affect costs; u_i is the inefficiency; and v_i is the random noise. See box in the next page for examples of cost functions and their properties.

Economic-based techniques can be classified in three groups, according to whether *a priori* assumptions are used to determine the cost function and the distribution of the inefficiency term. This is summarised in the table below. We describe each group of techniques in more detail below the table.

Table 3 Comparison of key characteristics of parametric, non-parametric, and semi-parametric techniques

	Parametric	Non-parametric	Semi-parametric
Cost function	Assumption of specific functional form required	No assumption on functional form required. Some basic axioms of production theory may be imposed.	No assumption on functional form required. Some basic axioms of production theory may be imposed.
Inefficiency	A priori assumption required regarding error term	All deviations assumed to be inefficiency	A priori assumption usually required regarding error term
Example	SFA	DEA	StoNED

Source: Frontier Economics

- Parametric techniques. These techniques require the assumption of specific functional forms for the cost function and the distribution of the error term. Once the cost function is parameterised different estimation techniques can be used to determine the underlying parameters from the observed data. Since the classical economic assumption of cost functions (mainly monotonicity and convexity – see box below for more details) are typically not explicitly imposed on the model, estimates might not always obey all of these axioms. An example of a parametric technique commonly used in benchmarking studies is Stochastic Frontier Analysis (SFA).
- Non-parametric techniques. These techniques do not require an a priori specification of the functional form of the cost function or the distribution of an error term. Non-parametric techniques might explicitly impose some axioms of production theory (e.g. monotonicity and

convexity) on the estimation, without restricting the precise functional form that has to be obeyed. Since these models generally lack a stochastic element, the level of inefficiency they determine is entirely deterministic. That is, all deviations from the estimated efficiency frontier are treated as inefficiency. A non-parametric technique commonly used in benchmarking studies is DEA. However, there are also variants of non-parametric techniques allowing for random noise.

- Semi-parametric techniques. These techniques try to combine some of the advantages of parametric and non-parametric techniques. For example, the cost function could be estimated following a non-parametric approach. Then, some distributional assumptions are imposed on the deviation from the frontier to determine inefficiency. An example of a semi-parametric technique is StoNED, which is used by the Finnish regulator to benchmark the electricity DSOs in Finland.

Cost functions and their properties in classical production theory

In production theory, a cost function is a function that describes the relationship between the costs and outputs or other cost drivers of a production process. Two cost functions often used for cost benchmarking using parametric models are the Cobb-Douglas function and the more general Translog function:

- The **Cobb-Douglas** cost function assumes a constant elasticity between costs and outputs, which implies a constant return to scale, as well as a constant elasticity of substitution between the outputs. For the example of two outputs y_1 and y_2, the Cobb-Douglas cost function of firm i can be expressed formally as

$$ln(c_i) = b_0 + b_1 \ln(y_i^1) + b_2 \ln(y_i^2) + \epsilon_i \ ,$$

  where c_i indicate costs, ϵ_i is the error term and b_0, b_1 and b_2 are cost function parameters.[6]

- The **Translog** cost function generalizes the Cobb-Douglas by incorporating second order (interaction and squared) terms. This allows heterogenous elasticities, substitution patterns and returns to scale. For the two output example, it can be expressed formally as

$$\ln(c_i) = b_0 + b_1 \ln(y_i^1) + b_2 \ln(y_i^2) + b_3 \ln(y_i^1)^2 + b_4 \ln(y_i^2)^2 + b_5 \ln(y_i^1) \ln(y_i^2) + \epsilon_i \ .$$

In theory, cost functions are usually assumed to have a number of properties (also referred to as axioms). Two key properties of production functions are monotonicity and convexity.

- **Monotonicity:** A vector of outputs can always be produced at a higher cost, holding input prices fixed. If input prices increase, then costs must increase for a fixed output vector.
- **Convexity:** The cost of producing a convex combination of outputs is no greater than the convex combination of costs.

We note that there are some exceptions that might require weaker axioms. For example, non-monotonicity (or weak disposability) may be required in some cases.

Engineering-based models

Engineering-based models rely more heavily on engineering or expert insights to develop an assessment of efficient costs. The methods can either be applied to aggregated costs (e.g. Reference Network Analysis) or to disaggregated costs for different activities (e.g. process benchmarking) or projects (e.g. engineering assessment). Hence, these techniques range in complexity, from simple relationship between cost and cost drivers, to more complex models for calculating an ‘optimal network’.

[6] It is straight forward to extend the Cobb-Douglas cost function to the case of three or more cost drivers. For a detailed discussion of the Cobb-Douglas cost function see: Biddle, J. (2012). Retrospectives: The Introduction of the Cobb-Douglas Regression. Journal of Economic Perspectives, 26(2), 223-236.

Engineering-based models can either be used to assess the efficiency of costs or as a tool to identify relevant cost drivers which can then be applied within other benchmarking techniques. For example, RNA is used in Germany to assess the efficiency of electricity TSOs; Model Network Analysis, an approach similar to RNA, but that assumes more stylised and homogenous structures (e.g. constant densities throughout a service area, ignoring specific routes), was used in Austria to calculate an environmental complexity parameter which was then used in the benchmarking model. [7]

Sometimes, these engineering-based models are also used to derive variables that are grounded in engineering logic and can be used for unit cost analysis or tested in economic-based models.

3.2 Descriptive techniques

Partial performance indicators (PPIs)

PPIs are a technique that allows one to compare the performance of businesses in delivering one specific output. PPIs are defined as the ratio of an input to an output. More complex PPIs could be defined using composite inputs and composite outputs. It is also possible to benchmark companies in relation to multiple PPIs.

A lower PPI indicates that less input is required to produce one unit of output indicating a more efficient production process. PPIs are useful to gain a first indication of the relative efficiency of different production processes or to conduct more detailed studies of the sources of identified inefficiencies. However, they do not provide a basis for a rigorous evaluation of the efficiency of two or more firms, as they lack a consistent way of evaluating the entire production process, which includes various outputs and inputs.

Formally, a PPI can be defined as follows:

$$PPI_i(x_i, y_i) = \frac{x_i}{y_i},$$

where x_i is an input (e.g. opex, capex, totex, number of full time employees, etc.) and y_i is an output (e.g. number of customers served, network lengths, etc.) of firm i, respectively. Inputs and outputs can be measured both in monetary terms and in physical terms. Since the PPI can be calculated for a specific year or over a period of time, it can be used for comparisons over time and across businesses. Unit costs are a subset of PPIs, where the inputs are expressed in monetary terms (e.g. cost per connection, cost per MWh served).

The virtue of the PPI is that it is easy to understand, calculate and interpret. Once relevant inputs and outputs are collected, PPIs can be used to understand at a high level how different companies

[7] See Section 3.4 for a description of this technique.

are performing across various dimensions. PPIs can also be used to identify potential data issues, for example by comparing a given PPI for a given company over time to identify potential data anomalies. More disaggregated PPIs can also be used to identify potential sources of inefficiencies.

Generally, the major caveat of the PPI approach is that it lacks a consistent way to evaluate firms' efficiency taking into account the effects of multiple inputs and outputs. For instance, if $PPI_i(x_i, y_i) > PPI_j(x_j, y_j)$ this could indicate firm i operates less efficiently than firm j, since it requires more of input x to produce output y resulting in higher costs. However, it could be that firm i produces more of an additional output y' that also requires x as an input. Moreover, there might be different inputs available to the firm that act as substitutes in the production of the output y. Making intensive use of input x in the production process can be the most efficient way of production if it allows a less intensive use of other inputs. An example could be firm i having higher capital costs that allow producing the same output with lower employment of human capital. A single PPI is not able to inform such trade-offs and therefore has to be treated with caution and can therefore not be used as a general measure of efficiency. Alternative methods like DEA and MTFP can be used to make this sort of assessment as they allow aggregation of multiple inputs and outputs to produce an overall measure of productivity. We describe these in greater detail in the next section.

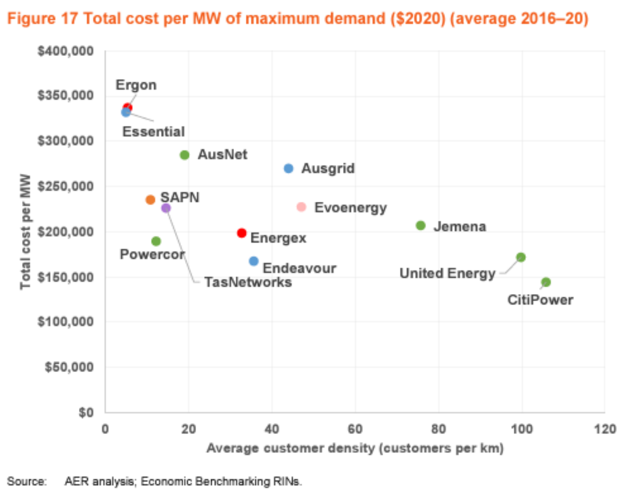
PPIs are used in a number of jurisdictions. For example, in Australia the AER estimates PPIs of key inputs and outputs for DSOs and TSOs.[8]

3.2.1 Australia – Application of PPIs for electricity TSOs and DSOs

The AER estimates PPIs based both on total opex and on disaggregated cost categories. Total opex PPIs show total opex per customer, circuit length, and maximum demand. The more disaggregated PPIs focus on costs for specific activities, like vegetation management, maintenance and emergency response. These disaggregated PPIs use a specific output for each cost category. The AER typically compares the PPIs against a measure of density to allow for better comparison across operators.

8 See for example the AER 2022 annual benchmarking report. https://www.aer.gov.au/networks-pipelines/guidelines-schemes-models-reviews/annual-benchmarking-reports-2022

Figure 3 AER's 2021 PPI of opex per maximum demand vs density for DSOs



Source: AER 2021 Annual benchmarking report, https://www.aer.gov.au/system/files/Distribution%20-%20Report%20-%20AER.pdf

Figure 3 plots the PPI of total costs per maximum demand (in MW) on the y-axis against the average customer density per kilometre on the x-axis for different Australian electricity DSOs. The figure provides a quick and clear indication for a negative relationship between the costs of providing the maximum demand and the population density in the area the DSO is operating in. This contributes to a general understanding of one the factors driving a DSO's costs. It also provides an indication that some DSOs provide the service at lower cost than others, even if they encounter similar levels of population density. For instance, TasNetworks has higher cost per maximum demand than Powercor, even though TasNetwork's population density is actually slightly higher. However, there could be factors other than population density that are specific to TasNetworks and that imply that it encounters higher costs according to this PPI even though it generally operates efficiently. Therefore, it is unlikely that a definitive conclusion on whether or not TasNetworks is actually operating less efficient than Powercor can be made by considering only the PPIs plotted in the figure.

3.3 Economic-based techniques

3.3.1 Parametric techniques

Ordinary Least Squares (OLS)

The OLS regression model is a standard econometric technique whose theoretical properties and practical application is generally well understood by many practitioners in the field of cost benchmarking. In the context of efficiency benchmarking, the model is used to estimate the relationship between firms' costs and cost drivers such as outputs and environmental variables. The estimated model allows to obtain predictions of the mean cost level conditional on the observable characteristics. It is then possible to estimate the efficiency of the individual electricity or gas network operator relative to the mean cost level by comparing the predicted costs to the actual costs

Let c_i denote the scalar cost level of firm i that depends on a column vector of outputs $\boldsymbol{y}_i$[9] and the error term ε_i.[10] OLS then estimates the intercept, α, and the row vector of slope coefficients, $\boldsymbol{\beta}$,[11] in the following model:

$$c_i = \alpha + \boldsymbol{\beta}\boldsymbol{y}_i + \varepsilon_i$$

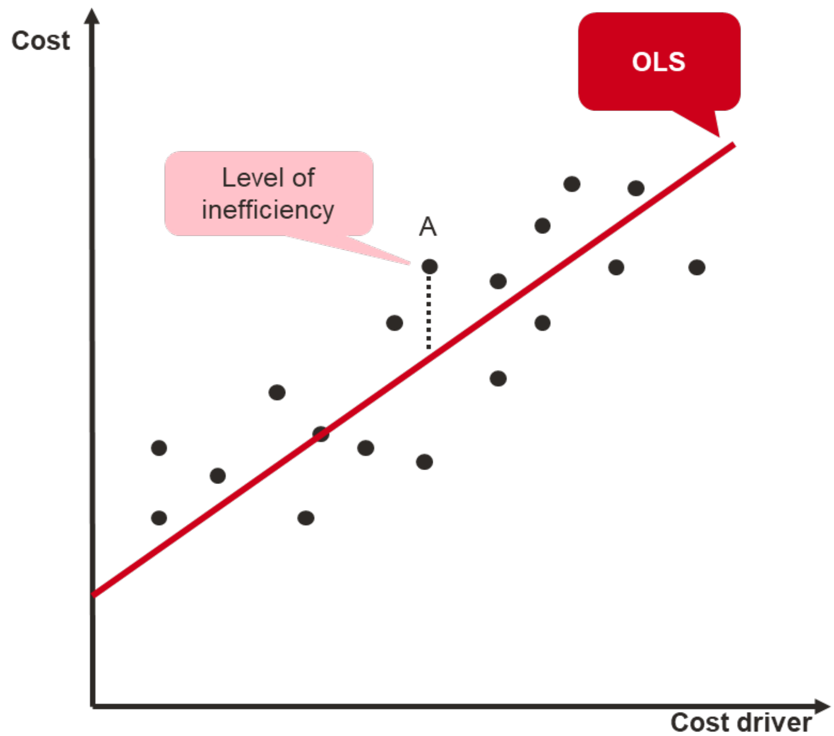
by minimising the sum of squared residuals, which are defined as the difference between the predicted and the actual value of the dependent variable c_i. In principle, the predicted cost for the individual firm, $\hat{c}_i$, can then be used as the cost benchmark against which the actual cost level, c_i, can be assessed. In this case, the estimated regression line (also called 'conditional average function') is considered the efficiency benchmark. Figure 4 illustrates the level of inefficiency that is attributed to a firm at point A according to the OLS estimate.

9 When presenting mathematical formula we use a bold font to indicate vectors.

10 Environmental variables are dropped from the regression equation for notational convenience.

11 For example, the slope coefficient could estimate how strongly cost (here: c_i) varies with changes of service levels (as embodied in the vector **y**).

Figure 4 Illustration of an OLS efficiency frontier



Source: Frontier Economics

Two key advantages of the OLS model are its relative simplicity and its widespread use in econometrics. As a consequence, the OLS model and its results are typically well understood by regulators and regulated companies. The model can be used to explore the relevance and statistical significance of cost drivers and estimate efficiency score. The model is also flexible and can be applied to both historical and forecast data. It can also easily be extended to a panel context.

A caveat of the OLS regression model in the context of efficiency benchmarking is that it does not separate out the distinct influence of firms' inefficiency from general stochastic noise in the error term, ε_i. In other words, any departure from the prediction is considered an inefficiency, even though it could be down to random noise.

Moreover, the conditional average function obtained from the OLS estimation represents average costs, which are unlikely to coincide with efficiency costs. Hence the line is unlikely to represent an efficiency frontier. Figure 4 indicates that many of the observations outperform the regression line (cost benchmark), sometimes quite substantially, which gives an intuitive impression that likely not all inefficiencies are captured by the benchmark.[12]

Several alternative parametric models attempt to address the caveats of the standard OLS framework for efficiency benchmarking.

Corrected OLS (COLS)

The COLS regression technique is an extension of the standard OLS regression that can be used to define a more ambitious efficiency benchmark. For this purpose, the conditional average function given by the OLS regression line is shifted down. The larger the downward shift the more ambitious the resulting efficiency benchmark against which firms' costs are compared becomes.

The size of the downward shift is based on some regulatory judgment. [13] One option is to shift the regression line downwards such that it passes through the observation pertaining to the firm exhibiting the lowest costs. Figure 5 illustrates such a shift. The result is an efficiency frontier that identifies some positive inefficiency for all firms but those setting the frontier. A caveat of this approach is that it is fully deterministic and does not account for random noise that could lead to deviations from the benchmark that are the result of unexpected events outside of the control of the firm rather than actual inefficiency. Some regulators have therefore taken more cautious approaches, for example by shifting the regression line downwards such that it passes through the observation pertaining to the firm representing a certain quantile of the cost distribution (e.g. choosing the efficient firm to have costs below 90% of all firms in the sample), rather than the most efficient firm.

Given its close relationship to the well know OLS regression framework, the technique is generally well accessible to regulators and regulated firms. Furthermore, it can also be applied quite easily in different contexts, e.g. using forward looking or historical data or in a panel context. The technique is therefore frequently considered by regulators when choosing a benchmarking method and has been used a lot in the past.

---

12 From a technical point of view, assume that the error term ε_i can be expressed as the sum of two components, v_i and u_i, representing stochastic noise and inefficiency, respectively. Since inefficiency only affects costs by increasing them, it has a mean larger than zero, which we denote by $\mu>0$. This implies that also the composite error ε_i has mean $\mu>0$, which will result in the intercept of the OLS model picking up the mean inefficiency in the estimation of the OLS regression equation. The intercept defines the OLS regression line. Therefore, the use of OLS for benchmarking likely requires additional steps if the regulator wants to set the efficient cost at a more ambitious level than the average level of efficiency in the sample.

13 We note that there is no consensus in the literature on the precise methods that are associated to the terms COLS and MOLS. In fact, both terms are used interchangeably depending on the interpretation of the author. The descriptions provided in sections 0 and 0 reflect Frontier Economics' view on the meaning of both terms . For a detailed review of the evolution of the terminology used in the academic literature see: Parmeter, C. F. (2023). Is it MOLS or COLS? In P. Macedo, V. Moutinho, & M. Madaleno (Eds.), Advanced Mathematical Methods for Economic Efficiency Analysis (pp. 229-249). Springer.

For example, the British energy regulator, Ofgem, assesses efficiency levels of electricity and gas DSOs by calculating individual efficiency scores as the ratio of submitted costs and the OLS prediction. It then uses a pre-defined quantile of the resulting efficiency score distribution (typically the 85th quantile) to define an efficiency frontier that is applied to all DSOs in the sample.

Modified OLS (MOLS)

Similar to COLS, the MOLS regression technique extends the standard OLS framework to define a more ambitious efficiency benchmark by shifting down the OLS regression line to account for the composite error structure (inefficiency and random noise), which is ignored in OLS. In contrast to COLS, the modified OLS determines the size of the shift of the regression line based on statistical analysis of the OLS residuals. Concretely, MOLS shifts the OLS regression line down by an estimate of the average inefficiency. The average inefficiency is estimated from the OLS residuals by making some assumptions around the distribution of the inefficiency and the random noise. The idea to estimate the mean inefficiency to correct the OLS regression line is closely related to SFA.

Figure 5 illustrates a potential shift of the OLS regression line in an application of MOLS. Since the shift of the regression line represents the mean efficiency level, excluding noise, the MOLS frontier does not intersect the firm with the lowest cost. It thus leaves room for random noise to explain deviations from the frontier that are outside the control of the firm and do not represent actual inefficiency.

While MOLS avoids the need for an ad hoc assumption by the regulator on the size of the shift of the OLS regression line (as this is estimated from the sample), its implementation requires assumptions on the distribution of the inefficiency term and random noise. The close relationship to OLS implies that the technique is well accessible to regulators and regulated firms, and can be easily applied in different contexts. Therefore, it is frequently considered as a potential benchmarking technique by regulators.

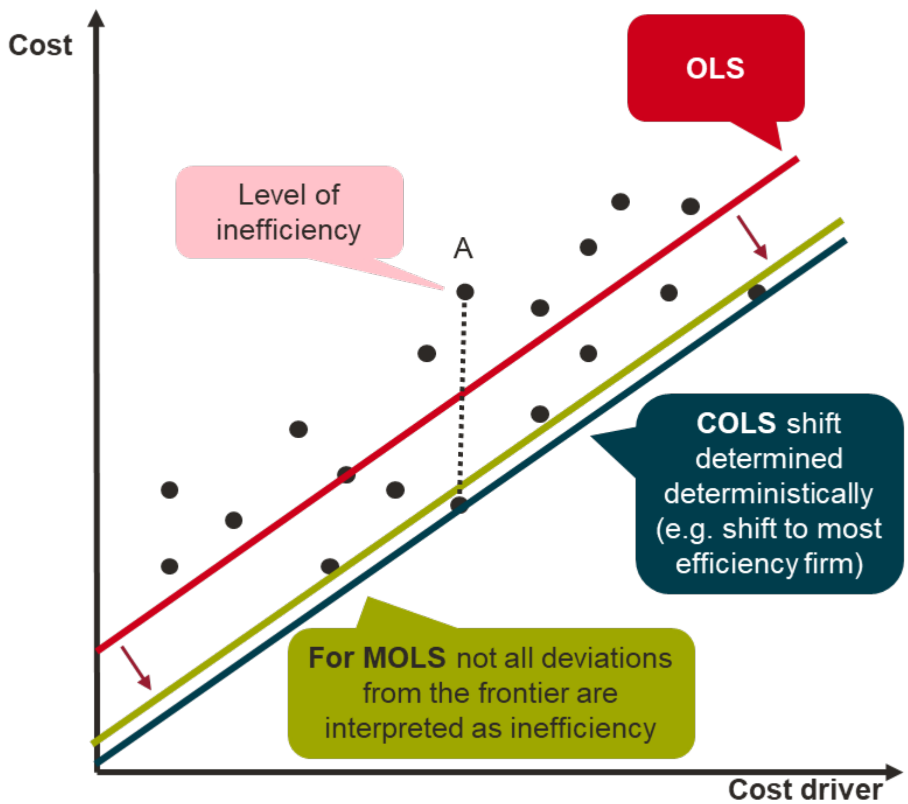
For example, the Austrian energy regulator, E-Control, applies MOLS (in combination with DEA) to assess the efficiency levels of the electricity[14] and gas[15] DSOs. E-Control uses a Cobb-Douglas log-linear functional form for estimating the costs function and assumes that the inefficiency term follows a Half-Normal distribution.

---

14 See E-Control report for electricity DSOs. https://www.e-control.at/documents/1785851/1811582/Regulierungssystematik_4_Periode_STROM_Dez+2018.pdf/a413df20-00b2-9dca-ba43-4ae52754b27e?t=1562139961156

15 See E-Control report for gas DSOs. https://www.e-control.at/documents/1785851/1811582/02_Finale+Regulierungssystematik+4_RP.pdf/40fcc26d-253d-0533-2d74-3774dce4e341?t=1668673860094

Figure 5 Illustration of a COLS and MOLS efficiency frontier



Source: Frontier Economics

Quantile regressions

Quantile regressions are an extension to standard linear regression models. Different from OLS, quantile regressions do not estimate the conditional *mean* of the dependent variable as a function of observed covariates, but rather a conditional *quantile* function that predicts the chosen quantile of the dependent variable based on the observed covariates.[16] The technique is an alternative to COLS that allows to directly estimate the cost benchmark as the quantile chosen by the regulator. Compared to COLS, it does not require two separate steps to first estimate the OLS regression line and then shift it to determine the chosen benchmark.

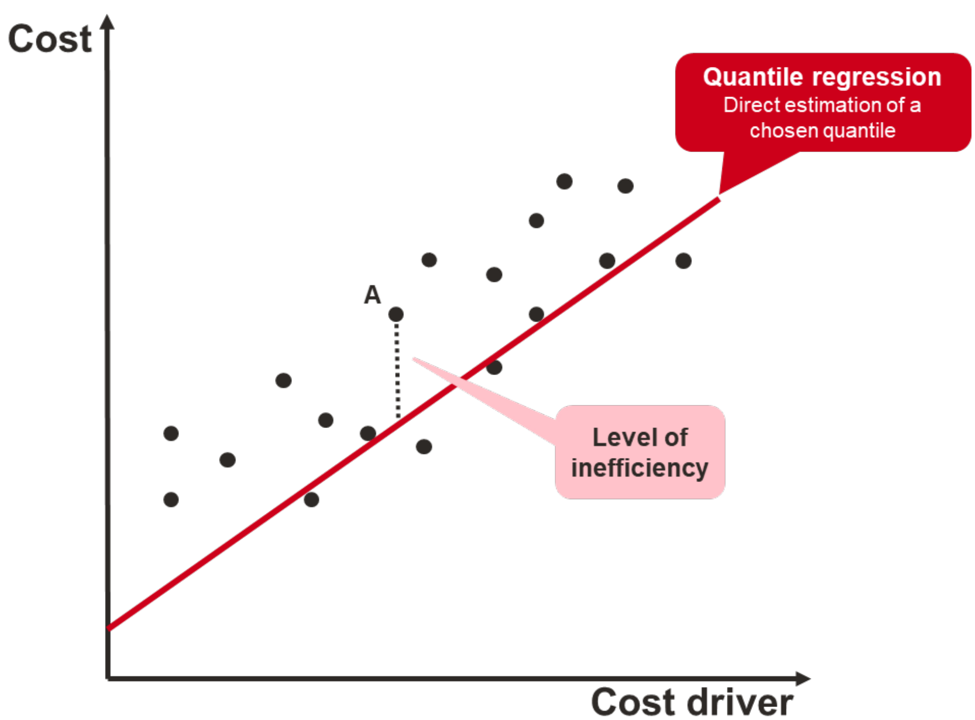
While OLS estimates the regression parameters by minimising the sum of *squared* residuals, quantile regression minimises the sum of *absolute* residuals. Different quantiles can be estimated by assigning different weights to positive and negative residuals. For example, a conditional median

16 For a short introduction to quantile regressions see: Koener, R., & Hallock, K. F. Quantile regression: An introduction.. http://www.econ.uiuc.edu/~roger/research/intro/rq3.pdf.

function can be estimated by assigning equal weights to positive and negative residuals. In general, if a quantile below the median is chosen in the estimation, negative residuals are weighted higher to push the regression line below the median and through the quantile of interest.[17]

Figure 6 illustrates the use of quantile regression for efficiency benchmarking graphically. In the example, the regulator set the efficiency frontier at the 10th quantile. This assumption implies that a firm that produces a comparable output at lower costs than 90% of the firms in the sample is considered efficient.[18] Costs above the regression line, such as at point A, are considered inefficient. Firms that operate at costs below the regression line are more efficient than the frontier defined by the regulator.

Figure 6 Illustration of a quantile efficiency frontier



Source: Frontier Economics

The application of quantile regressions for efficiency benchmarking is similar to the use of COLS, in that the regulator has to set how ambitious the cost frontier should be set. Compared to COLS, the

---

17 In practice, to estimate a conditional quantile τ the absolute value of the negative residuals is assigned a weight equal to $1 - \tau$, while the value of the positive residuals is assigned a weight equal to τ.

18 In this example, the quantile regression line is estimated by assigning to negative residuals a weight of 0.9 in the estimation and to positive residuals a weight of 0.1. This results in a regression line that reduces the presence of negative residuals.

regression line is estimated directly for the chosen quantile and does not require a shift of the conditional mean function.[19] It requires fewer assumptions to obtain consistent estimates of the quantiles and is less prone to outliers as it is based on a minimisation of a loss function of absolute distances to the regression line rather than the squared distances as in OLS.

While quantile regression is similar to OLS with respect to the data requirements and practical implementation, it produces less precise results with the same data (i.e. the estimator is less efficient) if the OLS assumptions are met. In addition, it has – to our knowledge – not yet been applied in a regulatory setting of efficiency benchmarking for energy network operators. Therefore, potential limitations of the approach have not been explored as carefully as for related approaches. Also, regulators and regulated firms are likely to be less informed on the use of the technique.

Stochastic Frontier Analysis (SFA) and variations

SFA is an extension of the standard OLS regression framework that explicitly attempts to differentiate random noise from technical inefficiency. By definition, random noise is assumed to affect costs randomly, i.e. it can both increase or decrease costs; technical inefficiency, in contrast, is assumed to increase costs. To implement the estimation, it is necessary to make some distributional assumptions on the inefficiency term and the random noise that reflect how costs are affected by these different components. By doing so, the SFA model allows to explicitly account for both the existence of random noise and inefficiency in the benchmarking of operators.

