Wales & West House
Spooner Close
Celtic Springs
Coedkernew
Newport NP10 8FZ

Tŷ Wales & West
Spooner Close
Celtic Springs
Coedcernyw
Casnewydd NP10 8FZ

Telephone/Ffôn: 0800 912 29 99
Fax/Ffacs: 0870 1450076
Email/Ebost: enquiries@wwutilities.co.uk
www.wwutilities.co.uk

RIIO Team
Network Price Controls
Systems & Networks
10 South Colonnade
Canary Wharf
London

By Email: RIIO2@ofgem.gov.uk

23rd August 2019

RIIO-2 tools for cost assessment Consultation – WWU response

Dear RIIO Team,

We welcome the opportunity to provide our views on the RIIO-2 tools for cost assessment Consultation.

The responses we provide in this document build on our response to the Ofgem RIIO-2 Framework consultation, and significant participation across the RIIO-2 workgroups throughout 2018 and 2019 to date. We will continue to support further RIIO-2 workgroups through to the conclusion of this consultation process.

Our response is structured as follows:

1. This cover letter
2. An executive summary of our key points
3. Responses to the RIIO-2 tools for cost assessment consultation

We have followed the numbering convention utilised within the consultation documents, responding where necessary. Our response is marked as not confidential and may be published in full. Should you have any queries on the responses please do not hesitate to contact me.

Yours sincerely

Steve Edwards
Director of Regulation & Commercial

Contents

WWU RIIO-2 tools for cost assessment response Executive Summary

We welcome and support the use of the cost assessment toolkit in setting RIIO-GD2 allowances. It is key we reflect and consider not only the statistical outputs of the cost assessment modelling but also include practical and industry knowledge to the results.

The proposed model selection criteria is appropriate and aligns with other utility cost assessment practices.

- Economic/technical rational
- Transparency and ease of interpretation for stakeholders
- Robustness of data

These three broad criteria agree to our internal assessment process and are the most important questions to be asked when setting the cost allowances.

Ensuring transparency for stakeholders helps to interpret results and understand the cost drivers of the gas distribution networks (GDNs). Thus, in turn allows us to drive efficiency within the network.

Given the small population of GDNs, panel data is important, this should reflect size and economies of scale of a network and helps to discount inefficiency of GDNs.

Considering the issue with data points, consistency of input data across GDN's is a large focus before even assessing cost model's suitability. If errors occur pre-input we could run the risk of 'cherry picking' cost categories. Using the Totex aggregated modelling helps to keep the risk lower if the composite scale variable is reflective of Totex cost drivers.

We would urge you to consider a detailed review of data consistency due to the impact it would have undermining the benchmarking and we will continue to have poor model fits.

Real price effects and ongoing efficiency should be reflective of relevant levels of materiality and always linked to external, relevant and independent evidence.

Regional factor claims need to be sufficiently justified and clearly incremental to cost base and assessed on a case by case basis.

Once responses are submitted and reviewed the working groups between September 2019 and June 2020 to further to improve data consistency and to help assess the correct tools for the cost allowance setting process will be critical.

Approach to econometric analysis

Q1 - What model estimation options should be considered for our cost assessment and why?

In its consultation document for RIIO-GD2,[1] Ofgem discusses four model estimation options:[2]

Corrected ordinary least squares (COLS):[3] Ofgem used COLS in GD1. It involves shifting a (pooled) OLS regression line to some point, in an attempt to separate noise (e.g. modelling errors or data errors) from inefficiency. Ofgem (and Ofwat) have used upper quartile and upper third benchmarks previously. This technique *assumes* the difference between the shifted regression line and the actual observation to be inefficiency and is based on judgement as to where to shift the regression line to account for modelling errors.

Random Effects (RE):[4] While comparable to the pooled OLS estimator, the RE estimator can take into account unobserved firm heterogeneity[5] in the sample when estimating model coefficients. Whether the estimated firm effect represents legitimate differences in efficient expenditure based on unobserved heterogeneity (differences in companies' operating environments), or differences in efficiency, requires judgement.[6] To that end, similar to pooled OLS, the RE estimator requires an assumption to derive efficient cost predictions. In this regard, Professor Andrew Smith observes that there is an inconsistency in regulatory applications in estimating inefficiency from RE compared to the academic literature.[7] Hence, the treatment of heterogeneity in determining the network operators' inefficiency requires careful review to ensure consistency with the literature and provide evidential support to regulatory discretion on the issue

Stochastic Frontier Analysis (SFA):[8] This technique is an extension of the standard econometric approach and attempts to account for modelling errors by separating the residual into: (i) a symmetric noise component; and (ii) an asymmetric inefficiency component. That is, the standard econometric model is extended with an additional term (the inefficiency term). In order to make this separation, distributional assumptions are typically made (e.g. the inefficiency term is half-normal).

---

[1] Ofgem (2019), 'Consultation – RIIO-2 tools for cost assessment', 28 June.
[2] Ofgem (2019), 'Consultation – RIIO-2 tools for cost assessment', 28 June, para 2.8, 2.18 and 2.19.
[3] Kumbhakar, S., H.-J. Wang, and A. Horncastle (2015), *A Practitioner's Guide to Stochastic Frontier Analysis using Stata*, Cambridge University Press, section 3.3.1.
[4] Baltagi, B. (2013), *Econometric Analysis of Panel Data*, John Wiley & Sons ltd.
[5] That is, differences in underlying characteristics.
[6] Greene (2005), 'Fixed and Random Effects in Stochastic Frontier Models', Journal of Productivity Analysis, Volume 23, Issue 1, pp 7-32.
[7] Smith, A. (2019), 'Note for Ofgem on Alternative Methodologies', June.
[8] Kumbhakar, S., H.-J. Wang, and A. Horncastle (2015), *A Practitioner's Guide to Stochastic Frontier Analysis using Stata*, Cambridge University Press.

