Quality measures in cost benchmarking

Prepared for
the Danish Competition and Consumer Authority (DCCA)

Final

www.oxera.com

oxera
compelling economics

Contents

Figures and tables

Executive summary

The Danish Competition and Consumer Authority commissioned Oxera to review other regulators' experience of including quality measures within cost benchmarking, with a particular focus on data envelopment analysis (DEA) and stochastic frontier analysis (SFA).

Oxera approached ten regulators for interviews across a range of sectors, including water, energy and healthcare. We were able to carry out seven such interviews. We also reviewed published and unpublished documents detailing the regulator's approaches to the issue, and provide a challenge to some of the perspectives based on work undertaken by us and our academic associates.

List of regulators approached and interviewed

Regulator	Country	Sector	Interviewed
ACM	The Netherlands	Electricity and gas distribution	✓
ARERA	Italy	Water and energy	
Bundesnetzagentur	Germany	Energy	
E-Control	Austria	Electricity distribution	✓
Energiavirasto	Finland	Electricity distribution	✓
ERSAR	Portugal	Water and wastewater	✓
Monitor	UK	Healthcare	
NVE	Norway	Electricity distribution	✓
Ofgem	UK	Gas distribution	✓
Ofwat	UK	Water and wastewater	✓

Source: Oxera. We specify more specific sectors where relevant depending on the interview.

Some key themes that emerged from this review are outlined below. We note that the views expressed are those of the interviewees, and not necessarily of the regulatory authority in question.

Motivation for including quality in benchmarking

In order to provide incentives for utilities to serve their customers at an efficient cost level, regulators benchmark utilities' costs against one another to emulate a competitive market. Typically, techniques such as econometric methods (e.g. panel data estimators, SFA) and DEA are used to identify the relationship between inputs (e.g. total expenditure, or TOTEX) and outputs at the industry frontier. This is then used to identify efficient costs for each company, which forms the basis for setting revenue caps.

An important drawback of a cost benchmarking approach that accounts for only outputs/activities and costs is that it does not provide incentives to maintain an optimal quality of service reflecting the stakeholder needs. This raises the concern that companies may achieve low costs by providing a poor quality of service or ignoring asset condition/health. For example, consider two water utilities that are required to provide the same outputs (e.g. serve the same number of households). If one utility spends less on maintenance, it would be identified as more efficient because it serves the same number of households at a lower cost. However, lower maintenance spending could lead to poorer asset condition, and therefore more frequent and longer service interruptions. Customers may prefer more reliable services and for companies to spend more (as long as the money is efficiently incurred) on achieving higher levels of quality.

Regulators have sought to overcome these issues by including quality measures within cost benchmarking. From our discussions with regulators, we understand that there are some key building blocks when designing a quality-integrated cost benchmarking approach:

- **choice of quality measures**—which quality measures should we use within cost benchmarking? What are the practical considerations when constructing these measures, and the incentive properties of using different quality measures?
- **use within benchmarking—**how should quality measures be included in cost benchmarking, and what are the specific considerations when doing so within DEA and econometric methods?
- **translating cost benchmarking results into a revenue cap**—do the results of a quality-integrated benchmarking approach need to be interpreted differently, and how should these results be translated into a revenue cap?

Choice of quality measures

We split the different quality measures into two broad categories. These are non-monetised measures of quality (e.g. the number of minutes of service interruptions), and monetised measures of quality (e.g. the value that customers place on avoiding service interruptions).

An important determinant of the choice of quality measures are the incentives that a regulator wants to provide to companies. NVE and E-Control's experts[1] stated that they wanted to incentivise companies to choose the optimal level of quality by trading off the costs of an improvement in quality against the benefits to customers. They therefore construct a monetised measure of quality (in their case, the value of lost load, or VOLL) based on stated preference surveys.[2] This monetised measure represents the benefits to customers of a marginal improvement in quality. When included within cost benchmarking in an appropriate way (discussed further below), regulators can incentivise companies to trade off the benefit to customers of an improvement in quality against the costs of implementing it.

However, NVE and E-Control explained that designing the survey was an involved process. NVE explained that its approach involved six customer categories, where a representative group of customers within each category were asked to provide estimates of their economic loss if they experienced an outage of various durations. NVE explained the challenges in carrying out such studies. In particular, when customers are asked hypothetical questions, it may not be possible to accurately value the consequences of an outage, especially if customers have not experienced extended outages before. Furthermore, customers may provide strategic responses since the results may indirectly influence their tariff or reliability of supply.

NVE made use of numerous pilot studies and focus groups, with re-designs of the survey at each iteration. An additional consideration is that the value of lost

[1] The Norwegian and Austrian energy regulators, respectively.
[2] Stated preference surveys are used to quantify the value that customers place on different measures based on customer responses to a survey. For example, the survey might ask what a customer is willing to pay to avoid a service interruption lasting one, two or four hours.

load may differ by customer category (e.g. industry versus residential), as well as within each category (some residential customers may experience a higher cost of interruptions than others). Furthermore, large sample sizes are required for such analysis. In its study of residential customers, NVE explained that, even with an initial size of about 100,000 possible respondents, the survey was submitted to a representative group of only 4,400. Of these, approximately 2,200 responded and only 1,350 responses provided reasonable results.[3] As a possible approach to mitigate the resource impact on the regulator (at least, at the valuation stage), we note that Ofwat places the responsibility of valuing different quality metrics on the water and wastewater companies.[4] Moreover, Including a monetised measure of quality such as VOLL in TOTEX involves certain strong assumptions that require validation.

Using non-monetised quality measures in benchmarking has been explored in less depth among the regulators that we interviewed. Given the potentially large resource investment required to construct monetised quality measures, non-monetised measures may be an attractive alternative. Nonetheless, regulators have raised a number of considerations about using non-monetised measures.

Ofwat explained that various quality measures could be used within cost benchmarking. However, many of these are not primary cost drivers, unlike other variables such as scale. When included alongside these primary cost drivers, the estimated relationship between quality measures and costs can be quite weak. This may lead to results that are not based on operationally intuitive relationships between costs and quality. As a result, Ofwat suggested selecting a few quality measures that have strong relationships with costs and constructing a composite variable by weighting them together as a possibility.

Another issue raised by E-Control is that quality measures included within DEA benchmarking should reflect company scale. For example, rather than using a measure such as the average number of minutes of service interruption per customer, using the total number of minutes of service interruptions may be more appropriate provided that companies do not end up being identified as efficient solely because of the value on a particular factor.[5] This is because DEA assumes that if there are two companies with different costs, outputs and quality levels, a hypothetical company formed from a weighted average of the costs, outputs and quality levels of the two companies should also be feasible.[6] In order for this assumption to hold, outputs and quality measures reflecting company scale should be used.

Use within benchmarking

Variables included within cost benchmarking can be broadly split into two types: inputs and outputs. In a regulatory context, cost benchmarking methods such as SFA and DEA estimate minimum input levels (e.g. TOTEX[7]) and for a given level of outputs (e.g. number of households served) for each company. The ratio of the estimated minimum inputs to the actual inputs used is a measure of each company's efficiency. The table below shows, for each regulator that has or is considering quality-integrated benchmarking, the

[3] There may be inconsistencies that invalidate a particular response. For example, in one part of the survey, a respondent might claim that they do not experience any costs when there are service interruptions, but responses on other parts of the survey may suggest otherwise.
[4] Ofwat provides quality incentives outside of its cost benchmarking models rather than within them.
[5] To mitigate this situation, appropriate weight restrictions can be introduced in the model.
[6] Due to the assumption of convexity.
[7] Total expenditure, or TOTEX, is composed of operating expenditure and capital expenditure.

estimation methods used and the inputs and outputs considered within the model specification.

Summary of inputs and outputs used by regulators who have considered or implemented quality-integrated benchmarking

	Method	Input	Output
NVE	DEA	TOTEX + VOLL	Number of customers Number of substations Length of high-voltage network[8]
Energiavirasto	2008 – 2011: average of DEA and SFA 2012 – 2019: StoNED	2008 – 2011: TOTEX + VOLL 2012 – 2015: OPEX + 0.5*VOLL 2016 – 2019: OPEX	Distributed energy Number of customers Network length 2016 – 2019: VOLL
E-Control	DEA and MOLS[9]	TOTEX + VOLL	Weighted sum of low, medium and high voltage modelled network lengths[10] Peak load at grid levels 4-7 Peak load at grid levels 6-7[11]
Ofwat	Random effects[12]	BOTEX (the sum of OPEX, capital maintenance and growth costs)	Various variables capturing scale, topography, density and treatment complexity[13]

Source: Oxera, based on regulatory and other published documents.

Note: E-Control and Ofwat have not officially implemented quality in their cost benchmarking approach.

Within DEA benchmarking, NVE, E-Control[14] and Energiavirasto[15] have considered including VOLL as an input directly within TOTEX. This is driven by the principle that both TOTEX and VOLL are costs, where VOLL is the cost to customers due to power interruptions. Therefore, DSOs/TSOs should minimise society's costs and not just their own costs. In other words, incentives should be equal for both reducing costs *and* improving service quality. By creating a composite variable combining both monetised quality and TOTEX, companies will be rewarded in the same way whether they reduce costs by €1 or reduce VOLL by €1. This provides optimal incentives because the company would be encouraged to improve quality only if the benefit to customers outweighed the costs of doing so.

This is in contrast to the approach where quality is included separately as an output (or as an input) in a DEA model. Intuitively, DEA works by calculating the ratio of a weighted average of outputs to the weighted average of inputs. In

[8] Amundsveen, R., & Kvile, H. M. (2016), 'Balancing incentives: The development and application of a regulatory benchmarking model', In Productivity and Efficiency Analysis (pp. 233-247). Springer, Cham, p.242.
[9] Modified ordinary least squares adjusts the constant estimated by ordinary least squares for the presence of inefficiency in the error term.
[10] E-Control uses a modelled network length rather than the actual network length, stating that actual network length is controllable by the company. E-Control argue that if actual network lengths are used, there is an incentive to oversize existing grid structures and/or not to disassemble any lines. See E-Control, 'Regulatory Regime for the Third Regulatory Period: Electricity Distribution System Operators. 1 January 2014 – 31 December 2018', p. 33.
[11] Grid levels refer to the transformer voltage level.
[12] Random effects accounts is similar to OLS, but accounts for the panel structure of the data.
[13] See Ofwat (2019), 'PR19 final determinations. Securing cost efficiency technical appendix', December, appendix 2.
[14] E-Control has not implemented quality-dependent cost benchmarking.
[15] The Finnish regulator now includes quality as an output rather than as an input.

its most basic form, DEA determines the weights by giving each company the best possible view of its efficiency. Consequently, a weight of zero could be placed on quality, and therefore companies could appear to be efficient by focusing on outputs other than quality. One way to avoid this issue would be to specify appropriate weight restrictions that link the weight on quality to one or more weights placed on other input/output measures. Determining these weight restrictions will require additional research and can be informed by econometric methods, operational rationale and expert judgement. The issue of a DEA model ignoring a particular output/input or placing unusually large or small values on some of the input/output factors is not specific to quality, and weight-restricted DEA models are considered in regulatory applications.[16,17]

Energiavirasto has since moved away from its approach of including VOLL as an input within TOTEX.[18] Due to extreme weather events in 2010 and 2011, energy distributors in Finland suffered significant outages, leading to very high costs of outages for some companies and counterintuitive modelling results. For one company, the regulator found that the value of outages was about 15 times the value of TOTEX (excluding VOLL). As a result, Energiavirasto now includes VOLL as an undesirable output alongside three other output variables rather than as an input, and uses OPEX as its input. It considers that: (i) including it as an undesirable output produces more intuitive results;[19] and (ii) power outages are better modelled as an undesirable output rather than as an input as outages are not an essential factor of production used in producing outputs.

While Ofwat does not include quality measures in its cost benchmarking as part of the recently completed price review (PR19), it has considered their inclusion as an output within an econometric model in the past, and has discussed with us the challenges of doing so.[20]

- First, cost drivers included in an econometric model should ideally be outside of management control (i.e. exogenous). For example, water companies need to make connections when requested by property owners or occupiers, and so the number of connections is not controllable by the company and is appropriate for including within benchmarking. However, many quality measures can be influenced by companies (i.e. they are endogenous). To take service interruptions as an example, companies can influence their frequency and duration based on the level of maintenance they choose to carry out.
- Second, Ofwat considers that there may be a non-linear relationship between quality and costs. Companies with a low level of quality might find

[16] There has been some exploration of the use of weight restrictions on geographical variables using NVE's benchmarking model. For further details, see Bjørndal. E. and Bjørndal. M, 'Weight Restrictions on Geography Variables in the DEA Benchmarking Model for Norwegian Electricity Distribution Companies'.
[17] The 2012 Pan-European TSO Benchmarking study has utilised weights on DEA. See Frontier, Consentec and Sumicsid (2013), 'E3GRID2012 — European TSO Benchmarking Study', July, for further details.
[18] Energiavirasto also now uses the StoNED method, which in principle combines the merits of both DEA and SFA. However, in its calculation of the efficiency score, this approach shares similarities with DEA in that it works by weighting outputs in such a way that it gives a best view of each company's efficiency score.
[19] For companies that are severely affected by storms, Energiavirasto explained that the StoNED method gave them a positive shadow price, allowing them to collect more costs, while for other companies the shadow price is negative, providing an incentive to reduce outages to improve their efficiency. Nonetheless, there may be limitations to such an approach. For further details, see section 3.3.2.
[20] Ofwat's views on cost-quality integration in the ongoing CMA PR19 water price redeterminations can be found in Ofwat (2020), 'Reference of the PR19 final determinations: Response to Yorkshire Water's statement of case', May. Ofwat (2020), 'Reference of the PR19 final determinations: Cost efficiency—response to common issues in companies' statements of case', May.

it cheaper to improve quality relative to another company that is at the frontier in its quality of service.

- Third, since quality measures are not primary cost drivers, it can be difficult to obtain models with statistically significant and intuitive coefficients. Therefore, the benchmarking results may not be based on operationally intuitive relationships between costs and cost drivers.

Nonetheless, Ofwat considers that some of these issues may be addressed by specifying flexible functional forms or using a composite indicator.[21] In the ongoing appeal by Yorkshire Water against Ofwat's final determinations in PR19, we have developed quality-integrated cost models, building on Ofwat's cost models, that provide incentive-compatible outputs.[22] These are currently being reviewed by the UK Competition and Markets Authority (CMA).

We note that Monitor, the erstwhile regulator of healthcare services in the UK, considered SFA to set tariffs where it included a number of service quality measures in its cost benchmarking approach. This is based on a survey distributed to NHS providers with questions to measure patient satisfaction. It then constructs a quality variable based on the percentage of respondents that answered 'strongly agree' to different questions on relevant quality of service questions. However, there are many different questions which could be used to measure quality. As a result, Monitor used principal component analysis[23] to construct a single composite index that captures the most variation possible across a range of questions. This quality index is then included as an output in its SFA model. Monitor notes that it obtained statistically insignificant coefficients on its quality index.[24] There is of course a need for a balanced consideration for the inclusion of factors (whether quality or non-quality measures) in a cost model, and statistical significance alone should not dictate the choice.

On the issue of endogeneity, Ofwat considers that, if a cost driver is able to account for important differences between companies, this would need to be balanced against any risk that a cost driver might be endogenous. It should also be noted that ignoring important factors in a cost model can also result in endogeneity.[25] In contrast, E-Control's experts included a monetised measure of quality within TOTEX when developing its econometric approach, with the aim of avoiding the issue of endogeneity. It states that this is another reason why it prefers including quality within TOTEX, in addition to the harmonising of incentives as discussed above. Nonetheless, this approach is equivalent to including quality as a cost driver under additional constraints. As a result, this places certain assumptions on the relationship between costs and quality that will need justification.

Translating cost benchmarking results into a revenue cap

[21] For example, Ofwat has included density and its squared term. Estimation results suggest that density has a 'U'-shaped impact on costs. This captures the idea that serving very sparse areas incurs additional costs of maintaining large networks, while serving very dense areas incurs additional costs of serving highly urban areas.
[22] Oxera (2020), 'Quality measures in cost benchmarking', May. Oxera (2020), 'Integrating cost and outcomes', prepared for Yorkshire Water Services, March.
[23] Principal component analysis is a statistical technique that constructs weighted combinations of a set of variables to capture the maximum amount of variation possible in a smaller set of variables, called principal components. The first principal component captures a higher proportion of the total variation compared to all the other principal components.
[24] Monitor (2016), '2016/17 National Tariff Payment System: A consultation notice. Annex B5: Evidence on efficiency for the 2016/17 national tariff', 11 February, section 4.1, paragraph 1.
[25] Oxera (2020), 'Quality measures in cost benchmarking', May. Oxera (2020), 'Integrating cost and outcomes', prepared for Yorkshire Water Services, March.