Formally, SFA estimates the same regression equation as OLS, but models the error term ε_i as the sum of two distinct components, v_i and u_i, representing stochastic noise and inefficiency, respectively. The regression model thus becomes:

$$c_i = \alpha + \boldsymbol{\beta}\boldsymbol{y}_i + u_i + v_i.$$

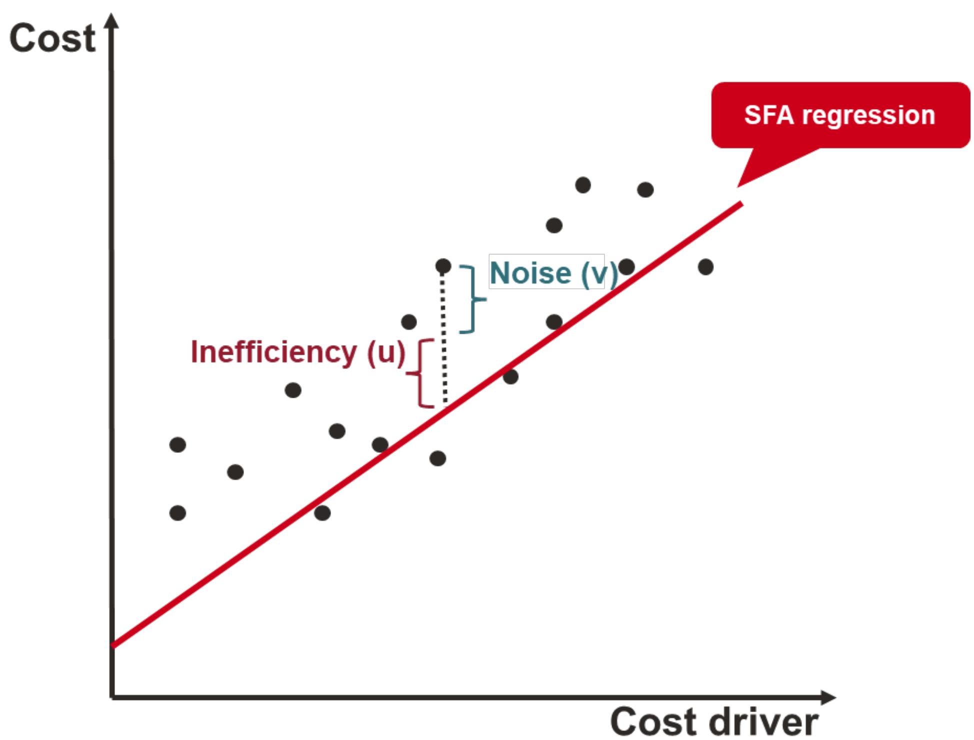
The model identifies v_i and u_i by assuming both error components follow distinct distributions that reflect the idea that the support of inefficiency, u_i, has to be positive (as it can only increase costs), while the stochastic noise term is distributed symmetrically around zero (as it can increase or decrease costs). Typical choices of distribution functions for the inefficiency term are the Half-Normal and the Exponential distribution; the stochastic noise is typically assumed to follow a Standard Normal distribution.

19 This also implies that the slope coefficients may change depending on the quantile chosen for setting the frontier. In contrast, the efficiency frontier estimated using COLS and MOLS is always based on the OLS slope coefficients, only the OLS intercept is adjusted..

Once assumptions around the distribution of the error term are made, the estimation can be implemented either in one step by maximum likelihood or in two steps by Generalized Methods of Moments (GMM) or the so called plug-in likelihood (PL) developed by Fan, Li & Weersink (1996).[20]

As mentioned above, the conditional average function resulting from the SFA regression represents an efficiency frontier that allows for the fact that deviations from the frontier can also be the result of random shocks v_i. While the actual level of inefficiency of the individual firm, u_i, is not observed, it is possible to estimate the expected value of u_i given the value of the composed error $\varepsilon_i = u_i + v_i$ following an approach developed by Jondrow et al. (1982).[21]

Figure 7 Illustration of a SFA efficiency frontier



Source: Frontier Economics

20 It is instructive to consider the functioning of the two-step approaches as both start by first estimating the above equation by OLS. As discussed in Section 3.3.1 this leads to an intercept that is biased upwards by the estimate of the mean inefficiency μ. The PL and GMM estimations use the estimated OLS residuals to estimate the parameters of the assumed error distributions (by maximum likelihood and Method of Moments respectively). It is then possible to obtain an estimate of μ and use this estimate to correct for the upward bias in the intercept.

21 Jondrow, J., Lovell, C. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics, 19(2-3), 233-238.

The main advantage of SFA is that it allows to account for random noise in a consistent framework and in an intuitively appealing way. However, a precise estimation of the parameters of both the efficiency and the random noise distribution requires larger datasets than the other parametric approaches and it requires sufficient variation in the data. In addition, the application of SFA requires the regulator to make assumptions on the specific distributional forms of the inefficiency and stochastic noise terms.

SFA is often considered as a benchmarking tool by regulators of electricity and gas networks. For example, Bundesnetzagentur in Germany uses it to benchmark electricity and gas DSOs; the Australian Energy Regulator uses it to benchmarking electricity DSOs; Ofgem, the UK energy regulation, uses SFA to conduct robustness checks of its preferred OLS based benchmarking approach.

In the academic literature various extensions to the SFA framework have been developed that allow to split the general error term, ε_i, into up to four random components. For example, the **four random components SFA** explicitly models both a, a time-constant inefficiency u_{i0} and a time-varying inefficiency u_{it} as well as a firm-specific heterogeneity b_i, and random noise v_{it}.[22] The regression equation of the model is:

$$c_{it} = \alpha + \boldsymbol{\beta y}_{it} + b_i + u_{it} + u_{i0} + v_{it},$$

where i is a firm, and t is time. To estimate time-varying inefficiency the model requires panel data. In fact, the panel dataset has to be sufficiently large and exhibit sufficient variation over time to identify the different stochastic components in addition to the firm fixed effect, b_i. In the context of cost benchmarking of electricity and gas networks such large dataset will often not be available to regulators. In addition, when estimating this model it is important to consider whether the assumptions made and data available allow to appropriately distinguish between inefficiency (and potential changes in efficiency) and heterogeneity. Besides the relative novelty of these approaches, these problems might help to explain why they have – to our knowledge - not yet been applied in practical work on efficiency benchmarking of electricity and gas networks so far.

3.3.2 Non-parametric techniques

Data Envelopment Analysis (DEA)

DEA is a non-parametric technique that is used to determine an efficiency frontier empirically, without the need to make ex-ante assumptions about the cost function (i.e. how inputs are turned

[22] The four random components stochastic frontier model nests many of the alternative SFA extensions developed in the literature. For a discussion see for instance Colombi, R., Kumbhakar, S. C., Martini, G., & Vittadini, G. (2014). Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. Journal of Productivity Analysis, 42, 123-136.

into outputs). At a high-level, the DEA cost efficiency frontier is a frontier that envelopes all data points and follows certain axioms of production (e.g. convexity). The frontier is calculated by solving a linear programming with some constraints. The linear programme identifies the optimal combination of input and output weights that maximise the productivity (ratio of weighted outputs to weighted inputs) of the firms under consideration given the constraints (this guarantees that each firm is efficient as possible given the assumptions taken).[23]

Figure 8 illustrates the use of DEA. It plots firms' costs on the y-axis against a cost driver on the x-axis. The latter can be thought of as a single output affecting costs or as a 'virtual' cost driver that is obtained as the weighted sum of cost drivers minimis in the case that costs are driven by multiple factors. The DEA frontier combines data points with lowest costs across the range of different cost driver quantities. The level of inefficiency of a given firm can then be determined by assessing its distance from the frontier.

For example, the part of the efficiency frontier associated to firm A in the figure is obtained by the linear interpolation of the cost levels of firms B and C, which are A's closest neighbours on the frontier line. More precisely, the line between points D and E defines the continuum of points that can set the scope of potential efficiency increases, as each point on this line represents the potential to decrease costs without the need to reduce cost drivers (e.g. outputs) or to extend use of one or more cost drivers without increasing costs (i.e. inputs). For example, the distance between points A and D indicates the potential for cost reduction . Given that outputs of network operators are typically fixed in the short term, this is the level of inefficiency typically applied for cost benchmarking of electricity and gas networks.

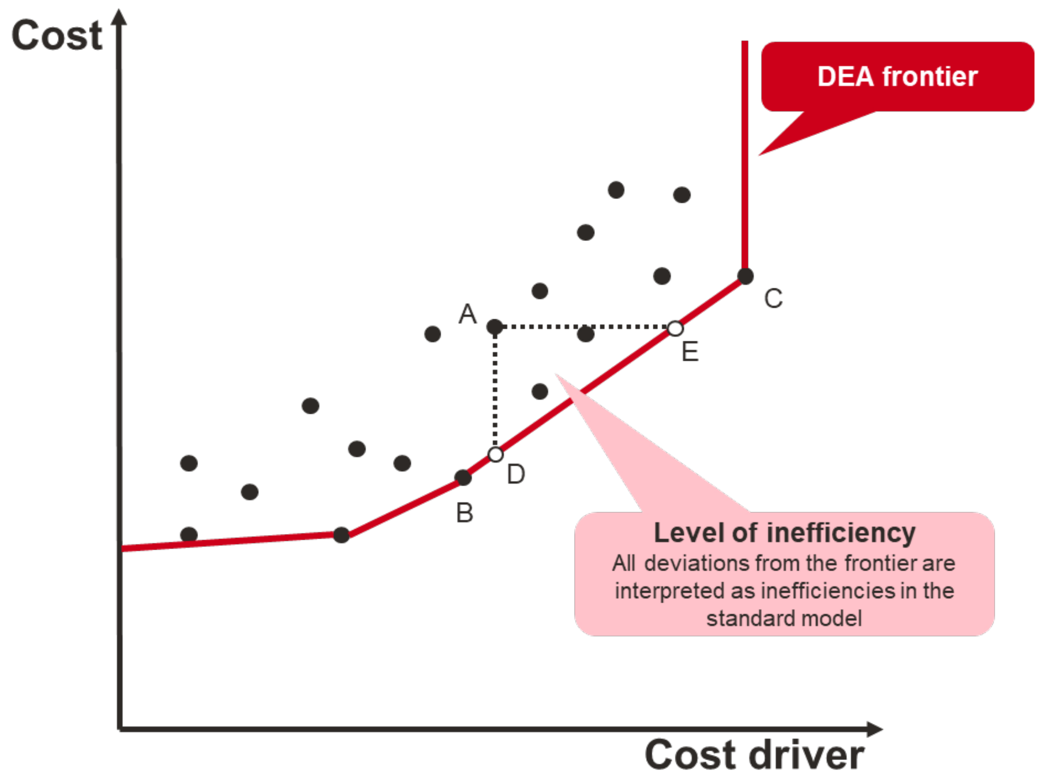
Other approaches can be used to determine the comparison point from which to infer the efficiency (e.g. a point on the frontier between D and E). Generally, the relative efficiency levels obtained from the DEA model are fairly robust to the alternative chosen to calculate the level of inefficiency.[24]

---

23 The ACM already explored DEA in detail in 2017. For details of the study prepared by advising economists at the time see Lawrence, D., Fallon, J., Cunningham, M., Zelenyuk, V., & Hirschberg, J. (2017). Topics in efficiency benchmarking of energy networks: Choosing the model and explaining the results. .https://www.acm.nl/sites/default/files/documents/2020-08/economic-insights-topics-in-efficiency-benchmarking-of-energy-networks-choosing-the-model-and-explaining-the-results.pdf

24 For a short discussion on the effects of the choice of the comparison point in DEA applications see page 133 in Banker, R. D., Charnes, A., Cooper, W. W., Swarts, J., & Thomas, D. (1989). An introduction to data envelopment analysis with some of its models and their uses. Research in Governmental and Nonprofit Accounting, 5(1), 125-163.

Figure 8 Illustration of a DEA efficiency frontier



Source: Frontier Economics

As mentioned above, a key advantage of DEA is that it does not require an ex-ante assumption around the cost function to estimate the efficiency frontier. DEA also allows for different returns to scale assumptions, e.g. constant returns to scale or variable returns to scale. Compared to other non-parametric techniques, such as partial performance indicators or MTFP, the weights used in the DEA estimation to aggregate multiple inputs and outputs are chosen such that every firm appears as efficient as possible for the given data.

These advantages explain why DEA is frequently applied to assess the efficiency of electricity and gas networks. For example, DEA is used in the European benchmarking study of TSOs in Europe (TCB18); in Germany for electricity and gas DSOs; in Norway for electricity DSOs; and in Austria for electricity and gas DSOs.

A disadvantage of the DEA estimator is its known bias in small samples, particularly if many inputs and outputs are modelled. In those situations, the DEA cost frontier is likely to be upward biased , as firms are estimated as being closer to the frontier than they actually are. A further limitation of the DEA model is its deterministic nature: all differences to the frontier are interpreted as inefficiencies, including those that result from general noise in the production process or general data and modelling issues, such as measurement error. The lack of a stochastic component also implies that it is not possible to calculate confidence intervals or other measures of uncertainty of the estimated inefficiency scores in a simple way. This bears the risk that overly strong conclusions might be made based on DEA efficiency estimates that could be quite volatile when estimated using a relative low number of observations relative to the number of cost drivers included in the modelling.[25] Given that DEA is deterministic, it is also not possible to use the technique to estimate and test the relevance or statistical significance of individual cost drivers. Therefore, unlike parametric approaches, a separate tool or technique is required to determine the set of relevant outputs to be included in the DEA model.

DEA variations

This section presents two non-parametric extensions of the standard DEA that address some of the disadvantages identified above: Free Disposable Hull (FDH) and Bootstrap DEA. We discuss a semi-parametric extension – Stochastic DEA – in Section 3.3.3.

Free Disposal Hull (FDH)

The Free Disposal Hull (FDH) is a non-parametric technique closely related to DEA. FDH differs from DEA as it does not require an assumption of convexity (i.e. a linear combination of inputs and outputs of two firms in the sample does not need to be enveloped by the efficiency frontier). In the simple example of one input/one output, the removal of this assumption results in a stepwise efficiency frontier. Figure 9 Illustrates the FDH frontier and the determination of the level of inefficiency graphically.

Generally the convexity of input and output sets is a common and well-founded assumption in economics.[26] Furthermore, the frontier set by the FDH is less ambitious than the DEA frontier. Dropping the assumption of convexity has the material effect that efficiency scores of inefficient companies tend to improve by construction

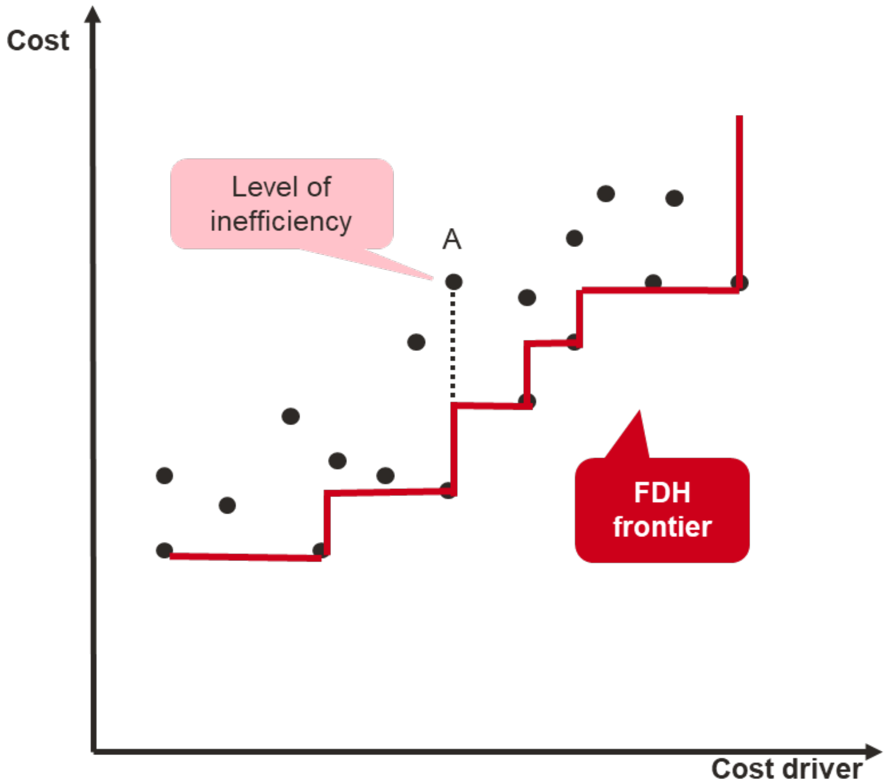
To our knowledge FDH is not applied in the context of cost benchmarking of gas and electricity network operators.

---

[25] We present a rule of thumb on the relationship between sample size and number of outputs in Section 5.3, Table 9.

[26] A short discussion of convexity in economic production theory is provided in Section 3.1.

Figure 9 Illustration of a FDH efficiency frontier



Source: Frontier Economics

Bootstrap DEA

An important extension to the standard DEA framework is the bootstrap DEA developed by Kneip et al. (2008).[27] It employs a bootstrap procedure to correct for bias in the efficiency scores (i.e. difference between estimated score and true efficiency score) and construct confidence intervals of the calculated efficiency scores, which can be used for hypothesis testing These confidence intervals account for uncertainties of the estimated inefficiency scores.

The bootstrap procedure consists of three steps: 1) randomly re-sampling a subset of the data; 2) estimating a DEA on the resampled data (also known as bootstrap sample); 3) deriving metrics of interest (e.g. mean of efficiency scores). This procedure is repeated a number of times and summary

[27] Kneip, A., Simar, L., & Wilson, P. W. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. Econometric Theory, 24(6), 1663-1697.

statistics of the metric of interest are calculated (e.g. the mean of the mean efficiency score). These summary statistics are used to correct for bias and calculate confidence intervals.

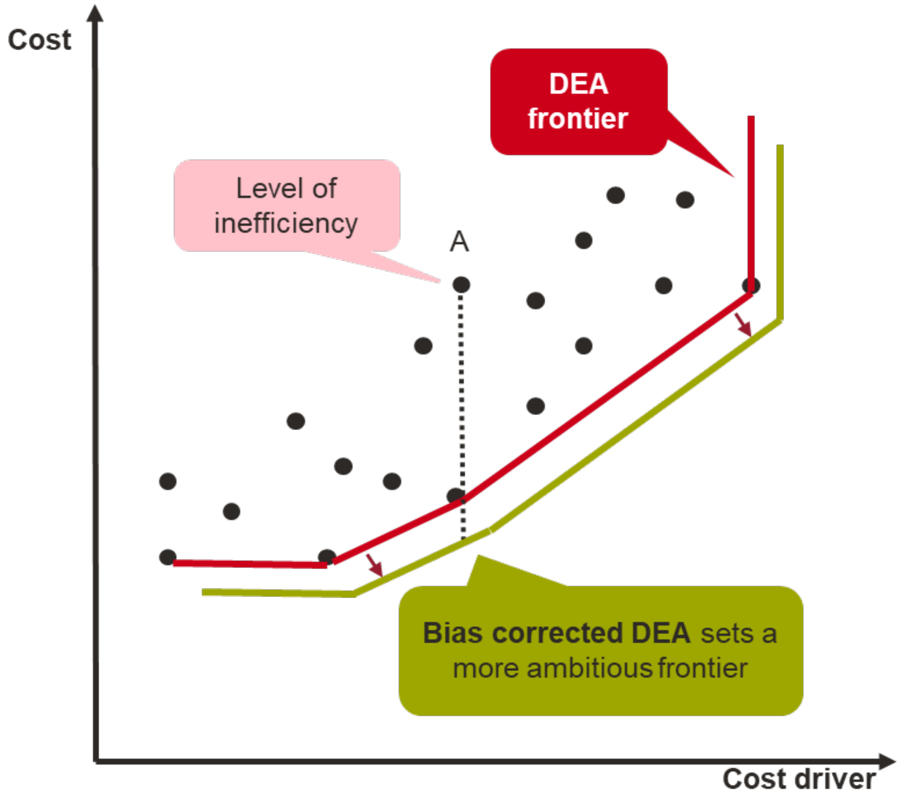
Kneip et al. (2008) show that its bootstrap approach can be used to correct for the known bias of the DEA efficiency scores. The bootstrap bias is estimated as the difference between the bootstrapped efficiency scores (average scores from the bootstrapped samples) and the estimated DEA efficiency score. Kneip et al. (2008) demonstrate that this bootstrap bias can be used to correct for the bias between the estimate DEA efficiency scores and the (unobserved) true efficiency score (after making some adjustments, including adjusting for number of bootstrap samples). Intuitively, the bias exists because of limited sample size, where firms that are not truly efficient are estimated to be on the DEA efficiency frontier. The bootstrapping is an approach designed to mimic this heuristic of gaining additional data so that the estimated cost frontier can be shifted out to account for this observable compression that arises from only have a limited sample.

Kneip et al. (2008)'s approach can also be used to estimate confidence intervals. Usually, a sufficiently large sample is required for the confidence intervals produced by the bootstrap to be sufficiently narrow to be useful in practical applications.

See Annex B.1 for a more detailed discussion including formal representations of the bias corrected/bootstrap DEA.

Other extensions attempt to also extend the DEA framework to explicitly allow for stochastic noise by combining it parametric approaches. These semi-parametric approaches, such as the stochastic DEA, are discussed below.

Figure 11 Illustration of bias corrected DEA



Source: Frontier Economics

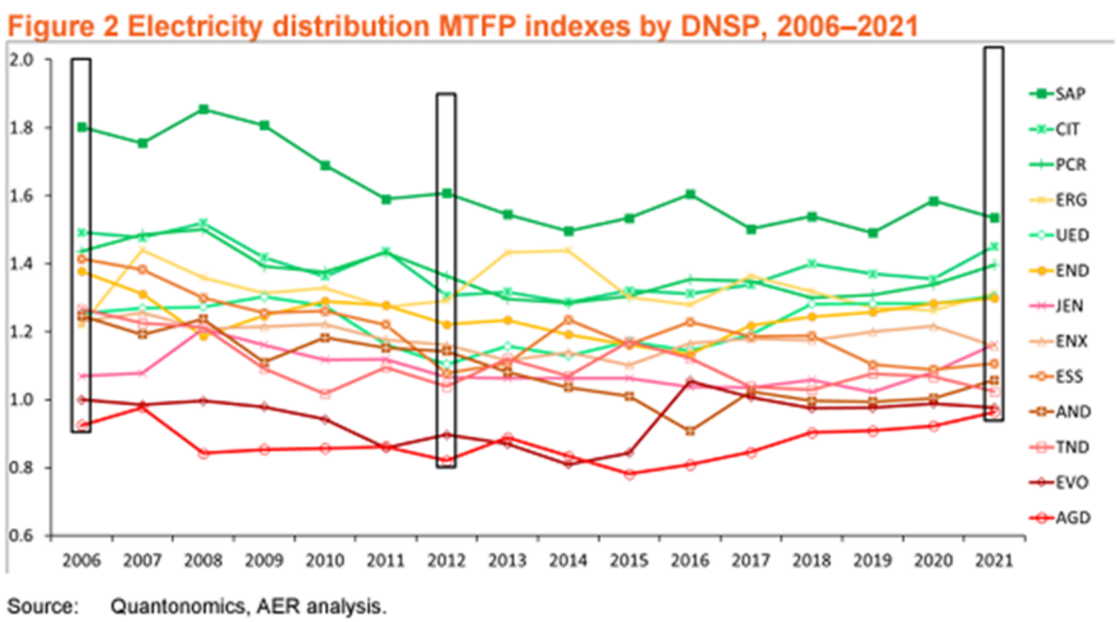
MTFP and MPFP

The multilateral total factor productivity (MTFP) is a non-parametric technique that determines relative efficiency of firms by comparing the output they produces to the inputs they used. Similarly to the DEA, outputs and inputs are aggregated using a set of weights. Different to the DEA, the weights used to aggregate inputs and outputs into a single variable are obtained from observed data such as the revenue and cost shares of outputs and inputs, respectively.

The comparison of the relative efficiency between any two firms and over time is then obtained via their relative performance against a hypothetical firm that represents a sample average. The multilateral partial factor productivity (MPFP) follows the same approach as the MTFP, but only considers one input in the estimation (e.g. opex) rather than all inputs.

There are different approaches to estimate a MTFP index. For example, the Australian Energy Regulator (AER) considers the Caves, Christensen and Diewert (1982)[28] (CCD) multilateral Törnqvist TFP in the Australian electricity network regulation as supplementary benchmarking techniques for both electricity TSOs and DSOs. However, the AER does not use these techniques to set cost allowances (we describe this index in detail in the annex). Figure 10 shows the MTFP that the AER estimated for the electricity DSOs over time. The AER used this analysis to comment on productivity improvements. For example, it found that SA Power Networks (SAP) productivity has decreased between 2020 and 2021, while AusGrid (AGD) productivity has increased. The chart also shows that in 2021 SAP is more productivity than AusGrid (although the AER notes that not all operating differences are taken into account in the analysis and that some DSOs may operate in more or less favourable environments than others and thus appear more or less efficient).

Figure 10 AER's 2022 assessment of MTFP for electricity DSOs over time



Source: AER's 2022 Annual benchmarking report – electricity distribution network service providers.

Usually, MTFP indices are designed to satisfy a number of desirable properties. For example, The CCD index satisfies the following two properties:

28 Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). Multilateral comparisons of output, input, and productivity using superlative index numbers. The Economic Journal, 92(365), 73-86.

- It can be considered an approximation of a cost function (and corresponds to a Translog production function)[29]; and
- The relative performance between any two firms is not dependent on the base firm against which the comparison is conducted.[30,31]

These properties are desirable. The first property indicates the MTFP/MPFP approach has a close link to economic production theory. The second properties makes it easy to compare efficiency across companies, compared to other indices (e.g. total factor productivity (TFP)). Other advantages of the MTFP index are that it can generally be applied quite flexibly, it is transparent and it is also applicable to small data samples. If the required input and output data is available, the calculation and updating of the index is also easily done.

There are also some challenges associated with the use of MTFP for efficiency benchmarking. One challenge is that the index (as most index techniques) is sensitive to the choice of weights used to aggregate outputs and inputs into univariate indices. Another challenge is that in line with other non-parametric approaches, the MTFP does not allow for noise in the production process.

3.3.3 Semi-parametric techniques

Stochastic DEA

Stochastic DEA attempts to mitigate concerns over the lack of noise in the classic DEA estimator while incorporating axioms of production in a nonparametric fashion.[32]

The standard version consists of two steps: first, a parametric stochastic frontier model is estimated and efficient inputs (costs) are derived from this model; second, a frontier is estimated by fitting a DEA on the set of outputs and efficient inputs from the first step.

29 Indices that can provide a second order approximation of a cost function are called ‘superlative’. A second order approximation means that the highest power in the series expansion used in the approximation is 2 (i.e. only linear and quadratic terms used in approximating the cost function).

30 This property is called ‘transitivity’.

31 See Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). Multilateral comparisons of output, input, and productivity using superlative index numbers. The Economic Journal, 92(365), 73-86.

32 See Annex B.1 for a more detailed discussion including formal representations of the stochastic DEA.

A key challenge with stochastic DEA is which parametric model should be used in the first step. Another challenge is that there are several different proposed versions of stochastic DEA,[33] but theoretical or empirical guidance as to which approach to deploy in practice is limited.

It is not clear at present if any form of stochastic DEA has been applied in a regulatory context.