Data Envelopment Analysis (DEA):[9] DEA is a mathematical optimisation approach that creates virtual comparators based on weighted combinations of actual comparators. Companies' efficiency is based on the performance of peer companies that have similar characteristics. The approach only requires inputs and outputs to be specified and the returns to scale assumption that determines how the virtual comparators are to be created, but not the functional form relating input (TOTEX) to outputs.

The main advantages and disadvantages of these methods are summarised in the following table.

Table 0.1 Comparison of model estimation options

	Advantage	Disadvantage
COLS	• Relatively simple and requires minimal technical knowledge	• Does not account for noise in the data • Does not account for heterogeneity that is not explicitly captured through cost drivers or pre-modelling adjustments • UQ benchmark arbitrary resulting in under/over-estimation of inefficiency
RE	• Can determine the company effect within the error term	• Consideration needs to be given as to whether this effect can be interpreted as inefficiency or whether it may be due to heterogeneity • Regulatory application involves an arbitrary correction for modelling errors similar to COLS. Hence the limitations noted above applies.
SFA	• Can separate inefficiency from noise empirically • Can provide confidence intervals around the predicted inefficiency	• Requires specifying distributional assumptions of the inefficiency term, but alternative assumptions can be used, and the distributions can be made quite flexible
DEA	• No explicit functional form needs to be imposed ex ante	• Similar to COLS and RE, does not account for the presence of noise in the data, but the standard model can be extended

From the perspective of academic rigour, SFA and DEA are superior to COLS as they do not rely on an arbitrary correction of the benchmark and more than 40 years of academic research has gone in to their development and applications in regulation and other contexts. COLS result in companies' efficiency being under or over-estimated.[10]

[9] Thanassoulis, E. (2001), *Introduction to the Theory and Application of Data Envelopment Analysis: A foundation text with integrated software*, Springer; Zhu, J. (2016), *Data Envelopment Analysis: A handbook of Empirical Studies and Applications*, Springer.

[10] Oxera (2013), 'Recommendations on cost assessment approaches for RIIO-ED1', February.

Depending on the data availability and quality,[11] Ofgem should therefore consider SFA and DEA.

Moreover, reliance on only one approach risks that the outcome may be due to the idiosyncrasies of the particular approach used. As such, multiple approaches should be consider. The results should then be compared and, based on an understanding of the approaches and their relative pros and cons, differences in the outcomes for individual companies reduced as far as possible.

As part of this question, we would like to respond to some statements made in Professor Andrew Smith's note to Ofgem on alternative methodologies particularly with regards to DEA and, to a lesser extent, SFA.

DEA

Professor Smith questions the efficacy of DEA as a credible alternative to econometric approaches for the reason that the approach does not provide cost elasticities and thus can be seen as a black-box. However, this is incorrect as it is possible to estimate elasticities from DEA.[12]

Moreover, Professor Smith states that DEA would not allow clarity and transparency with regards to the estimated relationship. However, greater transparency compared to the current approach could be achieved. For example, if, instead of combing the outputs into a composite scale variable (CSV), as Ofgem currently does, DEA is run with the outputs specified separately, it can reveal varying substitution rates between the outputs and elasticities with respect to cost for different companies. The current approach of combining outputs into a CSV conflates substitution rates and scale elasticities, and imposes the same impact on a heterogeneous sample of networks.

More generally, with regards to Professor Smith's view that DEA can be seen as a black box and thus does not meet the criterion in regulation of transparency, we note that DEA is widely used in regulatory contexts across sectors and jurisdictions.[13] In addition, DEA is also generally considered to be an extremely transparent approach that operational managers and non-technical audience can understand. In particular, when estimating a companies' inefficiency, DEA identifies the specific peer units that the company is being compared to and the weight given to each. Thus, a company can quite readily see why it is being considered inefficient and also identify whether its peers and weights attached to them are appropriate. DEA also provides a number of other parameters that aid transparency, including elasticities, the weights given to each of the inputs and outputs in terms of their relative importance in determining that company's inefficiency, the target input/output levels needed to be considered efficient, the economies of scale under which a company is deemed to be operating, etc.

[11] DEA requires relatively few independent observations and SFA has been shown to be feasible with the GB GDN data. For example, see: Deloitte (2016), 'Annex 4 - GD17 Efficiency Advice Relative efficiency of Northern Ireland Gas Distribution Networks Final report', March.

[12] For example, see: Forsund, F. R. and Hjalmarsson, L. (2004), 'Calculating scale elasticity in DEA models', *Journal of the Operational Research Society*, 55:10, pp. 1023–1038.

[13] See for example Frontier Economics and TU Berlin for Bundesnetzagentur (2018), 'Effizienzvergleich Verteilernetzbetreiber Gas (3.RP)', 21 December.

Another point Professor Smith makes against the use of DEA is lack of discriminatory power under variable returns to scale (his concerns appears to be that many companies can end up with 100% efficiency). However, we note that the academic and regulatory literature has explored this aspect extensively and there exists several methods for handling this issue.[14]. For example, weight restrictions can be considered (to include a priori operational or economic views about the relative importance of the outputs and inputs). Another alternative would be to use several years of data from a company as distinct observations (similar to the pooled OLS approach used by Ofgem).

Finally, Professor Smith argues that DEA assumes that all deviations from the efficiency frontier are due to inefficiency and that there is no allowance for the influence of data/measurement errors. We first note that this is no different to Ofgem and Ofwat's use of COLS and RE which suffer from the same limitation and similar regulatory judgement can and has been considered in regulatory contexts. Alternatively, there is extensive literature that have considered the uncertainty around the DEA estimates of efficiency and have presented options for accommodating data errors and undertaking robustness checks.