Once the results of a benchmarking exercise have been obtained, they need to be translated into a revenue cap. Whether the benchmarking results need to be treated differently depends on whether monetised quality has been included directly within TOTEX as an input. If this is the case, the predictions from the benchmarking model include not only the efficient level of costs, but also the monetised quality measure. Efficiency scores therefore cannot be interpreted in the usual way (as the ratio of efficient costs to actual costs).[26] As a result, NVE and Energiavirasto (when it included monetised quality in TOTEX) have explained that they calculate revenue caps in a different way to when quality is excluded from TOTEX.

NVE accounts for this within its calculation of the revenue cap by treating the VOLL as if it were the company's own costs. In practice, this is implemented by:

- using a single input (TOTEX + VOLL) in DEA benchmarking, which then determines the revenue cap. This means that reducing TOTEX by €1 or VOLL by €1 affects revenue caps in the same way;
- subtracting VOLL from the revenue cap.

As a result, each company's profits (outperformance) can be represented by

$$Profit = Revenue\ cap - (TOTEX + VOLL)$$

This means that reducing TOTEX by €1 or VOLL by €1 has the same effect on outperformance. Furthermore, customers would also be indifferent between an increase in TOTEX by €1 and a decrease in VOLL by €1. Therefore, this approach provides equal incentives to companies to reduce costs and to improve quality, which provides companies with the highest profit while providing the lowest costs for their customers.

While E-Control have not incorporated quality within its current regulatory framework, they describe that their ideal framework would consist of compensation payments to each network user who experiences a power outage with the VOLL. In order to reimburse operators for the VOLL, the regulatory cost base would include both VOLL and TOTEX. Network operators might receive a 'budget' for compensation payments for the first regulatory period when quality-integrated cost benchmarking is applied, or network operators could fund the payments in advance. Subsequently, they can either outperform their targets (with a quality improvement) or observe a lower return if the 'budget' is insufficient due to a quality deterioration.

However, if quality is not included in TOTEX, but is instead included as an output, benchmarking results can be translated into a revenue cap in the usual way.

Stakeholder reactions

Companies have raised a number of issues in response to quality-integrated benchmarking in different regulatory regimes.

- **Impact on company allowances and potential lobbying**—changing a model specification could result in winners and losers (i.e. companies that

[26] That is, efficiency scores are the ratio of efficient costs to actual costs in the standard case. However, if VOLL has been added to costs, efficiency scores should be interpreted as the ratio of (efficient costs + VOLL) to (actual costs + VOLL). One may interpret these efficiency scores as socioeconomic efficiency rather than business efficiency—i.e. efficiency that is based on society's costs rather than only on business costs.

benefit or suffer from the inclusion of a variable in the model). Indeed, these updated specifications may represent a more accurate assessment of each company's efficient costs if the impact of quality is captured. However, this creates the potential for different company reactions. E-Control published a working paper in 2014 with the proposal to integrate quality of service in cost benchmarking. Analysis revealed that, while most companies saw an improvement in their estimated efficient costs, some companies could challenge the adoption of quality-dependent cost benchmarking.

- **External factors affecting service quality performance—**in court cases involving NVE and Energiavirasto, companies have argued that interruptions are due to factors outside their control. Nonetheless, for Energiavirasto, the court found that company actions can still significantly influence outages. For NVE, the court concluded that the financial impact of the events on which the court case was based[27] tended to be relatively small. Since it did not pose a material risk to the financial health of the company, the company's claim was rejected.

  NVE emphasises that granting exceptions can remove or dilutes the economic incentives to perform well. NVE notes that, even if interruptions are caused by incidents that are outside a company's control, the *consequences* for consumers can still be controlled with appropriate planning and investments.[28]

  We note that some of these issues may need to be treated on a case-by-case basis. For example, a number of regulators including Ofwat and Ofgem allows companies to make cost adjustment claims if it believes that its modelling approach does not adequately account for unique exogenous factors affecting their costs.

We have also considered stakeholder reactions to quality regulation that falls outside of cost benchmarking, as these may also be applicable to the incorporation of quality within cost benchmarking.

- **Level of risk—**Energiavirasto has a separate quality incentive outside of cost benchmarking. This quality incentive consists of using the average realised outage costs over two previous regulatory periods (i.e. a period of eight years) as the reference level of outage costs. The difference between the reference level and actual level of outage costs is then used to adjust company profits. Energiavirasto applies a ceiling on the magnitude of the quality incentive, where the impact of the quality incentive is limited to a certain percentage of profits.

  Companies raised concerns when Energiavirasto increased the ceiling on the magnitude of the quality incentive, arguing that this represented an increase in risk that was not compensated for in other parts of the regulatory framework.[29] While the court found that the impact of raising the ceiling was not large enough to significantly increase risk to companies, this experience highlights the importance of examining the amount of risk that a regulatory approach places on companies.

[27] These were extreme weather events, specifically tornadoes, which NVE explained were very uncommon in Norway.
[28] For example, mitigating the consequences of incidents may occur before the incident (maintenance, controlling power flow, redundancy investments, contingency planning), during the incident (identifying faults and deploying crew and equipment), and after the incident (repairs, strategy, prioritising customers with high VOLL).
[29] In addition to quality within cost benchmarking models, it also has quality incentives within the wider regulatory framework.

- **Strength of incentives—**the Dutch ACM[30] regulates quality using a Q-element, which is a penalty or reward based on a company's realised level of quality relative to a benchmark.

  E-Control stated that its minimum quality requirements[31] had not been effective because violations of these minimum standards had not led to financial penalties.

  These experiences suggest that the strength of quality incentives needs to be considered carefully. Nonetheless, we note that, despite Ofgem providing relatively small financial incentives (around 0.5% of TOTEX), it still considers that these have been sufficient in maintaining a high level of customer satisfaction in the gas sector based on the survey evidence that it has collected.

[30] The Dutch energy regulator.
[31] The level of SAIDI (the average outage duration for each customer served) is regulated, such that it should not exceed the average over the last three years.

1 Introduction

The Danish Competition and Consumer Authority (DCCA) has commissioned Oxera to review other regulators' experiences of integrating quality performance within cost benchmarking models, with a particular focus on two frontier-based estimation approaches: Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA).

In its efficiency assessment of Danish water and wastewater utilities, the DCCA uses both DEA and SFA. To determine a final efficiency score, the DCCA takes the maximum of the DEA and SFA scores (known as the 'best-of-two' approach). The DCCA is exploring the possibility of integrating quality measures within its cost benchmarking of water and wastewater utilities, with a particular focus on DEA.

There can be significant benefits to incorporating quality directly within benchmarking. By benchmarking costs and quality across different companies, companies are encouraged to provide quality at efficient levels of costs. Furthermore, depending on how quality regulation is implemented (as discussed further in the rest of this report), it can provide incentives for companies to provide an optimal level of quality by making them trade off the benefits to customers from an improvement in quality and the costs required to achieve this improvement.

However, there are a number of factors to consider, including:

- what quality variables are typically considered, and how these are constructed for their inclusion in cost benchmarking;
- practical considerations such as the cost of constructing and maintaining robust quality measures for use in benchmarking;
- specifying models that have intuitive underlying relationships between costs and quality in DEA and SFA;
- the incentive properties of different methods of integrating quality;
- the translation of benchmarking results based on quality-dependent models into allowed revenues;
- stakeholder reactions and feedback.

The practical experiences of other regulators, whether or not they have been successful in incorporating quality within cost benchmarking, can be important in informing how these issues are addressed. Where quality has been incorporated within cost benchmarking, we consider the reasons driving a regulator's particular approach, the challenges and successes associated with its approach, and changes that it intends to make in future reviews. In other cases where it has not been integrated within cost benchmarking, we consider the challenges faced and work-in-progress material shared by regulators. Where relevant, we have sought insights from the academic literature and our work in the area to provide an additional perspective on the different approaches used by regulators.

The rest of the report is structured as follows.

- Section 2 provides further detail on the methodology that we have used in this report.

- Section 3 presents evidence from the stakeholder interviews that we have conducted, our review of published and unpublished regulatory documents, and insights from our work and the academic literature where relevant.
- Section 4 brings together the key messages across the whole report, and concludes.
- Appendix A1 provides a summary of the interviews that we conducted with regulators.

2 Methodology

Regulators determine efficient costs for companies typically by identifying two different sources of efficiency improvements that companies can make: catch-up and frontier shift. Catch-up efficiencies are associated with companies becoming more cost efficient at producing a given level of outputs by adopting industry best practice (i.e. 'catching up' with the industry frontier). In contrast, frontier shifts are cost reductions resulting from general productivity improvements in the industry (e.g. due to technological progress or cheaper inputs). Therefore, even efficient companies can reduce costs through frontier shift. We provide a brief overview of catch-up and frontier shifts in Box 2.1. This report considers the estimation of catch-up efficiencies using cost benchmarking models that include quality measures.

Box 2.1 Catch-up and frontier shift efficiencies

Catch-up efficiencies are cost reductions that companies can make by adopting industry best practices and moving towards the industry cost frontier. These efficiencies are assessed by comparing benchmarking a company's costs and outputs against other company's costs and outputs. There are two main types of approaches to estimating catch-up efficiencies: DEA and econometric techniques (such as SFA and MOLS). Both of these approaches seek to estimate the minimum cost required for each company to produce its outputs. By allowing companies to recover only their efficient costs, cost benchmarking approaches provide incentives to reduce their costs to increase profits. For a review of DEA and SFA methods, we refer the reader to Deuchert and Parthasarathy (2019),[32] Thanassoulis (2001)[33] and Kumbhakar *et*. al (2015) respectively.[34]

The focus of this report is on introducing quality measures into these cost benchmarking approaches. In the absence of quality measures in cost benchmarking (or elsewhere in the regulatory approach), companies may have incentives to provide a poor quality of service to reduce costs.

Frontier shifts (also known as ongoing efficiencies or dynamic efficiencies or Xgen) are cost reductions that can be made due to technological progress or cheaper inputs, allowing the same outputs to be produced with fewer inputs. When estimating efficient costs, a regulator needs to determine what level of frontier shift companies can be expected to achieve over the next regulatory period. A common method used by regulators (e.g. Ofwat, E-Control) to estimate frontier shift is to use the EU KLEMS data, which contains industry level productivity information. At a high level, this involves choosing a set of comparable sectors to the regulated sector in question, and taking a weighted average of the productivity growth rates.

There are many issues to consider when estimating frontier shift. These include, but are not limited to, the choice of comparator sectors, the weights used, the time period to average over, whether to use gross output and value-added productivity measures (or partial or TFP measures), whether any adjustments need to be made for any catch-up efficiencies in the comparator industries and how historical and future uncertainties can be accommodated. These issues are outside the scope of this report, and we refer the interested reader to Kumbhakar, Parthasarathy and Thanassoulis (2019), and Oxera (2016) for a detailed discussion of these issues.[35]

Source: Oxera.

[32] Deuchert, E. and Parthasarathy, S. (2018–19), five-part series of articles on the German energy regulator's benchmarking framework covering efficiency methods (DEA and SFA), functional form assumptions, cost driver analysis, outlier analysis and model validation, *ew–Magazin für die Energiewirtschaft*.
[33] Kumbhakar, S. C., Wang, H. J., & Horncastle, A. P. (2015) ,'A practitioner's guide to stochastic frontier analysis using Stata', Cambridge University Press.
[34] Thanassoulis, E. (2001), 'Introduction to the theory and application of data envelopment analysis', Dordrecht: Kluwer Academic Publishers.
[35] Kumbhakar, S., Parthasarathy, S. and Thanassoulis, E. (2019), 'A critical assessment of Bundesnetzagentur's approach to determining the German electricity grid sector's productivity factor (Xgen)', October. Oxera (2016), 'Study on ongoing efficiency for Dutch gas and electricity TSOs. Prepared for the Netherlands Authority for Consumers and Markets (ACM)', January.

We have considered a wide variety of information sources in undertaking this review of international experiences of integrating quality performance measures within regulatory cost benchmarking models.

To gain an initial understanding of each regulator's approach, we have reviewed publicly available material where available. This was based on Oxera's understanding of different regulatory approaches in the EU, desk research, as well as any additional material shared with us during interviews, as detailed further below.

We conducted interviews with EU regulators to develop a more complete understanding of the practical considerations and experiences underlying each regulator's decisions. Our approach to conducting interviews involved the following.

- We considered a list of regulated sectors and jurisdictions to contact for interviews. Where possible and relevant, we contacted EU water regulators. However, given the limited number of relevant examples in the water sector, we also contacted non-water regulators, particularly if they were relatively experienced in incorporating DEA and SFA within their cost benchmarking approaches.
- We developed an initial list of questions based on our understanding of regulatory approaches in the EU and the issues of interest to the DCCA. We then worked these questions into a final list with further comments and feedback from the DCCA.
- These questions were then circulated to regulators ahead of each interview. Each interview lasted approximately one hour, during which detailed notes were taken by the Oxera team.
- After the interviews, many regulators helpfully shared additional material on their current approach or work in progress. As this is a topical issue for all regulators across sectors and geographies, the regulators we spoke to in the course of the project were happy to make time in reviewing written material and clarifying questions.

In total, we approached ten regulators. We received responses from seven regulators, whom we were able to interview. These regulators covered a range of sectors and countries, as detailed further in Table 2.1. Some of these regulators covered multiple sectors. However, for the purposes of this report, we have listed the main sector discussed during our interviews.

Table 2.1 Summary of regulators and whether they integrated quality measures within cost benchmarking

Regulator	Country	Sector	Interviewed
ACM	The Netherlands	Electricity and gas distribution	✓
ARERA	Italy	Water and energy	
Bundesnetzagentur	Germany	Energy	
E-Control	Austria	Electricity distribution	✓
Energiavirasto	Finland	Electricity distribution	✓
ERSAR	Portugal	Water and wastewater	✓
Monitor	UK	Healthcare	
NVE	Norway	Electricity distribution	✓
Ofgem	UK	Gas distribution	✓
Ofwat	UK	Water and wastewater	✓

Note: For the regulators we interviewed, we specify specific sectors where relevant depending on the Interview.

Source: Oxera.

Different regulators are at different stages of developing their regulatory approaches, and so only a few have integrated quality within an econometric or DEA framework. Other regulators have considered introducing an integrated approach, but may have encountered issues such that further work is required before introducing quality-dependent cost benchmarking in regulation.

Table 2.2 Summary of regulators and whether they have integrated quality measures within cost benchmarking

	Quality-dependent DEA	Quality-dependent econometrics	Quality regulated separately from the cost benchmarking
Norwegian Water Resources and Energy Directorate (NVE)	Implemented		Yes
Energiavirasto	Implemented		Yes
E-Control	Considered	Considered	Yes
Ofwat		Considered	Yes
Ofgem			Yes
ACM			Yes
ERSAR			Yes

Source: Oxera.