StoNED and related approaches

A limitation of the SFA and related parametric approaches is that cost functions typically estimated are assumed to be linear in parameters (i.e. the function can be expressed as the sum of parameters multiplied by a variable. Parameters cannot appear as exponents or cannot be squared, however variables can). The Stochastic Non-parametric Envelopment of Data (StoNED) approach addresses this limitation by estimating the frontier non-parametrically under the constraint that certain axioms of production are satisfied (monotonicity and convexity). At the same time StoNED keeps the stochastic nature of the SFA by including an error term which can be decomposed into inefficiency and random noise. The decomposition is implemented by making some parametric assumptions on the distribution of the errors. The stochastic nature is also what differentiates StoNED from the group of purely non-parametric approaches such as DEA.

Let $m(x_i)$ denote the conditional mean function of the stochastic frontier model. The regression equation of interest for the application of StoNED and related approaches is:

$$c_i = m(x_i) + u_i + v_i.$$

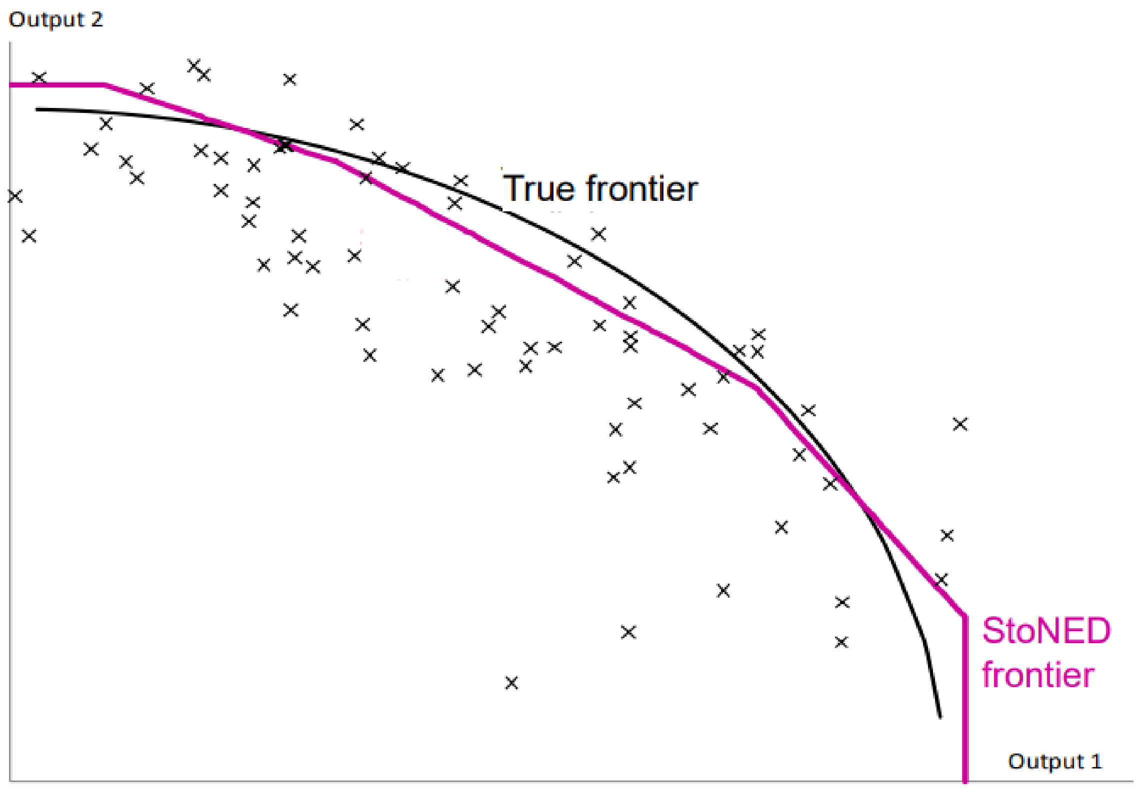
The estimation is similar to the SFA estimation with the use of the PI estimator.[34] First, StoNED estimates $m(x_i)$ non-parametrically without accounting for the non-zero mean of the error term $\varepsilon_i = u_i + v_i$ (this leads to an upward bias in the conditional mean function). StoNED estimates $m(x_i)$ non-parametrically by concave non-parametric least squares (CNLS) under the additional constraint that axioms of monotonicity and convexity are satisfied. Second, the model residuals are used to estimate the parameters of the distribution functions of u_i and v_i under the same type of distributional assumptions that are used in SFA.

33 See for instance Banker, R. D., & Maindiratta, A. (1992). Maximum likelihood estimation of monotone and concave production frontiers. Journal of Productivity Analysis, 3(4), 401-415.. Simar, L., & Zelenyuk, V. (2011). Stochastic FDH/DEA estimators for frontier analysis. Journal of Productivity Analysis, 36, 1-20.

34 In fact, the plug-in likelihood approach has first been developed and applied as part of the StoNED related literature (it was first introduced by Fan, Y., Li, Q., & Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models. Journal of Business & Economic Statistics, 14(4), 460-468.)

Figure 11 compares a StoNED frontier with a true frontier in a *simulated* sample with two outputs. For the specific simulation used in the example, StoNED is found to approximate relatively well the true efficiency frontier.

Figure 11 Illustration of StoNED in a two output example



Source: Timo Kuosmanen (2014), StoNED method in benchmark regulation, BNA conference. https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Netzentgelte/Evaluierung_ARegV/Wissenschaftskonferenz/ARegV_WK20140527_S2-2_Kuosmanen-StoNED-Method-in-Benchmark-Regulation.pdf?__blob=publicationFile&v=2

The StoNED model is closely related to two alternative semi-parametric models with similar properties. The first related model is the one by Fan, Li & Weersink (FLW).[35] The authors developed the two-step PI approach to the estimation of $m(x_i)$ using kernel smoothing, without imposing any axioms of production in the estimation. The second related models is Parmeter and Racine (2012).[36]

[35] Fan, Y., Li, Q., & Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models. Journal of Business & Economic Statistics, 14(4), 460-468.

[36] Parmeter, C. F., & Racine, J. S. (2013). Smooth constrained frontier analysis. In Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White Jr (pp. 463-488).

Parmeter and Racine also imposed axioms of monotonicity and convexity on the estimation, and used kernel based methods for the estimation of $m(x_i)$.

Generally, the three semi-parametric approaches are closely related and can be expected to perform similarly in practical cost benchmarking applications. Table 4 provides an overview of the approaches. The primary difference lies in the non-parametric method used to estimate $m(x_i)$:[37]

- FLW uses kernel regression. Kernel regression essentially runs weighted least squares, where the weights depend upon the distance of an observation to the point of interest. This distance is controlled by a bandwidth. For a larger bandwidth, points further away receive greater weight.
- StoNED uses convex non-parametric least squares (CNLS). CNLS is quite similar to DEA. Line segments (facets) are built such that the estimated function satisfies both monotonicity and convexity, while also ensuring that the resulting error term has mean 0. This last condition is what distinguishes DEA from CNLS. DEA encapsulates the data while CNLS estimates a conditional mean model.
- Parmeter/Racine uses constrained kernel regression. This procedure is very similar to FLW. The only difference is that a second weight is introduced that modifies the original estimator whenever various constraints (here monotonicity/convexity) are violated. The change in the weight is driven by how much the constraint is violated.

A notable difference is the imposition of axioms of production on the non-parametric estimation, which is not part of the FLW approach. There is a general decision regulators have to make, whether they want to maintain these axioms in the cost benchmarking study. An advantage is that the axioms represent plausible properties of cost functions. The assumptions are also commonly applied in other non-parametric techniques such as DEA to provide some basic structure to the production technology. However, the data can differ from the ideal cost function, for instance because measurement error results in the cost function being represented imprecisely in the data. In that case, the additional structure imposed on the non-parametric estimation can lead to inconsistent estimates of the cost frontier.

The major advantage of the non-parametric approaches compared to SFA is that no functional form has to be assumed on $m(x_i)$ (i.e. it is not necessary to decide a priori whether the relationship between cost and cost drivers follows a Cobb-Douglas or Translog function. It thus reduces the risk of obtaining inconsistent estimates due to functional form misspecification. The drawback from the greater flexibility is that the methods become more complex and difficult to understand. While SFA

[37] A more detailed discussion of the FLW, StoNED and Parmeter & Racine approach that also includes formal representations is provided in Annex B.1.

is still based within the standard OLS framework, the application of StoNED and related approaches requires techniques that are likely new for regulators and regulated operators.[38]

StoNED has been applied in Finland for benchmarking of electricity DSOs since 2012, so there is regulatory precedent. Given the high similarity between the three methods (FLW, StoNED, Parmeter/Racine), lessons learnt from the practical application of StoNED from Finland could be taken into account when applying the FLW and Parmeter/Racine approaches.

Table 4 Overview of StoNED and related approaches

	FLW	StoNED	Parmeter/Racine
General estimation approach consists of three steps	1. Estimate cost frontier non-parametrically. Estimation will lead to biased results as the conditional error does not have mean zero due to the one-sided distribution of inefficiency that has not been accounted for. 2. Make distributional assumptions on the distribution of error terms to estimate the distributional parameters on the non-parametrically estimated residuals. 3. Shift the conditional mean by the estimated mean of inefficiency to correct for the bias in the first step.		
Estimation method in Step 1)	Kernel regression	Concave non-parametric least squares	Constrained kernel regression
Axioms of production imposed on Step 1)	None	Monotonicity and convexity	Monotonicity and convexity
Regulatory precedent	None	Finland	None

Source: Frontier Economics

[38] For a discussion on this topic, refer to 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). Journal of productivity analysis, 41, 85-109.

3.4 Engineering-based techniques

Engineering models or assessments

Engineering assessments can be applied for an assessment of incurred or expected costs for capital-intensive investment projects. Engineering based approaches that involve aspects of network planning typically focus on the investment cost and capital expenditure required. To assess the efficiency of totex, assumptions would need to be made regarding the relationship between certain asset types and operating expenditures (e.g. for maintenance purposes).

The investment projects considered can differ in size and type. For example, these investment projects could consists of large load-related investment projects into new transmission assets (lines, transformer stations), large non-load related investment projects in replacement of specific assets (e.g. a maintenance program for replacement of a specific cohort of transformer stations), and non-load related investment projects in asset replacement.

The engineering assessment requires technical expertise to evaluate the optimal design, need, and costs for such investment projects. As energy regulators typically lack the specific knowledge, this assessment is normally undertaken by technical consultants having the practical knowledge from planning and implementing similar projects. The assessment includes various instruments. For example, unit costs can be used to assess the cost efficiency of the incurred/expected costs provided by the network operator; statistical analysis can be used to assess the need for replacement investments (e.g. assessing the ratio between fault of components and their age); and engineering planning can be used to assess the configuration of a specific large investment project (e.g. dimensioning of lines and transformers).

The advantage of engineering assessments is that if the technical consultant can draw on knowledge from similar projects, then no direct comparators are required . In regulatory applications (partly) relying on information from other network operators on comparable projects may be necessary for the acceptance of the results by network operators, for example when assessing certain unit costs Engineering assessments allow a disaggregated view on incurred/expected costs which should help in identifying potential inefficiencies for specific cost items.

Engineering assessments can be considered a regulatory tool related to micro-management as they require detailed data from the network operators which is necessary for the in-depth assessment of projects (e.g. a breakdown of volumes, unit costs used, justification of projects). Providing and assessing this data can be time-consuming for network operators and regulatory authorities. The results on allowed costs usually rely on the expertise from technical consultants, which may partly draw upon non-public proprietary information from other projects or their subjective assessment. Hence, there may be issues with transparency.

Ofgem used engineering assessments as part of the disaggregated benchmarking of the cost activities for electricity distribution operators in RIIO ED2.[39] A key component of the activity-level cost assessment was the engineering review of operators' so-called 'Engineering Justification Papers' (EJPs). The purpose of the EJPs was to provide justifications for load related and non-load related investments. Ahead of the assessment, Ofgem published guidelines[40] on the content and covered investment projects for EJPs. Each of the operators' EJPs was reviewed and cross referenced against other supporting documents. The engineering assessment of an EJP led to one of three outcomes: (i) justified; (ii) partially justified; and (iii) unjustified. Ofgem then decided whether to allow the investment based on this assessment.

Process benchmarking

Process benchmarking[41] is a technique for comparing processes of operators. These can be processes across departments, operators with similar objectives, or operators that have different tasks but use comparable processes. The key objective is to identify, and then implement, ways of improving an operator's processes. The efficiency of processes is assessed by key performance indicators (KPIs), which may include PPIs and unit costs.

Process benchmarking is implemented on a disaggregated cost level. It goes a step further compared to PPI as it not only compares the 'cost per activity' (i.e. maintenance costs / network length) but also the underlying processes for the activity. For example, attending an incident involves the following high-level processes: maintenance team is informed of the incident, the team approaches the incident, the team addresses the incident. For each of these processes an assessment could be made around costs incurred considering time required, number of full time employees involved, resources used, etc.

Process benchmarking is a regulatory instrument related to micro-management as it requires detailed data from the network operators. When applied to a particular operator, process benchmarking considers a group of processes of that operator. The group of processes has to be clearly separated from other processes to enable a proper comparison between companies. For processes that are comparable across operators, it may be possible to benchmark the processes by collecting data from the operators, using benchmarks of KPIs from publicly available sources

39 Ofgem. (2022). RIIO-ED2 Final Determinations Core Methodology Document. https://www.ofgem.gov.uk/sites/default/files/2022-11/RIIO-ED2%20Final%20Determinations%20Core%20Methodology.pdf

40 Ofgem. (2021, February). Engineering Justification Papers for RIIO-ED2. https://www.ofgem.gov.uk/sites/default/files/docs/2021/02/riio_ed2_engineering_justification_paper_guidance.pdf

41 For more details on process benchmarking we refer to: E-Bridge. (2012). Study of the feasibility of determining TenneT's cost efficiency via process benchmarking: Study for Nederlandse Mededingingautoriteit (NMa).

(potentially from other sectors), or asking specialised consultants to identify relevant benchmarks. The more sector-specific the processes the less the analysis can rely on data from other sectors.

The main advantage of process benchmarking is that it can provide in-depth information on operators' performance on a disaggregated level and where any inefficiency might be coming from. The drawback is that it can be time-consuming and complex to identify like-for-like operators, define like-for-like processes, and gather like-for-like data. For example, different accounting principles (e.g. allocation of costs to specific processes) can distort the results. If participants consider the benchmarking process to be non-transparent or unpredictable, they will tend to be hesitant to accept the outcome.

Reference Network Analysis (RNA)

RNA is an analytical cost model approach which consists of designing concrete, optimal networks for real transport and supply tasks.[42,43] It can be used to assess the price tag of existing and planned physical networks. RNA can be applied as a 'greenfield' or 'brownfield' approach. A pure 'greenfield' approach ignores the existing network and builds a new network from scratch. A 'brownfield' approach recognises the history of the network development and retains certain parts of the existing grid as a constraint in the optimisation. In either case, the costs of building the network required according to the RNA model can be used as a benchmark for the actual costs of an electricity or gas network.

The application of RNA can be divided in three stages:

- First, the researcher identifies the key elements of the transport and supply task at hand, which is typically understood to be, at a minimum, the location and size of infeeds and offtakes of the network.
- Second, based on these key elements, the researcher designs an optimal network specification (e.g. optimal amount of network length, distance between customers and substations, etc.). This is generally done using network planning models based on engineering rationale.
- Third, a price is assigned to each asset of the network specification, generally based on annuities. Total capital costs are then derived from these prices. Operating costs can sometimes be derived using a markup on capital costs.

The results from the RNA can be used for efficiency benchmarking in three ways:

42 RNA models in the telecommunication sectors are sometimes called BULRIC (bottom-up long-run incremental cost).

43 For more details on RNA we refer to: Frontier Economics/Consentec. (2012). The potential application of reference network modeling to TenneT: Feasibility study for NMa.

- **Absolute reference model**: the operator in question is compared directly against the cost derived from a reference model.
- **Relative reference model**: a number of operators are modelled and for each the ratio of actual cost to modelled cost is constructed. Operators are then assessed on the basis of this ratio, e.g. requiring all operators in the sample to get 'as close' to their optimal model as the best performing operator.
- **Input to another benchmarking technique**: in principle a reference model can be used as an approach to derive structural variables that capture the 'scale' of the transport and supply task. Interpreted in this way, these variables can be used as input (i.e. serving as operators' output parameter or structural parameter) to another benchmarking technique.

A key advantage of RNA for cost benchmarking is that it can be applied in situations in which there is a lack of information that would allow a comparative analysis. This can for instance be the case if there is only a low number of comparable firms or if a specific type of investment is so new that no comparable projects exist. The procedure can also be a useful approach to cover the heterogeneity of network operators and derive cost drivers explaining the supply tasks of operators. That is also the reason why results from RNA are used as input to other benchmarking techniques. For example, one of the key differences between network operators is density, i.e. whether they serve an urban or rural area. RNA can be used to identify the relationship between the network length that you would need to serve an area with a given density. This 'modelled' network length variable can then be used as a cost driver for other benchmarking technique.

A major caveat of the approach is that there can be high costs in terms of the amount and detail of data required for the analysis and the effort that has to go into the modelling. This is particularly true if only a few or no comparable firms exist and the analysis has to be carried out in a detailed way to obtain reliable numbers for every step of the modelling.

Currently, no European energy regulator uses RNA as a benchmarking technique for DSOs. The experiences from the Swedish regulator with the Network Performance Assessment Model (NPAM) were not that convincing as it has led to serious conflicts between the regulator and a number of utilities resulting in lengthy legal proceedings involving court rulings and appeal cases. The NPAM was only applied from 2003 to 2006.[44] However, RNA is currently applied for benchmarking of electricity TSOs in Germany[45].

RNA aims to construct a detailed model of the layout of the network (i.e. creating a detailed spatial model) making use of a considerable body of data (e.g. spatial demand and generation profiles,

---

[44] Jamasb, T., & Pollitt, M. (2007). Reference Models and Incentive Regulation of Electricity Distribution Networks: An Evaluation of Sweden's Network Performance Assessment Model (NPAM). EPRG Working Paper.

[45] For details we refer to the case study 'Germany – Electricity TSOs' in section 7.5.

transmission line routes, standardised network assets). Another approach is the so-called Model Network Analysis (MNA). The MNA approach makes a number of simplifying assumptions in order to reduce the complexity of the model and the volume of data that is required to undertake the work. Both approaches produce an optimal network specification, but differ with regards to the details of the network. In particular, MNA can be used to gain a deeper understanding of how the supply task of a given operator's service area might be expected to impact its network configuration and hence its costs.

For example, MNA[46] has been used successfully in this narrower role in Austria by E-Control for electricity DSOs, where stylised models of the networks were developed in order to provide proxy variables (modelled network length, by voltage level) that captured connection density. These proxy variables were then used in other benchmarking techniques (DEA, MOLS) to assess electricity DSOs' efficiencies.[47] An approach of this kind ensures that important structural effects can be captured robustly using variables that are not under the direct control of the operators. For this application the data requirements are less strict as compared to using RNA as a benchmarking technique. MNA was also applied by Bundesnetzagentur to identify and sense check the cost drivers for the benchmarking analysis for electricity and gas DSOs.

46 E-Control did not apply RNA for this task but the less complex Model Network Analysis (MNA).

47 See for example this E-Control report. https://www.e-control.at/documents/1785851/1811528/Entscheidungen-der-Regulierungsbehoerde-Ausgestaltung-3te-Periode-Strom.pdf/225b49e0-6534-40e4-afa1-97d83f8edbde?t=1413905499198

4 Evaluation of the long list of benchmarking techniques

In this section we summarise our approach to move from the long list of techniques to the short list. We first describe the approach that we use to short list benchmarking techniques and then present our evaluation of the long list of techniques.

4.1 Our evaluation approach

We have used four criteria to short list the long list of benchmarking techniques:

- Promotion of efficiency
- Transparency
- Applicability
- Robustness.

We identified these criteria by considering ACM's regulatory context and expected future challenges, described in Section 2.2, as well as benchmarking best practice.

Promotion of efficiency

This criterion is used to assess whether a technique is able to:

- Identify historical cost efficiencies;
- Potentially identify areas for efficiency improvements. This is to address one of the implications of CBb's ruling; and
- Account for future efficiency considerations, e.g. to ensure that necessary investments into the energy transition are not disincentivised.

The benchmarking techniques we included in our long list could all be used to determine the relative efficiencies of TSOs and DSOs. However, some techniques might be better suited at identifying historical cost efficiencies, and others better placed at identifying areas for efficiency improvements. The other components of the benchmarking analysis (costs, cost drivers, data) also play a role in this (e.g. future efficiency could be determined by estimating a COLS on a sample that includes forecast data or by using a historical dataset but including appropriate forward-looking cost drivers). We consider this combination of factors when applying this criterion to evaluate the long list.

4.1.2 Transparency

This criterion is used to assess whether a technique can be considered transparent along two dimensions:

- The **technique is well understood**, including its key assumptions and implications for the estimation of efficiency.
- The results from the techniques can be **replicated by interested third parties** (e.g. the TSOs and DSOs). This requires that the data used and potentially the implementation of the technique be made available to those third parties or their advisors. Simpler techniques are usually easier to replicate.

Transparency can increase the confidence that the regulators as well as the TSOs and DSOs place on the results of the techniques. For example, if the impact of key assumptions is well understood (e.g. because there is extensive regulatory precedent), the regulator is in a position to make better use of its results (e.g. if efficiencies are known to be biased the regulator could adjust them).

Transparency can also improve how TSOs and DSOs engage with benchmarking. This could ultimately lead to a better benchmarking process, from model development to understanding which conduct is rewarded and which is disincentivised.

Ensuring the transparency of the technique was one of the key implications of CBb's ruling.

4.1.3 Robustness

This criterion pertains to whether the results of a technique are likely to remain more or less the same as some of the underlying assumptions (e.g. choice of distribution of inefficiency) or data are changed slightly (e.g. use of different proxies for the same cost drivers; addition of one year of data; removal of one comparator).

Robust techniques are likely to be more useful to regulators and the sector, because regulators can be more confident about their findings and the acceptance by operators of results should increase. If the results of a technique strongly depend on a specific assumption or change significantly with small changes in the data, it is possible that these results will be questioned and potentially overturned on appeal.

4.1.4 Applicability

This criterion is used to assess the feasibility of implementation of a technique. We consider three dimensions:

- Data requirements. We consider whether the data required for the technique is publicly available or can be collected, and whether the quality of the data is sufficient.
- Sample size. We consider the sample of comparator firms that is required to implement the technique. Some techniques require a larger sample size than others.

Proportionate resource costs. We make some qualitative considerations around the resources that ACM will likely require to apply a specific technique and related components of the benchmarking analysis (e.g. collecting a sample of comparators, identifying cost drivers, testing robustness, deciding how to use the results of the benchmarking analysis).

4.2 Our short list of techniques

The table below shows our short list of benchmarking techniques. We first provide a summary of our rationale for selecting these techniques. We then present a table that explains for each technique the key reasons for its inclusion or exclusion from the short list.

Table 5 Our short list of benchmarking techniques

Group	Sub-group	Technique
Descriptive technique	Performance indicators	Partial Performance Indicators (PPIs)
Mainly based on economic theory	Parametric	Corrected OLS (COLS), Modified OLS (MOLS) Stochastic Frontier Analysis (SFA)
	Non-parametric	Data Envelopment Analysis (DEA) Bootstrap DEA
	Semi-parametric	Stochastic non-smooth envelopment data (StoNED)
Mainly based on engineering rationale	Engineering based	Engineering models Reference Network Analysis

Source: Frontier Economics

4.2.1 Summary of our reasoning for the short list

We selected a subset of techniques within each of the groups from Section 3. The key reason for this is that techniques within each group have some unique advantages compared to other groups:

- **Descriptive techniques.** These techniques can be used for an exploratory analysis that serves as input for more complex techniques as well as to identify where inefficiency is coming from.
- **Economic-based techniques**. These techniques can be used to make an assessment of overall efficiency in a relatively simple and transparent way. Within this group, each of the parametric, semi-parametric, and non-parametric techniques has its own advantages and disadvantages.
- **Engineering-based techniques**. These techniques could be particularly useful in understanding the efficiency of costs of new investment projects or costs related to new activities, especially when historical data is scarce.

Within each group, we apply our evaluation criteria to short list the techniques and select those techniques that, on balance, meet most of the criteria. We also give weight to whether there is precedent for the use of a technique in the regulation of TSOs and DSOs.

Amongst descriptive techniques, we shortlisted PPIs. We consider that PPIs complement the parametric and non-parametric techniques as they could be used to either identify where the inefficiency might be coming from or benchmark some specific large investment projects.

For the parametric, semi-parametric, and non-parametric techniques we have selected those techniques that we consider score relatively well on all of the evaluation criteria and that have been used in other regulatory contexts. There are pros and cons to each of these techniques and there is merit in combining results of techniques from different groups. We discuss this in Section 6.

Amongst engineering-based techniques, we selected engineering models and RNA. Similarly to PPIs, engineering models could be used to either identify where the inefficiency might be coming from or benchmark some specific large investment projects. RNA and process benchmarking require a detailed engineering and business understanding of the networks and their processes. We consider that the benefits that these two techniques might bring compared to other techniques in the short list are likely to be outweighed by the resources required to implement these techniques and validate the results. ACM is interested in exploring RNA in more detail. Therefore we included it in our short list.

Key reasons for inclusion or exclusion from short list

In the table below we have summarised the key reasons for inclusion in or exclusion from the short list.