SFA

Similarly, on SFA, Professor Smith makes a number of unjustified claims. As with DEA, we note that SFA is widely used in regulatory contexts across sectors and jurisdictions.[15] First, while it is true that OLS often yields unbiased and consistent estimates of the parameters in the model, this is not always the case. If, for example, inefficiency changes over time or with other factors, then OLS estimates of the model parameters and inefficiency will be biased.[16]

Professor Smith argues that SFA requires strong assumptions, but the supposition that some chosen benchmark, such as an upper quartile, provides an appropriate adjustment for noise for each company is a far stronger assumption. It is also the case that, with a panel data set, there are a number of panel SFA models that are more appropriate and can be applied instead of pooled SFA to which this comment is initially made.

Professor Smith also states that SFA does not overturn the rankings of companies compared to OLS and thus questions the value of SFA in separating noise from inefficiency compared to regulatory judgement applied to OLS.

---

[14] Dyson, R. G. and Thanassoulis, E. (1988), 'Reducing Weight Flexibility in Data Envelopment Analysis', *Journal of the Operational Research Society*, 39:10, pp. 563–576; Cooper, W. W., Seiford, L. M., Zhu, J. (2004), *Handbook on Data Envelopment Analysis,* Kluwer Academic Publishers. The Norwegian regulator also addressed this, see Bjorndal, E., Bjorndal, M. and Camanho, A, (2008), 'Weight Restrictions on Geography Variables in the DEA Benchmarking Model for Norwegian Electricity Distribution Companies', December.

[15] For example, ORR (2018), 'PR18 Econometric top-down benchmarking of Network Rail: A report' July; Deloitte (2016), 'Econometric benchmarking in the UK postal sector. Final report', produced for Ofcom, May;
https://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/Netzentgelte/Strom/EffizienzvergleichVerteilernetzbetreiber/effizienzvergleichverteilernetzbetreiber-node.html.

[16] Kumbhakar, S., H.-J. Wang, and A. Horncastle (2015), *A Practitioner's Guide to Stochastic Frontier Analysis using Stata*, Cambridge University Press, p 71.

While some rank correlation is expected between SFA and OLS in a pooled model context, this is not necessarily the case in other contexts. For example, Oxera, in a report that was published by Ofgem, showed that Ofgem's OLS with UQ correction results in over- and under-correction for noise and company heterogeneity compared to a robust SFA model.[17] That is, in a panel model context, the SFA results of inefficiency, including the rank order of companies, can be different to OLS. Moreover, this demonstrates that OLS can provide substantially incorrect inefficiency estimates for some companies, which is far more important that a company's estimated rank.

For these reasons, simplicity and transparency should be not be given disproportionate weight compared to the accuracy of estimating relative efficiency in deciding the appropriateness of a method.

As such, we consider it extremely unlikely that DEA or SFA would not yield new insights compared to COLS, as Professor Smith contends.

Q2 - Do you agree with our proposed criteria for developing potential cost pools? If not, what additional criteria do you propose and why?

Ofgem (based on CEPA) proposes the following criteria for developing a long list of potential cost pools:

- "**complementarity**: Is there a strong technical/economic reason to believe that activities or groups of expenditure are complementary and should be benchmarked together and a consistent set of cost drivers can be identified?
- **cost trade-offs**: Can GDNs make trade-offs in expenditure between the different activities/areas included in the cost pool, and so benchmarking those activities/costs together will help avoid biased relative efficiency results or unintended managerial incentives for the GDNs?
- **cost boundary complexity**: How complex is the boundary of cost reporting data that needs to be defined to benchmark the identified cost pool/activity (eg how well defined is the group of costs within Ofgem's regulatory reporting templates)?
- **risk of inaccurate/biased models**: Is there too much 'noise' in the data to be confident that including certain types of expenditure within aggregated regressions could lead to inaccurate model results, or coefficient estimates that are difficult to interpret using engineering/economic logic?"[18]

A key criteria in economic literature for modelling the costs or production is whether activities are separable. The inputs and outputs for the units being modelled must be exclusive and exhaustive for that unit and activity in question. That is, the inputs captured by the model are complete (i.e. no inputs are omitted) are only used by that unit or activity (i.e. no other activity uses any of the same inputs) in order to produce all the outputs associated with that activity (i.e. no outputs are omitted). The unit in question must also

[17] Oxera In association with Distinguished Professor Subal Kumbhakar (2013), 'Recommendations on cost assessment approaches for RIIO-ED1-An independent submission by Oxera to Ofgem', February.
[18] CEPA (2019), 'RIIO-GD2 cost assessment – econometric modelling & regional factors', June, p. 48.

have a degree of autonomy or independence. This appears to be similar to the first three criteria above. Clearly, the fourth criteria is also an important consideration.

Evidently, the availability and quality of the data at the level being considered (including the availability of specific outputs or cost drivers), is a minimum requirement.[19] In addition, the GDNs deliver workload and outputs in many different ways, so ensuring consistency is crucial in order to be able to make meaningful comparisons across companies. For instance:

- the differences between insourcing and outsourcing pose issues in terms of cost allocation as well as services included/excluded;
- emergency and repairs activities are also a common inconsistent area. Stranded time in these areas can be utilised across other activities offsetting other external spend, which can make these cost areas inconsistent. The non-formula metering contracts offset a material amount of stranded time from emergency but some networks will not have these benefits if suppliers in their network have a different approach to meter work. There is a similar issue with repairs work. The repair teams also carry out the connections and mains replacement workload in their downtime, which again could be inconsistent across GDNs depending to departmental set up and contractual agreements.
- there is a trade-off between capital rebuilds and refurbishment. Capital rebuilds can cost significantly more than a refurbishment option but both achieve the outputs required. These activities are currently reported differently across Opex and Capex and cannot be linked to demonstrate the trade-off across outputs or asset categories. GDNs can report these differently. Firstly, through the insourced/outsourced issue, depending on the capitalisation of overheads versus outsourced contracts the costs can be treated differently across activities. Secondly, through the grouping of activities into a work package, if you were to replace small parts at the same time as a major rebuild you may include in a capital unit, yet if this was carried out at the same time as a maintenance it could be included in an Opex unit.