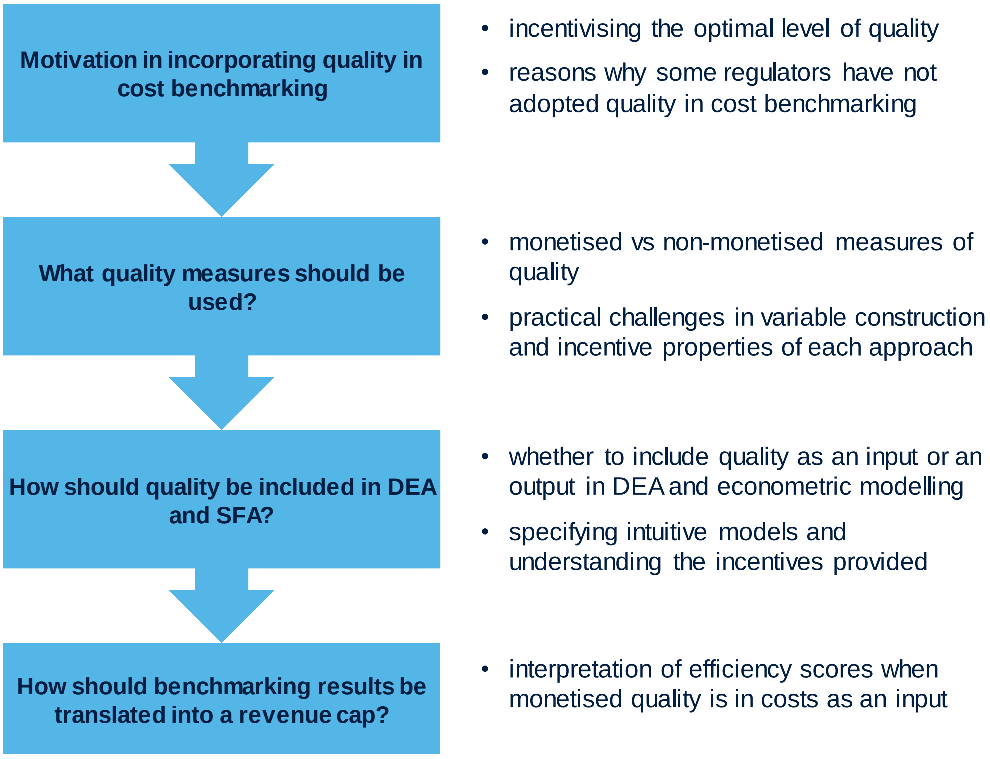
We have also reviewed some of the academic literature on where quality is incorporated into cost benchmarking models, and our work in the area which also involved discussions with Professor Emmanuel Thanassoulis and Professor Subal Kumbhakar, who specialise in DEA and SFA, respectively. This is particularly important as it provides alternative solutions that regulators could have considered or additional issues that have been overlooked and potential solutions, which could be helpful for the DCCA.

Based on our discussion with regulators, we consider that there are a number of key building blocks in designing quality-dependent cost benchmarking:

- **motivation behind adopting quality-dependent cost benchmarking**—why have some regulators decided to adopt or not to adopt quality in benchmarking?
- **choice of quality measures**—which quality measures should be used in benchmarking?
- **inclusion in DEA and SFA**—how should the quality measure be included in DEA and SFA?
- **translating results into a revenue cap**—how should results from benchmarking be translated into a revenue cap?

This is summarised in Figure 2.1, and explored further in section 3.

Figure 2.1 Building blocks of including quality in cost benchmarking



Source: Oxera.

In summary, our approach draws from a rich set of sources, including published and unpublished regulatory material, stakeholder interviews, academic literature and our work in the area that involved input from world-leading academics. We were therefore able to consider the issue of incorporating quality within cost benchmarking from multiple perspectives.

3 Evidence from stakeholder engagement

In this section, we present the results from the stakeholder engagement and our review of published and unpublished regulatory materials.

- Section 3.1 considers the motivation behind why some regulators integrate quality measures directly and why others have not done so.
- Section 3.2 considers the factors driving which quality measures are included in benchmarking, and the associated challenges.
- Sections 3.3 and 3.4 consider how quality should be included in DEA and econometric approaches respectively.
- Section 3.5 considers how the results from benchmarking can be translated into a revenue cap.
- Finally, section 3.6 provides further evidence on stakeholder reactions to the introduction of quality regulation, both within cost benchmarking as well as more generally within the regulatory framework given that these perspectives may be applicable to quality-dependent cost benchmarking.

3.1 Motivation for introducing quality-dependent regulation

In this section we examine why some regulators have included certain quality measures directly within cost benchmarking, and why other regulators have not.

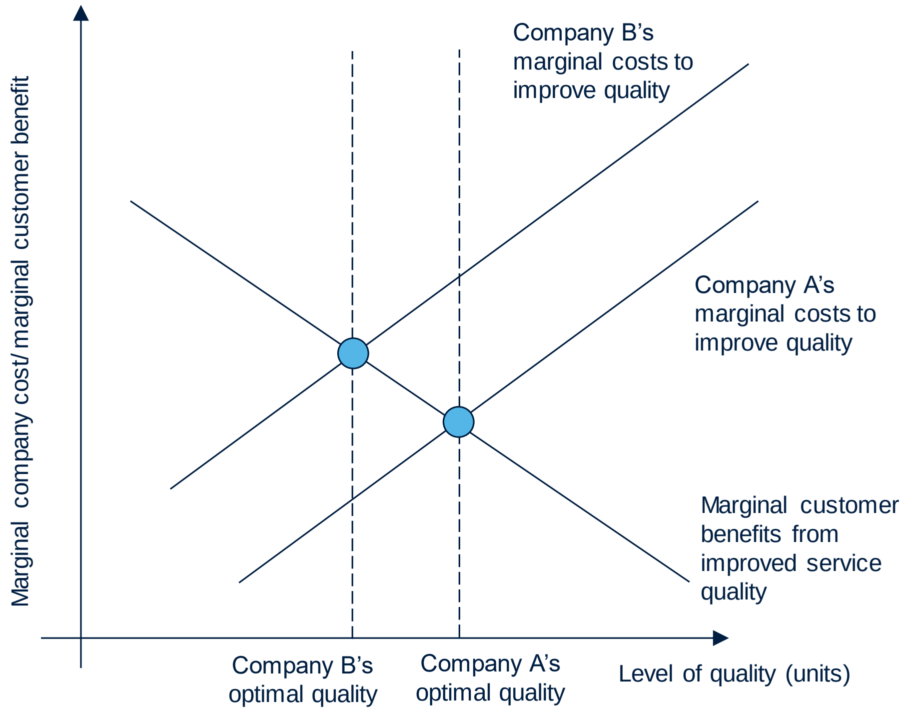
3.1.1 Motivation for including quality measures in cost benchmarking

NVE and E-control[36] explained that a key reason for including quality directly within their cost benchmarking approaches was to incentivise companies to achieve an *optimal* level of quality.[37]

The concept of an optimal quality level is shown in a stylised example in Figure 3.1.

[36] E-Control has not adopted quality-dependent cost benchmarking, but has stated that choosing an optimal quality level is a key reason why it would like to adopt it.
[37] Further details of how this is achieved is provided in section 3.3.

Figure 3.1 Optimal levels of service quality for two different companies



Note: To interpret the graph, consider company A. At levels of quality below company A's optimal quality level, the marginal costs of an improvement in quality are lower than the marginal benefits. It would therefore improve total customer welfare if company A improved its level of quality until it reached the optimal level (and vice versa). On the other hand, company B has a higher marginal cost of achieving a given level of quality (e.g. due to environmental factors), and so company B's optimal level of quality is lower.

Source: Oxera.

The upward-sloping lines represent the *marginal*[38] costs associated with a unit improvement in quality. At low levels of quality, companies will find it cheaper to improve the level of quality by one unit than if the level of quality is higher, represented by the upward slope of the lines. At the same time, customers will be willing to pay less and less for each additional unit of quality as quality rises, as reflected by the downward-sloping lines.[39] The regulatory framework should therefore incentivise these companies to invest in higher levels of quality until they reach the optimal level where the benefit derived by customers from a unit improvement in quality equals the cost of improving quality by one unit. However, in a regulatory framework that benchmarks only costs, companies may have perverse incentives to keep their quality of service low in order to keep their costs low and perform better on the cost benchmarking and this information (i.e. the cost-quality trade off) may not be revealed fully even if service quality is considered separately.

Nonetheless, it may not be optimal to provide an excessively high quality of service, as the costs of providing a higher level of quality may ultimately outweigh the benefits to customers. This leads to high prices for customers, and a reduction in quality in this case would improve customer welfare.

[38] Marginal costs are the costs required to improve quality by a given amount.
[39] Known as the marginal social benefits of quality, which are the benefits to society of an improvement in service quality.

However, one issue is that the regulator does not know the optimal quality level, and this level can differ by water company depending, for example, on the priorities of the customers of each company and their willingness to pay for quality improvements.

As illustrated in Figure 3.1, a hypothetical company B that finds it more costly to provide quality than company A also has a lower optimal level of quality. Equally, customers' marginal benefits may differ across companies such that the marginal customer benefits line in Figure 3.1 also differs by company (not illustrated).

It may be difficult for the regulator to gather the required information to determine the optimal level of quality and thus set the required quality level for each water company as well as the efficient costs required to achieve this level of quality.

The key principle is therefore that a regulatory framework should provide *incentives* for companies to choose optimally between quality and cost.

3.1.2 Motivation for not including quality measures in cost benchmarking

While adopting quality-dependent cost benchmarking may improve incentives for companies to provide optimal levels of quality, many regulators have not yet formally adopted this practice.[40] This is due to a variety of reasons, including the following (see sections 3.3 and 3.5 for a further explanation of these factors):

- potential challenges in specifying a model that leads to statistically and operationally robust results—e.g. statistical significance, addressing potential endogeneity, non-linear relationships (Ofwat; see section 3.4);
- possible challenges from companies due to the impact of including quality on estimated efficient costs for some companies (E-control; see section 3.6);
- the fact that quality incentives outside of the cost benchmarking framework may also be effective in maintaining a high level of quality (Ofgem; see section 3.6.6);
- a lack of understanding by companies of more sophisticated benchmarking methods (ERSAR, which currently uses sunshine regulation instead of an econometric or DEA benchmarking approach; see appendix 4A1).

Although not mentioned by regulators in response to this question, other reasons for not including a quality measure within cost benchmarking could include the following.

- The cost of implementation. Providing incentives to choose a socially optimal level of quality requires the construction of a monetised measure of social benefits from improvements in quality, which can be costly (see section 3.2).
- Not all measures of quality are suitable to be monetised. Regulators monetise quality measures by undertaking surveys. This relies on asking customers how much money they would be willing to pay for an improvement in quality (e.g. how many euros they would be willing to pay to

[40] Different regulators are at different stages of incorporating quality within cost benchmarking.

avoid a one-, two- or four-hour interruption). However, the value or monetary impact of some quality measures may be less easy for a consumer to quantify intuitively (e.g. the concentration of certain minerals in water), which potentially limits the effectiveness of the survey method and consequently the robustness of the monetised measure. However, see section 3.2.2 for a discussion of Ofwat's approach to outcome delivery incentives, which include a wide range of monetary incentives.

- Regulators may not want to incentivise companies to reduce certain measures of quality in order to reduce cost. The ACM highlighted the example of safety metrics in gas distribution.

These issues are considered in further detail in the following sections.

There are good reasons to adopt quality-dependent cost benchmarking—notably, to analytical determine the cost–quality trade-off and improve the incentives for companies to deliver more optimal quality of service levels.

However, not all regulators have been able to adopt this approach and not all measures of quality have been considered for inclusion within cost benchmarking. Nevertheless regulators continue to explore how best to achieve the integration as the drawbacks of treating quality outside of cost benchmarking is well understood.

3.2 Quality measures used

In this section, we consider different experiences in choosing and constructing the cost measures used in cost benchmarking. We note that the measure most commonly used by European energy regulators is a monetised measure of service interruptions (see in section 3.2.1). We also provide an overview of Ofwat's approach, which manages to integrate a wider range of monetised measures (albeit not within cost benchmarking—see section 3.2.2). Finally, we discuss some issues in choosing non-monetised quality measures (see section 3.2.3).

3.2.1 Experiences of EU regulators in constructing monetised measures of quality

Much of water and wastewater provision involves the delivery of non-market goods (goods that are not directly traded in markets, such as clean beaches), so customer valuations are not directly observed. In such cases, stated preference techniques are used, which involve surveying people to identify customer willingness to pay. Revealed preference techniques explore actual customer behaviour and prices paid for a related good (for example, travel costs).

NVE, Energiavirasto and E-Control include VOLL within their cost benchmarking (either in internal work or already implemented within the regulatory framework). This is a measure of quality monetised to reflect the marginal benefits of an increase in quality to customers. However, an important insight from our interviews with regulators is that constructing this monetised measure can be (but is not necessarily—see section 3.2.2) a difficult and expensive process.

NVE provided significant detail on the challenges that it faced when calculating this monetised variable.

First, the calculation of a monetised value is based on stated preference work, which inherently limits the range of quality measures that can be considered. The cost of interruptions, for example, is measured by asking those taking part in a survey: 'What are you willing to pay to avoid losing power for one, two or four hours?'.[41]

While the experiences presented apply to DSOs, they are also applicable to water. For example, the number and length of water supply interruptions may be more easily understood than another measure such as the concentration of iron in water.[42] However, see section 3.2.2 for Ofwat and the water industry's experience of monetising quality measures in England and Wales.

Second, surveys can be expensive and time-consuming to design, implement and analyse. In the design of its survey, NVE made use of pilot studies and focus groups, with re-designs of the survey at each iteration. Despite this, it found that a high proportion of the survey data did not yield consistent results. NVE provided an example from its study of residential customers. It found that some consumers might state that they did not value interruptions, whereas inference based on other questions showed this to be untrue. It also expressed concern that some respondents chose to respond strategically to the survey. As a result, even with very large sample sizes, with studies of an initial size of 100,000 possible respondents, approximately 2,200 responded and only 1,350 respondents were found to provide reasonable and usable survey results.

Third, there are other methodological issues. The survey was meant to capture the cost to customers of power interruptions. However, these costs are highly heterogenous and depend on a number of factors, including:

- the type of customer (e.g. industrial versus residential versus commercial versus agricultural);
- timing of the interruption (for example, the cost of interruptions differs by the hour of day, day of week and month of year);
- whether the outage was planned or unplanned.

Moreover, it can be difficult to obtain the required data to robustly capture these heterogeneities. For example, an ideal way to determine how the cost of interruptions varies by timing would be to repeatedly administer the survey to the same person at different times.[43] Nonetheless, over time, NVE has managed to differentiate VOLL according to six different categories of customers using the surveys.

Energiavirasto also stated that determining the factors required to convert outages into a monetary value was an involved process. It took the regulator 7–8 months to gather the relevant information, consult academics and engage with consultancies to come to a view on the right price of outages. Even so, this price is not distinguished between different customer demographics, which is an issue that it would like to address.

In summary, while a monetised quality measure reflecting the social benefits associated with an improvement in quality provides desirable

[41] This is the question posed to the 'residential' group of customers, which is one of six customer groups surveyed by NVE.
[42] This is one of the measures considered by Ofwat. Ofwat notes that 'iron and manganese are common in groundwater supplies and can lead to objectionable colour and turbidity (cloudiness) of drinking water as well as staining laundry and fixtures', which customers may find difficult to value. See Ofwat (2017), 'Delivering Water 2020: consultation on PR19 methodology. Appendix 2: Delivering outcomes for customers', 11 July, p. 25.
[43] This would allow the use of statistical techniques that isolate the time-varying dimension of the cost of interruptions to customers. If different customers were instead interviewed at different points in time, it would be more difficult to determine whether any differences in the cost of interruptions were due to the differences in time, or due to the fact that a different customer was being interviewed.

incentives for companies, the process of constructing this measure can be costly.

3.2.2 Ofwat's experience in constructing monetised measures of quality

In recent price control reviews, Ofwat has placed most of the burden of quality of service valuation on companies, mitigating, to some extent, the burden on the regulator discussed above. For PR19, Ofwat expected companies to employ a range of techniques to engage with customers on their priorities for service improvements, and their willingness to pay for these improvements. Ofwat wanted companies to:[44]

- use 'revealed preference' methods rather than relying solely on 'stated preference' methods to determine willingness to pay;
- make better use of customer, operational and social media data to target service improvements;
- use behavioural science to nudge customers into better behaviours (e.g. in relation to water efficiency and sewer blockages);
- involve customers and communities as knowledgeable active participants in co-developing and co-delivering solutions.

In PR14 (i.e. the previous price review), most companies employed stated preference techniques to determine customers' willingness to pay for improvements to service levels—such as reduced interruptions to supply, improved bathing water quality and reduced sewer flooding. Here, surveys were undertaken by companies that presented hypothetical trade-offs to customers with associated cost implications in order to derive the willingness to pay estimates. The main approaches were contingent valuation[45] and conjoint analysis.[46]

At PR19 (the recently concluded price review), Ofwat wanted companies to improve their stated preference approaches but also to implement revealed preference (and other) approaches. This was due to concerns around survey biases—including their hypothetical nature.

Evidently, there is no 'market' for issues such as supply interruptions or sewer flooding ('non market goods'). However, there are 'related goods' that are traded in other markets (both substitutes and complements). Revealed preference approaches impute a value from the market prices paid by users of these related goods. In practice four main approaches are used (and examples of these are provided below):

- averting (or 'defensive') behaviour (e.g., water discolouration)
  - cost of buying water filtration
  - cost of buying bottled water
- damage (e.g., sewer flooding incident)

[44] Ofwat (2017), 'Delivering Water 2020: Consulting on our methodology for the 2019 price review', July.
[45] Contingent valuation is a survey-based approach for determining the value a person places on a non-market good. It involves asking people to state their willingness to pay to obtain a good, or their willingness to accept to give up a good.
[46] Conjoint analysis is a survey-based approach where respondents are asked to choose between different packages that trade-off various product attributes. Based on the responses obtained, statistical analysis is then used to determine the valuation placed on the different attributes.