Table 6 Summary of our evaluation by technique

Technique	Short-list	Key reason for inclusion in/exclusion from short list
Parametric		
OLS	No	Does not allow to identify efficiency frontier.
COLS	Yes	No statistical uncertainty included when determining efficiency frontier. However, there is regulatory precedent (e.g. in the UK and Australia) and the technique is well understood.
MOLS	Yes	Compared to OLS and COLS, allows to determine efficiency frontier including a measure of statistical uncertainty based on the distribution of the error terms. There is regulatory precedent (e.g. E-Control applies MOLS for benchmarking of electricity and gas DSOs).
SFA	Yes	Compared to MOLS, it allows to split the error term between inefficiency and random noise. However, it is more data intensive than MOLS. There's regulatory precedent for this technique (e.g. Bundesnetzagentur applies SFA for benchmarking of electricity and gas DSOs; the AER applies SFA for electricity DSOs).
Quantile regression	No	From a conceptual point of view, it is similar to COLS. It not only shares the disdavantages of COLS (not allowing for statistical uncertainty and requiring some regulatory judgement to set the frontier), but there is also no regulatory precedent and it is less well understood. Given that COLS is included in the short list we therefore exclude quantile regression from the short list.
Four random component SFA	No	Compared to SFA, it allows to further split the error term between short and long run sources of inefficiency, heterogeneity, and random noise. It is more data intensive than SFA. It is necessary to make assumptions with regards to the four components (distribution of inefficiencies, etc.) which makes the technique more complex and assumption driven. As far as we are aware there is no regulatory precedent for this technique. Given these downsides we excluded it.
Non-parametric		

Technique	Short-list	Key reason for inclusion in/exclusion from short list
DEA	Yes	Compared to MOLS and SFA it does not require assumptions on functional forms. By construction, efficiency increases with number of inputs/outputs included and decreases with number of observations. In theory, it requires only a small sample, but in order to get reasonable results the technique requires a large enough sample. We included DEA as it addresses some of the disadvantages of parametric approaches (although it has its own disadvantages), is simple to understand and there is regulatory precedent for the application.
Bootstrap DEA	Yes	This is similar to DEA. Key difference is that it attempts to understand the distribution of efficiency scores and potentially adjusts for biases. The confidence intervals around the efficiency scores that this technique produces are potentially large. We included it in the short list as an additional technique for sensitivity checks to DEA.
MTFP/MPFP	No	Could be compared to unit cost models as MTFP/MPFP can be seen as a ratio of an output index and an input index. Index measures are sensitive to the choice of the weights used to determine the output and input index. The advantage is that it does not require a large sample. MTFP/MPFP is normally used as a method to determine dynamic efficiencies (i.e. changes in productivity) and not static efficiencies. Other techniques such as DEA and MOLS are designed to identify static efficiencies. There is no regulatory precedent of using MTFP or MPFP to set catch-up efficiencies (the AER in Australia considers it, but it does not use it).
Semi-parametric		
Maindiratta, FLW, Parmeter and Racine	No	These techniques are similar to StoNED/DEA, but – to our knowledge - have not been applied in a regulatory context yet. The models therefore are also not available in commercial statistical software. For these reasons, we excluded them from the short list (and included the related StoNED approach instead).
Stochastic FDH/DEA	No	This is a two step approach, where both stochastic and non-stochastic benchmarking models are applied in sequence to ensure

Technique	Short-list	Key reason for inclusion in/exclusion from short list
		that monotonicity and convexity are satisfied. There is no regulatory precedent. We excluded them from the short list.
StoNED	Yes	This is similar to FLW in terms of implementation. The difference is that there is regulatory precedent (Benchmarking of electricity DSOs in Finland) and it is available in commercial software.
Performance indicators		
Partial Performance Indicators (PPI) (which includes unit cost models)	Yes	We propose to define PPI in a wider sense, to include PPIs based on aggregated costs (e.g. total opex) or disaggregated costs (e.g. maintenance opex, planning opex, investments). PPI does not allow to identify overall efficiency. However, it can be useful to identify where inefficiency comes from and does not necessarily require an international sample. This technique could complement top-down techniques like MOLS and DEA.
Engineering-based models		
Reference network analysis (RNA)	Yes	This technique relies on detailed engineering models. Because of its complexity it may not be considered transparent by all stakeholders. Also, detailed data are required. These may not be made publicly available or shared with all interested parties. We consider it too resource intensive to collect data, implement and validate. We included this technique as ACM expressed interest in understanding the potential advantages and disadvantages of RNA in more detail .
Process benchmarking	No	This technique analyses the different processes for a certain activity. It uses a number of techniques, including engineering models, unit costs, and other PPIs to understand specific processes. We think it is too resource intensive to implement and validate to be used to inform high-level efficiency estimates.
Engineering models	Yes	Engineering models allow to assess the efficiency of specific processes or activities, e.g. the efficiency of specific large investments or whether a replacement investment program is cost-

Technique	Short-list	Key reason for inclusion in/exclusion from short list
		efficient. They do not allow to identify overall efficiency. However, this technique could complement top-down techniques like MOLS and DEA.

Source: Frontier Economics.

5 Comparative analysis of short-listed techniques

In this section we undertake an in-depth comparative analysis of the short-listed techniques against each of the four evaluation criteria from Section 4. In Table 7 we list the questions that we have considered and summarise our findings.

Table 7 Summary of findings of comparative analysis

Criterion	Question	Findings
Promotion of efficiency	Which technique should be used to benchmark a given cost category?	■ Econometric techniques likely more appropriate for high-level cost categories and business as usual activities ■ Engineering models and RNA likely more appropriate for lower-level business as usual activities, or new activities (especially where significant new investments can be anticipated)
	Which technique can be used to identify where inefficiency is coming from?	■ PPIs can provide a high level indication and are relatively simple to implement ■ Econometric techniques can be applied to disaggregated cost data, but usually the quality of disaggregated data is lower ■ Engineering based models can be used for specific activities
Transparency	Which technique is more transparent?	■ More transparent: PPIs, simple engineering models, econometric models (COLS, MOLS, SFA, DEA, StoNED). We consider that all these econometric models are transparent as the implementation of the model is clear. However, DEA and StoNED might be considered less transparent as they do not show the relationship between costs and cost drivers as explicitly as the other models. There is also limited precedent in the use of StoNED. ■ Less transparent: complex engineering models, RNA

Criterion	Question	Findings
Applicability	What is an appropriate sample size?	■ Econometric techniques require a larger sample size than engineering models, RNA and PPI (particularly relevant since sample sizes may be small in the Dutch context)
	How to account for heterogeneity?	■ All techniques can account for heterogeneity. This can be done within the technique or by adjusting cost data ex-ante or results ex-post.
	How to account for economies of scale?	■ Econometric techniques (e.g. DEA, SFA) and PPIs can be used to estimate economies of scale ■ Economies of scale can be assumed for all techniques
	Which technique can be used with forecast data?	■ All techniques can accommodate forecast data, which can either be included directly in the estimation or used to forecast efficient allowances
	Which technique is likely to be more resource intensive?	■ PPI is the least resource intensive, followed by economic-based models. Complex engineering models and RNA are the most resource intensive
Robustness	How do you ensure robustness?	■ By ensuring good data quality and testing results of models for small variations in data and assumptions

Source: Frontier Economics

5.1 Promotion of efficiency

Which technique should be used to benchmark a given cost category?

The technique to use to benchmark a given cost category depends on how the cost category is classified. Costs incurred by a business can be classified according to two key dimensions:

- Type of costs – whether the cost is due to business-as-usual activities or new/bespoke activities;
- Level of disaggregation – whether the cost represents a specific activity (lower level of aggregation) or a group of activities (higher level of aggregation).

For business-as-usual activities, the relationship between cost and cost drivers is well understood. This is because the operators have been undertaking the activity for some time. Hence, it is also likely that comparative data exists to benchmark the costs driven by these activities. Given that comparative data is likely to be available, it is possible to benchmark these activities using economic-based models.

For new or bespoke activities, the relationship between cost and cost drivers might not be well established. For example, because the operator is inexperienced. Even if the operators have an understanding of this relationship, if the activity is unique to a small subset of operators, information asymmetries might make it more difficult for the regulator to establish this relationship. Also, where a new activity only commenced recently, it may not yet show in the available data at all or only to such an extent that it cannot be distinguished from noise. Therefore, to benchmark the costs driven by these types of activities engineering-based models or RNAs might be more appropriate.

The level of disaggregation can also affect which techniques to use. For lower levels of aggregation, PPIs or simple econometric or engineering models might be better suited than more complex economic-based models. The reason for this is that economic-based models require a sample of comparative data and results depend on the quality of data. At lower levels of aggregation, it may be more difficult to collect precise data or ensure that operators allocate costs and cost drivers to disaggregated activities in a comparable way. For example, operators might follow different accounting practices to allocate overheads between different activities or some operators might prefer equally efficient opex solutions to capex solutions.

Figure 12 illustrates which techniques are usually better suited at benchmarking costs along the two dimensions (type of costs, level of disaggregation).

Figure 12 Techniques by cost category

Source: Frontier Economics

For a given cost category, the academic literature has investigated which of the most known economic-based techniques are likely to perform best at identifying efficiencies. However, there is no clear consensus. Findings seem to depend on the true relationship between costs and cost drivers which is assumed in those academic studies (see box below for more details).

Comparison of performance of different economic-based techniques using Monte Carlo analysis

A number of academic studies have undertaken Monte Carlo simulations to identify which techniques are better at identifying true inefficiencies. These studies take the following approach:

- The researcher generates a dataset of firms, with given inputs and outputs by specifying a data generating process (DGP). This allows the researcher to know what the 'true' efficiency frontier is
- Different techniques are applied to this data and the resulting efficiency scores estimated
- The scores are compared with the 'true' efficiency scores using some performance metrics (e.g. Mean Squared Error)
- The steps above are repeated for various types of DGPs

Based on the studies we have reviewed, performance depends on the assumptions underlying the DGP, so there does not seem to be a clear preferred option that performs best in all cases. The table below summarises the findings of these studies.

Study	**DGP**	**Performance**
Andor & Hesse (2014)	Variation of production functions (Cobb-Douglas, Cresh,[48] Translog), noise-to-signal ratios, distribution of inefficiency and inputs, sample size, heteroscedasticity, number of inputs, omitted variables, collinearity	SFA performs on average best, but StoNED is found to also perform relatively well, especially in noisy settings. DEA performs less well
Andor, Parmeter & Sommer (2019)	Variation of production functions (parametric functions, Cobb-Douglas and Translog, as well as non-parametric approaches combining the Translog with a DEA and using DEA only), noise-to-signal ratios, distribution of inefficiency, sample size	The method that performs the best, on average, is the combination approach involving taking the maximum of the DEA and SFA estimates. Regarding individual methods, SFA with a Translog production function outperforms the others, on average, followed by DEA with variable returns to scale, SFA with a Cobb-Douglas production function, and DEA with constant returns to scale.
Andor & Parmeter (2017)	Variation of production functions (Cobb-Douglas, Translog), noise-to-signal ratios, sample size	At times MOLS performs better than SFA.

Which technique can be used to identify where inefficiency is coming from?

In general, all of the techniques can be used to identify the root cause of inefficiency.

- **PPIs** (including unit costs) can be derived and compared across operators. This can provide an indication of where inefficiencies might be coming from. However, exogenous differences between operators also need to be taken into account. This is the approach that the AER uses when comparing PPIs of key unit costs across operators (e.g. opex per customer for operators serving areas of similar population density).
- **Economic-based models** can be applied at lower levels of disaggregation. Considering lower levels of disaggregation should enable the use of more specific drivers of costs at that level. The UK regulator, Ofgem, uses economic-based models on disaggregated costs as part of its benchmarking toolkit.[49] However, it is important to be mindful of the quality of the data available. It is unlikely that applying economic-based models to high level costs might indicate where inefficiency is coming from, as specific activities are not benchmarked separately at those higher level. A possible exception is DEA (and its variations). DEA allows to identify 'peer companies', i.e. operators that are located on the efficiency frontier and have similar outputs to the operator under consideration. If the sample is large enough (and therefore the peers are similar to the operator under consideration), it might be possible to further analyse these peers (i.e. unit costs for these peers for key scale drivers) against the benchmarked operator, to understand where inefficiencies might be coming from.
- **Engineering models** might be better placed to benchmark some specific activities, e.g. large capex projects, to attempt to understand whether these have been undertaken efficiently. The UK regulator, Ofgem, applies this approach when assessing costs for electricity TSOs and as part of the benchmarking toolkit for electricity DSOs.

5.2 Transparency

Which technique is more transparent?

The level of transparency of a technique can be assessed along the amount of data required, assumptions and implementation of the technique, and whether data is likely to be publicly available or confidential. The most complex techniques are those that require a large amount of data (particularly data that is confidential and difficult to verify) and a large number of assumptions.

[48] For a definition of a CRESH cost function, see for example Hanock. (1971). CRESH production functions. Econometrics, 29(5). https://www.jstor.org/stable/1909573

[49] For more details we refer to the case study 'Great Britain – Electricity DSOs' in Section 7.2.

The most transparent technique is PPI. This is because the calculation is straightforward and no (or a limited number of) assumptions are required for it. In addition, PPIs can be based on high-level cost and cost drivers, which are usually available to interested stakeholders.

The second most transparent group of techniques are economic-based models (and simple engineering models). While it can take some time to understand the nuances around their assumptions, these models are usually well known and understood. These techniques can be based on high-level cost and cost drivers.

The least transparent techniques are likely to be complex engineering-based models and RNA. This is because the amount of data used and the assumptions required can be large and they require specific engineering and detailed network modelling knowledge. Moreover, data used for these techniques could be considered confidential by some stakeholders (e.g. some RNA models would require geographic location of assets).

5.3 Applicability

What is an appropriate sample size?

Different techniques require different samples of comparator firms in order to be implemented.

Engineering-based models and RNAs can be implemented without a sample of comparator firms. These models estimate efficient costs of activities by using engineering rationale or network optimisation. However, some comparator data might be used to determine assumptions on unit costs that enter the models.

PPIs can also be calculated based only on data of the operator under consideration. In this case it is possible to compare the development of PPIs over time for this operator. However, in order to interpret the results of PPIs it is useful to compare the PPIs across comparator operators.

Economic-based models require a sample of comparator firms. This is because the relationship between costs and cost drivers is estimated empirically exploiting the variation in this relationship in the sample. Usually most of the variation is cross-sectional, i.e. between operators. Therefore, while extending the sample over time is useful to account for some year effects, adding more comparators is usually more useful.

Usually, the larger the sample size the better as more cost drivers can be taken into account. More complex parametric techniques (e.g. SFA, StoNED) require a larger sample as more parameters are estimated. However, the larger the sample the more difficult it is to ensure that data is comparable across operators and all exogenous differences have been accounted for. This is particularly true of an analysis that has to rely on international data, because the regulator might have less knowledge or control on data coming from other jurisdictions.

The choice of the sample is important when implementing economic-based models. Table 8 below summarises relevant regulatory precedents on sample size. Generally, the academic literature only provides very few rules of thumb for the minimum sample size required in an empirical application. Table 9 provides on overview of some approaches that could be used for different techniques. The applicability of a certain method with a given data size therefore has to be assessed for each specific context separately.

Table 8 Selected regulatory precedents on sample size

Jurisdiction	Model	Sample	Comment
Germany – gas DSOs	DEA, SFA	Around 190 observations	National cross-sectional sample from gas DSOs
Ofgem, UK – gas DSO	COLS	Around 104 observations	National balanced panel of historical and forecast data from two regulation periods
Ofgem, UK – electricity DSO	COLS	Around 78 observations	National balanced panel of historical and forecast data from two regulation periods
AER, Australia	OLS, SFA	Two estimation periods, 2006-2021 (1074 observations) and 2012—2021 (666 observations)	Use of international panel data from New Zealand and Ontario (Canada) to increase sample size
EA, Finland – electricity DSOs	StoNED	690 observations	National unbalanced panel from 89 electricity DSOs over an 8-year period
E-Control, Austria – gas DSOs	DEA, MOLS	20 observations	National cross-sectional sample from gas DSOs
E-Control Austria – electricity DSOs	DEA, MOLS	38 observations	National cross-sectional sample from electricity DSOs

Source: Frontier Economics

Table 9 Rule of thumbs on sample size for economic-based models

Technique	Rule of thumb
DEA	Banker et al. propose three times number of inputs plus outputs.[50]
SFA, COLS, MOLS	An appropriate sample size can be calculated using a target for the power of a statistical test to be applied and by making some assumptions around the data that could be used in the regression estimation (e.g. the distribution of costs and cost drivers). For example, if the statistical test of interest is the significance of the coefficients of the econometric model estimated, it is possible to calculate the minimum sample size required for such a test to have an 80% power and 5% level of significance, which will depend on distributional assumptions for the data.
StoNED	Given the nonparametric nature, a large sample is better. It is challenging to construct a definitive rule of thumb

Source: Frontier Economics

How to account for heterogeneity?

Usually there are exogenous differences between operators that affect their costs, for example differences in terrain or population density. Operators do not usually have control over these exogenous differences. Statistically controlling for heterogeneity is important to avoid that these differences are considered inefficiencies.

All the techniques under consideration allow to account for exogenous differences. There are three main approaches to do this, which can be combined:

- **Ex-ante adjustment**. This is done by adjusting the data before it is used in any model for exogenous differences. For example, Ofgem adjusts cost data of gas DNOs by differences in wages in different areas of the country.
- **In-model adjustment**. This is done by including relevant drivers of exogenous differences in the model. This approach is advantageous when the relationship between costs and cost drivers is not well known or the regulator wants to estimate it using the available data. However, the number of cost drivers that can be added is restricted due to limited sample sizes. Regulators adjust for some cost differences in this way, for example, the Finnish regulator

[50] Banker, R. D., Charnes, A., Cooper, W. W., Swarts, J., & Thomas, D. (1989). An introduction to data envelopment analysis with some of its models and their uses. Research in governmental and nonprofit accounting, 5(1), 125-163.

includes the ratio of metering points/connections points to account for differences between rural and urban electricity DSOs.

- **Ex-post adjustment**. This is done after the model has been estimated and the efficiencies have been derived. Adjustment of the efficiency scores or the implied allowances are made to account for factors unaccounted for in the model. For example, the AER adjusts the efficiency scores ex-post by a set of Operating Environment Factors (OEFs).51 Another example is the UK water regulator, Ofwat, which allows additional costs due to a number of operator specific circumstances through its cost adjustment claims process.

In principle, adjusting for cost differences ex-ante or in-model is preferred to ex-post adjustments. This is because the efficiency scores will be biased if not all material exogenous factors have been accounted for during the estimation of the model.

How to account for economies of scale?

If there are economies of scale, the larger the scale of operation of an operator, the smaller its marginal costs. When undertaking the benchmarking analysis it is important to determine whether there are economies of scale and then decide whether to account for them or not. Accounting for economies of scale (or not) can provide different incentives to operators:

- If economies of scale exist and are considered when estimating the efficiency frontier, then regulated operators are only compared to operators of a similar scale.
- If economies of scale exist but are not considered, then the regulated operators are compared to all operators of any scale. Therefore, the efficiency assessment includes a scale component. This means that particularly smaller operators (operating at a sub-optimal scale) may be considered inefficient because of their smaller scale. This may incentivise these operators to merge in order to improve their efficiency. However, it is worth nothing that DSOs and TSOs may have limited control on the scale of their operations in the short term (e.g. they may not be legally or politically permitted to merge).

Regulators could adopt different approaches to establish whether there are economies of scale. Some techniques can be used to estimate economies of scale. For example, econometric models can allow estimating (dis)economies of scale (e.g. by considering whether the coefficient of the scale variable differs from 1). Analysis of unit costs at different scale of operation can also reveal whether there are economies of scale for specific activities (e.g. if unit cost of delivering a certain activity changes as the scale of delivering the activity increases). The regulator could also decide to

[51] AER. (2022). Annual Benchmarking Report: Electricity Distribution Network Service Providers. https://www.aer.gov.au/networks-pipelines/guidelines-schemes-models-reviews/annual-benchmarking-reports-2022

impose specific economies of scale assumptions, informed by regulatory objectives, academic literature or expert input, or otherwise, as a cross-check.

All of the techniques in our short list allow to account for economies of scale, as indicated in Table 10 below.

Table 10 How to model economies of scale by technique

Technique	Approach
PPI	Economies of scale can be estimated empirically from the data by calculating the PPIs at different levels of scale
DEA	Economies of scale can be assumed (typically by choosing whether the data envelop should follow a constant return to scale assumption, variable return to scale assumption, or non-decreasing return to scale assumption). The variable return to scale assumption allows to model both increasing and decreasing return to scale.
COLS, SFA	Economies of scale can be estimated from the data. Alternatively, constraints can be imposed on the scale coefficients.
StoNED	Standard StoNED assumes variable returns to scale. It is possible to impose an additional restriction for other returns to scale assumptions.
Engineering models, RNA	Assumptions can be used when calculating costs from these models (e.g. by using decreasing unit costs)

Source: Frontier Economics

Which technique can be used with forecast data?

All the techniques in our short list can be used with forecast data. There are different ways in which forecast data can be used:

- **Inclusion of costs and cost drivers in estimation sample**. All techniques can be estimated using a sample that includes only historical data, or only forecast data, or a combination of both. Inclusion of forecast data in the sample allows to directly estimate the expected future efficiencies. For example, the UK's Ofgem estimates its models for gas DSOs using a sample that includes both historical and forecast data. If forecast data is included in the sample it is important to consider whether the operators have incentives to provide accurate forecasts. At RIIO1 Ofgem applied two approaches meant to get accurate forecasts: **(1) well-justified-business-plan**. Ofgem defined criteria for a well-justified-business-plan and assessed operators' business plans against those criteria. In case all criteria were met the operator was eligible for fast-tracking. This meant that the allowed revenues were set based on the business plan without any further evaluation of the reported forecast costs. **(2) Information-Quality-Incentive** (IQI). The IQI was used to set the strength of the upfront efficiency incentives each operators faces according to differences between its forecasts and Ofgem's assessment of its efficient expenditure requirements. It was meant to encourage operators to submit accurate expenditure forecasts to Ofgem, because 'truth-revealing' resulted in the highest incentives.52 In preparation of RIIO2 Ofgem commissioned an assessment of RIIO1 including well-justified-business -plans and Information-Quality-Incentive (IQI), which proposed different options for adjustments and/or improvements.[53] As a consequence in RIIO2 for electricity distribution operators Ofgem made several key changes to its approach including removing the fast-track-process, replacing the IQI with a new 'confidence-dependent incentive rate' approach and introducing a new **Business Plan Incentive**. The latter involves a four stage process including penalties e.g. for breaching minimum requirements for business plans and poorly justified cost proposals, and rewards for ambitious cost proposals and additional customer value.54

- **Use forecasts of cost drivers**. Parametric techniques, engineering-based models, and PPIs can be used to determine the relationship between cost and cost drivers. This relationship can be determined using historical data. Then, forecasts of cost drivers can be fed to the estimated model to forecast future efficient costs.[55] The static efficiency can be determined from the historical sample or from the forecast sample or a combination of both. The latter is the approach adopted by the water regulator in England and Wales, Ofwat, at the current price control of water and wastewater companies, PR19. For setting allowances for wholesale water and wastewater, Ofwat uses econometric models to estimate the historical efficiency and catch-up. It then forecasts the costs using its models and forecasts of cost drivers, and applies the historical catch-up. For setting allowances for retail, Ofwat estimates econometric models over the historical sample and uses the model to predict the costs. The catch-up efficiency is derived as an arithmetic average of historical and forecast catch-up efficiencies.

52 For an example of the IQI for electricity DSOs in RIIO1 we refer to Ofgem. (2014, November). RIIO-ED1: Final determinations for the slowtrack electricity distribution companies: Final decision (pp. 37ff).https://www.ofgem.gov.uk/sites/default/files/docs/2014/11/riio-ed1_final_determination_overview_-_updated_front_cover_0.pdf

53 CEPA (2018, March), Review of the RIIO framework and RIIO-1 performance, Final Report for Ofgem.

54 For details on the Business Plan Incentive we refer to Ofgem. (2022, November). RIIO ED2 Final Determinations Overview document (pp. 64ff). https://www.ofgem.gov.uk/sites/default/files/2022-11/RIIO-ED2%20Final%20Determinations%20Overview%20document.pdf

55 For PPIs, this for example could mean deriving historical efficient unit costs, and then assuming these unit costs remain constant in the future.

Which technique is likely to be more resource intensive?

We have assessed how much resources a technique is likely to involve. Resources are related to number of data points to be collected, as well as assumptions required and complexity in implementation. For this reason our assessment of resource intensity is in line with our assessment of transparency.

- PPIs are the least resource intensive as no (or limited) assumptions are required to implement these descriptive techniques.
- Simpler engineering models and economic-based models are likely to be more resource intensive than PPIs. This is because more data needs to be collected, and relationships between costs and cost drivers need to be established (using engineering expert input and/or empirically). Some time will be required to test and validate assumptions. We note that a considerable amount of effort might be required to ensure that data is comparable and of sufficient quality.
- More complex engineering models and RNA are likely to be more resource intensive. This is because the number of data points, variables considered, and assumptions is likely to be significantly larger than for the other techniques considered.

5.4 Robustness

How do you ensure robustness?

We consider a model (and its results) to be robust if they do not change significantly when making small changes to the data or the assumptions used. Robustness is usually assessed empirically during the model development phase. A well specified model that relies on high quality data should allow one to identify a robust relationship between cost and cost drivers which does not depend on specific assumptions or small changes in the data used.

Robustness of a model can be tested during the development phase, e.g. by removing some comparators or changing assumptions regarding cost functions. An important step to ensure robustness is to ensure that the data is comparable across operators and consistently defined over time. PPIs over time can be used as part of this process as they can be used to identify changes in key drivers over time. As part of the model development the regulator can investigate whether the changes are genuine changes in costs (and if so what drives them) or driven by accounting practices or reporting errors (and if so, how to adjust for them).

6 Assessment of potential benefits of combining different techniques

Benchmarking is a valuable instrument of the regulatory toolbox. However, when applying this tool, it is unlikely that there exists a unique right benchmarking model. Hence, there may be merit in combining different benchmarking methods that complement each other in order to offset possible shortcomings of a single benchmarking technique. There are two main ways of using different techniques and then combining the results, either to benchmark

- A given cost category; or
- Different cost categories.

In this section we present our assessment on combining benchmarking techniques and conclude by indicating which combinations of techniques could be appropriate in the Dutch context.

6.1 Combining techniques to benchmark a given cost category

For a given cost category (e.g. opex or totex), a range of different benchmarking techniques could be used to assess the efficiencies of the TSOs and DSOs against that cost category. Combining different techniques can help the regulator increase the robustness of its findings by addressing some of the uncertainties around three key factors:

- Choice of a specific technique;
- Assumptions underlying a specific technique;
- Other aspects of the benchmarking analysis (e.g. cost drivers, sample).

We expand on these factors below and summarise the practical approaches adopted by regulators to combine techniques.

Choice of a specific technique

As discussed in Section 3, each benchmarking technique has strengths and weaknesses. Using different techniques (e.g. DEA and SFA) can help mitigate some of the weaknesses. If the assessment of efficiency is similar across different techniques this can increase the confidence that the regulator has in its findings (as these are not driven by the assumptions and weaknesses of a specific technique). For economic-based techniques, the same set of inputs and outputs can be used for different techniques. Therefore once data is collected, the additional resources required to use different techniques are relatively low.

This approach has been adopted by a number of regulators. For example, the AER assesses the efficiency of the Australian electricity DSOs using both SFA and COLS models. Bundesnetzagentur uses DEA and SFA for the German electricity and gas DSOs, and E-Control uses DEA and MOLS in Austria.

Assumptions of a given technique

The regulator can decide to combine techniques when there is uncertainty around assumptions underlying the technique, for example around the choice of functional form. This is the approach followed by the AER for electricity DSOs. Not only does the AER consider different techniques (SFA and COLS), but for each technique it estimates both a Cobb-Douglas and a Translog functional form.

Other aspects of benchmarking analysis

When undertaking a benchmarking analysis, there could be uncertainties around some of the other aspects beyond the choice of techniques and its assumptions, such as which cost drivers to use, which comparators to include, and which year to consider (or years in a panel data context). If there are no clear ways to resolve these uncertainties, it might be useful to test a range of models with different model specifications for cost drivers.

For example, there could be two equally valid proxies for the same cost driver (i.e. good data quality, both supported by engineering and economic rationale, good fit to the data). In this case, the regulator might decide to consider combining results from two models that both use a different proxy. For example, this is the approach adopted by Ofwat when benchmarking the efficiency of retail costs of water and wastewater operators. For the current price control PR19, Ofwat developed a number of top-down and bottom-up retail models of costs and assigned equal weights to models that contain proxies of propensity to default. This is also the approach Ofgem used for the electricity DSOs, where benchmarking models with different cost drivers were estimated.[56]

Regulatory precedent on combining techniques

Regulators have combined results from different techniques in a number of ways: choosing the highest efficiency scores between different models; taking (weighted) averages; in-the-round assessment, i.e. forming a high-level view of efficiency without using results from models mechanistically; and using efficiency scores as cross-checks for the results of the preferred model(s). Table 11 summarises relevant regulatory precedents.

[56] For more details we refer to the case study 'Great Britain – Electricity DSOs' in Section 7.2.