Overall, as well as CEPA's criteria, consistency of cost reporting is crucial to ensure quality of the data at the required level. Moreover, there also need to be clear cost drivers at the level of cost examined. When granular cost categories are used, Ofgem needs to have consistent ways of aggregating the results that take into account trade-offs between activities and avoid cherry-picking.

Q3 - Should we continue to use the Cobb-Douglas functional form? If not, why?

At GD1 Ofgem used a Cobb-Douglas functional form in its cost models, which takes the following general form:

$$\ln(cost) = \beta_0 + \beta_1 \ln(cost\ driver) + \epsilon$$

Such a functional form is relatively simple, and the coefficients provide the cost elasticity for each company, so interpretation is straightforward. Whether or not it is appropriate

[19] Other issues such as how the relationships between activities can be captured within the method are discussed elsewhere in this response.

depends on the underlying relationship between costs and cost drivers. Indeed, it may be an over simplification and result in over/under-estimated each company's inefficiency.

More flexible models have been used by other regulators, e.g. in the water sector and in the postal sector, translog or semi-translog models have been used. However, the models used in PR14 were criticised by the CMA in Bristol Water's appeal of PR14 as the resulting implied elasticities were not always in line with engineering or economic logic.[20] This is not to say that translog, or indeed other functional forms, are not appropriate, but with more flexible function forms it is essential that checks are undertaken to ensure that the implied elasticities for *each* company are of the correct sign and order of magnitude based on engineering or economic logic (this does not mean that an expected outcome should be imposed, but the model parameters should be 'sense-checked'). Ofwat took this into account in its model development process for PR19, stating they "have made sure the coefficients align with expectations".[21]

For some cost drivers, a more flexible functional form may align with operational insights. For instance, if a sparsity/density measure were to be included in the models directly, then the expectation might be that particularly sparse and particularly dense companies are faced with higher costs. This could be captured by including a linear and a quadratic term of the measure to capture a u-shaped impact.

Overall, whether or not this functional form is appropriate is an empirical question that depends on how well the data fits different functional forms and most importantly, whether the interpretation of the models are in line with engineering knowledge.

Q4 - Do you agree with the proposed model selection criteria and model development phases?

Ofgem lists three criteria to be considered when selecting models:[22]

- economic/technical rationale;
- transparency and ease of interpretation for stakeholders;
- robustness.

Overall, these three broad criteria seem appropriate for the purpose of model selection but further explanation of their relative importance could be useful. For example, we consider that the economic/technical rationale (i.e. models aligned with economic and operational insight) to be the most important criteria. We would then consider the robustness of the models developed to be the second most important criteria.
This is in line with Ofwat's model development and assessment criteria for PR19, namely:[23]

[20] CMA (2015), 'Bristol Water plc. A reference under section 12(3)(a) of the Water Industry Act 1991', 6 October, para 4.
[21] CEPA (2018), 'PR19 Econometric Benchmarking Models', March, p. 10.
[22] Ofgem (2019), 'Consultation – RIIO-2 tools for cost assessment', 28 June, para 2.39.
[23] Ofwat (2018), 'Cost assessment for PR19: a consultation on econometric cost modelling', March, p. 8.

- Use engineering, operational and economic understanding to specify an econometric model, and form expectations about the relationship between cost and cost drivers in the model.
- Assess whether the estimated coefficients are of the right sign and of plausible magnitude.
- Consider if the estimated coefficients are robust. For example, are they stable and consistent across different specifications? Are the estimated coefficients statistically significant?
- Assess the consequences of cost drivers under management controls, in particular, the risk of any perverse incentive.
- Consider the statistical validity of the model more widely – does the model perform well in terms of statistical tests and diagnostics?
- Consider the appropriate estimation method.

While Ofgem's 'transparency and ease of interpretation for stakeholders' is also important this should not be at the sake of getting as close as possible to the true estimate of inefficiency, for which Ofgem's other two criteria are critical. For example, while translog models may be slightly more complex than simpler models, their alignment with economic and operational insight can be tested to ensure they are robust and their interpretation can be explained to stakeholders, as has been done in other cases.

We also consider that Ofwat's additional criteria of avoiding perverse incentives is also important, though it can be overplayed, i.e. it is highly unlikely that a company will build a longer network just to look better on the cost assessment models (see discussion below).[24]

Economic/technical rationale

Given the relatively small dataset, as well as potential heterogeneity amongst companies we consider the economic/technical rationale to be particularly important. Ofgem's consultation document notes that this is the first step when selecting cost drivers and a functional form. However, it should be emphasised that even after the cost drivers and type of model are selected, any results should be cross-checked against economic/technical rationale. In particular, operational insights can help to validate the sign and magnitude of coefficients.[25] This is also a key lesson to be learned from the CMA's Bristol Water Decision.[26]

Transparency

Transparency and replicability is clearly important in a regulatory context. We note that all the approaches discussed in Question 1 are implementable in standard econometric packages (e.g. Stata, R).