- lost income and higher insurance cost following sewer flooding
- travel costs (e.g., beach closure)
  - cost of inconvenience and time of travelling to nearest clean beach
- hedonic pricing (e.g., house prices)
  - higher house prices near clean beaches (amenity)
  - lower house prices near sewage treatment works (odour)

At PR19 Ofwat emphasised that companies should undertake a range of these types of approaches. There was also an emphasis on 'triangulation' of the results obtained, both between the different willingness to pay approaches and between these results and other sources of customer and operational data.

In essence this triangulation involved sense checking the results obtained across a broad array of approaches and data sources. For example, South West Water noted.[47]

> We have undertaken Stated and Revealed preference studies and a comprehensive triangulation process to provide accurate valuations covering the breadth and depth of water and wastewater services, such as water restrictions and resource options, sewer flooding and sustainable drainage options, pollution incidents, bathing water quality, water disruption and drinking water quality.

In PR19, Ofwat also made several changes to its outcome delivery incentives (ODI) approach, making it more challenging relative to PR14:[48]

- Ofwat adopted a number of common performance commitments (PCs) with standard definitions that cover areas such as leakage, supply interruptions, environmental performance, resilience and asset health;
- Ofwat required companies to set their performance-level targets at least at the forecast upper quartile in 2024–25 for four of these common performance commitments;[49]
- Ofwat removed the cap on ODI rewards and penalties (which for PR14 was set at +/-1% to +/-2% of return on regulated equity, RORE), and set an indicative range for the overall value of ODIs equal to +/-1% to +/-3% of RORE;
- in-period ODIs were the default (rather than adjustments at the end of the regulatory period);
- Ofwat introduced new measures of customer experience for household customers (the customer measure of experience, C-MeX) and developers (the developer services measure of experience, D-MeX). Ofwat proposed that +/-12% of household retail revenues would depend on C-MeX performance.

[47] South West Water (2018), 'An engaged WaterFuture: Engaging customers', Business Plan annex, September.
[48] Ofwat (2019), 'PR19 draft determinations: Delivering outcomes for customers policy appendix'.
[49] Those that will be assessed on a comparative basis at PR19—water quality compliance, supply interruptions, sewer flooding, and wastewater pollution incidents.

Ofwat stated that the above measures would mean that an average-performing company would expect to incur penalties on its ODI package.[50]

In terms of push-back from the industry, some companies argued that others in the sector had put forward overly ambitious forecasts. Others argued that applying single industry-wide performance levels for the three upper-quartile performance commitments was inappropriate due to company-specific issues (and that different PCs or deadbands should apply). Ofwat made few adjustments in its draft determinations:

- Ofwat accepted the transition argument for supply interruptions and that a common glide path to upper quartile performance was necessary. However, a glide path was ruled out for pollution incidents and internal sewer flooding;
- on the basis of the evidence that it had reviewed, Ofwat rejected virtually all claims that company-specific factors make the targets unachievable;
- Ofwat ruled out using deadbands (given the company-specific evidence). Deadbands are a 'neutral zone'—a specified range of performance levels where the ODI underperformance or outperformance payment is zero.[51]

The Final Determinations did however represent a significant movement compared with the Draft Determinations: reduced stretch, greater protections, and more symmetry in ODI downside/upside risk:[52]

- Ofwat reduced the stretch target for water supply interruptions. At the Draft Determinations, the PC for 2024–25 was set at three minutes (the forecast UQ), with a glide path. However, in the Final Determinations Ofwat adjusted the 2024–25 target to five minutes, with an amended glide path in the first four years. In addition, underperformance payment rates were now symmetrical with outperformance payment rates.
- In the case of internal sewer flooding, Ofwat reduced the downside ODI payments for three companies it regards as performing poorly, by introducing collars for the early years of 2020–25. Moreover, ODI payment rates were now symmetrical (having previously been downward-skewed).
- On leakage, Ofwat made the leakage PC less stretching for some companies (while there remained a requirement for a minimum 15% improvement). In addition, for companies that are already performing very well on leakage, the regulator provided some additional funding for enhancement expenditure that takes leakage beyond the forecast UQ.
- On water quality compliance (Compliance Risk Index), at the Final Determinations Ofwat amended the deadband to a score of 2.00 throughout the 2020–25 period, in part to allow more flexibility in performance to take into account the uncertainty created by the ban on the use of metaldehyde (which has been overturned by the High Court).
- On asset health, Ofwat reduced the stretch on mains repairs for all companies in all years. The underperformance payment rate on mains

[50] Ofwat (2017), 'Delivering Water 2020: Consulting on our methodology for the 2019 price review', July, p. 72.
[51] See: Ofwat (2019), 'PR19 initial assessment of plans: Glossary', January.
[52] See: Ofwat (2019), 'PR19 final determinations: Delivering outcomes for customers policy appendix', December; and Ofwat (2019), 'PR19 final determinations: Overall level of stretch across costs, outcomes and allowed return on capital appendix', December.

repairs for all companies was set to the industry average, to provide a more balanced spread of incentives and risks.

- Ofwat also applied caps and collars to financially material and/or highly uncertain PCs and allowed caps and collars on other PCs where company proposals are supported by high-quality customer engagement. Where the vast majority of companies have caps and collars on a common PC, Ofwat has applied caps and collars to all companies.

3.2.3 Specification of non-monetised quality measures to include in cost benchmarking

Given the resources required to construct a monetised measure of quality, using non-monetised quality measures may be an attractive alternative. When incorporating quality measures within benchmarking, the underlying assumptions driving the modelled relationship need to be carefully considered.

For example, E-Control's experts stated that a consideration when using quality measures within a DEA model is that the quality measure used needs to reflect the company's scale. This is because there are some challenges in accommodating ratios in DEA. Further details are provided in appendix A3. As a result, ratio measures that do not reflect a company's scale (e.g. the proportion of water delivered meeting a certain level of quality) may need to be converted to a measure that reflects the company.[53] For example, instead of using the proportion of water meeting a particular standard, one could specify a model where the volume of water not meeting standards is an input and the volume of water successfully meeting standards is an output. Of course, to avoid a situation where a company ends up being determined as efficient purely on the 'value' of a particular measure without consideration of other relevant factors, it may be necessary to include weight restrictions.

Careful examination of the relationship between cost drivers and costs is not a consideration specific to DEA models. The functional form between costs and their drivers also needs to be carefully considered in econometric models such as SFA (e.g. how does a unit change in a ratio measure affect cost?).

Another consideration (based on Ofwat's experience of attempting to incorporate quality within an econometric approach) is that quality is just one of many cost drivers, and not a primary cost driver (such as scale). As such, it may be difficult to obtain desirable statistical properties such as an intuitive sign and statistical significance (and general robustness of the model outputs). This is further complicated by the fact that a very large number of quality measures are available, not all of which have a significant influence on costs.

To mitigate these issues, alternative approaches can be used to create summary measures of quality, which can take account of a wider range of quality measures than using a variable that captures only one specific aspect of quality. Ofwat suggests that stakeholder engagement could be used to identify the most relevant quality measures. Another method is to use principal component analysis,[54] which aims to capture as much of the variation in a set of measures as possible in a small number of variables. A third method would be to adopt an approach similar to Ofwat's service incentive mechanism (SIM)

---

[53] E-Control suggested that in order to capture interruptions, a measure such as the total number of customer minutes lost may be appropriate.
[54] Principal component analysis is a statistical procedure for reducing the dimensionality of a set of variables by representing it with a few uncorrelated variables that capture most of its variability.

scores for customer service performance, which broadly captures customer service into an index.[55]

A large number of quality measures can be included in a general regulatory framework. However, how quality measures to be included in cost benchmarking should be constructed needs to be considered carefully.

- If regulators want to incentivise an optimal level of quality that takes into account the trade-offs between customer benefits and the costs of higher quality, a monetised measure of the social benefits of quality needs to be constructed—which can be a highly involved process. However, Ofwat's approach places the onus on the companies to determine customers' priorities for quality and their willingness to pay. Ofwat then challenges the rewards and penalties that companies propose for delivering or failing to deliver these outcomes.
- When using non-standard data such as ratio or ordinal data (quality measures are commonly measures on these scales), care needs to be taken that the models specified are consistent with the underlying assumptions of the estimation approach and that the model specification matches economic intuition.
- As there are many potential quality of service measures, one approach may be to consider a composite measure. This composite measure could be constructed through consultation with companies, using existing composite measures (such as the SIM in the case of Ofwat) or a statistical approach such as principal component analysis. This might result in models with more desirable statistical properties.

3.3 Incorporation of quality into DEA models

An important issue to consider is *how* a particular quality measure should be included in a DEA model. In particular, should a monetised quality measure be added directly to company costs, or should quality be included as a separate output?

The way in which quality is incorporated within cost benchmarking may influence the results of the benchmarking process, and will therefore affect companies' incentives to invest in quality and maintain optimal trade-off.

3.3.1 Quality as an input

While E-Control does not currently implement quality within its cost benchmarking approach, it has undertaken work exploring the implications of doing so. E-Control's experts point out that an input-oriented DEA model calculates efficiency scores by considering how far inputs can be reduced while holding the level of output constant. They consider that there are two reasons why it is more appropriate to include quality on the input side. First, there is a methodological issue, since certain quality measures such as interruptions represent an 'undesirable' output that should be minimised rather than maximised. Second, outputs should be exogenous quantities, but quality measures are likely to be influenceable by companies.[56]

Furthermore, E-Control's experts consider that including quality measures as a separate input or output may be problematic as it would allow some companies to appear 100% efficient by specialising in reducing costs or improving quality

[55] Ofwat's SIM is a measure of how customers feel about the services that companies provide. It consists of two components—the number of complaints that companies receive, and customers' satisfaction with their company's handling of queries and resolving issues. See Ofwat, 'Service incentive mechanism', https://www.ofwat.gov.uk/regulated-companies/company-obligations/customer-experience/service-incentive-mechanism/, accessed 02/01/2019 for further details.
[56] E-Control (2014), 'Optionen zur Einbeziehung der Versorgungsqualität in derzeitige bzw. künftige Regulierungsrahmen für Stromverteilernetzbetreiber', pp. 39–40.

instead of trading off the two, in the absence of weight restrictions (see Box 3.1).

Box 3.1 Specifying quality as an input or output within a DEA model

DEA works by calculating input and output weights in a way that depicts companies in the best possible light. As additional inputs and outputs are added, companies that are exceptional on one input or output are likely to receive efficiency scores at or close to 100%.[57] This has important implications for how quality should be included in a DEA model, given that a regulatory framework should provide incentives for companies to optimally trade off the costs of increasing quality against the marginal benefits to customers of doing so. If quality is included as an input or output that is separate from costs, and weight restrictions are not placed on DEA that define sensible trade-offs between the factors, it is possible for companies to specialise in quality and ignore or place unusual importance to costs/other outputs (and vice versa) to obtain a high efficiency score. Note also the earlier point that weight restrictions can help to ensure intuitive and incentive-compatible outputs.

Note: For a further discussion of these issues, we refer the reader to Thanassoulis, E. (2001). ‘Introduction to the theory and application of data envelopment analysis’, Dordrecht: Kluwer Academic Publishers, chapter 8.

NVE and E-control expressed a preference for an alternative approach.[58] This approach would add the monetised measure of quality[59] (in their case, VOLL) directly from TOTEX, rather than introduce additional inputs or outputs are introduced in the DEA model to capture quality.

NVE specified an input-oriented DEA model with three outputs and TOTEX (with VOLL added), as detailed in Table 3.1 below. It accounts for environmental characteristics outside of a company’s control that can affect the likelihood of having an outage, such as strong winds and sloped areas, using a second-stage regression approach.[60]

NVE is currently considering if and how companies are serving different types of customer with different demands for security of supply. For example, one company could serve customers with a high demand for security of supply, which would lead to a high VOLL per unit of outage (and vice versa). This company would need to incur additional expenditure on ensuring that its grid is more strongly reinforced and maintained more regularly. These would represent outputs that might not be captured by the DEA modelling.

As a result, NVE is exploring the inclusion of an output variable that captures customers’ demand for security of supply. In practice, it would calculate this variable by considering the VOLL had there been an outage of a particular length of time (e.g. five hours for all companies). In this case, companies would have a higher value of this variable if their customers had a higher demand for a secure power supply. NVE states that this approach produced reasonable results, pointing out that there were no unintuitively large changes in efficiency scores.

For a discussion of how the results from a benchmarking approach (either DEA or econometric) that includes quality directly within TOTEX can be translated into a revenue cap, and the incentive properties of such an approach (i.e. the

[57] Depending on the returns to scale assumptions.
[58] E-Control has not officially incorporated this approach into its regulatory framework.
[59] See sections 3.1 for incentive properties of such an approach and 3.2 for the practical difficulties associated with constructing a monetised measure of quality.
[60] In the second stage, they regress DEA scores from the first stage on environmental (or Z-) variables. See appendix A1.1 for further details.

balance of risk and reward, and incentives to trade off quality and costs), see section 3.5.

NVE also noted that interruptions is just one aspect of quality, and that there are other aspects of quality that may not need to be monetised. For other measures that have not been monetised, it adopts other forms of regulation such as mandating a minimum level of quality.

Table 3.1 Specification of NVE's DEA model

Input	Outputs
Cost base composed of the sum of: operation and maintenance costs, cost of energy losses, VOLL, depreciations, and the regulatory rate of return on the regulatory asset base.	Number of customers (proxy for demand) Number of substations (proxy for the distribution of demand) Length of high-voltage network (proxy for transport distance)

Source: Amundsveen, R., & Kvile, H. M. (2016), 'Balancing incentives: The development and application of a regulatory benchmarking model', In Productivity and Efficiency Analysis (pp. 233-247). Springer, Cham.

However, we note that that this approach is not without its problems, especially when using a single input variable 'TOTEX'. The issue is that the model implicitly assumes that €1 of VOLL has the same effect in terms of supporting the cost drivers (e.g. the number of customers) as €1 of TOTEX. This is a strong assumption. It can be overcome by using multiple inputs (TOTEX, VOLL) and introducing weights restrictions on the trade-offs between the modelled 'expenditure' and VOLL (and the cost drivers). These would need to be based on estimates of how much more investment is needed to improve quality.

Nonetheless, NVE comments that imposing weight restrictions involves assumptions that need to be validated. This is because it is difficult to obtain estimates of how much investment is needed to improve quality, as this is likely to be specific to each company and also to different activities within each company. Furthermore, by having TOTEX and VOLL as separate inputs, it would allow companies with low TOTEX but high VOLL to seem efficient, which would remove the incentive to provide a good quality service (unless they are addressed through weight restrictions).

3.3.2 Quality as an output

Energiavirasto had also initially included the cost of outages within TOTEX in the regulatory period 2008 to 2011, in the same way described above for NVE and E-Control. In particular, Energiavirasto estimated a DEA model where the input variable was TOTEX (including the cost of outages) and three output variables.[61]

However, Energiavirasto explained that quality incentives already existed elsewhere in the regulatory framework when it first included quality within cost benchmarking. This led to objections from companies that it amounted to double regulation of quality. That is, the impact of a rise in outage costs on profits would manifest itself twice within the regulatory framework—first through the efficiency incentive, due to the inclusion of quality within TOTEX in cost

[61] From 2008 to 2011, Energiavirasto applied an average of DEA and SFA efficiency scores.

benchmarking models; and second through the separate quality incentive framework.[62]

As a result, in the next regulatory period, from 2012 to 2015, the input variable consisted only of half of the value of outage costs. Energiavirasto also moved to using the StoNED method,[63] adding half the outage costs to controllable OPEX (having switched away from a TOTEX regime). The rationale behind including half of outage costs within cost benchmarking was because the other half of outage costs was regulated via the quality incentive that existed outside of cost benchmarking.[64]

However, in 2010 and 2011, there were large storms leading to significant power outages for several companies. Consequently, the cost of outages was up to fifteen times that of the efficient cost level for the most heavily affected company.[65] As a result, in the following regulatory period, from 2016 to 2019, Energiavirasto moved from including the cost of outages as part of the input costs to treating it as a separate (undesirable) output.