Table 11 Regulatory precedents on combining techniques for a given cost category

Approach	Regulatory precedent
Maximum efficiency score across models	BNetzA uses a best-of-four approach to determine the individual efficiency score for electricity and gas DSOs by taking the maximum of the DEA (non-standardised/standardised capital costs) and SFA (non-standardised/standardised capital costs) estimates.
Arithmetic average of efficiencies from different models	■ The AER benchmarks energy DSOs based on the **arithmetic average of the efficiency score** of four econometric models that differ with respect to functional form (Cobb-Douglas vs Translog) and estimation technique (SFA vs OLS with dummies for DSOs). Cost drivers are the same across models. ■ The Finnish regulator used the arithmetic average of DEA and SFA for the regulatory period 2008-2011. ■ Ofgem takes the arithmetic average of disaggregated (activity level) benchmarking results on one side and the average of the results of its three totex econometric models to determine the cost allowance of electricity DSOs in the RIIO-2 period on the other side. The average of the results of the three totex econometric models is derived by assigning equal weights to each of the three models (hence, each econometric model has an overall weight of 16.67%).
Weighted average of efficiencies from different models	■ E-Control weights results of DEA and MOLS. The regulator chose the weights to balance pros and cons: Gas DSO: 50% DEA and 50% MOLS; Electricity DSO: 50% MOLS, 25% DEA (with aggregated model network length for LV, MV, and HV) and 25% DEA (with separate model network length for LV, MV, and HV). ■ Ofwat estimates efficiency scores as a weighted average of the scores from its econometric models. Ofwat estimates models using Random Effects. The models differ because both different levels of aggregation of costs are considered and, for a given level of aggregation, different cost drivers are used. Weights reflect the confidence that Ofwat places on the results stemming from its models (e.g. due to data quality issues). For example, when assessing retail costs Ofwat placed 25% weight on models based on disaggregated cost data and 75% weight on models based on total costs.

Approach	Regulatory precedent
In-the-round assessment	In Ireland, for the RC3 price control the Irish Water regulation, the CRU, assessed the efficiency of Uisce Éireann (formerly Irish Water) by considering an efficiency assessment across a range of econometric models (different weights used to estimate the Composite Scale Variable, different samples (inclusion/exclusion of Uisce Éireann from the sample)). The CRU did not mechanistically translate the efficiency scores from the model into an efficiency target. It used the results of the modelling to decide whether to apply an efficiency challenge or not. The CRU ultimate decided to apply an efficiency challenge based on its assessment of what has been achieved in other jurisdictions following the introduction of economic regulation.[57]
Cross-checks	The AER complements the regression models used to determine the cost allowance of electricity DSOs with detailed MTFP, MPFP and PPI analyses that allow for identification of potential sources of inefficiencies and to cross-check results from the regression approaches.

Source: Frontier Economics

We note that for economic-based techniques, there is some evidence from the literature that to identify the true efficiency scores the maximum of the score across different DEA and SFA models tends to perform better than the average score across those same models.[58] The implication is that individual methods tend to systematically underestimate the efficiency score on average. Tsionas (2021) *'propose[d] and implement[ed] a formal criterion of weighting based on maximising proper criteria of model fit (viz. log predictive scoring) and show how it can be applied in Stochastic Frontier as well as in Data Envelopment Analysis models'.*[59]

The results from this literature depend on the underlying characteristics of the data used in the study, and therefore they may not be applicable to the Dutch context. Further, the results of those studies are based on empirical observations from the simulations, no formal theoretical analysis has been conducted to assess the merits of such approaches. However, a challenge with a theoretical analysis of combination approaches is that the theoretical rate of the bias (either for SFA or DEA) depends on the underlying data generating process which makes further scrutiny difficult.

57 See CRU. (2019, December). Irish Water Revenue Control, Revenue Control 3 (2020-2024). https://www.wateradvisorybody.ie/wp-content/uploads/2020/04/Irish-Water-Revenue-Control-3-Decision-Paper.pdf

58 See for instance Andor, M. A., Parmeter, C., & Sommer, S. (2019). Combining uncertainty with uncertainty to get certainty? Efficiency analysis for regulation purposes. European Journal of Operational Research, 274(1), 240-252.

59 Tsionas, M. G. (2021). Optimal combinations of stochastic frontier and data envelopment analysis models. European Journal of Operational Research, 294(2), 790-800.

A benefit of the combination of SFA and DEA approaches is that they are also transparent and easy to implement. We consider that combining different benchmarking techniques will likely involve some regulatory discretion.

6.2 Combining techniques across different cost categories

Different techniques can also be used to benchmark different cost categories (or levels of aggregation of the same cost category). This is useful in a number of situations:

- To make assessments more robust;
- To cross-check results;
- To identify sources of inefficiencies;
- To circumvent a lack of comparative data.

To make assessments more robust

Different techniques could be applied to different levels of aggregation of the same cost category. Results can then be combined to obtain an overall assessment of efficiency. For example, a high level assessment of totex can be combined with lower level assessments of the components of totex. The same or different techniques can be used.

This is the approach used by Ofgem which uses a benchmarking toolbox of bottom-up and top-down models on different costs categories and at various levels of aggregation.[60]

For PR19 Ofwat adopted a similar approach when assessing water and wastewater wholesale costs. Ofwat estimated top-down models of total expenditures and bottom-up models of disaggregated expenditures (for water these costs included water resources costs and treatment water costs; for wastewater sewage collection costs and sewage treatment costs). It then determined the efficiency of the companies by first aggregating the results of the bottom up models and then calculating an average of the efficiencies estimated from top-down models and bottom-up models.

To cross-check results

Use of different techniques can also be employed to cross-check the results of the 'preferred' approach. For example, the AER's 'preferred' suite of models to set the efficiencies of the electricity DSOs consists of top-down econometric models of total opex. The AER sets the efficiencies based on the average scores across four econometric models that differ by estimation technique (SFA or

60 For more details we refer to the case studies 'Great Britain – Gas DSOs' in Section 7.1 and 'Great Britain – Electricity DSOs' in Section 7.2.

Fixed Effects[61]) and cost function specification (Cobb-Douglas or Translog). The AER cross-checks the validity of its results by using two different sets of techniques applied to different cost categories:

- **PPIs**. The AER estimates unit costs on both a total cost and an opex basis, using the two main cost drivers (circuit length and customer numbers) for the denominator of those ratios. It then compares the unit costs across operators taking density of the networks into account and makes some inference on whether the relative performance of each operator is in line with the findings from its econometric models (acknowledging that PPIs are simpler than other techniques used and rankings may be affected by factors not controlled for in the PPI analysis).
- **Opex MPFP.** The AER uses this technique to estimate opex productivity across operators and over time. It then compares the relative rankings of the operators at a given point in time, and how these have changed over time (i.e. which company has improved its productivity). The AER also checks whether the relative rankings from the Opex MPFP are consistent with the findings from its econometric models

To identify sources of inefficiencies

Benchmarking techniques can also be used to potentially identify where inefficiency comes from. Techniques used to identify where inefficiency comes from do not necessarily need to coincide with the techniques used to set the overall efficiency challenge, as these two objectives can differ Therefore, two sets of techniques can be combined to achieve these two different objectives.

For example, the overall efficiency challenge could be determined by assessing the performance of operators at a high level of cost aggregation using more complex techniques (e.g. SFA). Instead, the sources of inefficiencies could be determined by assessing the performance of operators at a lower level of cost aggregation, which can help identify which operators are relatively efficient at undertaking certain activities. In this regard, PPIs could be used to provide some high level indications of which operators might have lower unit costs than others.

To circumvent a lack of comparative data

For some cost categories, such as routine costs driven by business as usual type of activities, the regulator might have access to comparative data. However, for other cost categories (e.g. bespoke operator-specific capex investments or costs driven by new activities) comparative data might not be available.

In these situations, different techniques are likely to be needed to benchmark the two categories of costs as some techniques require comparative data while others do not. For example, economic-

61 Some regulators use Fixed Effects and Random Effects panel data models to benchmark costs. These models are used to estimate the effects of company-specific characteristics which do not change over time. These effects are often interpreted as inefficiency, although they could also capture some persistent heterogeneity.

based techniques usually require comparative data, while engineering-based techniques do not (or to a lesser extent).

A number of regulators apply different techniques to different cost categories for the reasons outlined above. For example, for PR19 Ofwat used different techniques to benchmark water and wastewater operators' expenditures.

- For routine year-on-year expenditures, which operators incur in the normal running of their businesses, Ofwat uses a suite of econometric models (based on Random Effects) estimated using historical data from the English and Welsh operators.
- For more ad hoc expenditures aimed at increasing the level of services or at providing new customers with the current service level (enhancement expenditures), Ofwat uses a range of techniques that vary by the particular enhancement expenditure under consideration. These techniques include unit costs, simple econometric models largely based on forecast data, and engineering 'deep dive' assessments.

Implications for benchmarking the Dutch electricity and gas TSOs and DSOs

Based on the assessment above and in the previous section, we considered whether each of the technique in our short list could be used in the Dutch context to benchmark electricity and gas TSOs and DSOs, and what the potential challenges could be. Our findings are the following:

- PPIs could be used to provide descriptive statistics for all operators. They could also be used to develop a better understanding of where inefficiency is coming from (e.g. by comparing unit costs of specific activities or cost categories).
- COLS, MOLS, SFA, and DEA can also be used for all operators. The key challenge would be to identify an appropriate sample of comparators. For DSOs, it may be possible to estimate simple econometric models (like COLS) on the Dutch sample alone. This would need to be tested empirically. For the TSOs, given that there is only one gas TSO and one electricity TSO, it would be necessary to enlarge the sample with international data. The Pan-European study of TSOs and the AER's international benchmarking of DSOs are two useful case studies that indicate which factors should be taken into account when conducting international benchmarking.
- Engineering based models can be used for all operators. Those models might be particularly useful to benchmark new large investments that are due to the energy transition, so potentially more useful for the electricity sector than the gas sector.
- RNAs could be used for TSOs. However, we do not think they can be used to benchmark total costs of DSOs because of the large number of different distribution assets that would need to be modelled.

7 Case studies

In this section we present six case studies which are useful for understanding how different benchmarking techniques are used in regulatory applications. Table 12 lists the six case studies and summarises the reasons for selecting them. Some of the lessons learnt from DSOs are also applicable to TSOs, for example around combining different techniques, use of an internal sample, and use of PPIs.

Table 12 List of case studies

Country	Sector	Regulator	Reason for selecting case study
Great Britain	Gas DSOs	Ofgem	Use of different techniques to benchmark different cost categories (economic and engineering-based techniques) Use of forward looking data Estimation of econometric models with a small sample
Great Britain	Electricity DSOs	Ofgem	Use of totex regression models and disaggregated benchmarking to benchmark a given cost category (totex) Use of forward looking data Estimation of econometric models with a small sample
Finland	Electricity DSOs	Energy Authority	Development of benchmarking (from DEA, to DEA and SFA, to StoNED) Application of StoNED in a regulatory context
Australia	Electricity DSOs	AER	Combination of different benchmarking techniques (SFA, OLS with fixed effects), and PPIs as cross-checks Use of international sample to address challenges with small samples
Germany	Electricity TSOs	BNetzA	Application of RNA
Germany	Electricity/ Gas DSOs	BNetzA	Different benchmarking techniques applied (DEA, SFA) Outlier analysis and cost driver analysis

Source: Frontier Economics

7.1 Great Britain – Gas DSOs

7.1.1 Introduction

Gas DSOs in Great Britain are regulated by Ofgem through the RIIO regime (Revenue = Incentives + Innovation + Outputs). The starting point for determining operators' allowed revenues are the forecasts of costs that the DSOs submit to Ofgem as part of their detailed business plans. Forecast costs are assessed by Ofgem using a toolbox of comparative analyses. Some costs are assessed using OLS regressions that rely on historical and forecast data of all DSOs; other costs (for which the available data does not allow a comparative assessment) are assessed through DSO-specific approaches, which include technical assessments and expert reviews. The total baseline cost allowance obtained from the assessment of the submitted cost forecasts ('benchmarking efficiency assessment') is also subject to additional ongoing efficiency adjustments.

7.1.2 Benchmarking technique – Complementary use of OLS, non-regression techniques and technical assessments

The RIIO-GD2[62] framework developed by Ofgem sets out how the efficiency analysis of gas DSOs is conducted in the UK.[63] The method used to assess the individual cost items is chosen based on the data that is available for comparative analysis. Approximately 85% of forecast totex is assessed by an OLS (top-down) regression model that is conducted using normalised and adjusted controllable totex.[64] About 10% of forecast totex is assessed using non-regression based analyses, which include individual reviews as well as comparative methods such as unit cost models. The remaining 5% of forecast totex is related to operator specific activities or new projects (e.g. bespoke outputs). These are assessed using technical assessments or expert reviews of the submitted costs.

Ofgem's OLS model is estimated by regressing adjusted controllable totex on two types of variables assuming a Cobb-Douglas functional form:

- **A single composite scale variable (CSV).** Ofgem used a CSV given the small sample size. The CSV is a weighted average of different cost drivers. The weights are based on industry spend proportions for the disaggregated cost activities to which the drivers apply. The residual weight is assigned to the scale driver Modern Equivalent Asset Value (MEAV). Ofgem

62 RIIO-GD2 = RIIO Gas Distribution 2nd regulatory period

63 For a full documentation of Ofgem's approach and methods for the cost benchmarking of gas DSOs in the RIIO-2 period see Ofgem's final determination decision: Ofgem (2020). RIIO-2 Final Determinations for Transmission and Gas Distribution network companies and the Electricity System Operator. https://www.ofgem.gov.uk/publications/riio-2-final-determinations-transmission-and-gas-distribution-network-companies-and-electricity-system-operator.

64 Controllable totex are defined as the sum of controllable opex, capex and repex (replacement expenditures). Totex adjustments cover regional labour market conditions, urbanity and sparsity, among others.

undertakes a robustness analysis where it adds a squared term of CSV (which is found to be statistically insignificant) to the cost function.

- **Two separate linear time trends for historical and forecast data**. The trends control for time effects and allow for a structural break between the two time periods.

The estimation uses clustered robust standard errors at the DSO level. Ofgem considered a random effects and a SFA model for robustness analyses, which yielded similar results as the OLS regression. Ofgem decided to use the results from the OLS model to set the allowances as it is less data intensive, requires fewer assumptions, and is more conservative and more transparent than the other approaches.

7.1.3 Calculation of efficient cost levels

Ofgem uses the OLS model to estimate predicted costs and compare them with the DSOs' submitted costs.

As discussed in Section 3.3.1, the conditional average function estimated by the OLS regression does not represent a cost frontier, as it includes the average level of inefficiency in the sample. Ofgem therefore follows a COLS approach to define a more ambitious efficiency benchmark than the OLS regression line. For this purpose it calculates individual efficiency scores as the ratio of submitted costs and the OLS predictions. It then uses a pre-defined quantile of the resulting efficiency score distribution (typically the 85th quantile) to define the efficiency frontier used for benchmarking. This is then applied to the modelled costs of all DSOs in the sample.

7.1.4 Data sample

The total cost allowance defined for gas DSOs in the RIIO-2 regulation period is based on data from all 8 gas DSOs. The benchmarking makes use of a combination of historical data (2014-2020) and forecast data (2021-2026) to maximize sample size and explicitly account for potential changes in the types of costs that DSOs experience.

7.1.5 Conclusions

This case study shows how to combine different techniques to assess different cost categories; how to use forward looking data, and also how to address some potential challenges with a small sample size.

Although the majority of totex is assessed using a top-down econometric model, other approaches are used to assess the remaining costs. Forward looking data is included in the econometric model (we discuss Ofgem's incentives around forecast data in Section 5.3).

The relatively small number of gas DSOs in the UK limits the benchmarking methods that Ofgem can apply in its assessment of DSOs' submitted costs. Therefore, Ofgem relied on a COLS technique, which is less data intensive than other more complex models. Ofgem also chose to include a composite scale variable because of the limited sample size (although Ofgem made some assumptions to derive this variable).

Some aspects of Ofgem's RIIO-GD2 were appealed by a number of companies on a number of grounds, including on setting an 85th percentile efficiency target.[65] The Competition and Markets Authority (CMA) found that Ofgem was not wrong in setting the 85th percentile efficiency target. We note however that the grounds of appeal are narrow and the CMA needs to apply specific legal standards.[66] Because of this, operators may decide not to appeal aspects of benchmarking if they believe they will be dismissed under those specific legal standards. Therefore, we consider that the fact that other aspects of Ofgem's benchmarking has not been appealed does not necessarily mean that the operators agree with those aspects.

7.2 Great Britain – Electricity DSOs

7.2.1 Introduction

Electricity DSOs in Great Britain are also regulated by the RIIO regime. As for gas DSOs, the starting point for determining operators' allowed revenues are the forecasts of costs that the DSOs submit to Ofgem as part of their detailed business plans. Forecast costs are assessed based on a combination of totex regressions and disaggregated (activity level) benchmarking. Operators have to meet a catch-up efficiency challenge defined by a glide path moving the benchmark efficiency score from the 75th to the 85th quantile over the first three years of the regulatory period. In addition, operators are subject to an ongoing efficiency challenge of 1% per year.

65 For a summary of the CMA's decision see CMA (2019, September). Northern Powergrid (Northeast) Plc and Northern Powergrid (Yorkshire) Plc v Gas and Electricity Markets Authority: Final determination.. https://assets.publishing.service.gov.uk/media/61791296d3bf7f55ff1fc099/Energy_appeals_-_Summary_of_final_determination_28.10.21.pdf

66 *Ibid.* para 9

7.2.2 Benchmarking technique – Combining results from totex regressions and disaggregated modelling

The RIIO-ED2[67] framework sets out how the efficiency analysis of electricity DSOs is conducted in the UK.[68] Ofgem combines two different approaches to assess the costs submitted by the electricity DSOs: totex regressions and disaggregated (activity level) benchmarking.

The **totex regressions** used for cost benchmarking consist of three different models. These models primarily differ in the choice of explanatory variables used. Similar to the regression specification in Ofgem's gas distribution framework discussed in Section 7.1, CSVs are used to aggregate information from various cost drivers while keeping the number of estimated parameters in the regression low.

- Model 1 uses a bottom-up CSV aggregating a long list of cost drivers, which were identified through an activity-level analysis.[69]
- Model 2 uses a top-down CSV that only covers key cost drivers.[70] In addition, the model includes capacity released as another demand driver.
- Model 3 uses the same top-down CSV as Model 2, but additionally includes the number of heat pumps and electric vehicles to model the (expected) uptake of low carbon technologies (LCT).

All three models are estimated assuming a Cobb-Douglas functional form; pooled OLS is used for the estimation. While Model 1 and 2 are applied to historical and forecast data, Model 3 is restricted to forecast data, as the uptake of electric vehicles and heat pumps has been too low historically to allow estimation of the associated parameters. Model 1 and 2 include a dummy variable for the RIIO-ED2 regulation period,[71] to account for structural changes resulting from the change to the RIIO-ED2 period and the increased importance of the transition to net zero emissions.

The **disaggregated benchmarking** consists of a range of techniques depending on applicability to the various cost categories. The techniques employed include regressions, unit cost analysis, and qualitative assessments based on engineering knowledge.

---

67 RIIO-ED2 = RIIO Electricity Distribution 2nd regulatory period

68 For a full documentation of Ofgem's approach and methods for the cost benchmarking of electricity DSOs in the RIIO-2 period see Ofgem's final determination decision: Ofgem (2022). RIIO-ED2 Final Determinations. https://www.ofgem.gov.uk/publications/riio-ed2-final-determinations.

69 The bottom-up CSV includes MEAV, customer numbers, faults driver, peak demand, capacity released, length OHL, total network length, and spans affected ONI drivers.

70 The top-down CSV includes MEAV, network length, customer numbers, total faults, and peak demand .

71 Recital 7.138, page 249 in the RIIO-ED2 Final Determinations Core Methodology. https://www.ofgem.gov.uk/sites/default/files/2022-11/RIIO-ED2%20Final%20Determinations%20Core%20Methodology.pdf

Ofgem considers the totex regressions and the disaggregated benchmarking approaches as *“complementary since they seek to capture different characteristics of the DNOs’ Business Plans and explore the efficiency and justification of the plans using different tools and techniques”*.[72]

7.2.3 Calculation of efficient cost level

Ofgem calculates the efficient cost level as follows:

- First, for each of the totex regression models and for the disaggregated benchmarking Ofgem calculates an operator’s specific efficiency score as the ratio of submitted costs and modelled costs. For the totex regression models, modelled costs are estimated from the predictions of the models; for the disaggregated benchmarking, modelling costs are estimated from the predictions of the disaggregated benchmarking models.
- Second, Ofgem calculates the model-specific benchmark efficiency scores as a given quantile of the efficiency score distribution over all operators. The quantile applied in the calculation is increased along a linear glide path from the 75% to the 85% quantile over the first three years of the regulatory period.
- Third, Ofgem calculates the weighted average of the model-specific efficiency scores from all totex regressions (with weight equal to 16.67% for each of the three totex models) and the disaggregated benchmarking (with 50% weight) to obtain an overall efficiency score.
- Fourth, the modelled cost obtained using the four benchmarking models are multiplied with the overall efficiency score. This results in four model-specific cost allowances.
- Finally, Ofgem sets the cost allowance by taking again the weighted average of the cost allowances of the four models (with weight of 16.67% for each regression model and 50% for the disaggregated benchmarking).

7.2.4 Data sample

The total cost allowance defined for electricity DSOs in the RIIO-2 regulation period is based on data from all 6 electricity DSOs. The benchmarking makes use of a combination of historical data (2016-2021) and forecast data (2022-2028) in regression Models 1 and 2, but is restricted to forecast data in Model 3.

[72] Recital 7.580, page 345 in the RIIO-ED2 Final Determinations Core Methodology. https://www.ofgem.gov.uk/sites/default/files/2022-11/RIIO-ED2%20Final%20Determinations%20Core%20Methodology.pdf

7.2.5 Conclusions

This case study shows how to use techniques at different level of cost aggregation to benchmarking totex, how to use forward looking data, and also how to address some potential challenges with a small sample size.

Totex is benchmarked by combining a top-down assessment of totex with a bottom-up assessment of disaggregated costs. As for RIIO-GD2, forward looking data is included in the econometric models (we discuss Ofgem's incentives around forecast data in Section 5.3).

The relatively small number of electricity DSOs in the UK limits the benchmarking methods that Ofgem can apply in its assessment of DSO's costs for the RIIO-2 regulatory period. In contrast to the RIIO-GD2 regulation, Ofgem uses both a totex top-down approach and a disaggregated bottom-up approach to obtain a final cost allowance.

The two case studies on DSO regulation in the UK provide good examples of ways to combine different techniques for cost benchmarking.

Some aspects of Ofgem's RIIO-ED2 were recently appealed by an electricity DSO, including on allocation of allowances between cost categories. The CMA found that Ofgem made an error in allocating allowances between cost categories.[73] As mentioned in the previous case study, we consider that the fact that other aspects of Ofgem's benchmarking has not been appealed (potentially due to the specific legal standards and narrow grounds of appeal) does not necessarily mean that the operators agree with those aspects.

7.3 Finland – Electricity DSOs

7.3.1 Introduction

The Energy Authority (EMV), the Finnish regulator, has been in charge of setting revenue allowances and monitoring and evaluating network operators since 1995. The amendment of the Electricity Market Act in 2004 caused significant changes in the supervision system. The EMV has applied benchmarking to all electricity distribution operators on a regular basis since 2005. The benchmarking techniques and their regulatory application changed over time:

73 CMA (September 2019), Northern Powergrid (Northeast) Plc and Northern Powergrid (Yorkshire) Plc v Gas and Electricity Markets Authority, Final determination. https://assets.publishing.service.gov.uk/media/650b0b1527d43b000d91c321/21_September_2023_Final_determination_-_RIIO-2_ED2_Appeal_-_version_for_publication_A.pdf

- In the 1st regulatory period (2005-2007) the general productivity factor for operating costs was based on a DEA Malmquist Index using total costs.
- For the 2nd regulatory period (2008-2011) EMV also applied an operator specific productivity factor when setting the cost targets for operating costs. The parametric SFA was used in parallel with DEA and the operating cost targets were calculated based on the average of DEA and SFA efficiency scores estimated based on total costs.
- In the 3rd regulatory period (2012-2015) EMV introduced key methodological changes. Operating costs only (and not total costs) were used in the efficiency analysis and StoNED was used as benchmarking technique. EMV made some changes on the initially recommended model specification based on feedback from the electricity distribution operators. Despite these changes distribution operators sued EMV, demanding changes to the regulatory model as such (e.g. level of efficiency improvements). However, none of the operators contested the general principles of the StoNED model.74
- In the 4th (2016-2019) and 5th (2020-2023) regulatory period the StoNED method was further developed by including a proxy for capital costs as a fixed input.

EMV is currently preparing the 6th/7th regulatory period and is considering further developments of efficiency benchmarking using StoNED. We understand that the gradual evolution of benchmarking in the Finnish regulatory regime resulted in a high level of acceptance by stakeholders.

7.3.2 Benchmarking technique – StoNED for operating costs and unit costs for capital costs

In the current regulatory period EMV has used StoNED to calculate the efficient opex of electricity DSOs.

The efficiency of capital expenditures (i.e. investment costs) is implicitly assessed by using replacement values instead of historical costs when determining operators’ regulated asset base and corresponding capital costs (depreciation and financing costs). Replacement values are calculated using unit costs per asset class published by EMV. Hence, if operators’ outturn investment costs are lower than implied by these unit costs, operators profit from the difference (and vice versa). However, there is no assessment of whether the number of operators’ ‘physical assets’ is efficient.

Separating the assessment for operating and capital costs ignores the potential trade-off between opex/capex and provides adverse incentives regarding the opex/capex mix. In contrast to the StoNED model from the 3rd regulatory period, which only included one input (operating costs), the

74 Timo Kuosmanen; Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model; Energy Economics 34 (2012) 2189–2199.

StoNED model for the 4th/5th regulatory period includes two types of inputs: 1) operating costs as a controllable input; and 2) replacement value of the network as a fixed input, i.e. where no efficiency target applies.

Including the fixed replacement value should (partly) address the opex/capex trade-off. The results indicate that operating costs can be offset to some extent by capital investments. However, for most operators the impact is low and close to zero. The average of the impact is 0.0076 and the median only 0.0002. This means that on average or at the median 1€ replacement values reduces operators' operating costs by 0.0076 € or 0.0002€.[75]

7.3.3 Calculation of efficient cost levels

EMV applies ex-post regulation, i.e. a 'reasonable return' is calculated ex-post, and it compares operators' outturn profits with an allowed 'reasonable profit'. When calculating the 'reasonable profit' EMV uses operators' 'efficient' operating costs and capital costs (depreciation and WACC*RAB) based on replacement values.

The reference value for the 'efficient' operating costs is calculated using the StoNED estimation model. This benchmark is the regulator's best approximation for what the efficient level of controllable costs is for each DSO, given its outputs.

The efficiency frontier derived from StoNED can be presented as shadow price profiles for operators' outputs and fixed inputs corrected for environmental factors and the expected (industry wide) inefficiency. In the StoNED method shadow prices are set in such a way that the performance of operators is seen in the most favourable light. The shadow price profiles of the efficiency frontier derived from the StoNED method differ with respect to the level of the shadow price they allow for different output variables. Some shadow price profiles put more weight, for example, on the output 'number of customers', some on the output 'network length'. Hence, the 'efficient' operating costs for a specific operator are calculated by multiplying this operator's outputs with the shadow profile which maximises the value of its operating costs. These operating costs are then corrected for an environmental factor (ratio of the number of connections to metering points) and expected industry's inefficiency. Hence, in the absence of a competitive market, the shadow prices can be interpreted as a yardstick market in which operators compete in terms of cost-efficiency with other network operators.

These 'efficient' operating costs are adjusted annually to reflect that if DSOs' outputs increase over time, so should efficient operating costs. The StoNED model is not re-estimated every year: the

---

75 Sigma-Hat Economics, Tehostamiskannustin sähkön jakeluverkkoyhtiöiden valvontamallissa: Ehdotus Energiaviraston soveltamien menetelmien kehittämiseksi neljännellä valvontajaksolla 2016 – 2019 (The efficiency incentive in the electricity distribution system operator control model: Proposal for the development of the methodology applied by the Energy Agency in the fourth control period 2016 – 2019), p. 22, Final report, 2014.

parameter estimates (shadow price profiles, impact from environmental factors, expected inefficiency) established in the StoNED estimation at the start of the regulatory period are retained, and efficient operating costs are only updated annually to reflect output changes and to account for inflation.