Moreover, transparency is linked to the above point, as compliance with the economic/technical rationale is aided if stakeholders can readily interpret the results. However, as stated above, we consider that this is the least important of Ofgem's three

[24] Ofwat (2019), 'Supplementary technical appendix: Econometric approach', January, p. 12.

[25] Assessing whether the magnitude is in line with operation insights is likely to be more challenging than determining the expected sign of the coefficient.

[26] CMA (2015), 'Bristol Water plc. A reference under section 12(3)(a) of the Water Industry Act 1991', 6 October, para 4.50.

criteria, as transparency can be improve if the interpretation of more complex model, such as translog, is explained to stakeholders. We note that translog model, SFA and DEA have been used in numerous other regulatory contexts, for example in Germany[27] and Scandinavia.[28]

Robustness

The robustness tests listed in Ofgem's consultation document[29] are discussed here in turn. We noted that in DEA, these statistical tests are not necessarily relevant, even though these could be important in the model development process where econometric methods can be employed to inform an appropriate DEA model. The robustness tests required and undertaken for DEA are different and are not covered in Ofgem's consultation document or any of the commissioned reports.

Statistical significance of the coefficients: In addition to significance, operational insight should also be considered here. For instance, a variable that is not statistically significant but important from a technical point of view should not necessarily be rejected if the model is sensible otherwise. Similarly, a significant coefficient which has a counter-intuitive sign or magnitude may mean that a variable should not be included (in this form).

RESET test: This is a test of whether the functional form of the model is correct, in particular, whether squared or cross-product terms are missing. It can be useful to help assess the appropriateness of the functional form. However, overall we do not consider that this is an overly important test.

Normality of errors: This is not an overly important test. In fact, the assumption is likely to be violated in the presence inefficiency, as well as noise, in the error term. The error term in an OLS model is therefore likely to be skewed. SFA, on the other hand, uses this skewness and allows for separation between noise (normally distributed) and inefficiency (e.g. one-sided normal distribution).

Correlation: This is likely to not be a major concern for the purpose of determining efficiency gaps. However, correlation would affect the interpretation of individual effects and should therefore be kept in mind when assessing the effect of specific cost drivers (i.e. the sign and magnitude of coefficients), especially if being used to forecast efficient costs.

Heteroscedasticity: We agree with Ofgem's statement that heteroscedasticity is of limited concern, as its impact can be addressed by using robust standard errors.

Panel effects: This can be critical. GDNs may be quite heterogeneous, in that they may have may unique factors that impact upon their costs that cannot be captured within the modelling (e.g. due to a relatively small ample size, data on the factors not being collated, etc). Panel approaches enable such heterogeneity to be accounted for. For example, in

---

27 https://www.bundesnetzagentur.de/DE/Sachgebiete/ElektrizitaetundGas/Unternehmen_Institutionen/Netzentgelte/Strom/EffizienzvergleichVerteilernetzbetreiber/effizienzvergleichverteilernetzbetreiber-node.html.

28 Bjorndal, E., Bjorndal, M. (2016), 'Evaluation of the StoNED method for benchmarking and regulation of Norwegian electricity distribution companies'.

29 Ofgem (2019), 'Consultation – RIIO-2 tools for cost assessment', 28 June, Table 2.3.

the true random effects model (see answer to question 1), [30] the effect estimated from a random effects model is assumed to represent heterogeneity and nothing to do with inefficiency.

Endogeneity: Some endogeneity is likely to be present in the cost models considered here, as not all explanatory variables are truly independent but can be determined by companies' decisions. This could, for instance, be the case with workload, length of mains, or measures of asset heath were to be included in the models. However, as stated above this issue can be overplayed. Ofwat addressed this issue in the context of length of mains, stating the following:

> "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."[31]

As such, we would agree with Ofgem's assessment that this issue needs to be considered on a case-by-case basis.

Additionally, Ofgem mentions adjusted R-squared as a measure of model fit.[32] This is the standard measure to use here and we agree that it should feed into the model selection process. However, especially given the limited number of data points, small changes in R-squared values should not be given too much weight compared to operational insights. Given heterogeneity, we would recommend that Ofgem tests for the relationships between cost and the outputs at different quantiles of the dataset where possible and/or tests the impact of certain data points on results by including/excluding some companies, rather than solely relying on the average relationship identified through OLS. DEA can also help to gain insights as it allows for more flexible relationships between inputs and outputs.

Model development phases

Ofgem sets out two model development phases[33]. While we broadly agree with these, we have set out some additions in the below table:

[30] Greene (2005), 'Fixed and Random Effects in Stochastic Frontier Models', Journal of Productivity Analysis, Volume 23, Issue 1, pp 7-32.
[31] Ofwat (2019), 'Supplementary technical appendix: Econometric approach', January, p. 12.
[32] Ofgem (2019), 'Consultation – RIIO-2 tools for cost assessment', 28 June, para 2.49.
[33] Ibid, Figure 2.1.

Table 0.2 Ofgem's model development phases and proposed alterations

Phase	Ofgem's proposition	Proposed alterations
Phase 1	predictive power (adjusted R-squared); statistical robustness; economic/engineering rationale; are the results consistent with the rest of the price control?	Move economic/ engineering rationale to the top Add sub-bullet: Are the key cost drivers accounted for? Are the coefficient in line with operational insights in terms of sign (positive/negative) and, where possible to assess, magnitude?
Phase 2	removal of years/companies from panel; random effects; within sample forecasting (when appropriate)	Add: how do the coefficients of the remaining variables change when cost drivers are removed/added/alternative definitions used? Is this in line with economic/ engineering rationale? What are the results using other approaches (e.g. SFA and DEA)?

Source: Ofgem (2019), 'Consultation – RIIO-2 tools for cost assessment', 28 June, Figure 2.1 and Oxera.