Energiavirasto and its advisers explained that, in their StoNED model, shadow prices of the cost of outages (included as an output) can be either positive or negative. Companies with low costs of outages tended to have negative shadow prices, where an increase in the cost of outages leads to a fall in efficient costs. However, companies with exceptionally high costs of outages, due to storms for example, will have a positive shadow price, and can justify higher operating costs in that year due to storms with the high level of outages experienced.[66]

Furthermore, in 2016–19, Energiavirasto limited the magnitude of the efficiency challenge to 20% of the efficient level of profits. As a result, this limited the impact of outage costs on DSOs' allowed revenues.

Energiavirasto's advisers explained that there are also other reasons why the cost of outages should not be interpreted as an input. First, a company cannot supply a greater level of outputs by increasing its cost of outages. Second, the cost of outages is not an essential component of producing outputs.[67] Instead, they viewed the cost of outages as an undesirable output.

A summary of the changes to Energiavirasto's approach to incorporating quality within cost benchmarking is provided in Table 3.2.

[62] The quality incentive consists of using the average realised outage costs over two previous regulatory periods (i.e. a period of eight years) as the reference level of outage costs. The difference between the reference level and actual level of outage costs is then used to adjust DSOs' profits. Energiavirasto places a floor and ceiling on the impact of the quality incentive at no more than 15% of the DSO's efficient level of profits in that year (i.e. the quality-adjusted profit cannot be less or more than the quality-unadjusted profit by 15%).
[63] Stochastic Non-parametric Envelopment of Data (StoNED) is a method that aims to combine the merits of both DEA and SFA. StoNED shares similarities with DEA in that the frontier is determined by choosing weights (which can be interpreted as shadow prices) on outputs such that the company's efficient costs are maximised. See Energiavirasto (2015), 'Regulation methods in the fourth regulatory period of 1 January 2016 – 31 December 2019 and the fifth regulatory period of 1 January 2020 – 31 December 2023', 30 November, p. 87.
[64] Sigma-Hat Economics Oy (2013), 'Keskeytyskustannusten enimmäismäärän rajaaminen sähkön jakeluverkkotoiminnan valvontamallin tehostamiskannustimessa', 26 April.
[65] Sigma-Hat Economics Oy (2013), 'Keskeytyskustannusten enimmäismäärän rajaaminen sähkön jakeluverkkotoiminnan valvontamallin tehostamiskannustimessa', 26 April, p. 16.
[66] Sigma-Hat Economics (2014), 'Tehostamiskannustin sähkön jakeluverkkoyhtiöiden valvontamallissa: Ehdotus Energiaviraston soveltamien menetelmien kehittämiseksi neljännellä valvontajaksolla 2016 – 2019', 21 October, p. 17.
[67] Sigma-Hat Economics (2014), 'Tehostamiskannustin sähkön jakeluverkkoyhtiöiden valvontamallissa: Ehdotus Energiaviraston soveltamien menetelmien kehittämiseksi neljännellä valvontajaksolla 2016 – 2019', 21 October, p. 16.

Table 3.2 Energiavirasto's benchmarking approach over time

Year	Estimation method	Inputs	Output
2008–11	DEA and SFA	OPEX + CAPEX + cost of outages	Distributed energy Number of customers Network length
2012–15	StoNED	OPEX + 0.5*cost of outages	Distributed energy Number of customers Network length
2016–19; 2020–23	StoNED	OPEX	Distributed energy Number of customers Network length Cost of outages

Source: Presentation shared by Energiavirasto with Oxera on 25/11/2019.

Energiavirasto notes that its methods give the best view of each company's efficiency. Due to its decision to include quality as an output when there were already three other outputs in the model, Energiavirasto found that, on average, the difference in efficiency scores due to the inclusion of quality was relatively small. Indeed, it found that for about 70% of companies the shadow price for the cost of outages was negative and small in magnitude, meaning that the cost of outages had a limited impact on efficient costs.[68] This was in contrast to approaches where the cost of outages was included directly within TOTEX so that there was a 1:1 relationship between revenue caps and the cost of outages.

The low impact of quality when considered as an output can indicate that most companies found it more beneficial to be assessed on the other outputs than on quality (especially if no explicit weight restrictions are applied). This in turn could imply that only a few companies had a high level of quality for their level of OPEX compared to the other companies. This may have the effect of 'discouraging' companies from improving their quality levels to improve their efficiency rating as they have a low weight placed on the quality measure.

We also note that while Energiavirasto's implementation of StoNED does permit both positive and negative weights on quality, this is not necessarily based on operational reasoning. It can be shown that within a DEA framework, allowing for both positive and negative weights is equivalent to including quality both as an input and as an output. The DEA model then chooses weights such that it maximises each company's efficiency score. This means that the model may treat quality as an input for one company, and as an output for another company, depending on whichever provides a higher efficient cost. As a result, companies may not always receive incentives to minimise interruptions (as more interruptions leads to a higher cost prediction of interruptions is included as an output).

Nonetheless, Energiavirasto emphasises that there are quality incentives outside of cost benchmarking, and that, as a consequence of the large-scale storms in early 2010, DSOs were set gradual requirements to have a weatherproof network coverage by law (Electricity Market Act). Therefore, DSOs have strong incentives to strive for a high-quality service outside cost benchmarking, which is why its role has not been highlighted within cost

[68] Sigma-Hat Economics (2014), 'Tehostamiskannustin sähkön jakeluverkkoyhtiöiden valvontamallissa: Ehdotus Energiaviraston soveltamien menetelmien kehittämiseksi neljännellä valvontajaksolla 2016 – 2019', 21 October, p. 20. For example, the regulator explains that even when the shadow price of the cost of outages takes positive values, it is at most +0.1. This implies that the change to efficient costs will be only 10% of the cost of outages.

benchmarking. Indeed, Energiavirasto considers that the legislated security of supply obligations have had the greatest effect on improving quality and reducing outages.

On whether it would like to include other measures of quality within its cost benchmarking, Energiavirasto stated that it was too early to say what approach it would ultimately adopt, but it would use the current methodology as a starting point. The changes to be made would depend on discussions closer to the next regulatory cycle, where they may make changes such as considering a TOTEX framework.

While not adopted by the regulators that we have interviewed, we note that quality of service measures could be included as a direct (non-monetised) measure within DEA. However, we note that a number of regulators' experiences have highlighted the importance of accounting for environmental factors. For example, we note that Energiavirasto accounts for higher operating costs of operating in rural areas, and that NVE has made a correction for a variety of environmental variables, including coastal environments, forests, mountains and cities. As a result, if non-monetised variables are used, contextual variables would also need to be included alongside measures of quality. This is because the costs required to achieve a given quality level are likely to differ due to differences in operating environments. For example, maintaining energy supplies would depend on topology (as outages are easier to resolve in accessible locations).

Several regulators have included quality within benchmarking by including monetised quality within TOTEX. This is to avoid companies specialising in either quality or cost to gain an efficiency score of 100%, and to incentivise companies to optimally trade off the costs of reaching higher levels of quality and the benefit to customers of doing so.

However, Energiavirasto has switched from including monetised quality within TOTEX to including it as an (undesirable) output. This was due to extreme weather conditions leading to excessively high costs of outages if the cost of outages were included as an input. The regulator has also placed a limit on the magnitude of the efficiency challenge that can be applied to companies.

3.4 Incorporation of quality in econometric models

As with DEA, the question of whether quality should be included as an input or an output, and the specification of operationally intuitive models, is a consideration for econometric approaches such as SFA or corrected/modified OLS.[69]

3.4.1 Quality as an input

In an econometric approach, from a technical perspective it is important—but not essential (and it may be impractical)—that cost drivers included are exogenous. E-Control's experts noted that quality measures may be controllable by the company and thus may cause estimation problems (while ignoring them leads to other statistical issues, notably omitted variable bias, as flagged by Ofwat). In our discussions with the regulator, it has emphasised the importance of ensuring that cost drivers in an econometric approach are exogenous and not influenceable by the company. Note that omitted variable bias (due to the omission of relevant service quality measures) is also likely to result in an endogeneity bias, given the likely correlation between service quality measures and some of the cost drivers that are typically included such

[69] Ordinary least squares is a method of estimating a linear regression model.

as scale and topographical characteristics. Hence, omitting service quality measures from the cost models because service measures are endogenous can result in the same statistical issue that it is seeking to mitigate.[70]

On this issue, the CMA[71] expressed endogeneity concerns about Ofwat's cost modelling at PR14, where explanatory variables for mains renewal, leakage or various quality of service measures may be under company control. However, the CMA noted that 'given limitations in the available data, it may be better, *in some cases* to include an explanatory variable that carries risks of endogeneity than to fail to take any account of potentially important differences between companies'.[72]

While Ofwat has also expressed concerns over endogeneity at the current price review,[73] it balances these risks of endogeneity against other concerns. For example, when using the length of mains as a driver of scale, it notes that 'while companies have a degree of control over the length of mains, we consider that it remains substantially determined by exogenous factors, and the benefit it brings in terms of providing a good proxy of scale outweighs any concerns around endogeneity'. (Note the above point that both statistical issues (omitted variable bias and endogeneity) can result in biased model outputs, and that omitting service quality measures because they can be endogenous can also result in endogeneity.)

E-Control's experts have explored ways of including quality within benchmarking by specifying a log-linear model where quality is accounted for by adjusting inputs (i.e. the dependent variable) rather than outputs (the explanatory variables). As its dependent variable, it uses TOTEX plus VOLL.[74] The independent variables are selected based on explanatory power and ease of implementation.[75] The model is then estimated using modified OLS.[76]

In arriving at an estimate of each company's efficiency scores, E-Control's experts combined DEA with econometric results. They compared these results against an approach that excludes quality from TOTEX in the modelling. These results, along with stakeholder reactions, are discussed in further detail in section 3.6.

However, similar to the case of DEA, our understanding is that this approach may impose strong assumptions on the relationship between TOTEX and cost drivers. Consider, for example, a specification where:

$$(TOTEX + cost\ of\ interruptions) = \beta_0 + \beta_1 \cdot connections + \epsilon$$

This model assumes that the number of connections is related to the costs of interruptions and TOTEX in the same way. Implicitly, this assumes that a one-

[70] Oxera (2020), 'Quality measures in cost benchmarking', May. Oxera (2020), 'Integrating cost and outcomes', prepared for Yorkshire Water Services, March.
[71] The Competition and Markets Authority (CMA) is a UK government department responsible for strengthening and protecting competition. Companies can refer to the CMA to report on and determine disputed company determinations.
[72] Competition and Markets Authority (2015), 'Bristol Water plc. A reference under section 12(3)(a) of the Water Industry Act 1991. Final report', p. 73.
[73] In the current price review, Ofwat noted that if the error term contains managerial inefficiency, it is possible that the quality measure is correlated with the error term. This leads to an issue of endogeneity, resulting in biased estimates of the model coefficients, and hence distorted error terms on which efficiency estimates are derived in the econometric method. See Ofwat (2018), 'Cost assessment for PR19: a consultation on econometric cost modelling', March, p. 10.
[74] Energiavirasto has adopted a similar approach in earlier regulatory periods.
[75] E-Control includes the length of medium-voltage network.
[76] OLS is able to consistently estimate coefficients associated with each cost driver, but is unable to consistently estimate the intercept as it includes the expected value of inefficiency across all companies. Modified OLS accounts for this by making an adjustment to the intercept.

unit fall in the cost of interruptions requires one unit of TOTEX to achieve for all companies.

In contrast, including quality as an output (or input) allows the relationship between quality and costs to be estimated rather than assumed. In particular, if a non-monetised measure is used as an output, it can provide incentives for companies to achieve a given level of quality at a cost that is efficient relative to its peers. Nonetheless, we note that this may not provide incentives to choose an optimal level of quality, and the model design, including weight restrictions, should be carefully considered.

3.4.2 Quality as an output

Ofwat stated that it would like to include quality directly within its cost benchmarking. In the ways that it has explored to do this, quality measures have been included as a cost driver rather than as an adjustment to costs (the dependent variable in the econometric specification).

The regulator justified the inclusion of service quality by explaining that its inclusion within an econometric specification may help to improve the robustness of an econometric model. If the quality of service is excluded but is correlated with other independent variables in an econometric specification, this will lead to an omitted variable bias because quality of service is a cost driver. This means that the relationship between costs and cost drivers cannot be estimated correctly, leading to distorted residuals and hence biased efficiency scores. Furthermore, as Ofwat’s models are used to set future allowances by using forecast cost drivers and estimated model parameters, biased coefficients will also lead to incorrect levels of future allowances (see Box 3.2 for further details on Ofwat’s cost benchmarking approach).

Box 3.2 Overview of Ofwat's cost benchmarking approach

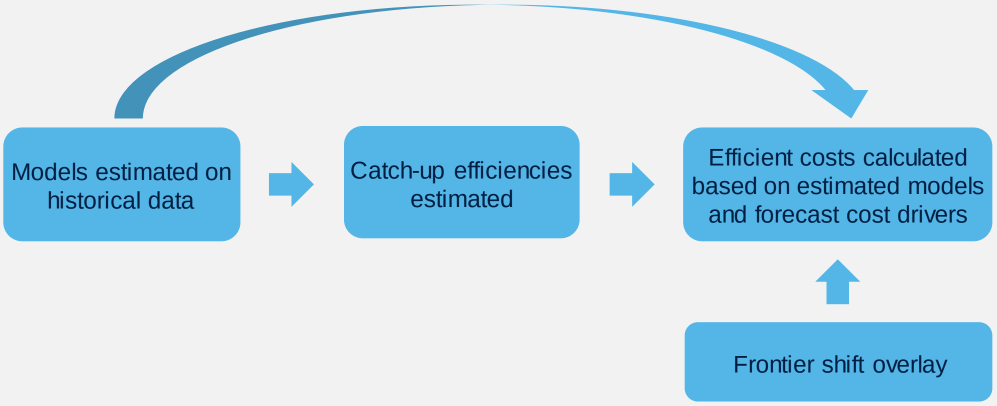
The figure below provides a stylised depiction of Ofwat's approach to calculating allowances for base expenditure (i.e. OPEX, capital maintenance and growth costs).

Econometric models are estimated on historical data on wholesale activities. Efficiency scores are then calculated as the ratio of predicted to actual costs, and the upper-quartile efficiency score is used to set a catch-up efficiency challenge for all companies.

Ofwat then uses forecasts for each cost driver in the model to predict future efficient cost levels. The modelled cost predictions over the next regulatory cycle are calculated by applying the estimated model coefficients to the forecast cost drivers. Efficient costs are then calculated by applying the catch-up challenge and a frontier shift (net of real price effects) overlay to these predicted costs.

Therefore, if the models are incorrectly estimated (e.g. if there is considered to be a significant endogeneity or omitted variable bias problem), the modelled cost predictions over the next regulatory cycle will also be biased.

Stylised depiction of Ofwat's approach to cost benchmarking and setting allowed revenues



Source: Oxera.

We also note that there has been push-back from England and Wales water and wastewater companies on the separation between cost benchmarking and quality, where companies have argued that: (i) companies are being set too low a cost level, as they are being benchmarked against companies with a lower quality of service; and (ii) there may not be sufficient funding to achieve future quality targets. The first issue is still being debated,[77] but in its final determinations Ofwat presented evidence showing that some of its estimated upper-quartile cost-efficient companies had also achieved strong quality of service performance.[78] On the second issue, Ofwat's cost benchmarking models assume that quality of service improvements have been made over the historical period, and hence that the cost of service improvements are already implicitly included within cost benchmarking. Again, the issue is still being debated, but in its final determinations Ofwat made some amendments.[79]

Separately from its final determinations, Ofwat stated that it had undertaken some initial modelling with quality included directly within an econometric framework. However, Ofwat stated that it had encountered a number of challenges when incorporating quality in cost benchmarking.

[77] Subsequent to the completion of the report in January 2020, four companies have appealed Ofwat's decision to the CMA.
[78] Ofwat (2019), 'PR19 final determinations. Securing cost efficiency technical appendix', December, p. 56.
[79] Ofwat (2019), 'PR19 final determinations. Securing cost efficiency technical appendix', December, p. 25.

First, as described above, quality measures may be controllable by the company, leading to an endogeneity issue, potentially resulting in biased efficiency scores and forecasts. However, as noted above, Ofwat considers that, if a particular variable is an cost driver that captures important differences between companies, the benefits of using such a driver can outweigh the concerns around endogeneity.