Although EMV applies an ex-post regulation, the StoNED parameter estimates[76] allow operators to easily calculate ex-ante the 'efficient' operating costs at different levels of input and output variables, and hence anticipate how the benchmark changes as a result of changes in the outputs, the capital stock, or the operating environment.

7.3.4 Data sample and cost drivers

The StoNED estimation for the 4th/5th regulatory period was based on an unbalanced panel model with 89 operators over an 8-year period, with a total of 690 observations. The study from ECKTA Oy (2022)[77] in preparation of the 6th/7th regulatory period uses an unbalanced panel model with 86 operators over the years 2008-2020.

Despite the large sample EMV does not apply econometric cost driver analysis to determine the outputs and environmental factors for the benchmarking analysis. The StoNED model for the current 4th/5th regulatory period uses two input variables (operating costs and replacement value of the network as fixed input), four output variables and one environmental variable.

The four output variables in the StoNED estimation are:

- The number of customers (i.e. metering points);
- The total length of the network;
- The volume of transmitted energy; and
- Regulatory outage costs.

Broadly, the objective is for these output variables to capture differences in the scale of the DSOs used in the model – a larger DSO will have higher operating costs and therefore any relative comparison across them needs to account for this. The environmental variable is the ratio of the number of connections to metering points.

---

76 EMV provides a publicly available spreadsheet application for this exercise.

77 Timo Kuosmanen, Natalia Kuosmanen, Sheng Dai; Kohtuullinen muuttuva kustannus sähkön jakeluverkkoyhtiöiden valvontamallissa: Ehdotus tehostamiskannustimen kehittämiseksi 6. ja 7. valvontajaksoilla vuosina 2024-2031 (Reasonable variable cost in the electricity distribution system operator control model: Proposal for the development of the efficiency incentive 6. and 7. for the monitoring periods 2024-2031); Report for EMV, 2022.

7.3.5 StoNED going forward

EMV is currently preparing the 6th/7th regulatory period and commissioned ECKTA Oy (2022) on possible development of the StoNED model specification. ECKTA Oy (2022) proposed inter alia to limit the range of shadow prices for outputs and to use current use value (depreciated replacement value) instead of replacement values as fixed input.

7.3.6 Conclusions

The Finnish case study shows how to gradually develop the application of different benchmarking techniques, how to extend the inputs to take into account possible trade-offs between opex and capex, and how to use the results from the benchmarking analysis to determine allowed opex during the regulatory period taking into account operators' changing supply task during the regulatory period.

The Finnish case study provides a good example of a gradual evolutionary development of benchmarking analysis. EMV started only with DEA. Taking into account the possible disadvantages of DEA (e.g. deterministic approach), EMV decided to apply an additional benchmarking technique, SFA, to DEA. EMV also recognised that both approaches have advantages and disadvantages and took this into account by taking the average of DEA and SFA efficiency scores when setting allowances for operators. EMV then developed an approach combining properties of DEA and SFA into the StoNED approach. We understand that the gradual development resulted in a good acceptance by regulated operators on the benchmarking techniques, as such, although there were discussions about the appropriate model specification with regards to the used output parameters.

The Finnish regulatory approach treats operating and capital costs in a different way. EMV recognised that this may results in distorted incentives with regards to an optimal opex/capex choice. Hence, in addition to operating costs EMV extended the inputs by a measure for capital intensity. In the current regulatory period EMW uses the replacement value of operators' network assets to control for the capital intensity, i.e. higher capital intensity reduces operating costs. In preparation of the forthcoming regulatory period EMV is considering to adjust the measure for capital intensity by using "depreciated" instead of "undepreciated" replacement values. This should better reflect the trade-off between opex and capex, because the impact from an older network on operating costs could be higher than a newer one.

Finally, the Finnish case study provides an example of how to deal with the impact from operators' variations in the supply task on efficient operating costs during the regulatory period. The 'efficient' operating costs for a regulatory period vary with operators' outputs (supply task). As the 'reasonable profit' is calculated ex-post, the outturn outputs (rather than the forecast output) of the relevant regulatory period are used for calculating the 'efficient' operating costs. This has the advantage that the impact of output variations on costs can be taken into account without the drawback of having to assess operators' forecasts. The disadvantage is that the relationship between outputs and costs are based on historical data and a historical production technology. If the production technology changes in the future because of new challenges from the energy transition than this is not appropriately reflected in the historical output-cost relationship. The estimation of future 'efficient' costs may then be distorted.

7.4 Australia – Electricity DSOs

7.4.1 Introduction

The AER is responsible for regulating electricity DSOs in Australia, which includes the determination of the revenues that DSOs are allowed to recover from their customers.

For this purpose, the AER assesses costs submitted by network operators using a variety of different cost benchmarking methods. Opex is assed used regression approaches (OLS and SFA); capex is assessed a combination of top-down and bottom-up models. In addition, on an annual basis, the AER checks the evolution of the operators' relative efficiency using MTFP, opex and capex MPFP, and PPIs. To increase the sample size available for the econometric analysis, the AER combines data of 10 Australian DSOs with data from 19 operators in New Zealand and 38 operators in Ontario (Canada).

7.4.2 Benchmarking techniques

In 2022, the AER assessed the costs of the electricity DSOs using a combination of three approaches:[78]

- OLS and SFA regressions for the determination of opex allowances;
- A combination of top-down and bottom-up assessments for the determination of total capex allowances; and
- A combination of alternative approaches, including MTFP, opex and capex MPFP and PPIs, for a more detailed study of the evolution and sources of operators' efficiency levels. These approaches are not directly used to set the cost allowances.

---

78 See the following documents for a detailed description of the AER's approach and methods. Cunningham, M., Hirschberg, J., & Quack, M. (2022). Economic Benchmarking Results for the Australian Energy Regulator's 2022 DNSP Annual Benchmarking Report https://www.aer.gov.au/system/files/Quantonomics%20-%20Benchmarking%20results%20for%20the%20AER%20-%20Distribution%20-%20November%202022.pdf; AER (2022) Annual Benchmarking Report. Electricity distribution network. https://www.aer.gov.au/networks-pipelines/guidelines-schemes-models-reviews/annual-benchmarking-reports-2022

The **opex benchmarking** is based on four regression models. The models differ with respect to the technique used (OLS with fixed effects for the Australian DSOs[79] and SFA) and the functional form assumed (Cobb-Douglas and Translog). The estimation is conducted on two different time periods using historical data: a long period from 2006 to 2021, and a short period from 2012 to 2021. DSOs from New Zealand and Ontario (Canada) are included to increase the sample size. The cost drivers considered include a range of outputs, some environmental variables,[80] and a linear time trend to account for systematic changes over time. All regressions include country fixed effects.

The **benchmarking** of operators' **capex** is based on a combination of top-down and bottom-up assessments of the total forecast capex submitted by the DSOs. Typically, the top-down review is conducted first and determines whether further detailed analysis is required. It also acts as a benchmark to evaluate the results of the bottom-up analysis. The assessments include the analysis of capex drivers, programmes, and projects. However, the AER does not determine individual capex drivers or whether certain programmes or projects should be undertaken. It thus ultimately leaves the investment decision to the network operators, consistently with its incentive-based regulatory framework.

The AER uses a combination of MTFP, MPFP and PPI calculations to analyse the evolution of various dimensions of operators' efficiency over time and across operators. The results from these models do not directly enter the calculation of the final efficiency scores used to determine cost allowances. Yet, they provide additional information that is used to gain a deeper understanding of the benchmarking outcomes and to perform consistency checks of the results from the econometric analysis.

7.4.3 Calculation of efficient cost level

The AER determines efficient costs for each DSOs by following these steps:

- First, for each company the AER determines an overall efficiency score as the arithmetic average of the scores from four economic models. Scores estimated can range between 0% and 100%.
- Second, the AER rebase the efficiency scores derived at the first step by an efficiency target, such that all companies that have a score above the target are considered 100% efficient. The efficiency target is initially set to 75%. The target is then adjusted by an operator-specific factor

79 The two OLS regressions also include fixed effects for each Australian DSO. These are the basis for the efficiency score calculation based on the OLS results.

80 Output variables included in the opex regression are customer numbers, circuit length and ratcheted maximum demand. Environmental factors accounted for in the econometric analysis include customer density, maximum demand density and the degree of network undergrounding.

that correspond to the AER's estimate of the impact of the operating environment on the company's operations which are not captured in the models.

- Finally, the AER sets efficient costs by multiplying the average predicted costs from the four econometric models by the score derived at the third step.

These steps are repeated over two periods of time, a long period starting in 2006 and a shorter period starting in 2012. The efficient costs from the two periods are averaged. The AER only determines efficient costs for those companies that are considered inefficient when comparing scores from the first step with the 75% target, otherwise it uses actual costs as the basis of its forecast allowance.

The four regression models used at the first and final step differ according to the technique used (OLS and SFA) and functional form assumed for the specification of the cost function (Cobb-Douglas and Translog). The efficiency scores are determined as follows:

- For the **OLS models** the efficiency score of operator i is determined based on the difference between the estimated fixed effect of operator i and the smallest estimate of the fixed effects of all operators in the sample. The efficiency score represents the long-run level of efficiency of operator i estimated from the data. It is 1 if operator i determines the frontier and is smaller than 1 otherwise.
- For the **SFA models** the efficiency is estimated by making distributional assumptions on the error terms. Even for SFA, the AER assumes a time-invariant inefficiency, therefore the efficiency score represents the long-run level of inefficiency.

7.4.4 Data sample

As mentioned above, the econometric analyses of the opex of Australian electricity DSOs is conducted on a short and long sample that both include international data from New Zealand and Ontario. In early stages of the development of the Australian benchmarking regime one of the AER's consultants had found that there was insufficient variation over time in the Australia sample to allow for reliable estimations using the econometric models.[81] The consultant suggested to add data from electricity DSOs in Ontario and New Zealand to the sample.

Some advantages of the inclusion of the DSO data from these two jurisdictions are that the data is publicly available, its quality has been checked, and has been used by national authorities, (the

81 Economic Insights. (2015). Response to Consultants' Reports on Economic Benchmarking of Electricity DNSPs. https://www.aer.gov.au/system/files/Economic%20Insights%20-%20Response%20to%20consultants%20%20reports%20on%20AER%20economic%20benchmarking%20-%20April%202015_1.PDF

Ontario Energy Board undertakes a benchmarking study of Ontario energy DSOs using the same data).

The AER's econometric analysis attempts to capture any potential systematic differences between the DSO operating and regulatory environment in the different jurisdictions included in the sample by including in each model a country fixed effect for New Zealand and a country fixed effect for Ontario. In addition, DSOs that are considered small compared to a typical Australian DSO are dropped from the international sample to reduce the influence that they would have on the estimation results.

7.4.5 Conclusions

The Australian case study provides an example of how different techniques can be combined to assess opex and how international data can be used to mitigate problems of small sample sizes. The AER addressed potential problems arising from a lack of comparability across an international sample by dropping the type of small operators that are not present Australian and including country fixed effects into the regression (we note that there could be other differences that have not been accounted for, such as differences in variable definitions or the regulatory environment).

While it is generally unclear whether this is sufficient to control for the effects of systematic differences between the countries, the AER's approach to complement its econometric analysis with extensive MTFP, MPFP and PPI analyses is a potential way to increase the credibility of the regression results as they are based on the Australian data only and can be used to explore sources of changes in the benchmarking results.

Finally, this case study provides an example of combining benchmarking techniques on a given cost category as well as across different cost categories.

The AER applied an earlier version of the methodology summarised in this section to set revenue allowances of the New South Wales DSOs and the Australian Capital Territory DSO in 2015. These networks appealed the AER's decisions in a number of areas, including opex benchmarking. The Australian Competition Tribunal found that the AER had erred in its application of opex benchmarking (and other aspects). The Tribunal ordered the AER remake its opex decision. The AER lodged applications with the Federal Court seeking review of the Tribunal's judgement, but the Court upheld the Tribunal's decision that the AER's approach to forecast opex was in error.

The Tribunal found several errors made by the AER on opex benchmarking. For example, the Tribunal found that the AER had placed too much weight on the results of a single econometric model (SFA Cob-Douglas) and relied too much on the data used for the benchmarking analysis (some of which was estimated or backcasted. It also found that the AER's use of country fixed effects would not correct properly for country-specific differences and analysis underpinning operating environment factors was inadequate. For more details about the outcome of the Tribunal's decision refer to Frontier Economics Australia's report 'Outcome of merits review of AER reset decisions for NSW and ACT networks'.[82]

Since the Tribunal's decision the AER has changed some aspects of its benchmarking approach, for example by considering four econometric models. The AER also worked with he industry to improve the data used in benchmarking.

7.5 Germany – Electricity TSOs

7.5.1 Introduction

Electricity TSOs in Germany are regulated using a revenue cap. The revenue cap is based on historical total costs, which are indexed during the regulatory period by CPI, a generic productivity factor, and an operator-specific efficiency dependent factor. The operator-specific factor is based on an efficiency analysis.

There are four electricity TSOs in Germany (Amprion, TenneT DE, 50Hertz, TransnetBW). In the 1st regulatory period (2009-2013) and 2nd regulatory period (2014-2018), the operator-specific factors were based on an international benchmarking analysis. The analysis used DEA on a sample of European TSOs. In the 3rd regulatory period (2019-2023) a relative RNA was applied to assess the efficiency of the four TSOs. Bundesnetzagentur had some concerns with regards to the transparency of the international benchmarking analysis which could not be resolved with the necessary certainty.[83]

Bundesnetzagentur is currently preparing the 4th regulatory period (2024-2028) for the electricity TSOs where the relative RNA will be applied again.

---

[82] See Frontier Economics' report. https://www.frontier-economics.com.au/documents/2016/04/outcome-merits-review-aer-reset-decisions-nsw-act-networks.pdf

[83] See Bundesnetzagentur's report. https://www.bundesnetzagentur.de/DE/Beschlusskammern/BK08/BK8_05_EOG/59_BesonderhUENB/592_Effizienzvgl/BK8_Effizienzvgl_basepage.html?nn=909818

7.5.2 Benchmarking technique – Greenfield and brownfield approach

Bundesnetzagentur commissioned BET/IFHT (2018)[84] for a study of the RNA for German electricity TSOs. After considering different variants of RNA,[85] BET/IFHT (2018) proposed two approaches:

- **Greenfield approach**. In a pure greenfield approach a reference network required to fulfil a defined supply task is built from scratch with no legacy structure. In practice, BET/IFHT (2018) applied a conditional greenfield approach using the existing networks nodes, existing transmission line routes and existing transmission line corridors as starting point, i.e. the number of network nodes and any routing of lines was exogenously given.

  To avoid discussions about the allocation of national congestion costs to the four electricity TSOs, the greenfield reference network for each TSOs was not compared to the real networks of the TSOs. Instead, it was compared to a 'real network plus' which included additional network components large enough to avoid any congestion. The necessary number of additional components was calculated using the same method as applied when calculating the reference network. Figure 13 illustrates this.

  The efficiency score is calculated as the ratio of total costs of the greenfield reference network to total costs of the real network plus. Both networks' costs are measured by summing up number of components weighted by standard unit costs. (See section below for more details)

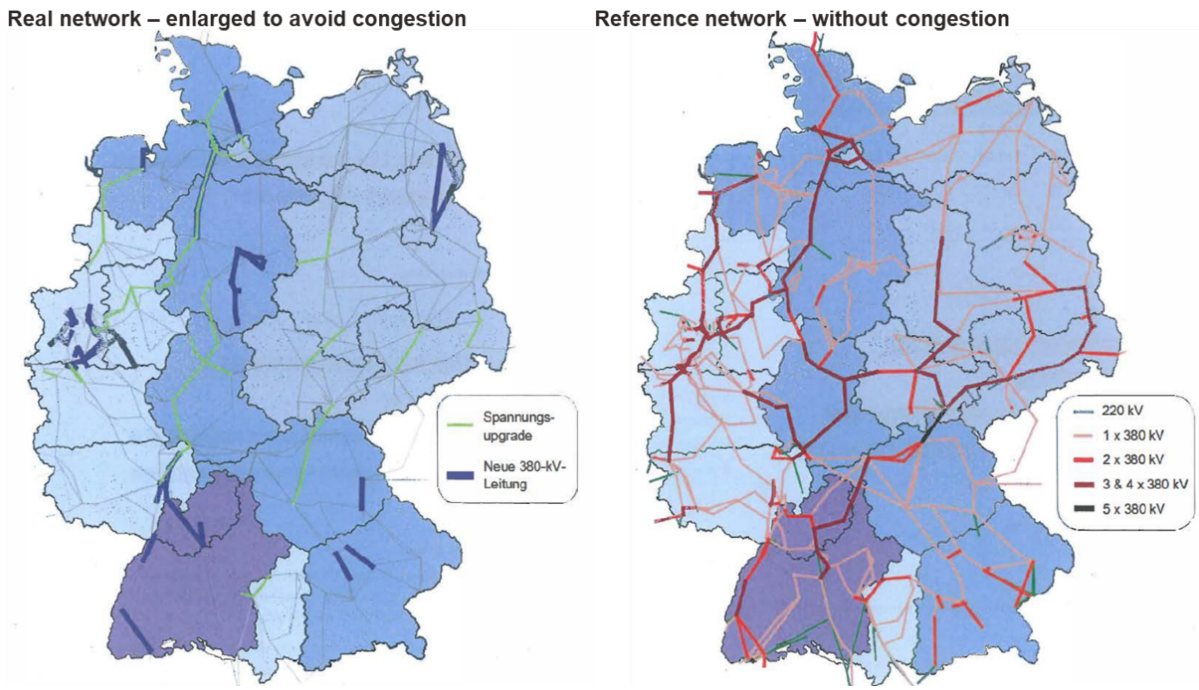
- **Brownfield approach**. The starting point of the brownfield approach is the existing real network of the TSOs. The real network is then analysed to identify which and how many network components can be removed from the real network without impairing technical and operational safety. The efficiency score is derived from the ratio of the costs of the theoretically reduced network to the costs of the real network. To calculate this ratio, the network components are weighted by standard unit costs.

---

84 BET/IFHT. (2018). Gutachten zur Referenznetzanalyse für die Betreiber von Übertragungsnetzen im Auftrag der Bundesnetzagentur. Retrieved from https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Netzentgelte/Strom/GutachtenReferenznetzanalyse.pdf?__blob=publicationFile&v=1

85 Discarded variants of the RNA were the comparison of real with greenfield reference networks having the 'same level' of congestion or comparison of real and greenfield reference networks with different level of congestion but including a price for congestion.

Figure 13 Comparison of Germany's electricity tranmsission network and a reference network using a 'greenfield' approach



Source: Gutachten Referenznetzanalyse' by BET and RWTH Aachen, 2018

7.5.3 Calculation of efficient cost level

Bundesnetzagentur used RNA to assess the

- **'structural efficiency' of the network**: the optimal amount of 'physical assets' to fulfil a defined supply task;
- **But not 'cost efficiency'**: the comparison of 'efficient' costs for reference network with actual costs of the TSOs).

This means that the 'physical network components' of the reference network and the real networks were weighted by the same standard unit costs. The total costs for the reference and real networks are calculated by the sum of depreciation[86] and operating expenditures.[87]

[86] Depreciation = (Quantity of physical assets*Standard unit costs)/depreciation period

[87] Opex = (Quantity of physical assets*Standard unit costs)*0.8%

The method for calculation of TSO's efficiency scores differed between the Greenfield and the Brownfield approach:

- **Greenfield approach**. TSOs' efficiency was calculated by a relative comparison of efficiency scores among the four TSOs applying two steps. First, the TSOs' total costs (based on standard unit costs) of the real network plus are compared to its reference network. The TSO with the highest ratio 'Real network/Reference network' was defined as 100% efficient. Second, the other TSOs' ratios are scaled to the 100% efficient TSO. BET/IFHT (2018) derived efficiency scores between 83.7% and 100%.[88]

- **Brownfield approach**. The efficiency scores are calculated by ratio of the total costs (based on standard unit costs) of the reference network with 'removed network components' and the real network. In BET/IFHT (2018) only one network component was identified as being redundant, resulting in efficiency scores of 100% for three TSOs (Amprion, 50Hertz, TransnetBW) and 99.92% for one TSO (TenneT DE).

BET/IFHT (2018: 58) included a disclaimer with regards to methodological assumptions made within the Greenfield approach having an impact on efficiency scores. The first assumption refers to the calculation of the congestion free "real network plus" which was compared to the reference network. This may have introduced uncertainties into the analysis. The second assumption referred to the restriction of technology choice to only use 380-kV asset when calculation the optimal reference network (even if 220-kV assets would be the better choice). And the third assumption referred to dealing with network structures where there was an overlap of network ownership in a supply region for two TSOs. BET/IFHT noted these assumptions could not be fully clarified from a regulatory and legal perspective. Given this uncertainty Bundesnetzagentur decided to use the efficiency scores from the Brownfield approach when setting the allowed revenues for the 3rd regulatory period. Hence, the impact from the operator-specific efficiency dependent factor on the total costs of the electricity TSOs was negligible.

7.5.4 Data sample and data requirements

The RNA was undertaken with 4 TSOs. This allowed to make a 'relative' comparison between the TSOs, in contrast to only comparing 1 TSO with its reference network ('absolute comparison').

The data requirements for the RNA were complex and consisted of:

- Network data, i.e. limits for line-monitoring, phase-shifter conditions;
- Physical assets (and their location);

[88] The individual scores for the four German TSOs based on the Greenfield approach were not disclosed.

- Market data, i.e. time series for generator production, RES production, and load profiles.
- Allocation of generation/RES and load to network nodes;
- Generators' maintenance schedules, must-run requirements for congestion management;
- Standardised network components used for reference network;
- Standardised unit costs for network components;
- Routes for lines.

Due to the complexity of the data requirement and the RNA calculations the project lead time for the analysis was around 1.5 years (July 2017 to November 2018).

7.5.5 Conclusions

The German case study for electricity TSOs illustrates how to deal with a small sample size by using an engineering based RNA. At the same time the German case study shows the challenges of RNA in the context of benchmarking and how to deal with these challenges.

Bundesnetzagentur decided to replace the DEA based European benchmarking analysis by a national RNA due to data transparency reasons. RNA is an interesting benchmarking technique to deal with a small sample size, because it compares real TSOs' with optimal networks. In principle, RNA is possible only with one TSO. However, as there are 4 TSOs in Germany a relative RNA, which allowed a more "light touched" efficiency assessment, was applied.

Bundesnetzagentur decided to restrict the efficiency assessment on the optimal size of the physical assets necessary to fulfil TSOs' supply task. So Bundesnetzagentur answered only the question, if the size of the physical assets is efficient, but not if the physical assets are operated, maintained or constructed at efficient costs. This can be interpreted as a "light-touched" efficiency assessment, as well.

Bundesnetzagentur decided to apply two RNA approaches: Greenfield and brownfield approach. The reason for the latter was some uncertainty on the results from the greenfield analysis and the resulting inefficiencies for a specific TSOs. The result from the brownfield approach was that all 4 TSOs were efficient. Bundesnetzagentur decided to use the efficiency scores from the brownfield approach, which meant no impacts from RNA efficiency results on allowed costs.

The German example highlights some caveats when using RNA. The decision on the reference network approach (Greenfield vs. Brownfield), i.e. accounting for the historical legacy of the network structure and effects of the 'optimal size' of the reference network. The data requirements are complex and a considerable amount of time to implement the methodology is required. The result from an RNA depends on the applied algorithm and underlying assumptions. This may imply low transparency and high uncertainty on what really drivers the results. Hence, Bundesnetzagentur decided to be cautious in using its efficiency scores because of the high degree of uncertainty surrounding RNA results with all TSOs treated as fully efficient.

7.6 Germany – Electricity and Gas DSOs

7.6.1 Introduction

Electricity and gas DSOs in Germany are regulated using a revenue cap. The revenue cap is based on historical total costs, which are indexed during the regulatory period by CPI, a generic productivity factor, and an operator-specific efficiency dependent factor. The operator-specific factor is based on an efficiency analysis. The large sample of German electricity and gas DSOs allows the application of two benchmarking techniques (DEA and SFA) and, in addition, econometric cost driver analysis and outlier analysis, when undertaking the benchmarking analysis for the electricity and gas operators.

Electricity and gas DSOs challenged various decisions from Bundesnetzagentur including also the benchmarking analysis for electricity and gas DSOs. However, we are not aware that the combination of efficiency scores using the “Best-of-Four” approach was challenged.[89]

7.6.2 Benchmarking techniques – DEA and SFA

The German Incentive-regulation ordinance (‘Anreizreguierungsverordnung – ARegV’) sets the same frame for the efficiency analysis of electricity and gas DSOs. It prescribes the application of two specific benchmarking techniques – DEA and SFA – and outlier analysis in DEA and SFA.

- **DEA**. For DEA, ARegV prescribes the application of constant-returns-to-scale, i.e. operators are responsible for the optimal size, and specific outlier analysis. Outliers in DEA are identified in a sequential approach using the dominance and super-efficiency criteria. The former criterion excludes operators from the sample which have a dominant impact on the efficiency scores of other operators; the latter excludes operators with a DEA efficiency score above a certain threshold.
- **SFA**. For SFA, ARegV only prescribes that outlier analysis shall be applied and provides certain instruments which may be used for that. The efficiency analysis for electricity and gas DSOs applied Cooks’ distance for the outlier analysis. SFA as a parametric approach requires choosing a functional form for the cost function and an assumption on the distribution of the inefficiency term. The SFA for electricity and gas DSOs uses exponential distribution of the

89 We note that in a decision from September, 26th, 2023 the Federal Court of Justice annulled the Bundesnetzagentur decision on gas DSO benchmarking analysis for the 3rd regulatory period. When writing this report no reasoning from the Federal Court of Justice was available.

inefficiency term. The SFA for electricity DSOs[90] uses a norm-linear functional form while for the gas DSOs a Translog functional form is applied. [91] [92]

In addition, ARegV prescribes that total costs (opex, depreciation, financing costs) shall be assessed by using two variants of capital costs:

- Non-standardised capital costs mainly based on operators' financial accounts (i.e. different depreciation periods possible); and
- Standardised capital costs based on annuities with standardised depreciation periods and financing costs.

7.6.3 Calculation of efficient cost levels

Operators' efficiency scores are evaluated against the efficiency frontier (i.e. 100% efficient operator). The operators have to reduce the inefficient part of total costs within one regulatory period (5 years). Hence, efficiency scores are mechanistically translated into regulatory total cost targets and there is no distinction between operating and capital costs. In addition, the efficiency analysis does not specify which (disaggregated) costs are inefficient. However, the German regulatory approach includes different safety nets for the operators:

- **Outlier analysis**. Operators identified as outliers are excluded from the efficiency analysis. This tends to increase the efficiency scores of the other operators;
- **Minimum efficiency score**. Operators with an efficiency score below 60% are set at 60% when calculating the operator-specific efficiency dependent factor.
- **Best-of-four approach.** ARegV prescribes the application of two benchmarking techniques – DEA and SFA – using two variants for capital costs (non-standardised/standardised). This results in four efficiency scores from four efficiency model specifications.[93] ARegV prescribes

---

90 For further details we refer to: Swiss Economics/Sumicsid/IAEW. (2019), https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Netzentgelte/Strom/Effizienzvergleich_VNB/3RegPer/Gutachten_EVS3_geschw.pdf?__blob=publicationFile&v=3.