Aggregated economic analysis

Q5 - Should the cost driver of the totex regression model be determined by the cost drivers of the 'bottom-up' models, or should the totex regression model account for different explanatory variables? Why?

Ofgem's GD1 TOTEX model used a composite scale variable (csv) as a cost driver that consisted of the workload drivers as they were used in the bottom-up regressions. Specifically, the csv was constructed as 38% MEAV, 43% REPEX workload, 2% connections workload, 2% mains reinforcement workload, 6% number of external conditioning reports, 5% maintenance MEAV and 4% emergency service csv.

Overall, the bottom up operational insights are a useful guide for selecting TOTEX cost drivers. However, Ofgem should not necessarily stick to it too stringently if other TOTEX models perform well. In particular, drivers of relatively smaller activities could be difficult to pick up at the aggregate level. The overall aim and criteria for the model development at the aggregate level should be the same for all models and additional 'artificial' constraints should not also be applied. DEA could be a useful technique in this context as it does not require pre-selecting the weights of different cost drivers and allows them to carry different weights for different companies

There are reasons for differences in cost drivers between top-down and bottom–up models. Bottom-up models focus on a specific activity/expenditure, so there is likely to be a very relevant cost driver from an operational perspective.

At the Totex level, trade-offs across the different cost areas need to be considered and the material drivers could be different to the bottom-up ones.

- Totex is more interested in the whole life cost of assets not just scale or condition;
- Bottom-up models do not consider trade-offs, but a Totex model needs to ensure these are reflected in their cost drivers e.g. emergency/repairs including utilisation on other activities.

Q6 - What could be appropriate cost drivers in middle-up models for opex, capex and repex? Why?

Ofgem's approach in GD1 to base cost drivers of the middle-up models on the individual cost drivers from the disaggregated regressions remains reasonable. However, as with TOTEX, we do not consider it necessary to stringently stick to the exact proportions used in the disaggregated models. In addition, Ofgem needs to ensure that trade-offs between activities are sufficiently captured even if they do not fall within the same 'middle-up' category (e.g. maintenance versus CAPEX replacement).

The following table summarises the main cost drivers from an operational perspective.

Table 0.3 Fundamental cost drivers of different activities

Cost activity	Ofgem's cost drivers in GD1	Fundamental cost drivers identified
REPEX	REPEX synthetic costs	• number of main to main connections required (diameter of main, road surface, location, material of main); • number of services to be replaced (diameter and depth, service frequency and length, road surface, location, entry type to property); • size of projects (number of connections, length of mains runs).
OPEX - Work management	MEAV	• size and complexity of network; • customer numbers; • work carried out locally and on site; • asset condition.
OPEX -Repairs	External condition reports	• number of repairs (quantity of metallic mains, condition and operating pressure of mains (LP 80-90% of workload but cheapest)); • size and depth of mains to be repaired; • geography (valleys)/sparsity (travel time, especially out of hours)
OPEX -Emergency	CSV of external condition reports and number of customers	• number of external reports (quantity of metallic mains, condition and operating pressure of mains); • number of internal reports (number of consumers, condition of consumers' internal pipework, level of CO awareness); • geographic density of calls and travel times.
OPEX -Maintenance	Maintenance MEAV	• volume of above ground and LTS assets; • condition of assets (age sometimes used as a proxy); • company policies, and CAPEX and REPEX activity (endogenous)
CAPEX -Connections	Connections synthetic costs	For each new load:[1] • location relative to local network; • size of local mains; • size of new load; • local ground conditions; • inclusion of mains work (re infill point below).
CAPEX -Mains reinforcement	Mains synthetic costs	• Changes to pattern of supply and demand to the network

Note: [1] Similar to synthetic costs, a new measure could be constructed that takes into account these points.

Q7 - For which opex activities are there trade-offs that support the rationale for testing 'totex and opex plus' modelling?

Many Opex activities have trade-offs which sit elsewhere in Totex not just trade-offs within Opex.

The consultation discusses GDNs combining the Repair and Emergency activities, whilst in theory this seems sensible given the relationship between repair deferrals (D2 rechecks) and the repair costs, it would also need to include other Emergency/Repair follow on jobs (Repex – Relay following escape).

However it does not consider the relative fixed nature of the emergency service or the utilisation across numerous Totex categories, which commonly can be a reason for fluctuating emergency costs.

The GDN's alternative insource/outsource operating models can cause different utilisation solutions between Emergency, Repair and Repex/Capex hence making it difficult to identify the 'Opex plus' activities.

Q8 - Are there other particular costs that we should aggregate and test in our analysis?

Reviewing cost categories and given the specific drivers the current aggregates Ofgem use for the middle up, we would not suggest alternative groupings given issues discussed in question 7.

Please see question 9 for the suggestion on Totex aggregating for asset related 'Opex plus' modelling.

Disaggregated economic analysis

Q9 - Are there trade-offs between opex and capex activities that support the rationale for considering 'opex plus' modelling?

Opex plus modelling for the trade-off of asset Maintenance, refurbishment and rebuilds would give a more consistent view of costs across years and asset condition cycles, but these interventions sit across all three expenditure types making it difficult to strip out specific 'Opex plus' costs for modelling.

In theory we support 'Opex plus' modelling for the asset intervention work but further work on BPDT identification and definitions of these trade-offs would need to be completed to enable consistent data across GDNs.

There are other trade-offs which currently fall outside of disaggregated economic analysis, IT spend should be considered as part of this trade off, given the move toward cloud storage solutions which are now included within Opex, where traditional storage options have been Capex. We highlight these costs in Q14 under Non economic analysis.

Q10 - Which cost areas should be assessed using workload drivers as opposed to other cost drivers? Why?