Second, there may be non-linearities in the relationship between quality and cost. Companies at the frontier of quality may find it very expensive to achieve further quality improvements, while companies with room to catch up may not find it very costly to improve quality. This means that flexible relationships may need to be considered (where relevant) in the econometric modelling.

Third, quality measures are not primary cost drivers—unlike scale, for example. The relationship between cost and quality may therefore not be as sizeable as the relationship between cost and other variables, especially if there is not a significant variation in quality of service across companies. Therefore, when quality measures are included in an econometric model, care should be taken to ensure that the model has desirable properties (e.g. statistical significance; though in itself statistical performance needs to be balanced against other criteria such as operational intuitiveness).

In fact, during Ofwat's model consultation for PR19,[80] one company proposed to incorporate Ofwat's SIM (service incentive mechanism) score within Ofwat's models for retail services. However, the coefficient on the variable was of an unintuitive sign (critical requirement in regulatory contexts) and lacked statistical significance.[81]

Fourth, Ofwat sets an allowance for each price control period by forecasting cost drivers and calculating the prediction of the econometric model using these forecast cost drivers (see Box 3.2 for further details). We note that Ofwat's regulatory framework encompasses a large number of performance commitments for which forward-looking targets are set. However, Ofwat pointed out that there may be instances where setting a forward-looking target may be more challenging. For example, policy may influence assumptions. Ofwat provided an illustrative example that in future price controls other quality of service metrics such as carbon reduction targets could be a common performance commitment.

While we have not managed to interview Monitor, we also consider their experience of incorporating quality in benchmarking in the past. Monitor benchmarks the cost efficiency of the providers of NHS services[82] (called NHS trusts) using random effects (i.e. a panel data estimator) and stochastic frontier analysis. They include a set of variables that captures:

- total activity (adjusted for type of cases);
- quality of service;
- degree of specialisation;
- local factors including demographics and disease prevalence; and
- size and type of trust.

[80] The price review for the period 2020–25.
[81] Ofwat (2018), 'Cost assessment for PR19: a consultation on econometric cost modelling. Appendix 1 — modelling results', March, p. 90, template 79.
[82] The National Health Service (NHS) is the UK Government funded medical and healthcare service.

Monitor includes quality of service based on patient satisfaction. They produce an index for the quality of care that NHS trusts produce to account for the additional costs this extra value may acquire. This index is constructed by considering the percentage of respondents that answered “strongly agree” from relevant questions in a survey.[83] However, there are many questions that measure of quality of service, and including each separately in an econometric specification may not lead to intuitive results. To produce a composite index, they use principal component analysis.[84] They find that the first principal component captures 63% of the total variation in the responses. This index is then included in SFA as an output.

Monitor stated that these quality variables are not statistically significant.[85] Furthermore, we note that the coefficient appears to be negative, suggesting that it may require further operational validation.[86]

We also consider ARERA’s, the Italian water regulator, approach to OPEX efficiency benchmarking for the next regulatory period (2020-2023). Their approach is still being developed and undergoing consultation. The proposed approach includes:

- using a panel data set of 98 companies over four years (2014-2017), covering a population of around 42 million, or around 70% of the population.
- using cost models with a Cobb Douglas functional form and the following variables:
    - inputs: operating costs;
    - input prices: cost of electricity supply, labour costs (Personnel costs over the total resident population), wholesale water purchase cost;
    - outputs: volume of water invoiced, length of network, resident population, population equivalent, availability and reliability of measurement data, compliance with the legislation on urban wastewater management, water losses.
- using SFA models of Battese and Coelli (1992), [87] Battese and Coelli (1998),[88] and Pitt and Lee (1981).[89,90]

[83] This is the NHS Staff Survey, which is an annual questionnaire completed by staff in NHS providers. Questions cover a range of questions, including the standard of training the NHS trust provides to staff, the quality of patient care and whether there are any errors that may have harmed patient outcomes. For a full list, see Monitor (2016), ‘2016/17 National Tariff Payment System: A consultation notice. Annex B5: Evidence on efficiency for the 2016/17 national tariff’, 11 February, table 2.
[84] Principal component analysis is a statistical technique that constructs weighted combinations of a set of variables to capture the maximum amount of variation possible in a smaller set of variables, called principal components. The first principal component captures a higher proportion of the total variation compared to all the other principal components.
[85] Monitor (2016), ‘2016/17 National Tariff Payment System: A consultation notice. Annex B5: Evidence on efficiency for the 2016/17 national tariff’, 11 February, section 4.1, paragraph 1.
[86] See table 4 of the Monitor (2016) report.
[87] Battese, G. E., & Coelli, T. J. (1992), ‘Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. Journal of productivity analysis’, 3(1-2), 153-169.
[88] Battese, G. E., and Coelli, T.J. (1988), ‘Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data’, Journal of Econometrics, vol. 38, pp. 387-399.
[89] Pitt., M. M., and L. F. Lee, 1981, ‘The measurement and sources of technical inefficiency in the Indonesian weaving industry’, Journal of Development Economics, vol. 9, pp.43-64.
[90] The three papers cited above impose different assumptions on inefficiency. For a description of these approaches, see Kumbhakar, S. C., Wang, H. J., & Horncastle, A. P. (2015), ‘A practitioner's guide to stochastic frontier analysis using Stata’, Cambridge University Press.

We note that the water losses, which is the quality variable included in the model is not statistically significant at the 5% level a number of specifications.[91]

ARERA's and Monitor's experience indicate that it can be challenging to specify a model that shows an operationally intuitive relationship between costs and quality and has desirable statistical properties. (See the earlier discussions on our work in the ongoing CMA price redeterminations where we developed quality-integrated models with appropriate outputs.)

Quality can be included in an econometric approach in one of two ways.

The first approach adjusts TOTEX for the monetised value of quality. However, the theoretical properties of such an approach need to be carefully considered.

The second approach treats quality as an output to be included as an explanatory variable. An advantage of this approach is that it can potentially improve the robustness of modelling by accounting for omitted variables. However, there are a number of technical challenges, such as non-linear relationships, endogeneity and obtaining desirable statistical properties such as statistical significance, which need careful consideration.

3.5 Translating benchmarking results into a revenue cap

As discussed in sections 3.3 and 3.4, one approach to incorporating quality within cost benchmarking is to include VOLL and TOTEX as a single input.

When the TOTEX input (in the rest of section 3.5, when we refer to TOTEX, we emphasise that this excludes VOLL) is combined with quality in either a DEA or an econometric model, it is important to note that the efficiency scores resulting from these models have a different interpretation from a quality-unadjusted model. Taking the VOLL adjustment as an example, efficiency scores in a quality-adjusted model are interpreted as:

$$\frac{Efficient\ (TOTEX + VOLL)}{Actual\ (TOTEX + VOLL)}$$

rather than the standard interpretation of:

$$\frac{Efficient\ TOTEX}{Actual\ TOTEX}$$

As a result, the efficiency scores from a quality-adjusted model cannot directly be used to calculate efficient TOTEX by simply multiplying the efficiency score by TOTEX.

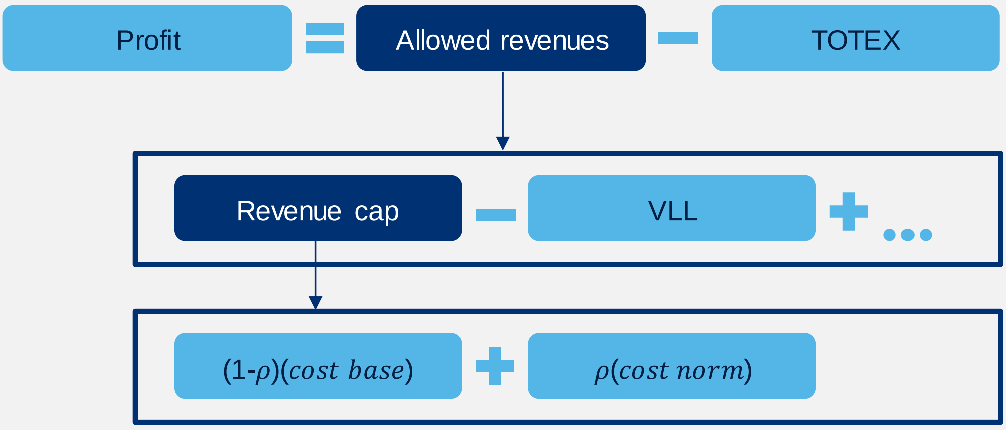
NVE has accounted for this difference in the way that it calculates allowed revenues for companies. This approach is integrated in a regulatory framework that achieves a balance of risk between companies and customers.[92] The regulator explained that a guiding principle in its approach was to ensure that the companies had balanced incentives to either improve quality or reduce costs. That is, if a 100DKK reduction in costs leads to a 60DKK rise in profits, then a 100DKK fall in VOLL should also lead to a 60DKK fall in profits. In the absence of such a balance, companies have high-powered incentives to focus on cost rather than quality (or vice versa), even if the social value generated from each is the same. We provide further details of this approach in Box 3.3.

[91] The coefficient on quality is statistically insignificant on the Battese and Coelli models, while it is statistically significant for the Pitt and Lee model.
[92] A regulator may not want to design a regulatory framework where there is a 1:1 correspondence between profits and costs or quality, because this places excessive risk on companies and there is insufficient sharing of benefits with customers.

Box 3.3 NVE's experience in translating results from a DEA model into allowed revenues

This box explains NVE's approach in translating the results of a DEA model with quality (included by adding VOLL into a DSO's TOTEX) into allowed revenues. A key feature of this approach is that a 1% reduction in VOLL will lead to the same rise in profits as a 1% reduction in TOTEX. As a result, companies do not receive a disproportionate incentive to focus on either TOTEX or the quality measure.

Stylised depiction of NVE's allowed revenues



Note: '…' is the sum of pass-through costs, which are considered to be outside of company control, and a mechanism for removing the time lag for investments. We do not consider these issues as they are outside the scope of this report, and we refer the interested reader to the cited paper.

Source: Amundsveen, R., & Kvile, H. M. (2016), 'Balancing incentives: The development and application of a regulatory benchmarking model', In Productivity and Efficiency Analysis (pp. 233-247). Springer, Cham.

The figure above provides a visual depiction of the regulator's approach. The revenue cap is calculated as the weighted sum of the cost base (which is the DSO's actual TOTEX + actual VOLL) and cost norm (which is the DSO's efficient TOTEX + efficient VOLL predicted by the DEA model). We note that there are other elements in the cost base, such as a regulatory rate of return. We do not consider these in further detail for the purposes of this report, and refer the interested reader to Amundsveen and Kvile (2016) as cited.

The parameter ρ represents the extent of risk-sharing between the company and its customers. The higher ρ is, the greater the risk borne by the company, and the stronger the incentives. For example, if ρ is 100%, the companies' allowed revenue is completely independent of their own costs, giving very strong incentives for cost efficiency.[93]

The important point to note here is that the efficiency challenge is applied to both TOTEX and the VOLL because the VOLL is treated as if it were part of TOTEX. As a result, the company receives the same profit from reducing VOLL by one unit as it does from reducing TOTEX by one unit. The company can choose to achieve the 10% efficiency challenge on (TOTEX + VOLL) in a way which is most profitable. For example, if a unit fall in VOLL is very expensive to achieve, then this improvement in quality is not worth the costs needed to achieve it, and the company will instead reduce TOTEX. As a result, the company's profit motive is aligned with its customers' willingness to pay for quality.

As a general principle, it is important that allowed revenues are calculated in such a way that companies have equal incentives both to reduce costs and to improve quality. This is to avoid companies having unbalanced incentives and choosing to focus only on costs or on quality or both but sub-optimally.

E-Control have also considered including VOLL directly within TOTEX (as described in sections 3.3.1 and 3.4.1). While they have not implemented this

[93] NVE has set the value of ρ to be 60%.

into their regulatory framework, E-Control's experts described that their ideal framework would consist of compensation payments to each network user who experiences a power outage with the VOLL. The regulatory cost base would then consist of TOTEX + VOLL (also known as SOTEX) in order to reimburse operators. Network operators might receive a "budget" for compensation payments for the first regulatory period when the quality dependent system is applied. Subsequently they can either outperform their targets (with a quality improvement) or observe a lower return if the "budget" is insufficient due to a quality deterioration.

If quality is included separately in a DEA model or an econometric approach, the interpretation of efficiency scores does not differ between models including or excluding quality. As a result, the translation of benchmarking results with quality to a revenue cap can be done in the same way as benchmarking results without quality. This is consistent with Energiavirasto's experience, where it explained that, in earlier regulatory periods where the cost of outages was included directly in TOTEX, a separate method of calculating the revenue caps is used compared with the current approach, where quality is included as a separate output.

When quality is included in benchmarking by adding it to TOTEX, the interpretation of the efficiency score differs relative to a quality-unadjusted approach. NVE has accounted for this in its calculation of allowed revenues.

However, when quality is included separately from TOTEX, efficiency scores can be interpreted in the usual way as if quality had not been included, and so the calculation of the revenue cap is unaffected.

3.6 Stakeholder reactions

Some discussion of stakeholder reactions was included in the above sections, particularly where these reactions have influenced the regulator's current choice of benchmark. This section presents further evidence of stakeholder reactions to different forms of quality regulation.

3.6.3 E-Control

While E-Control does not currently implement quality within its cost benchmarking approach, it has undertaken significant work exploring this as an option.

E-Control's experts combined DEA and econometric results (MOLS[94]) by weighting each approach in order to arrive at a view of each company's efficiency score and efficient costs.[95] They compared these results against models where quality had not been accounted for. Table 3.3 presents the mean, median, minimum and maximum efficiency scores across all electricity distribution network operators in Austria.

[94] Modified ordinary least squares, where estimated OLS parameters are adjusted to account for the presence of inefficiency in the error term.
[95] A weight of 45% is placed on the MOLS approach, 40% on one DEA model, and 15% on another DEA model.

Table 3.3 Efficiency scores estimated by E-Control from combining DEA and SFA: comparing cost benchmarking approaches that exclude and include quality

	Excluding quality	Including quality
Mean	91.34%	92.27%
Median	94.30%	94.92%
Minimum	70.90%	74.95%
Maximum	100.0%	100.0%
Observations	36	36

Source: E-Control (2014), 'Optionen zur Einbeziehung der Versorgungsqualität in derzeitige bzw. künftige Regulierungsrahmen für Stromverteilernetzbetreiber', Figure 10.

The results suggest that inclusion of quality has an overall positive impact on efficiency scores, as the mean and median efficiency scores have both increased. Indeed, 24 DNOs have a higher efficiency score when quality is included, while only 12 have lower efficiency scores. As expected, it finds that companies with an above-average reliability of supply benefit from the inclusion of quality in benchmarking.

However, when considering the total *monetary* value of inefficiencies, E-Control's experts find that when quality is included, total inefficiency had actually increased. Indeed, the largest fall in efficient costs was €6.6m, while the largest rise was only €1m. Across the entire industry, total inefficiencies in the approach including quality were higher by €13.8m.

E-Control's experts stated that the largest and most influential companies experienced the largest losses. These companies have the almost common opinion not to adopt quality-dependent cost benchmarking and hence the industry remains against any quality regulation framework.

3.6.4 NVE

In implementing its quality-dependent DEA approach, NVE has faced some opposition from companies that argued that power interruptions caused by exogenous events outside of the company's control, and therefore the social costs associated with these interruptions, should not be deducted from allowed revenues. NVE allows companies to apply for exemptions if there is an exceptional event. Nonetheless, NVE does not grant exemptions unless the cost of VOLL to the company is so high that it puts the financial health of the companies at risk.

For instance, in a court case, one company argued that an exemption should be granted for interruptions caused by tornadoes, which are an exceptionally rare phenomenon in the area served by the company. However, the court ruled against the company, citing that while the event was very rare, the economic impact was not sufficiently material. This is because the regulatory guidance is that companies should be able to earn a reasonable return on their investments over time if they are operating efficiently. Therefore, the economic impact in a single year needs to be very high for this not to hold.

3.6.5 Energiavirasto

Energiavirasto considers that its approach has been effective in leading to a higher quality of service. It has tracked the cost of outages over time, and finds that it has been falling year on year. Energiavirasto considers that this is due to the combined effect from the security of supply obligations set in legislations, the quality incentives outside benchmarking, and the benchmarking approach itself. Energiavirasto notes that the legislative security of supply obligations are likely to have the greatest effect in improving quality and decreasing outages.