91 For further details on the benchmarking analysis for the 3rd regulatory period we refer to: Frontier Economics/TU Berlin. (2019). Effizienzvergleich Verteilernetzbetreiber Gas (3. Regulierungsperiode); Report for Bundesnetzagentur. https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Netzentgelte/Gas/Gutachten_Effizienzvergleich_final.pdf?__blob=publicationFile&v=2

92 For further details on the benchmarking analysis proposed fort he 4th regulatory period we refer to: Frontier Economics/TU Berlin. (2023). Effizienzvergleich Verteilernetzbetreiber Gas (4. Regulierungsperiode); Report for Bundesnetzagentur. https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Netzentgelte/Gas/Gutachten_Entwurf2_4RP.pdf?__blob=publicationFile&v=2

93 DEA with non-standardised capital costs; DEA with standardised capital costs; SFA with non-standardised capital costs; SFA with standardised capital costs.

that the efficiency score used to set operators' cost targets is determined by the maximum of the four efficiency scores.

7.6.4 Data sample and cost driver analysis

The efficiency analysis for electricity and gas DSOs can draw on a data sample of around 200 operators each. The large data sample allows the application of econometric cost driver analysis, when specifying the model for the efficiency analysis. The data used in the cost driver analysis are also published by Bundesnetzagentur, which allows operators to replicate the results from the efficiency analysis.

7.6.5 Conclusions

The German case study for electricity and gas DSOs illustrates how to combine different benchmarking techniques and cost definitions. It shows the advantage of having a large data sample for operators regarding the choice of benchmarking techniques, econometric cost-driver analysis and outlier analysis. The German case study also shows how legal requirements shape the model specification of a benchmarking analysis.

The German regulatory regime recognises the different properties of benchmarking techniques. Hence, a combination of benchmarking techniques characterised by different properties (DEA and SFA) is applied. In addition, the German regulatory regime takes into account the possible impact of different depreciation policies and investment cycles on efficiency scores by using two definitions for capital costs: non-standardised and standardised. The latter are based on annuities using uniform depreciation periods and financing costs for all operators. When combining the different benchmarking techniques and cost definitions the German regulatory regime prescribes using the maximum of the efficiency scores in order to set cost targets for gas DSOs ("Best-of-Four"). This can be interpreted as a cautious approach.

The large number of electricity gas DSOs allows the application of SFA. In addition, the selection of appropriate outputs covering operators' supply task can be based on econometric cost-driver analysis. The analysis also allows to take into account outliers in the cost-driver analysis. The large data sample would also allow other benchmarking technique like StoNED. However, the current legal frame precludes this option.

One feature of the German regulatory framework is that the ARegV includes detailed legal requirements on the regulatory model, e.g. which benchmarking techniques to use. This limits the regulatory discretion of Bundesnetzagentur. However, the European Court of Justice decision from 2nd September 2021 (C-718/18)[94] claimed that this is in conflict with European law and more

[94] See the European Court of Justice's decision. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62018CJ0718

competences must be assigned to Bundesnetzagentur. The detailed implementation of the EuGH decision in German law and how Bundesnetzagentur will use the new assigned competences is still in an early phase.

8 Summary of key findings

In this report, we identified a short list of benchmarking techniques that could be used to determine the efficient costs of gas and electricity TSOs and DSOs in future regulatory periods. We found the following:

- There is a large set of benchmarking techniques that can be used to address the challenges of the energy transition and implications from the CBb's ruling. These techniques are part of the academic literature and have been used by regulators in other jurisdictions. These techniques include both descriptive techniques (PPIs), economic-based techniques (COLS, MOLS, DEA, SFA, StoNED), and engineering-based models (including RNA).
- A combination of techniques can be used to i) improve the robustness of the assessment of efficiency of a given cost category and ii) benchmark different cost categories. On point i), it is possible to combine results from different techniques applied at the same level of cost aggregation to mitigate some of the weaknesses of specific techniques (e.g. benchmarking totex using both SFA and DEA). It is also possible to combine results from techniques applied at different level of cost aggregation (e.g. results from a top-down benchmarking of totex are combined with results from a bottom-up benchmarking of components of totex). On point ii), different techniques can be applied to different cost categories, for example. SFA might be better suited for benchmarking business as usual activities, while engineering models might be better for bespoke large capex investments.
- Some of the techniques we identified can be used to understand where inefficiency is coming from. For example, even if the overall efficiency of costs is assessed using a top-down econometric model, it would be possible to apply specific models to disaggregated costs to understand where inefficiency comes from. It is also possible to use some descriptive statistics like PPIs to understand how unit costs might different between companies and use this information to attempt to understand the source of inefficiencies.
- Other aspects of the broader benchmarking framework are as important as the choice of technique to address the challenges set out above. For example, when undertaking a benchmarking analysis it is important to also consider the set of comparators, how the cost drivers are defined, whether the data is consistent across operators and over time, how the results of benchmarking are used (e.g. mechanistically or not) and which incentives are in place (e.g. whether the operators are incentives to provide accurate forecasts; what the implications of benchmarking opex and capex separately are).

Annex A – Technical details on the Caves, Christensen and Diewert (1982) multilateral Törnqvist TFP

Consider the index proposed by Caves, Christensen and Diewert (1982) (henceforth CCD): Let x_i and y_i indicate the vector of inputs and outputs for firm i, respectively. To aggregate all inputs and outputs into a scalar variable that can be used to obtain a univariate measure of efficiency the CCD multilateral Törnqvist TFP defines the output and input weights based on their revenue and cost shares. That is, the k-th element in the vector w_y^i containing the weights associated to the outputs of firm i is defined as

$$w_y^{i,k} = \frac{r_i^k}{\sum_1^N r_i^k y_i^k},$$

where r_i^k refers to the revenue associated to the k-th element in the output vector y_i. Likewise, the k-th element in the input weight vector w_x^i is defined as

$$w_x^{i,k} = \frac{c_i^k}{\sum_1^M c_i^k x_i^k},$$

where c_i^k refers to the cost associated to the k-th element in the input vector $\boldsymbol{x}_i$. A crucial feature of the MTFP approach is that every observation is set in relation to a sample average, before it is compared to any other observation in the sample. Let $\overline{w}_y^k$, $\overline{w}_x^k$, $\overline{\ln\,(x^k)}$ and $\overline{\ln\,(y^k)}$ refer to the sample averages of the individual quantities. The CCD's MTFP then compares the output of firm i to the sample average based on the following formula:

$$\theta_i^{CCD}(\mathbf{y}_i) = \frac{1}{2}\sum_i^N \left(w_y^{i,k}+\overline{w}_y^k\right)\left(\ln(y_i^k)-\overline{\ln\,(y^k)}\right)$$

Analogously, inputs are compared to the sample average based on

$$\theta_i^{CCD}(\boldsymbol{x}_i) = \frac{1}{2}\sum_i^N \left(w_x^{i,k}+\overline{w}_x^k\right)\left(\ln(x_i^k)-\overline{\ln\,(x^k)}\right)$$

The CCD MTFP between two firms i and j is defined as

$$MTFP_{i,j}^{CCD} = \theta_i^{CCD}(\mathbf{y}_i) - \theta_j^{CCD}(\mathbf{y}_j) - \theta_i^{CCD}(\boldsymbol{x}_i) + \theta_j^{CCD}(\boldsymbol{x}_j)$$

Annex B – Detailed discussion of semi-parametric techniques

This annex provides further details on some of the semi-parametric techniques discussed in the main report. Section B.1. covers constrained stochastic approaches such as FLW, Parmeter and Racine and StoNED. The section also includes a detailed discussion of the importance of homoskedasticity of the inefficiency distribution when applying these methods that is beyond the scope of the main text. Section B.2 deals with improvements of standard DEA/FDH methods, covering stochastic DEA/FDH as well as bias corrected/bootstrap DEA.

Different to the discussion provided in the main text the description of the different methods, and particularly their formal representations, are stated in terms of production functions instead of cost functions. It thus follows the academic literature more closely. However, all methods discussed in this annex can easily be translated into a cost function context.

B.1 Constrained stochastic approaches

We begin the discussion with the general cross-sectional stochastic frontier model:

$$y_i = m(\boldsymbol{x_i}, \boldsymbol{z_i}) + v_i - u_i = m(\boldsymbol{x_i}, \boldsymbol{z_i}) + \varepsilon_i,$$

where $m(\cdot)$ is the production frontier of technology that can be used to transform a traditional vector of inputs, $x \in R^{qx}$, and a vector of environmental or contextual variables $\boldsymbol{z} \in R^{qz}$ into scalar output y_i, distorted by some noise $v_i \sim D(0, \sigma_v(\boldsymbol{x_i}, \boldsymbol{z_i}))$ that is beyond the production process (as well as serving to capture measurement error and other potential model misspecification) and by technical inefficiency $u_i \sim D_+(\gamma(\boldsymbol{x_i}, \boldsymbol{z_i}), \sigma_u(\boldsymbol{x_i}, \boldsymbol{z_i}))$. Traditional estimation of the model begins by assuming a specific parametric functional form for the production technology $m(\boldsymbol{x}, \boldsymbol{z})$, as well as making distributional assumptions on both v_i and u_i as well as parametric specification of the distributional parameters, $\sigma_v(\boldsymbol{x}, \boldsymbol{z}), \gamma(\boldsymbol{x}, \boldsymbol{z})$, and $\sigma_u(\boldsymbol{x}, \boldsymbol{z})$. The unknown parameters are then estimated via the maximum likelihood (ML) estimator. For example, Aigner, Lovell & Schmidt (1977) assume that $m(\boldsymbol{x}, \boldsymbol{z}) = m(\boldsymbol{x}) = \boldsymbol{x}\beta$, that $\sigma(\boldsymbol{x}, \boldsymbol{z}) = \sigma_v$, $\gamma(\boldsymbol{x}, \boldsymbol{z}) = 0$, and $\sigma_u(\boldsymbol{x}, \boldsymbol{z}) = \sigma_u$, coupled with the assumption that $v_i \sim N(0, \sigma_v)$ and $u_i \sim N_+(0, \sigma_u)$. The models discussed in this paper have focused on relaxing many of the parametric assumptions deployed in Aigner et al. (1977) and a majority of the applied frontier analysis literature when SFA is used.

The semiparametric approaches that ACM is interested in all deal with estimation of the production frontier, $m(\boldsymbol{x}_i, \boldsymbol{z}_i)$, without placing parametric restrictions on the frontier itself. The approach is semiparametric since, after estimation of the production frontier, efficiency is calculated via distributional assumptions on the composed error. In the setup of Fan, Li & Weersink (1996) (FLW), they did not consider contextual variables, and the model becomes

$$y_i = m(\boldsymbol{x_i}) - u_i + v_i.$$

FLW's approach was to first ignore the composite nature of the error term and simply estimate the production frontier as though it were a conditional mean. However, this first step goes beyond straight application of kernel regression given that the estimated conditional mean is a biased estimator when ignoring the inefficiency term; the key condition required for consistent estimation of the production frontier in a regression setting is $E[\varepsilon|x] = 0$. However, given the one-sided nature of u, this condition is not satisfied, rather $E[\varepsilon|x] = \gamma$. Thus, the location of the production frontier cannot be identified in the regression setup as it is the case that

$$y_i = m(\boldsymbol{x_i}) + \varepsilon_i = m(\boldsymbol{x_i}) + \gamma + (\varepsilon_i - \gamma) \equiv g(\boldsymbol{x_i}) + \varepsilon_i$$

FLW proposed a solution to correct the (downward) bias in the estimation of $m(\boldsymbol{x})$ by retaining standard distributional assumptions from the SFA literature (e.g., Normal noise, Half Normal inefficiency) and estimating the corresponding distributional parameters via maximum likelihood on the nonparametrically estimated residuals from a local-constant (Nadaraya-Watson) regression estimator. Once these parameters have been estimated, the estimated conditional mean can be shifted (bias-corrected) by the estimated mean of in- efficiency (the mean correction factor). Under fairly weak conditions FLW show that the parameters of the composed error distribution can be consistently estimated at the parametric $\sqrt{n}$ rate.[95]

The approach just described is what is known as plug-in (PI) likelihood. The PI approach of FLW begins by making assumptions on the error components in the stochastic production frontier model. They assume that noise follows a Normal distribution (as per usual, with zero mean and constant, but unknown, homoskedastic variance, σ_v) and that technical inefficiency stems from a Half Normal distribution, with the unknown finite and homoskedastic variance to be estimated, σ_u.

Given these distributional assumptions and the (biased) nonparametric estimator of the frontier, one would estimate the semiparametric stochastic frontier model and the unknown distributional parameters as follows.

Step 1: Estimate the conditional expectation of (4.3), $E[y_i|x_i]$, using nonparametric methods; FLW deploy kernel smoothing but a regression spline or sieve estimator would work equally well. Call this $\widehat{m^*}(x_i)$and let the residuals be denoted $\widehat{\varepsilon_i^*} = y_i - \hat{g}(x_i)$.

95 Martins-Filho & Yao (2015) showed that these parameter estimators are actually biased and inefficient.

Step 2: Define the concentrated variance of the composed error term $\sigma^2(\lambda)$ as a function of $\lambda = \sigma_u/\sigma_v$ and $\sigma^2 = \sigma_u^2 + \sigma_v^2$, as follows:

$$\widehat{\sigma^2}(\lambda) = \frac{n^{-1}\sum_{i=1}^{n}\left(\widehat{\varepsilon_i^*}\right)^2}{1 - \frac{2\lambda^2}{\pi(1+\lambda^2)}}.$$

Step 3: Define the mean correction factor $\gamma = \sqrt{2/\pi}\sigma_{\mathrm{u}}$ as a function of λ, i.e.,

$$\hat{\gamma}(\lambda) = \frac{\sqrt{2}\hat{\sigma}(\lambda)\lambda}{\left(\pi(1+\lambda^2)\right)^{1/2}}.$$

Step 4: Estimate λ by maximizing the concentrated log likelihood function consistent with the Normal, Half Normal distributional assumptions which is

$$\hat{\lambda} = argmax\left(-n\ln\hat{\sigma}(\lambda) + \sum_{i=1}^{n}\ln\Phi\left(-\hat{\varepsilon}_i\lambda/\hat{\sigma}(\lambda)\right) - \left(2\hat{\sigma}^2(\lambda)\right)^{-1}\sum_{i=1}^{n}\widehat{\varepsilon_i^2}\right),$$

where $\widehat{\varepsilon_i} = \widehat{\varepsilon_i^*} - \hat{\gamma}(\lambda)$.

Step 5: The stochastic production frontier $m(x_i)$ is consistently estimated by

$$\hat{m}(x_i) = \widehat{m^*}(x_i) - \hat{\gamma}(\hat{\lambda}),$$

where $\hat{\gamma}(\hat{\lambda}) = \sqrt{2}\hat{\sigma}(\hat{\lambda})\hat{\lambda}/\left(\pi\left(1+\widehat{\lambda^2}\right)\right)$ and $\hat{\sigma}(\hat{\lambda}) = \sqrt{\widehat{\sigma^2}(\hat{\lambda})}$.

The concentration of the Normal-Half Normal likelihood function is for simplicity. One could also maximize the traditional ML function. If alternative distributional assumptions were made on u, for example exponential or truncated normal, a concentrated version of the log- likelihood function may not exist. The reason is that a simple closed form expression for $\hat{\sigma}^2$ may not exist with alternative distributional assumptions.

The PI-likelihood approach of FLW has been modified to impose axioms of production using a variety of approaches. For example, Parmeter & Racine (2012) and Noh (2014) use kernel based methods to impose monotonicity and concavity while Kuosmanen & Kortelainen (2012) use convex nonparametric least-squares (CNLS). Both of these approaches follow the same steps to estimate the distributional parameters, the only difference lies in how the conditional mean is estimated.

The Importance of homoskedasticity of the inefficiency distribution

A key assumption that warrants further discussion as it pertains to the use of the methods of FLW, Parmeter & Racine (2012), Kuosmanen & Kortelainen (2012), and Noh (2014) is the constancy of all the parameters of the distribution of ui. The reason for this is twofold. First, it is quite common to model 'heteroskedasticity' of ui on a set of variables, commonly termed determinants of inefficiency, typically denoted as the vector $\boldsymbol{z}$. In this case, the first stage regression ignores this structure entirely. Textbook econometrics shows that traditional heteroskedasticity in the zero-mean, symmetric error term leads to no impact on the consistency of the estimator of the conditional mean (in this case the production frontier, $m(\cdot)$). However, when ui depends on these determinants, the type of heteroskedasticity that arises is quite different from traditional heteroskedasticity in the two-sided error term due to the truncated nature of the distribution of ui at 0. In this case, all of the moments of ui depend on the determinants, including the non-zero mean, which implies that the first stage estimation of the mean is either misspecified (an input also influences ui), there is an omitted variable bias (a new variable was left out of the estimation of the conditional mean), or worse, both of these occur. Second, given the truncation of the distribution of ui, the moments of ui will depend on all the parameters of the inefficiency distribution, so, for example, if ui were assumed to be Truncated Normal with pre-truncation mean parameter $\mu(\boldsymbol{z})$ and pre-truncation variance parameter $\sigma u(\boldsymbol{z})$, even if one of these two parameters is constant (does not depend on $\boldsymbol{z}$), the post-truncation mean of ui will depend on $\boldsymbol{z}$.

One might be tempted to follow Kuosmanen & Kortelainen (2012), Parmeter & Racine (2012), or Noh (2014), imposing the desired constraints first, and then recover $E[u|\boldsymbol{w}]$. However, these methods only work as intended when the distribution of inefficiency is independent of $\boldsymbol{x}$ and $\boldsymbol{z}$, i.e. when $\boldsymbol{u}$ is homoskedastic. The issue the applied researcher faces here is much more subtle. When heteroskedasticity is present in u, one must recognize that what is being estimated is a conditional mean, and not a production frontier. Thus, it is not necessarily the case that the axioms of production should be expected to hold when estimating the conditional mean.

Consider the case of the estimation of a production frontier. The conditional mean of output could be non-monotonic in w if $\widehat{E}[u|\boldsymbol{w}]$ was non-monotonic, even though the production frontier is monotonic in w. Further, it is well known that adding two concave functions might not produce a concave function, so even if $\widehat{E}[u|\boldsymbol{w}]$ was concave in $\boldsymbol{w}$, subtracting it from the production frontier may not produce a concave production function in $\boldsymbol{w}$. And therein lies the danger of imposing constraints when estimating the conditional mean, it is not necessarily the case that they should be satisfied. This might seem innocuous except for the fact that imposing constraints on a conditional mean which are incorrect will not produce a consistent estimator and typically, consistent estimates in the first stage are needed for the second stage to produce valid estimates of inefficiency.

Take for example the discussion in Kuosmanen, Johnson & Saastamoinen (2015, pg. 233), who consider the estimation of a production frontier nonparametrically, while also allowing $\boldsymbol{u}$ to depend on x. In this case they state (in our notation) "…Note that the shape of function g can differ from that of frontier m because $E(u|\boldsymbol{x})$ is a function of inputs x . . . It is also worth noting that function g is not necessarily monotonic increasing and concave even if the production function m satisfies these axioms because $-E(u|\boldsymbol{x})$ can be a non- monotonic and non-concave function of inputs…". To apply CNLS in step 1, we need to assume that the curvature of the production function m dominates and that function g is monotonic increasing and concave (at least by approximation)." Unless the conditional mean of output satisfies the axioms of production, it is recommended that the axiomatic restrictions be enforced after consistent unrestricted estimation of the conditional mean as this will ensure that the first stage estimator of the conditional mean is consistent. In the context of both Kuosmanen & Kortelainen (2012) and Parmeter & Racine (2012), the constraints would be enforced prior to the recovery of the distributional parameters. In the Kuosmanen & Kortelainen (2012) setting, CNLS works as an estimator because of the constraints enforced, whereas Parmeter & Racine's (2012) approach is identical to FLW if no constraints are enforced, and in fact if one has heteroskedasticity then the approach of Simar, Van Keilegom & Zelenyuk (2017) is really the best alternative provided prior to any constraints being enforced.

Ensuring axioms of production

As standard nonparametric estimation methods are known to overfit the in-sample data, there is the tendency that the estimated production frontier could be either non-monotonic or non-concave, both of which may run contrary to the usual axioms of production.[96] The approaches of Kuosmanen & Kortelainen (2012) and Parmeter & Racine (2012) are essentially identical to FLW except that they require the estimated production frontier to obey traditional axioms of production theory in economics, such as monotonicity and concavity, something that FLW (and a majority of other studies), did not accommodate in their approach, although the idea of imposing constraints is mentioned in passing.[97]

Parmeter & Racine (2012) follow the framework of FLW closely, using the same kernel smoothing methods to estimate the conditional mean, but deploying constrained kernel estimation (Hall &

96 Monotonicity or free disposability of inputs and outputs are usually assumed or accepted as axioms in production theory, though there are also exceptions, e.g., cases with bad outputs or input congestion, which may require non-monotonicity (or weak disposability). Concavity and convexity are the key properties of cost and revenue functions, respectively, implied by economic theory.

97 We note here that the approach of FLW, which is firmly entrenched as a stochastic frontier model, is more general than the approaches of Kuosmanen & Kortelainen (2012) and Parmeter & Racine (2012) as these methods rely on axioms of production to produce an initial, consistent first stage estimator. Without these assumptions being true, the nonparametric estimators which they propose are inconsistent.

Huang 2001, Du, Parmeter & Racine 2013) to ensure that monotonicity and concavity hold. Kuosmanen & Kortelainen (2012) use an entirely different nonparametric estimator than that of FLW, known as concave nonparametric least-squares (CNLS), which also enforces monotonicity and concavity. One benefit of CNLS is that it has less bias than traditional kernel methods because it does not rely on smoothing the data, instead fitting the unknown frontier linearly and relying on the axioms of production to extract statistical gains (in terms of bias and mean square error) from the data (Seijo & Sen 2011). As yet, a detailed comparison of these two competing methods has not been considered.

To detail the approach of Parmeter & Racine (2012), in what follows we let $\{(x_i, y_i)\}_{i=1}^{n}$ denote sample pairs of inputs and outputs and x a point of support at which we evaluate the frontier. The goal is to nonparametrically estimate the unknown production frontier $m(x)$ subject to constraints on $m^{\boldsymbol{s}}(x)$ where s is a k-vector corresponding to the dimension of x. The elements of s represent the order of the partial derivative corresponding to each element of x. Thus $\boldsymbol{s}$ = (0, 0, . . . , 0) represents the function itself, while $\boldsymbol{s}$ = (1, 0, . . . , 0) represents $\partial m(x)/\partial x_1$. In general, for $\boldsymbol{s} = (s_1, s_2, \ldots, s_k)$ we have

$$m^{(s)}(x) = \frac{\partial^{s_1} m(x)}{\partial x_1^{s_1}}, \ldots, \frac{\partial^{s_k} m(x)}{\partial x_k^{s_k}}.$$

We consider the class of kernel regression smoothers that can be written as linear combinations of the output y_i, i.e.,

$$\widehat{m}(x) = \sum_{i=1}^{n} n^{-1} A_i(x) y_i,$$

which is a very broad class. For instance, the local constant or Nadaraya-Watson estimator (which is what is used by FLW) uses

$$A_i(x) = \frac{n K_\gamma(x_i, x)}{\sum_{j=1}^{n} K_\gamma(X_j, x)},$$

where $K_\gamma(\cdot)$ is a generalized product kernel that admits both continuous and categorical inputs, and γ is a vector of bandwidths; see Racine & Li (2004) for details.

In order to impose constraints on a nonparametric frontier, we shall require a nonparametric estimator that satisfies constraints of the form

$$l(x) \leq \widehat{m}^{(s)}(x) \leq u(x)$$

for arbitrary $l(\cdot)$, $u(\cdot)$, and $\boldsymbol{s}$, where $l(\cdot)$ and $u(\cdot)$ represent (local) lower and upper bounds, respectively.

The constrained estimator is obtained by introducing an n-vector of weights p chosen so that the resulting estimator satisfies the constraint above. We define the constrained estimator to be

$$\hat{m}(x|p) = \sum_{i=1}^{n} p_i A_i(x) y_i.$$

Construction of $\hat{m}(x|p)$ proceeds as follows. Let p_u be an n-vector with elements $1/n$ and let p be the vector of weights to be selected. In order to impose our constraints, we choose $p = \hat{p}$ to minimize the distance from p to the uniform weights $p_i = 1/n \; \forall i$ using the distance metric $D(p) = (p_u - p)'(p_u - p)$. The constrained estimator is then obtained by selecting those weights p that minimize $D(p)$ subject to constraints such as those given in the restrictions defined below, which can be cast as a general nonlinear programming problem. For the constraints we need to impose (frontier behaviour, monotonicity and concavity) we will have inequalities that are linear in p, which can be solved using standard quadratic programming methods and off-the-shelf software.[98] The appropriate bandwidth(s) for our unknown function can be estimated using any of the commonly available data-driven procedures and require estimation of the unrestricted function only. For notational simplicity we shall drop the '$|p$' notation with the understanding that the constrained estimator is that defined before.

An interesting feature of the application of constrained kernel methods is that it selects the weights in a manner that minimizes differences between the uniform weights (i.e. p_u) and the amount necessary to enforce the constraints. Conceptually there is nothing wrong with this approach, but it has the wrong focus. What one should be more interested in is minimizing the difference between the unrestricted production frontier and the theoretically constrained frontier. This approach was proposed by Li, Liu & Li (2017), whose criterion function is

$$min \sum_{i=1}^{n} \left(\tilde{m}(\boldsymbol{x_i}|\boldsymbol{p}) - \hat{m}(x_i)\right)^2.$$

With weights chosen in this fashion, Li et al. (2017) have shown that the resulting estimator significantly outperforms weights selected via $(\boldsymbol{p} - \boldsymbol{p_u})'(\boldsymbol{p} - \boldsymbol{p_u})$. As with Parmeter & Racine (2012) and Kuosmanen & Kortelainen (2012), direct enforcement of the constraints may not be correct if the object being restricted is a conditional mean, as opposed to the production frontier.

Let our constrained estimator, $\hat{m}(x|p)$, satisfy the following restrictions:

$$\sum_{i=1}^{n} p_i A_i(x_i) y_i - y_i \geq 0,$$

98 For example, in the R language it is solved using the quadprog package, in GAUSS it is solved using the qprog command, and in MATLAB the quadprog command. Even when n is quite large the solution is computationally fast using any of these packages.

$$\sum_{i=1}^{n} p_i \left[\sum_{s \in \boldsymbol{S_1}} A_i^{(s)}(x) \right] y_i \geq 0,$$

$$\sum_{i=1}^{n} p_i \left[\sum_{s \in \boldsymbol{S_2}} A_i^{(s)}(x) \right] y_i \leq 0,$$

where $\boldsymbol{S}_1$ is

$$[(1,0,\dots,0)\&\ (0,1,\dots,0)\&\ \cdots \&\ (0,0,\dots,1)]_k,$$

while $\boldsymbol{S}_2$ is

$$[(2,0,\dots,0)\&\ (0,2,\dots,0)\&\ \cdots \&\ (0,0,\dots,2)]_k.$$

These three conditions guarantee that the estimated frontier lies (weakly) above all observed output while respecting monotonicity and necessary conditions for concavity. One can follow the approach of Kuosmanen & Kortelainen (2012) and impose the necessary and sufficient conditions for concavity, however, this leads to $n(n-1)$ constraints that need to be satisfied and not all applications will necessarily lead to a solution. A simpler approach would be to impose the necessary conditions for concavity, then check to see which of the sufficient conditions are violated, adding them to the optimization problem and repeating until no sufficient conditions are violated.