Asset Condition workload drivers should be considered, due to the inherited state of the networks all GDNs are at different stages of the asset investment cycles.

Asset condition has a correlation to the maintenance and interventions activities we carry out and could be combined with the Scale drivers to produce a CSV. This would be advantageous to the regulator as the condition cannot be influenced by the GDN unless an intervention has been carried out.

Q11 - Should repex (or some categories of repex) be excluded from our regression analysis and assessed using other techniques?

Repex mains replacement should be included in the regression analysis, but consideration for the size of programme, technique and HSE policy levels will need to be factored into the synthetic unit costs.

Some other costs within Repex that would be excluded from the current regression analysis are Special crossings and MOBs.

Special crossings – the bespoke nature of these replacement jobs require specialist equipment and skills depending crossing structures (Rivers and listed structures can be more technically challenging). Given these structures are different across GDNs and do not have a consistent workload I would recommend these are dealt with as part of the Non-economic analysis.

Multi occupancy buildings – these need their own synthetic unit costs (Below 6 floors), specifically split between above and below 6 floors. The above 6 floor buildings require a specific structural design which can make the costs of different buildings bespoke, which include technical services.

Q12 - Are there other approaches to disaggregated benchmarking that we should consider?

We consider that the approaches discussed in our response to question 1 provide a reasonable directional guidance to GDNs' relative efficiency. However, estimated differences may be due to the inadequacy of such an approach. As such, it will be important to consider reasons for observed differences between companies, rather than just assume the differences are due to inefficiency. As mentioned above, it is essential that Ofgem considers trade-offs between activities to account for companies choosing different business models to achieve outcomes. This could be done within more aggregated models or as part of the aggregation process.

In addition to disaggregated benchmarking models, Ofgem could also consider engineering justification documents (which are being submitted with Business Plans), which include

- unit costs;
- justification for investment and cost variances;

- cost benefit analysis for certain areas, e.g. mandated work with poor payback periods and inconsistent costs across GDNs.

Non-economic analysis

Q13 - Should we assess business support costs at a group level in order to address cost allocations across companies within groups?

We would agree the business support costs should be assessed at group level, this will address cost allocations issues and should highlight expected economies of scale not achievable by stand-alone companies. These costs include HR, IT and Finance and can be utilised as a centralised team for group companies, stand-alone GDN's will always have a level of fixed overhead that will look more inefficient than a group.

There should be an adjustment made to these business support costs across companies to consider the expected efficiencies they can achieve with centralised support costs.

Q14 - Which types of business support costs should be benchmarked, and how should they be benchmarked?

Business support costs consist of:

- non-operational IT and telecoms;
- property management;
- finance;
- audit and regulation;
- HR and non-operational training;
- insurance;
- procurement;
- stores and logistics;
- CEO and group management.

In GD1, business support costs were benchmarked separately from direct activities. After concerns with the disaggregated benchmarking proposed in its Initial Proposals, Ofgem switched to a top-down approach to benchmark overall business support costs on a unit-cost basis.[34] To determine the unit cost it calculated a composite cost driver that corresponded to cost drivers of the individual activities consisting of revenue, end-users, employees and total spend. The benchmark was then set based on the upper quartile metric from the Hackett database. Efficient business support costs for this proxy

[34] Ofgem (2012), 'RIIO-GD1: Final Proposals – Supporting document – Cost efficiency', 17 December, Annex 6.

company were then calculated using the aggregate composite cost driver across electricity transmission and gas distribution.

The overall approach to benchmarking business support costs remains reasonable. Additional insights could be generated by investigating disaggregated activities.
However, cherry-picking remains a concern if these were to be combined for the purpose of setting the benchmark. We also agree with the approach of using comparator companies within the energy sector for the purpose of benchmarking, as companies in other sectors may be structurally different and therefore less suitable comparators. However, when choosing comparators differences in scale and scope should be taken into account to ensure a realistic benchmark.

Q15 - Which types of business support costs should be excluded from benchmarking?

In GD1, insurance was excluded from the benchmarking and but assessed separately (by examining overall industry trends). This was because differences in risk appetite and appropriate levels of coverage made comparisons across GDNs difficult. We consider that this remain the case.

We also note that, in ED1, IT&T costs were also examined separately, using a combination of ratio analysis and consultant's qualitative views, with most weight on the latter. This was because high level top down comparisons cannot account for justifiable differences between individual companies' IT strategies. We consider that this is also true for GDNs.

Regional factors and company specific effects

Q16 - How should we estimate and model the impact of regional factors?

Regional factor adjustments will be required if there are regional- or company-specific issues resulting in material cost differences between GDNs that are not already captured by the cost drivers used in the cost models. Only material regional factors would need to be adjusted for, which would suggest there will only be a need for a relatively limited number of factors.

As such, modelling the impact of regional factors is closely connected to the general model development process because:

- if the models already contain cost drivers that capture the factor in question, to some extent, then additional adjustments may not be necessary;
- any pre- or within-modelling adjustments would affect the model development process, so the economic/technical rationale and model fit would need to be re-assessed.

Any regional factor modelling would therefore need to be assessed relative to the models and methods used overall.

The more detailed steps involved in accounting for regional factors could be broken down as follows:

1. Identify the regional factor based on an economic/technical rationale. It needs to be justifiable based on the criteria set out in the next section

2. Construct a measure or measures to proxy the factor based on available data:

- the measure should be as relevant/specific as possible without becoming controllable. For instance, when calculating regional wage differences, sectoral as well as occupational wage data should be considered, rather than the wage levels paid by the GDNs themselves.[35] Our previous analysis has shown that accounting for the sector can lead to significantly different results than only considering occupation.[36]

- the measure should be readily interpretable in order to assess whether the direction and magnitude of any potential adjustments are plausible.