Energiavirasto noted that companies found it difficult to understand the impact of the cost of outages on their costs when it was included in the StoNED model, and in particular, the intuition behind whether there was a positive or negative shadow price.

A number of companies appealed to the Market Court in Finland concerning the 2012–15 price control period. In particular, several network companies criticised the inclusion of interruption costs, arguing that they were outside of a company's control. However, Energiavirasto instead takes the view that companies should be able to prepare for the likelihood of interruption events. The Market Court found in favour of Energiavirasto. In its decision, it stated that, even though interruption costs can be affected by factors such as extreme weather, the actions of the companies still have a substantial effect on them.

Another issue explored in the court case was that Energiavirasto raised the cap of the quality incentive *outside* of cost benchmarking from 10% to 20%. Companies considered that this represented an increase in risk that was not compensated for elsewhere within the regulatory framework. However, the Market Court found that the impact of raising the cap on the quality incentive affected the companies' efficient level of profits by only -0.15%, and therefore could not be considered to increase risk for the sector as a whole.[96]

3.6.6 Experiences of regulators with quality regulation outside of cost benchmarking

This section considers the experiences of regulators that have included quality within the regulatory framework but not within their cost benchmarking approach.

Ofgem regulates quality for gas distribution networks in two ways: by implementing a minimum standard of quality, and by providing rewards and penalties based on selected performance measures (see appendix A1.7 for further details). Ofgem acknowledges that minimum standards do not provide incentives for companies to go beyond the minimum required standard. It also noted that the overall reward/penalties associated with quality incentives tend to be quite small, at approximately £20/customer. However, Ofgem has nonetheless used evidence on quality measures to challenge GDNs that have been performing poorly. Ofgem stated that survey evidence suggests that customers believe that, overall, UK GDNs provide a high level of service.

E-Control explained that in Austria a minimum standard is imposed on SAIDI.[97] However, the regulator's experience is that this has been ineffective because violations of this minimum standard have not led to any penalties.

In England and Wales, Ofwat currently uses cost benchmarking models that exclude quality of service measures. As a result, companies have argued that this approach funds only the average industry performance on quality. Ofwat has stated that it expects companies to achieve higher standards in future (e.g. on leakage) without any additional allowance.[98] Some companies that provide

[96] MAO 427-501 / 12.
[97] The average outage duration for each customer served.
[98] Ofwat (2019), 'PR19 draft determinations: Delivering outcomes for customers policy appendix', July, p. 23.

high levels of quality of service and target further improvements (e.g. Anglian Water) have objected that this approach has does not recognise the costs of providing these high levels of service.[99]

The introduction of quality within cost benchmarking (as with any framework changes) has been controversial among companies. In the case of the two regulators that have successfully introduced quality into benchmarking—NVE and Energiavirasto—companies have brought the regulators to court. In both cases, the controllability of interruption costs was an important issue, and the courts found in favour of the regulators.

While E-Control has explored the possibility of quality-dependent benchmarking, there was opposition from the industry.

For regulators that have implemented quality regulation outside of cost benchmarking, there has been mixed experience. Ofgem stated that it had found quality incentives and minimum standards effective in leading to a high level of customer satisfaction. However, other regulators, such as E-Control, have found these to be less effective. Companies in England and Wales have argued that Ofwat's approach does not recognise the costs associated with high levels of quality.

[99] Anglian Water (2019), 'PR19 Draft Determination. Leakage cost adjustment claim', August, p. 1.

4 Summary and conclusion

Our discussions with regulators, our review of their documents, and our work that also involved input from our academic associates, have provided insight on the different methods that regulators have used to integrate quality measures in cost benchmarking.

A key consideration is **whether monetised or non-monetised measures of quality should be used**. The main advantage of using monetised measures is that it reflects the value that customers place on quality (the measure used is in practice is service interruptions). This can then be used to incentivise companies to trade-off the costs of improving quality against the benefits of doing so, leading companies to choose the **optimal quality level**. However, constructing quality dependent measures can be a highly involved process to construct such a measure.

Therefore, **non-monetised measures** may be a viable alternative as they are more readily available. One way of accounting for the large number of measures available might be to create a composite measure, either through statistical methods such as principal components analysis or through consultation with companies. Another consideration is that the measures included need to be considered in light of the model specification used. Modelled relationships should align with operational insight e.g. ensuring that scale is captured in the quality measure used.

We can distinguish the approaches used by regulators to integrate quality in benchmarking in two ways—either as an **input directly within the cost base** or as an **output** (which, if in ratio form, may include an input component)**.** In both DEA and econometric methods, including monetised quality directly into costs incentivises companies to treat the value of quality to customers (measured as the cost of service interruptions) as if it were their own costs. This provides **equal incentives for companies to provide high quality and reduce costs**. In contrast, when including quality as an output in a DEA model, there is a risk that companies can choose to **'specialise' on either outputs or quality to obtain a high efficiency score** (in the absence of weight restrictions in DEA). This is because DEA provides the best view of each company's efficiency.

However, we note that including monetised quality in costs imposes strong assumptions on the relationship between outputs, costs and quality. For example, it would require that providing an additional connection would require a particular cost *or* an equivalent in lost quality.

Some regulators have also explored including non-monetised measures as outputs in benchmarking. When specifying econometric models, it is important that estimated relationships are **operationally intuitive**. This has proved challenging for the regulators we considered.[100] However, there are approaches that can be explored, such as carefully constructing composite quality measures. Furthermore, an additional issue is the risk of endogeneity, which needs to be balanced against the benefits including quality in benchmarking (e.g. in accounting for important differences between companies).

The translation of benchmarking results into a revenue cap only **needs to be modified if monetised quality is included with costs as a single input within cost benchmarking**. The key principle is that the overall regulatory

[100] Ofwat, ARERA and Monitor.

framework needs to treat monetised quality as if it were the company's own costs.[101] However, if quality is not included as a single input with costs, translating results into revenue caps can be done in the usual way.

In response to the introduction of quality in regulation, companies have raised a number of different considerations depending the regulator's specific approach. However, the themes that emerged include the controllability of quality measures, palatability of outcomes, balance of risk and transparency of the proposed approach. Therefore, these may be useful issues to carefully consider when integrating quality within cost benchmarking.

[101] NVE does this by deducting the VOLL from allowed revenues

A1 Summary of interviews

We provide summary of interviews with each regulator below. We emphasise that these views are those of the interviewees and not necessarily the views of the regulators in question.

A1.1 NVE

The NVE regulates water and energy in Norway. Our discussions with NVE focused on the regulation of energy distributors. NVE uses an input-oriented DEA model with three outputs (number of customers, number of substations and the length of high voltage network) and one input (TOTEX). It then applies a second stage environmental correction, where DEA scores[102] from the first stage are regressed on the environmental (or Z-) variables.[103] It notes that many environmental variables are highly correlated and therefore, utilise factor analysis to create a composite variable from these variables. The adjusted DEA results are then used to determine allowed revenues as described in section 3.5.

NVE explained the history leading up to the introduction of quality in regulation. In the 1990s, it was concerned that because there were only incentives to be cost effective without consideration of quality measures, companies may fail to re-invest or spend enough on maintenance, leading to poor levels of quality. NVE notes that it would like each company to choose the optimal level of quality, and sought to design a regulatory approach that encourages companies to choose the optimal level of outages (from the perspective of society as a whole).

Its approach relies on calculating VOLL, which a monetised measure of service interruptions capturing the welfare losses experienced by customers in the event of a power interruption. Over time, this approach has developed from using a single cost per kWh within each of four customer categories independent of the duration of the interruption and times to an approach where VOLL is based on a variety of factors, such as the duration, time of day, and the day and month of the year.

NVE noted that it has undertaken a significant amount of work in developing the VOLL. Before its introduction, the regulator undertook three public consultations. It noted that there was pushback from the industry, which saw it as a punishment for outages. As a result, in 2001, it adopted an approach such that revenues were only lost if the actual cost of outages was above what was expected for the company, while if the actual cost of outages was below expectations, companies could collect more revenues. The expected cost of outages is based on two components:

- 50% based on a historical average;
- 50% based on an expected value based on a regression model.

[102] In Amundsveen and Kvile (2016), NVE note that they correct DEA results for bias using bootstrapping, which they state meets some criticism of the serial correlation of DEA-scores by Simar and Wilson (2007). See Amundsveen, R. and Kvile, H.M. (2016), ‘Balancing incentives: The development and application of a regulatory benchmarking model’, Productivity and Efficiency Analysis, Springer, pp. 233–247.
[103] They further note that since 2013, the second stage approach was improved such that the independent variables are not the Z-variables themselves, but the difference in the Z-variable for the DSO itself and its shadow company in stage 1. The shadow company’s Z-variables are calculated by applying the weights from DEA.

This benchmark was set for five years and was updated each year based on whether the companies underperformed or outperformed relative to this benchmark. This approach was used until 2007.

Its subsequent approach after 2007 treats VOLL as if it were the company's own costs. It stated that one reason driving this change is the change in regulatory approach before and after 2007, where previously if a company's costs reduced by €1m, its profit would rise by €1m.[104] After 2007, a €1m reduction in costs would raise profits by only €0.6m. However, it notes that the previous approach to regulating quality gave incentives such that a €1m fall in VOLL would still lead to a €1m rise in profits, leading to stronger incentives to focus on quality rather than costs. In order to avoid providing companies with incentives to focus only on either costs or quality, it integrated VOLL directly within benchmarking.

It notes that its approach relies on having a monetised measure of quality. While it acknowledges that there are likely weaknesses in estimating VOLL, it states that the resulting model outputs are ultimately more reliable than if VOLL had not been adjusted for.

Although there has been significant discussion on the impact that other quality factors have on costs, NVE states that it has not been able to find an appropriate estimate of the value of these other quality measures. As a result, it has imposed a minimum level of quality for some quality measures. NVE stated that while quality measures as an output within DEA is another alternative, it was not straightforward to do so.

NVE discussed some further considerations with its approach. First, NVE notes that whether or not quality is included as an input or output is only important insofar as it is viewed in the context of incentivising companies to make an optimal choice.

Second, NVE raised that there are environmental factors that can affect the possibility of having an outage, such as strong winds and sloped areas. NVE states that its second stage accounts for these differences in environmental and operating conditions. However, this is an area that there is still ongoing work to better capture these differences between companies.

Third, NVE states that another issue is that its approach to calculating VOLL companies may have highly heterogenous customer mixes. VOLL currently distinguishes between six types of customers, with some having a higher demand for a secure supply than others. It is possible for one company to have significantly more of a customer type that demands a high security of supply (and therefore, has a higher VOLL) compared to another company which may have customers that has a lower demand for security of supply. To meet a higher demand for security of supply, there is a greater need frequently maintain and more strongly reinforce the grid, leading to additional costs.

As a result, NVE is currently exploring the inclusion of an output parameter which reflects the demand for a secure supply within the DEA model. This output parameter is constructed using the function used to calculate VOLL (which distinguishes between different customer types, time of days, etc.). However, instead of using the actual duration of interruptions, it calculates the value assuming some fixed length of interruptions for all companies (e.g. what the VOLL would be had there been 12 hours of interruptions). This output

[104] In the absence of any other changes.

parameter would be included while keeping the *actual* VOLL (i.e. based on the actual duration of outages) within TOTEX as an input.

It finds that this approach works relatively well, with results remaining reasonable and that the magnitude of the change in results are intuitive.

NVE further discussed a number of challenges associated with constructing a monetised measure of quality.

- **Some quality measures are more intuitive than others—**since constructing the monetised measures rely it found that some quality measures are more difficult to intuitively understand compared to others. For example, it is easier for those customers to place a value on the cost on a power outage of a particular duration, rather than other quality measures such as an optimal level of voltage. Furthermore, it notes that while companies can better value the impact of voltage fluctuations, it notes that these are highly company specific, and therefore difficult to obtain figures that apply generally.
- **Cost of designing a survey**—the design and testing of the survey is costly, involving pilot studies, focus groups and further re-designs.
- **Large sample sizes required—**a high proportion of responses may not contain reasonable results. Some respondents might state that they do not experience any costs associated with lost loads, but answer other questions in a way that is inconsistent, suggesting that outages are costly. As a result, a fairly large sample is required at the initial stages of the survey to ensure that a reasonably large sample of usable responses remain. The regulator stated that during the 2017 update of the VOLL, they began with a representative sample of 4,400 drawn from a database of approximately 100,000 people to contact, but received only 2,200 responses where only 1,350 had reasonable answers.

NVE then discussed stakeholder reactions to its approach. It noted that companies initially argued for an arrangement where if the VOLL was caused by factors outside of the companies control, they would not suffer losses in revenues. However, NVE's approach was to not provide exemptions unless the cost to the companies was so high such that it would put the financial position of the company at risk. It notes that if it started to grant exemptions, then companies would submit a large number of claims.

A1.2 Energiavirasto

Energiavirasto is the Finnish energy regulator. It has introduced quality within DEA and SFA cost benchmarking since 2008 using DEA and SFA, before using the StoNED method after 2011. A summary of Energiavirasto's benchmarking approach is given in Table 3.2 in section 3.3.2.

From 2008 to 2011, Energiavirasto used both DEA and SFA in determining efficient costs for companies. These models consisted of three outputs (distributed energy, number of customers and network length) and one input, which is the sum of TOTEX and a monetised value of outage costs. This approach was also used by NVE, as described above.

However, Energiavirasto already had other quality incentives in place, leading companies object that Energiavirasto was double regulating quality. Therefore, in the regulatory period lasting from 2012 to 2015, Energiavirasto changed the definition of costs used in benchmarking to include only half of outage costs, rather than the full value of the cost of outages.[105] The regulator also switched to using the StoNED method, which in principle combines both the merits of SFA and DEA.

However, in 2010 and 2011, there were large storms that caused significant power outages. This increased the cost of outages significantly such that it led to unintuitive results. The worst affected company experienced an outcome where the total value of the cost of outages exceeded its efficient level of costs by a factor of 15.

Therefore, in the current and following regulatory period, Energiavirasto switched to including the cost of outages as an undesirable output rather than as an input. It explained that the StoNED method automatically assigns weights the cost of outages. It finds that the weights assigned are intuitive such that in years where there are many storms causing outages, the weight on cost of outages is positive, leading to additional allowances for companies to cope with these large storms. However, for years and companies with few storms, the weights are negative, providing an incentive for companies to reduce their costs.

It also provided intuitive reasons for using the cost of outages as an output rather than an input. Since avoiding outages increases costs for DSOs and is not an essential factor of production used to produce outputs, it is better modelled as an output rather than an input.

However, since there are already three other outputs in its model specification, Energiavirasto explains that adopting cost of outages as a fourth output variable limits the impact of including quality measures within cost benchmarking. This is because the model places a low weight on the quality measure relative to the other output measures.

Despite these considerations, Energiavirasto notes that the cost of interruptions has been reducing over time. This is because of the combined effect of legislated security of supply obligations, its quality incentives, and its cost benchmarking approach. Energiavirasto notes that the security of supply legislations are likely to have had the greatest effect in improving quality and reducing outages.

It notes that this approach has generated debate among companies. In particular, companies expressed that it is not clear how the StoNED method calculates the weights, and why these weights take on positive or negative values. Energiavirasto anticipate that there will be further discussion on this topic in the next price control period.

Energiavirasto discussed the construction of its monetised cost of outages that it uses in its quality incentive (i.e. outside of cost benchmarking). It explained constructing this measure was a process lasting 7–8 months, which involved getting the relevant data and coming to a view what of what the unit cost of outages should be. It notes that this unit cost only accounts for one set of customers, and does not differ by customer demographics (e.g. industry vs residential). Nonetheless, it stated that it has been challenging to construct other viable alternatives. Energiavirasto will be undertaking a methodological review next year, where it will consider these issues further.

Finally, we discussed whether Energiavirasto is considering introducing further service metrics within benchmarking. Although Energiavirasto's current

[105] It also began to use OPEX rather than TOTEX.

methodology is applicable until the year 2023, it aims to continue developing its methodology in line with changing market conditions. For example, in 2018, Energiavirasto ordered a survey concerning regulatory methods looking at demand responses in 2018. The regulator will comprehensively review its methodology for upcoming regulatory periods in 2020.

A1.3 E-Control

E-Control regulates electricity and gas networks in Austria. It states that while it has considered including quality measures within cost benchmarking, it has not formally introduced it within the regulatory framework in Austria. E-Control’s experts published a study/opinion on the introduction of quality regulation in the regulatory system. However, this is not an official E-Control position.