B.2 Improvements to DEA/FDH

Stochastic DEA/FDH

As is well known, one of the advantages of both DEA and FDH, inter alia, is that they handle the case of technologies with multiple outputs and multiple inputs in a simple manner. The stochastic versions of these approaches proposed in Simar (2007) and elaborated in Simar & Zelenyuk (2011), while combining non-parametric SFA with DEA or FDH also aimed to preserve this important advantage. Here we briefly summarize their approach as one of the ways to impose constraints of convexity or/and monotonicity (or free disposability) on technology, for the cases of technologies with multiple outputs and multiple inputs, and allowing for statistical noise or outliers, thus improving upon standard DEA/FDH.

To facilitate further discussions, let $\mathbf{y} \in R_{+}^{M}$ denote the vector of outputs (and, as before, $\mathbf{x} \in R_{+}^{q_x}$ denote the vector of inputs) and Ψ denote the production technology set of all the feasible input-output combinations, defined in general terms as

$$\Psi = \left\{(x, y) \in \mathbb{R}_+^{q_x+M} \mid \mathbf{x}\, can\, produce\, \mathbf{y}\right\}.$$

The Farrell-Debreu output oriented (technical) efficiency measure (Debreu 1951, Farrell 1957) of a DMU evaluated at an allocation $(\mathbf{x}, \mathbf{y})$ is defined as

$$\delta(\boldsymbol{x}, \boldsymbol{y}) = sup\{\delta | (\boldsymbol{x}, \delta\boldsymbol{y})\} \in \Psi.$$

Therefore $\delta(\boldsymbol{x}, \boldsymbol{y}) = 1$ means that the DMU with $(\boldsymbol{x}, \boldsymbol{y})$ is technically efficient (from an output oriented perspective), while $\delta(\boldsymbol{x}, \boldsymbol{y}) > 1$ indicates inefficiency in the sense that it is possible to increase all the outputs $\delta(\boldsymbol{x}, \boldsymbol{y}) > 1$ times in order to reach the technology frontier. Thus, the efficient level of output for an allocation $(\boldsymbol{x}, \boldsymbol{y})$ is then given by

$$y^{\partial} = \delta(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{y}.$$

Because Ψ is usually unobserved, and so are its frontier $y^{\partial}(\boldsymbol{x}, \boldsymbol{y})$ and its efficiency score $\delta(\boldsymbol{x}, \boldsymbol{y})$, they need to be estimated using a sample of DMUs, which we will denote with $\boldsymbol{S} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i) | i = 1, \dots, n\}$. Specifically, recall that the FDH estimator for Ψ is given by [99]

$$\widehat{\Psi}_{FDH,n} = \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{q_x+M} | \boldsymbol{x}_i \leq \boldsymbol{x}, \boldsymbol{y}_i \geq \boldsymbol{y}, \forall (\boldsymbol{x}_i, \boldsymbol{y}_i) \in \boldsymbol{S}\right\}$$

while the DEA estimator for Ψ can be written as

$$\widehat{\Psi}_{DEA,n} = \left\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^{q_x+M} | \sum_{i=1}^{n} \zeta_i \boldsymbol{x}_i \leq \boldsymbol{x}, \sum_{i=1}^{n} \zeta_i \boldsymbol{y}_i \geq \boldsymbol{y}, \zeta_i \geq 0, i = 1, \dots, n, \rho_1 \leq \sum_{i=1}^{n} \zeta_i \leq \rho_2\right\}$$

where $\rho_1 = 0, \rho_2 = +\infty$ if one is willing to assume constant returns to scale (CRS) model (see Farrell 1957, Charnes, Cooper & Rhodes 1978), $\rho_1 = 0,\ \rho_2 = 1$ if one is willing to assume decreasing returns to scale (DRS), and $\rho_1 = \rho_2 = 1$ if one assumes variable returns to scale (VRS) (e.g., see Färe & Grosskopf 1983, Banker, Charnes & Cooper 1984). Also note that $\widehat{\Psi}_{DEA,n}$, under VRS, is a convex closure of $\widehat{\Psi}_{FDH,n}$. The asymptotic properties of these estimators have been well studied and well summarized in the recent reviews of Simar & Wilson (2013, 2015).

In a nutshell, the stochastic DEA/FDH approach can be summarized as follows (see Simar & Zelenyuk (2011) for further details):

Step 1: Transform the data $(\boldsymbol{x}_i, \boldsymbol{y}_i), i = 1, \dots, n,$ into polar coordinates for the outputs, denoting the transformed data as $(\boldsymbol{x}_i, \omega_i, \xi_i)$, so that $(\boldsymbol{x}, \boldsymbol{y}) \Leftrightarrow (\boldsymbol{x}, \omega, \xi)$, where $\omega \in R_+$ is the modulus and $\xi \in [0, \pi/2]^{M-1}$ is the amplitude (angle) of the vector $\boldsymbol{y}$ describing the output-mix, i.e.,

---

[99] Free disposability of inputs and outputs is defined as (x,y)∈ Ψ implies that (x',y')∈Ψ for all $x'≥ x and y'≤ y.

$\omega = \omega(\boldsymbol{y}) = \sqrt{\boldsymbol{y}'\boldsymbol{y}}$ and $\xi = (\xi^1, \ldots, \xi^{M-1}) = \xi(\boldsymbol{y})$, where for $j = 1, \ldots, M-1$, $\xi^j = \arctan\left(\frac{y^{j+1}}{y^1}\right)$ if $y^1 > 0$ or $\xi^j = \frac{\pi}{2}$ if $y^1 = 0$.

Step 2: For each point of a selected grid $(\boldsymbol{x}_k, \boldsymbol{y}_k)$, $k = 1, \ldots, \tilde{n}$, use a non-parametric or semi-parametric methods described above to estimate the frontier in terms of $\omega(y)$, using the original sample $\boldsymbol{S} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i) : i = 1, \ldots, n\}$. For example, adapting the SVKZ approach, this will amount to estimating $m(\xi_k, \boldsymbol{x}_k)$, using model

$$\log \omega_i = m(\xi_i, \boldsymbol{x}_i) - u_i + v_i, \quad i = 1, \ldots, n$$

where $(v|\boldsymbol{x}, \xi) \sim f_{v|x,\xi}$ is a zero-mean symmetric noise and $(u|\boldsymbol{x}, \boldsymbol{z}, \xi) \sim f_{u|\backslash bmx,\xi}$ is asymmetric inefficiency term (e.g., from Half-Normal distribution, conditional on (x, ξ)), and denoting these estimates as $\widehat{m}(\xi_k, \boldsymbol{x}_k)$.

Step **3**: Project the grid points $(\boldsymbol{x}_k, \boldsymbol{y}_k)$, $k = 1, \ldots, \tilde{n}$ onto the estimated frontier to obtain $(\boldsymbol{x}_k, \boldsymbol{y}_k^*)$, where $\boldsymbol{y}_k^*$ are the fitted (or pre-whitened or filtered) values of outputs for a given $(\xi_k, \boldsymbol{x}_k)$, i.e.,

$$\boldsymbol{y}_k^* = \frac{ex\,p\big(\widehat{m}(\xi_k, \boldsymbol{x}_k)\big)}{\omega_k} \boldsymbol{y}_k$$

Step 4: For any given fixed value of interest $(\boldsymbol{x}, \boldsymbol{y})$, run the desired DEA or FDH programs using the pre-whitened sample $\boldsymbol{S}^* = \{(\boldsymbol{x}_k, \boldsymbol{y}_k^*) : k = 1, \ldots, \tilde{n}\}$ to compute the SFDH or SDEA estimate of output oriented technical efficiency, denoted as $\widetilde{\delta_{SFDH}}(\boldsymbol{x}, \boldsymbol{y})$ or $\widetilde{\delta_{SDEA}}(\boldsymbol{x}, \boldsymbol{y})$, i.e.,

$$\widetilde{\delta_{SFDH}}(\boldsymbol{x}, \boldsymbol{y}) = max\,\{\delta | \boldsymbol{x}_k \le \boldsymbol{x}, \boldsymbol{y}_k^* \ge \delta \boldsymbol{y}, \forall (\boldsymbol{x}_k, \boldsymbol{y}_k) \in \boldsymbol{S}^*\}$$

$$\widetilde{\delta_{SDEA}}(\boldsymbol{x}, \boldsymbol{y}) = max\left\{\delta | \sum_{k=1}^{\tilde{n}} \zeta_k \boldsymbol{x}_k \le \boldsymbol{x}, \sum_{k=1}^{\tilde{n}} \zeta_k \boldsymbol{y}_k^* \ge \delta \boldsymbol{y}, \zeta_k \ge 0, k = 1, \ldots, \tilde{n}, \rho_1 \le \sum_{k=1}^{\tilde{n}} \zeta_k \le \rho_2\right\}$$

In turn, one can obtain the SFDH and SDEA estimates, for any fixed point $(\boldsymbol{x}, \boldsymbol{y})$, of the efficient frontier that obeys free disposability (or monotonicity) and, for the case of SDEA, convexity of $\boldsymbol{\Psi}$, using

$$\tilde{y}^{\partial}(\boldsymbol{x}, \boldsymbol{y}) = \tilde{\delta}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{y}$$

Bootstrap DEA

DEA is well known to produce a biased estimator of the frontier, and correspondingly a biased estimator of firm level inefficiency. The true frontier lies somewhere above the DEA estimated frontier. This leads to estimated efficiency scores that are overly optimistic (or in statistical parlance have an upward bias – with the estimated frontier having a downward bias). Simar & Wilson (1998, 2000) proposed a bootstrap algorithm to correct for this bias. This procedure was further enhanced by the theoretical underpinnings of Kneip, Simar & Wilson (2008). The general idea is that the known distribution of the difference between the estimated and bootstrapped efficiency scores is capable of mimicking the unknown distribution of the difference between the true and estimated efficiency scores. This distribution then facilitates estimation of both the bias and confidence intervals for the estimated individual efficiency scores.

Our main setup is the technology set Ψ, defined as

$$\Psi = \{(\boldsymbol{x}, \boldsymbol{y}) | \boldsymbol{x} \text{ } can\, produce\, \boldsymbol{y}\}.$$

Here we have p inputs $(\boldsymbol{x})$ and q outputs $(\boldsymbol{y})$. Output-oriented technical efficiency θ_j (Farrell 1957) for the point $(\boldsymbol{x}_j, \boldsymbol{y}_j)$ is defined as

$$F^O(\boldsymbol{x}_j, \boldsymbol{y}_j) = sup\{\theta_j | (\boldsymbol{x}_j, \boldsymbol{y}_j / \theta_j) \in \Psi\},$$

which is what is calculated by the DEA estimator using standard linear programming. At issue in practice is that Ψ is unknown (which then makes $F^O(\boldsymbol{x}_j, \boldsymbol{y}_j)$ unknown) and so we must estimate it. The standard approach to conducting statistical inference on $F^O(\boldsymbol{x}_j, \boldsymbol{y}_j)$ is to rely on the asymptotic theory found in Kneip et al. (2008) and to use bootstrapping. If the true technical efficiency score F^O (we drop the dependence on $(\boldsymbol{x}_j, \boldsymbol{y}_j)$ in what follows) comes from the data generating process $\mathcal{P}$ and the estimator $\widehat{F^O}$ comes from the known $\hat{\mathcal{P}}$, then for a **valid** bootstrap estimator $\widehat{F^{O^*}}$ we have

$$\left(\widehat{F^{O^*}} / \widehat{F^O} - 1\right) \Big| \hat{\mathcal{P}} \sim \left(\widehat{F^{O^*}} / F^O - 1\right) \Big| \mathcal{P}.$$

The consistent bootstrap of Kneip et al. (2008) uses subsampling across two steps. First, let $m = n^{\kappa}$ for $\kappa \in (0,1)$ where n is sample size and m is the size of the subsample (κ in this case controls how large or small the subsample is). The bootstrap is

- **Step 1**: Generate a bootstrap subsample of size m, $S_m^* = \{(\boldsymbol{x}_j^*, \boldsymbol{y}_j^*)\}_{j=1}^m$ by drawing randomly **with replacement** from the original sample data, $S_n^* = \{(\boldsymbol{x}_j, \boldsymbol{y}_j)\}_{j=1}^n$.

- **Step 2**: Apply the DEA estimator where the technology is defined for the subsample drawn in the previous step to obtain the bootstrap estimator $\widehat{F^{O^*}}$.

These two steps need to be repeated a large number of times (B). The bias-corrected DEA efficiency score is then given by $\widehat{F^O}_{BC} = \widehat{F^O} - \widehat{bias_B}$, where the bias is calculated as

$$\widehat{bias}_B = \frac{m^{2/(p+q+1)}}{n}\left(B^{-1}\sum_{b=1}^{B}\widehat{F_b^{O}}^* - \widehat{F^O}\right).$$

Essentially for each subsample we calculate the DEA estimator using that subsample only, then calculate efficiency scores from the full sample. We then calculate the average efficiency for each observation across all B bootstraps and subtract from this the original efficiency score (the one that used the full sample). This is then adjusted for the rate of the subsample size (to accord with theory) prior to being subtracted from the original efficiency score.

We can also calculate quantiles of this distribution to use to construct confidence intervals for the technical efficiency scores. Here we present the equal-tailed percentile approach to construct confidence intervals. For $\alpha \in (0,1)$, the equal-tailed quantiles for the mth subsample are $\delta_{\alpha/2,m}$ and $\delta_{1-\alpha/2,m}$, such that

$$P\left(m^{2/(p+q+1)}\left(\widehat{F^O}^*/\widehat{F^O} - 1\right) \leq \delta_{\alpha/2,m}\right) = \alpha/2$$

$$P\left(m^{2/(p+q+1)}\left(\widehat{F^O}^*/\widehat{F^O} - 1\right) \leq \delta_{1-\alpha/2,m}\right) = 1 - \alpha/2.$$

These quantiles lead to the equal-tailed $100(1-\alpha)\%$ confidence interval for F^O where the bounds are

$$\left[\frac{\widehat{F^O}}{1 + n^{-2/(p+q+1)}\delta_{1-\alpha/2,m}}, \frac{\widehat{F^O}}{1 + n^{-2/(p+q+1)}\delta_{\alpha/2,m}}\right].$$

Two natural questions are the size of the subsample and the number of bootstraps to take. The number of bootstraps can easily be handled by simply increasing B until a desired level of accuracy is retained. There are several theoretical results for the optimal size of the subsample.

Annex C – List of references

- Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. American Journal of Political Science, 59(2), 495–510.

- Abadie, Alberto. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. Journal of Economic Literature, 59(2), 391–425.

- Abadie, Alberto, Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of CaliforniaAs tobacco control program. Journal of the American Statistical Association, 105(490), 493–505.

- AER. (2022). Annual Benchmarking Report - Electricity distribution network service providers, November 2022. https://www.aer.gov.au/system/files/AER%20-%202022%20Annual%20Benchmarking%20Report%20-%20Electricity%20distribution%20network%20service%20providers%20-%20November%202022_2.pdf

- AER. (2022). Annual Benchmarking Report - Electricity transmission network service providers, November 2022. https://www.aer.gov.au/system/files/AER%20-%202022%20Annual%20Benchmarking%20Report%20-%20Electricity%20transmission%20network%20service%20providers%20-%20November%202022_0.pdf

- Aigner, D. J., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production functions. Journal of Econometrics, 6(1), 21–37.

- Andor, M. A., & Parmeter, C. (2017). Pseudolikelihood estimation of the stochastic frontier model. Applied Economics, 49(55), 5651-5661.

- Andor, M. A., Parmeter, C., & Sommer, S. (2019). Combining uncertainty with uncertainty to get certainty? Efficiency analysis for regulation purposes. European Journal of Operational Research, 274(1), 240-252.

- 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). Journal of Productivity Analysis, 41, 85-109.

- 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.

- Banker, R. D., Charnes, A., Cooper, W. W., Swarts, J., & Thomas, D. (1989). An introduction to data envelopment analysis with some of its models and their uses. Research in Governmental and Nonprofit Accounting, 5(1), 125-163.

- Banker, R. D., & Maindiratta, A. (1992). Maximum likelihood estimation of monotone and concave production frontiers. Journal of Productivity Analysis, 3(4), 401-415.

- Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The Augmented Synthetic Control Method. Journal of the American Statistical Association.

- BET/IFHT. (2018). Gutachten zur Referenznetzanalyse für die Betreiber von Übertragungsnetzen im Auftrag der Bundesnetzagentur. https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Netzentgelte/Strom/GutachtenReferenznetzanalyse.pdf?__blob=publicationFile&v=1

- Biddle, J. (2012). Retrospectives: The introduction of the Cobb-Douglas regression. Journal of Economic Perspectives, 26(2), 223-236.

- Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). Multilateral comparisons of output, input, and productivity using superlative index numbers. The Economic Journal, 92(365), 73-86.

- CEPA (2018, March), Review of the RIIO framework and RIIO-1 performance, Final Report for Ofgem.

- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of Decision-making units. European Journal of Operational Research, 2(6), 429–444.

- Colombi, R., Kumbhakar, S. C., Martini, G., & Vittadini, G. (2014). Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. Journal of Productivity Analysis, 42, 123-136.

- CRU. (2019). Irish Water Revenue Control, Revenue Control 3 (2020-2024). https://www.wateradvisorybody.ie/wp-content/uploads/2020/04/Irish-Water-Revenue-Control-3-Decision-Paper.pdf

- Cunningham, M., Hirschberg, J., & Quack, M. (2022). Economic Benchmarking Results for the Australian Energy Regulator's 2022 DNSP Annual Benchmarking Report. https://www.aer.gov.au/system/files/Quantonomics%20-%20Benchmarking%20results%20for%20the%20AER%20-%20Distribution%20-%20November%202022.pdf

- Debreu, G. (1951). The coefficient of resource utilization. Econometrica, 19(3), 273–292.
- Du, P., Parmeter, C. F., & Racine, J. S. (2013). Nonparametric kernel regression with multiple predictors and multiple shape constraints. Statistica Sinica, 23(3), 1347–1371.
- E-Bridge. (2012). Study of the feasibility of determining TenneT's cost efficiency via process benchmarking, Study for Nederlandse Mededingingautoritie (NMa).
- Economic Insights. (2015). Response to Consultants' Reports on Economic Benchmarking of Electricity DNSPs. https://www.aer.gov.au/system/files/Economic%20Insights%20-%20Response%20to%20consultants%20%20reports%20on%20AER%20economic%20benchmarking%20-%20April%202015_1.PDF
- E-Control. (2013). Regulierungssystematik für die dritte Regulierungsperiode der Stromverteilernetzbetreiber 1. Jänner 2014 - 31. Dezember 2018. https://www.e-control.at/documents/1785851/1811528/Entscheidungen-der-Regulierungsbehoerde-Ausgestaltung-3te-Periode-Strom.pdf/225b49e0-6534-40e4-afa1-97d83f8edbde?t=1413905499198
- E-Control. (2018). Regulierungssystematik für die vierte Regulierungsperiode der Stromverteilernetzbetreiber 1. Jänner 2019 - 31. Dezember 2023. https://www.e-control.at/documents/1785851/1811582/Regulierungssystematik_4_Periode_STROM_Dez+2018.pdf/a413df20-00b2-9dca-ba43-4ae52754b27e?t=1562139961156
- E-Control. (2022). Regulierungssystematik für die vierte Regulierungsperiode der Gas-Verteilernetzbetreiber 1. Jänner 2023 - 31. Dezember 2027. https://www.e-control.at/documents/1785851/1811582/02_Finale+Regulierungssystematik+4_RP.pdf/40fcc26d-253d-0533-2d74-3774dce4e341?t=1668673860094
- Fan, Y., Li, Q., & Weersink, A. (1996). Semiparametric estimation of stochastic production frontier models. Journal of Business & Economic Statistics, 14(4), 460-468.
- Färe, R., & Grosskopf, S. (1983). Measuring Output Efficiency. European Journal of Operational Research, 13(1), 173–179.
- Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society, Series A General, 120(3), 253–281.
- Frontier Economics. (2010). RPI-X@20: The future role of benchmarking in regulatory reviews. https://www.ofgem.gov.uk/sites/default/files/docs/2010/05/rpt-benchmarking.pdf

- Frontier Economics/Consentec. (2012). The potential application of reference network modeling to TenneT, Feasibility study for NMa.

- Frontier Economics/TU Berlin. (2019). Effizienzvergleich Verteilernetzbetreiber Gas (3.Regulierungsperiode). https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Netzentgelte/Gas/EffizienzvergleichVNB/Gutachten_Effizienzvergleich_final.pdf?__blob=publicationFile&v=2

- Frontier Economics/TU Berlin. (2023). Effizienzvergleich Verteilernetzbetreiber Gas (4. Regulierungsperiode). https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Netzentgelte/Gas/EffizienzvergleichVNB/Gutachten_Entwurf2_4RP.pdf?__blob=publicationFile&v=2

- Hall, P., & Huang, H. (2001). Nonparametric kernel regression subject to monotonicity constraints. The Annals of Statistics, 29(3), 624–647.

- Hanock, G. (1971). CRESH production functions. Econometrics, 29(5). https://www.jstor.org/stable/1909573

- Jamasb, T., & Pollitt, M. (2007). Reference Models and Incentive Regulation of Electricity Distribution Networks: An Evaluation of Sweden's Network Performance Assessment Model (NPAM). EPRG Working Paper.

- Jondrow, J., Lovell, C. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. Journal of Econometrics, 19(2-3), 233-238.

- Kneip, A., Simar, L., & Wilson, P. W. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. Econometric Theory, 24(6), 1663-1697.

- Koener, R., & Hallock, K. F. (n.d.). Quantile regression: an introduction. http://www.econ.uiuc.edu/~roger/research/intro/rq3.pdf

- Kuosmanen, T. (2012). Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model. Energy Economics, 34, 2189-2199.

- Kuosmanen, T., Johnson, A., & Saastamoinen, A. (2015). Stochastic nonparametric approach to efficiency analysis: A unified framework. In J. Zhu (Ed.), Data Envelopment Analysis (pp. 191–244). Springer Science.

- Kuosmanen, T., Kuosmanen, N., & Dai, S. (2022). Kohtuullinen muuttuva kustannus sähkön jakeluverkkoyhtiöiden valvontamallissa: Ehdotus tehostamiskannustimen kehittämiseksi 6. ja 7. valvontajaksoilla vuosina 2024-2031 (Reasonable variable cost in the electricity distribution system operator control model: Proposal for the development of the efficiency incentive 6. and 7. for the monitoring periods 2024-2031); Report for EMV.

- Lawrence, D., Fallon, J., Cunningham, M., Zelenyuk, V., & Hirschberg, J. (2017). Topics in efficiency benchmarking of energy networks: Choosing the model and explaining the results. https://www.acm.nl/sites/default/files/documents/2020-08/economic-insights-topics-in-efficiency-benchmarking-of-energy-networks-choosing-the-model-and-explaining-the-results.pdf

- Li, Z., Liu, G., & Li, Q. (2017). Nonparametric Knn Estimation with Monotone Constraints. Econometric Reviews, 36(6–9), 988–1006.

- Martins-Filho, C. B., & Yao, F. (2015). Semiparametric stochastic frontier estimation via profile likelihood. Econometric Reviews, 34(4), 413–451.

- Noh, H. (2014). Frontier estimation using kernel smoothing estimators with data transformation. Journal of the Korean Statistical Society, 43, 503–512.

- Ofgem. (2014). RIIO-ED1: Final determinations for the slowtrack electricity distribution companies: Final decision. https://www.ofgem.gov.uk/sites/default/files/docs/2014/11/riio-ed1_final_determination_overview_-_updated_front_cover_0.pdf

- Ofgem (2020). RIIO-2 Final Determinations for Transmission and Gas Distribution network companies and the Electricity System Operator. https://www.ofgem.gov.uk/publications/riio-2-final-determinations-transmission-and-gas-distribution-network-companies-and-electricity-system-operator.

- Ofgem. (2021). Engineering Justification Papers for RIIO-ED2. https://www.ofgem.gov.uk/sites/default/files/docs/2021/02/riio_ed2_engineering_justification_paper_guidance.pdf

- Ofgem (2022). RIIO-ED2 Final Determinations. https://www.ofgem.gov.uk/publications/riio-ed2-final-determinations.

- Ofgem. (2022). RIIO-ED2 Final Determinations Core Methodology Document. https://www.ofgem.gov.uk/sites/default/files/2022-11/RIIO-ED2%20Final%20Determinations%20Core%20Methodology.pdf

- Ofgem. (2022). RIIO ED2 Final Determinations Overview document. https://www.ofgem.gov.uk/sites/default/files/2022-11/RIIO-ED2%20Final%20Determinations%20Overview%20document.pdf

- Parmeter, C. F. (2023). Is it MOLS or COLS? In P. Macedo, V. Moutinho, & M. Madaleno (Eds.), Advanced Mathematical Methods for Economic Efficiency Analysis (pp. 229-249). Springer.

- Parmeter, C. F., & Racine, J. S. (2012). Smooth Constrained Frontier Analysis. In X. Chen & N. E. Swanson (Eds.), Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White Jr. (pp. 463–489). Springer-Verlag.

- Racine, J. S., & Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous Data. Journal of Econometrics, 119(1), 99–130.

- Seijo, E., & Sen, B. (2011). Nonparametric least squares estimation of a multivariate convex regression function. Annals of Statistics, 39(3), 1633–1657.

- Sigma-Hat Economics. (2014). Tehostamiskannustin sähkön jakeluverkkoyhtiöiden valvontamallissa: Ehdotus Energiaviraston soveltamien menetelmien kehittämiseksi neljännellä valvontajaksolla 2016 - 2019 (The efficiency incentive in the electricity distribution system operator control model: Proposal for the development of the methodology applied by the Energy Agency in the fourth control period 2016 - 2019), p. 22, Final report.

- Simar, L. (2007). How to improve the performances of DEA/FDH estimators in the presence of noise. Journal of Productivity Analysis, 28(2), 183–201.

- Simar, L., & Zelenyuk, V. (2011). Stochastic FDH/DEA estimators for frontier analysis. Journal of Productivity Analysis, 36(1), 1-20.

- Simar, L., Van Keilegom, I., & Zelenyuk, V. (2017). Nonparametric Least Squares Methods for Stochastic Frontier Models. Journal of Productivity Analysis, 47(3), 189–204.

- Simar, Wilson, P. W. (1998). Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. Management Science, 44(1), 49–61.

- Simar, Wilson, P. W. (2000). A general methodology for bootstrapping in non-parametric frontier models. Journal of Applied Statistics, 27(6), 779–802.

- Simar, L., & Wilson, P. W. (2009). Inferences from cross-sectional, stochastic frontier models. Econometric Reviews, 29(1), 62–98.

- Simar, L., & Wilson, P. W. (2010). Inferences from cross-sectional, stochastic frontier models. Econometric Reviews, 29(1), 62–98.

- Simar, L., & Wilson, P. W. (2013). Estimation and Inference in Nonparametric Frontier Models: Recent developments and perspectives. Foundations and Trends in Econometrics, 5(2), 183–337.

- Simar, L., & Wilson, P. W. (2015). Statistical Approaches for Nonparametric Frontier Models: A Guided Tour. International Statistical Review, 83(1), 77–110.

- Sumicsid. (2019). CEER-TCB18 Pan-European cost-efficiency benchmark for electricity transmission system operators, Main Report for CEER. https://www.ceer.eu/documents/104400/6742745/TCB18_final_report_elec_190717.pdf/559c7df0-9cf3-2153-07bd-855bdf9a6a13

- Sumicsid. (2019). Pan-European cost-efficiency benchmark for gas transmission system operators, Report for CEER. https://www.ceer.eu/documents/104400/-/-/90707d6c-6da8-0da2-bce9-0fbbc55bea8c

- Swiss Economics/Sumicsid/IAEW. (2019). Effizienzvergleich Verteilernetzbetreiber Strom der dritten Regulierungsperiode; Report foru. https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Netzentgelte/Strom/Effizienzvergleich_VNB/3RegPer/Gutachten_EVS3_geschw.pdf?__blob=publicationFile&v=3