E-Control’s experts suggest that a ‘total social cost’ approach to incorporating quality measures should be adopted when considering quality-dependent cost benchmarking. This would involve monetising a quality measure (specifically, VOLL) and adding these losses to the cost base of the utility. It states that within DEA, including quality as an output is problematic because it provides incentives for companies to specialise in either outputs or quality in order to gain an efficiency score of 100%.

An additional consideration is that for quality measures to be included in DEA, they need to reflect the size of the utility. As a result, E-Control’s experts state that ratios should not be used as an input or output.

E-Control’s experts have also considered econometric approaches to estimating efficiency. Ideally, all outputs included as cost drivers in an econometric model should not be controllable by the utility (i.e. it should be exogenous). Since quality can be influenced by management, it has decided to include monetised quality measures within costs rather than as an output.

E-Control’s experts have explored the impact of including quality within cost benchmarking on companies. They note that 24 companies receive an improved efficiency score while 12 others see a worse outcome. While there are more ‘winners’ than ‘losers’ from the inclusion of quality in benchmarking, they note that if these changes in efficiency scores are monetised, the total level of allowances across the industry reduces. In other words, the inclusion of quality within cost benchmarking leads to an overall increased level of inefficiency across the industry.

Furthermore, it found that the largest and most influential companies would experience the largest fall in cost allowances in the industry and had nearly all of these companies were against the introduction of quality-dependent cost benchmarking.

E-Control explained that the industry remains against any quality regulation framework.

E-Control’s experts note that while they have minimum quality regulations, these have not been effective. In principle, the level of SAIDI is regulated (the average outage duration for each customer served) such that it should not exceed a normative set value equal to all DSOs on average over the course of the last three years. Despite DSOs violating the requirement, E-Control’s experts note that there have been no financial penalties.

We also discussed whether E-Control is considering using other measures of quality. E-Control's experts note that another candidate measure would be the frequency of outages, but that this is difficult to quantify monetarily, and that quantifying interruption costs is already a difficult task.

A1.4 ERSAR

ERSAR does not currently use either a DEA or an econometric approach to benchmark company costs. ERSAR defines efficiency metrics for the ten regional companies that operate in Portugal for the main OPEX cost lines.[106] On service quality, currently uses a sunshine regulation approach with 14 quality of service indicators each for water and wastewater (see appendix A2 for further details). Using these indicators, it assesses companies' performance using a traffic light system (i.e. green corresponds to outstanding performance, while red corresponds to poor performance).

Going forward, ERSAR intends to combine its quality measures into a composite variable in order to set an incentive. If a company performs well across a number of quality measures, then ERSAR would allow the company to collect a premium from its customers. However, it explained that the exact details of this incentive is still subject to further ongoing work and internal discussions.

We discussed whether ERSAR is considering implementing cost benchmarking using a DEA and/or an econometric approach. The Portuguese regulator is looking to implement cost benchmarking using a DEA and/or econometric approach in the near future. It notes that even if it introduces cost benchmarking, it will still maintain its current traffic light system used in assessing quality.

ERSAR explained that it has encountered a number of issues in attempting to introduce cost benchmarking. First, the industry structure has changed significantly over time, starting with 18 companies, which then merged into five and later split into ten. As a result, obtaining consistent data is a challenge. Second, it notes that companies are not sufficiently familiar with more sophisticated approaches, which is why ERSAR has opted to use an intuitive and easy to understand traffic light system (and will continue to do so). Third, it finds that companies have highly heterogenous operating environments, and that it is difficult to control for these heterogeneities.

A1.5 Ofwat

Ofwat explained that it currently keeps the evaluation of costs and performance outcomes separate. The current approach relies on the assumption that cost allowances provide sufficient funds for companies to reach quality targets. If companies outperform these targets, they will be able to receive outperformance payments.

However, Ofwat notes that there are legacy factors that might limit the ability of certain companies to reach common performance commitments (such as the effect of asset age on leakage). For these companies, Ofwat sets a company specific target. As a result, while the evaluation of costs and outcomes are separate, they remain linked at the policy level.

[106] These are the cost of purchasing reagents, energy consumption per cubic meter of water (water and sanitation), personnel costs, infrastructure maintenance costs, and an indicator for annual OPEX savings as a whole.

Ofwat states that it would like to explore the possibility of incorporating quality within cost benchmarking.

However, Ofwat notes that there are a number of challenges associated with incorporating quality within an econometric cost benchmarking model:

- there are a large number of potential quality measures. These tend not to be primary cost drivers, and therefore it is difficult to obtain statistically significant results. Ofwat suggest that one potential approach may be to create a composite variable capturing a range of factors. The measures to be included could be identified through a consultation with companies, and then weighting them together to form an index. Ofwat consider that the difficulty in obtaining statistically meaningful models is the main reason why quality measures have not yet been included in its econometric models;
- there is likely a complex relationship between quality and costs. Companies that provide a high quality of service would like need to spend significant resources to improve quality further, while a company with poor levels of quality may only require fewer resources to achieve the same increase in quality. These may be captured within an econometric model using a non-linear specification;
- endogeneity of quality measures is an issue. Ideally, cost drivers included in a model should be exogenous. However, quality measures can be influenced by company management and therefore their inclusion in a model can create an endogeneity issue. This needs to be balanced against the risk of an omitted variable bias that can result from failing to include these within cost benchmarking models. Both endogeneity and the omission of relevant cost drivers can lead to biased residuals and distorted efficiency scores.

A1.6 ACM

The ACM regulates energy in the Netherlands using revenue cap regulation. While the ACM includes quality within its regulatory framework, it does not include quality directly within cost benchmarking. The ACM regulates quality in two ways.

First, the ACM incentivises provides quality incentives using a Q-factor. The ACM carries out surveys to place a money value on outages. If companies perform better than the average,[107] then it receives a positive Q-factor, leading higher revenues. However, if it performs worse than average, then it loses a share of revenues through a negative Q-factor.

The ACM notes that surveys are expensive to carry out. The VOLL calculated has been updated twice, with 5–6 year intervals between each survey. They note that customers' willingness to pay for quality is unlikely to change significantly each year, allowing a trade-off to be made between the costs of carrying out the survey and the frequency with which the surveys are carried out.

Second, the ACM regulates quality using minimum requirements. A number of different measures are monitored, including how quickly companies respond to complaints, voltage levels, voltage dips, number and length of outages. Depending on the potential harm to customers, some measures are monitored

[107] The average value that consumers and small and medium-sized enterprises attach to quality is used as the benchmark for quality performance of system operators. See ACM (2017), 'Incentive regulation of the gas and electricity networks in the Netherlands', May, p.12.

more closely than others. For example, power quality has an additional monitoring process, where the results of the monitoring are discussed with TSOs and DSOs. Even if these indicators remain within the acceptable range but are deviating from the desired values, the ACM challenges companies to ensure that the quality measures are close to the desired values.

We discussed whether the minimum level of quality may be ratcheted i.e. increased over time. The ACM notes that the minimum level of quality associated with some quality measures may change over time. For example, it explained that newer appliances are more sensitive to fluctuations in power quality, and so regulation needs to be updated to reflect this. However, on other measures, it states that is less clear that standards need to improve over time (providing an example of whether percentage of methane and hydrogen in gas distribution need to be updated over time).

The ACM explained that the method of quality regulation to be used depends on the specific quality measure. They suggested that it is easier to determine a minimum level of standard for some quality measures rather than others. For example, it is difficult to pin down a minimum standard on a measure such as the number of outages. As a result, the ACM allows DSOs to trade off costs and quality on these measures. In gas distribution, the ACM prioritised having a safe network, and therefore did not design a regulatory framework that encouraged companies to trade off costs and safety.

Furthermore, it notes that other quality measures may have large fluctuations from year to year, making it difficult to include it directly within the tariff calculation using a Q-element.

The ACM states that it has no plans to integrate quality within cost benchmarking for practical reasons. They considered that its current method was easier than inclusion within cost benchmarking. Furthermore, its view is that the DSOs in the Netherlands are relatively similar, and so quality can simply be accounted for using a scaling factor rather than integrating it within cost calculations. If there are very atypical events that affect specific companies, the ACM will separately investigate (for example) if the company had taken appropriate actions in response to such events.

On stakeholder reactions to the introduction of quality regulation, the ACM noted that companies generally did not react positively. Companies did not feel that money invested in improving quality could be regained through the regulator’s approach. Furthermore, the ACM notes that companies likely prefer business-as-usual circumstances and so were opposed to changes brought about due to the introduction of quality regulation.

A1.7 Ofgem

Ofgem explained that it primarily incentivises high levels of service quality in gas distribution through incentive mechanisms. As a backstop for these incentive mechanisms, Ofgem also implements guaranteed minimum standards of performance. If these standards are not met, then companies would have to pay consumers directly.

Ofgem has found the approach of using minimum standards to be reasonably effective, with positive feedback from industry and customers. On certain quality measures, such as interruptions, where some GDNs have performed particularly poorly, Ofgem has been able to use these service quality metrics to make a case against these GDNs. However, it notes that the sizes of the payments to consumers are quite small, at only approximately £20 per person, or £20m in total over the previous price control period. Over the next price control, Ofgem is looking to tighten standards and increase payment levels.

The primary way that Ofgem incentivises quality is through the broad measure of customer satisfaction. This comprises of three measures, namely:

- **customer satisfaction**—companies that perform well or poorly receive a reward or penalty of $\pm$0.5% of base revenues. Targets for RIIO-1 are set using the industry upper quartile, based on a survey undertaken before RIIO-1;[108]
- **stakeholder engagement incentive**—companies that perform well receive a reward of +0.5%;
- **complaints handling—**companies that perform poorly receive a penalty of -0.5% of TOTEX.

Overall, Ofgem told us that its survey results indicate that customers are generally happy with the quality of service provided.

[108] RIIO-GD1 is the price control period for gas distribution networks, running from 2013 to 2021.

A2 List of quality measures

Ofwat has set 14 common performance commitments for PR19 (the price control for the year 2020–25), shown in Table A2.1.

Table A2.1 Ofwat's common performance commitments, and measures used

Performance commitment	Measure
Customer experience	Customer measures of experience (C-MeX) based on survey information
Developer experiences	Developer services measures of experience (D-Mex) based on survey information
Water quality compliance	The DWI's Compliance Risk Index (CRI)[109]
Leakage	Leakage in megalitres per day, three-year average
Per capita consumption (PCC)	Average amount of water used by each person that lives in a household property (litres per head per day)
Internal sewer flooding	The number of internal flooding incidents per year (sewerage companies only)
Pollution incidents	Category 1 – 3 pollution incidents per 10,000km of sewerage network, as reported to the Environment Agency and Natural Resources Wales
Risk of severe restrictions in a drought	Percentage of the population the company serves that would experience severe supply restrictions (for example, standpipes or rota cuts) in a 1 in 200 year drought
Risk of sewer flooding in a storm (new risk-based resilience metric – wastewater)	Percentage of population at risk of sewer flooding in a 1-in-50 year storm (medium and high risk properties).
Mains bursts as a measure of asset health for water	Mains bursts per 1,000km of mains
Unplanned outages as a measure of asset health for water	Reduction in peak week production capacity multiplied by the duration in days, divided by 365.
Sewer collapses as a measure of assert health for wastewater	Sewer collapses per 1,000km
Treatment works compliance	A composite variable consisting of a number of indicators, such as total pollution incidents, serious pollution incidents and delivery of environmental outcomes.

Source: Ofwat. See https://www.ofwat.gov.uk/outcomes-definitions-pr19/ for further details.

ERSAR's indicators used to asses drinking water supply services are shown in Table A2.2.

[109] See Drinking Water Inspectorate (2018), 'DWI Compliance Risk Index (CRI)', August' for further details.

Table A2.2 Quality measures considered by ERSAR

Drinking water supply services	Wastewater management services
Service coverage	Service coverage through sewerage networks
Service affordability	Service affordability
Service interruptions	Flooding occurrences
Safe water	Reply to written suggestions and complaints
Reply to written suggestions and complaints	Cost recovery ratio
Cost recovery ratio	Connection to the service
Connection to service	Sever rehabilitation
Non-revenue water	Sewer collapses
Mains rehabilitation	Adequacy of human resources
Mains failures	Standardised energy consumption
Adequacy of human resources	Accessibility to wastewater treatment
Real water losses	Emergency discharge control
Standardised energy consumption	Compliance with discharge permit
Proper sludge disposal	Proper sludge disposal

Source: ERSAR, 'Assessment of the quality of service provided to users by water utilities in Portugal: 3rd generation of the assessment system'.

A3 Ratio data in DEA

This appendix provides a numeric example of issues with DEA when using 'non-standard' data such as ratios.

Consider two efficient companies A and B which form the frontier in the wastewater industry. These companies produce two outputs, total treated sewage load and % of sewage treatment works complying with a particular standard, using one input, TOTEX. The percentage of compliant works is calculated as the ratio of the total number of works and the number of compliant works. This is summarised in Table A3.1 below.

Table A3.1 Hypothetical data illustrating underlying DEA calculations with ratio data

		Company A	Company B	Virtual company formed by standard DEA (0.5 * (A+B))	Appropriate virtual company
Input	TOTEX	1	1	1	1
Output 1	Total sewage load	100	50	75	75
Output 2	% of compliant load	40%	70%	55%	50%
Underlying quality data	Volume of compliant load	40	35		37.5
	Total load	100	50		75

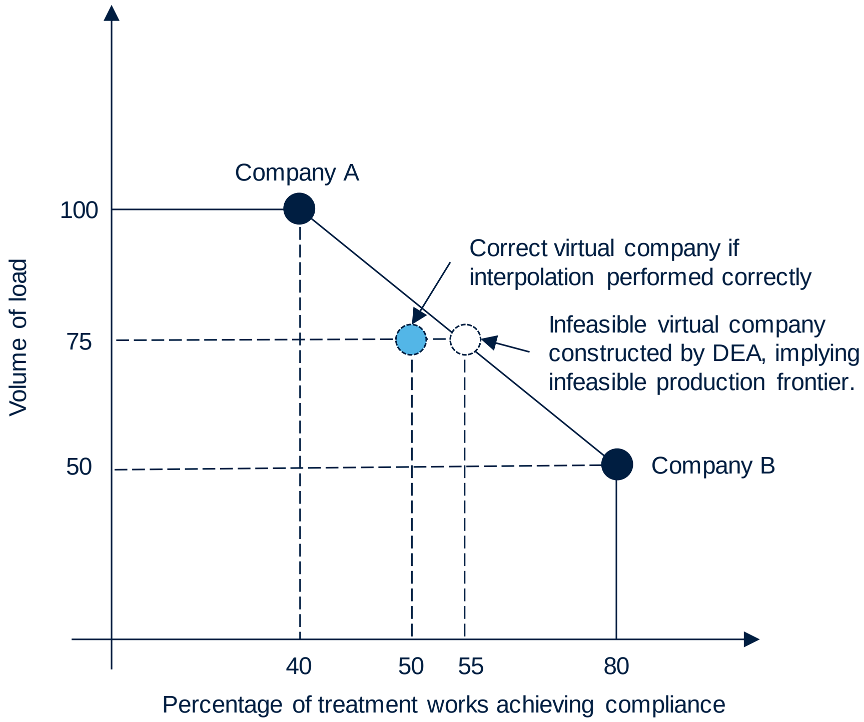
Source: Oxera.

DEA relies on the assumption of convexity, which allows virtual peers to be formed through convex combinations[110] of existing companies. That is, if companies A and B are feasible, then DEA assumes that the company formed from creating a weighted average of company A and B's inputs and outputs should also be feasible. This is shown in the column labelled 'virtual company (0.5*(A+B)) formed by standard DEA', where the virtual company has 55% of compliant works, load of 75 and TOTEX of 1.

However, this assumption does not hold with ratio data. Interpolating companies A and B correctly would yield a virtual company with 50% of compliant works with the same level of load and TOTEX. This appropriate virtual company has a lower level of quality than the virtual company identified by using the standard DEA approach. Therefore, using the standard DEA approach, an infeasible frontier can be identified, as shown in Figure A3.1.

[110] That is, weighted averages whose weights are non-negative and sum to one.

Figure A3.1 **Illustration of infeasible frontier estimated by DEA with ratio data**



Source: Oxera, based on Emrouznejad, A. and Amin, G.R. (2009), 'DEA models for ratio data: Convexity consideration', *Applied Mathematical Modelling*, **33**:1, pp. 486–498.

www.oxera.com

oxera
compelling economics