- for pre-modelling adjustments constructing a summary measure may not be necessary, but the size of the impact needs to be estimated directly in order to remove it from costs prior to the modelling.

3. Determine which part of the cost base the adjustment should apply to, based on economic/technical rationale

4. Carry out pre- or within-modelling adjustment:

- which one is more suitable to account for a regional factor depends on: whether the issue is company specific or industry wide, whether the accuracy of the modelling is affected (e.g. pre-modelling adjustments can add more noise to the data set and affect the robustness of the subsequent cost modelling); and whether the impact of the effect is known or can be reasonably estimated using available data.

- even if pre-modelling adjustments are used, regression analysis may be helpful to test for significance of a regional factor (in addition to other cost drivers).

- If pre-modelling adjustments are used, then a bottom-up or engineering quantification can be used. While within-modelling would involve appending the cost model with the associated measure.

5. Assess the plausibility of estimated impacts:

- cross-checks with economic/technical rationale;

- jointly assess plausibility of regional factors.

A within modelling approach ensures that the impact is incremental to the specific cost models used by Ofgem and can be tested for its significance. If such an approach is not possible then a bottom-up or engineering based assessment will be required to account for the regional differences prior to modelling.

[35] In which case the measure would become controllable and include any potential inefficiencies.
[36] Oxera (2019), 'Regional factors in the cost assessment for GD2', 28 June.

Q17 - Do you agree with the proposed criteria for justifying regional cost factors that we have outlined?

Any regional factors claims need to be sufficiently justified. Ofgem/CEPA set out the following criteria for justifying regional cost factors:

- "the regional or company-specific factor in question is clearly defined (i.e. there is a clear technical / economic rationale for why it would be expected to impact company costs);
- the relevant factor, and the subsequent costs it drives, are beyond the control of an efficient company (having taken all the feasible measures to mitigate the costs);
- the company (or a small number of companies) are impacted by a significant amount, and in a materially different way to others."[37]

In addition to the above points, we also consider it important to assess whether a regional factor is clearly incremental to, and not already captured by, Ofgem's cost assessment. This would need to be assessed on a case-by-case basis, taking into account the other cost drivers used and cost adjustments carried out in a particular model. For instance, Ofwat maintained that regional wage adjustments were not needed in models that already included density (and density squared) as a cost driver, because it considered that these effects are correlated.[38] Ofgem would need to examine whether similar issues occur in the context of GD2 when assessing how and whether to adjust for regional wages differences and density/sparsity.

Real price effects and ongoing efficiency

Q18 - What RPEs should we account for, how should we gauge materiality, and what criteria should we use for index selection?

How to gauge materiality?

The materiality of RPEs can be determined based on the materiality of the cost categories in the efficient GDN cost structure and the relative value of the input price indices over and above inflation used to index revenues.

If a particular cost category does not consist of a significant proportion of a GDN's cost base (for instance, if it is less than 1% of the GDN cost base), then it can be assumed to increase in line with inflation.[39]

[37] CEPA (2019), 'RIIO-GD2 cost assessment – econometric modelling & regional factors', June, section 7.2.1.

[38] Ofwat (2019), 'Supplementary technical appendix: Econometric approach', January, p. 15.

[39] Ofgem (2012), 'RIIO-T1/GD1: Initial Proposals – Real price effects and ongoing efficiency appendix', para. 2.35 and Table 2.2.

Similarly, if the input price indices do not differ materially from inflation (the difference between the input price and inflation is on average close to zero), the real price effect can be assumed to be zero.

What RPEs to account for?

We consider the significant cost categories and RPEs to consist of labour and materials.[40] Therefore, RPEs should be accounted for these categories.

What criteria should be used for index selection?

The selection of relevant indices within the cost categories would depend on the following best practice criteria. The indices that best fulfil these criteria should be selected for indexation.

- **Relevant to gas distribution**—ideally, the index should map only those inflationary pressures that have an impact on the distribution companies' costs and revenues. This requires the index to be representative of the organisation's cost base in delivering its services. Due to the diverse nature of these costs, a bespoke weighted index may be preferable to a single generic index given that the weightings of the components of generic indices are unlikely to match the weightings of the distribution network costs. When examining a weighted index, the component indices chosen should be reflective of the typical inflationary cost drivers (e.g. labour, rent, materials and fuel costs) that affect a GDN's cost base.
- **Transparent**—the index needs to be well-understood and transparent so that consumers can analyse its movements. If a bespoke weighted index is chosen, the components and weights used to calculate the index need to be clearly defined such that all stakeholders are able to interrogate the index, understanding its past and predicted future movements. This consideration is of particular importance in the implementation of a composite index. In this context, it is necessary to select indices from reliable, independently published sources that use industry-recognised formulae.
- **Forecastable**—this is partly related to the underlying characteristics of the index, since it is likely to be beneficial to the organisation and its customers to adopt an index that is not subject to a high degree of volatility. As such, the statistical properties of the selected index (e.g. its volatility) need to be well-understood. The most forecastable indices are likely to be those that have been in existence for a reasonable period of time, since their behaviour can be observed across differing economic circumstances. This would also enable consumers/companies to foresee changes in prices/revenues so as to better plan their expenditures.
- **Non-controllable**—movements in a GDN's costs (which stem from its own purchasing decisions) should not have a material impact on movements in the index. This avoids the creation of a potential distortionary effect on incentives, whereby the organisation loses its incentive to be efficient in purchasing its inputs.
- **Produced in a timely manner**—the index should be published soon after the relevant cost/price change, in order to enable speedy adjustments to tariffs.

[40] Oxera (2019), 'RIIO GD-2 support: real price effects', 28 June.

- **Subject to limited revisions, not affected by compositional changes and well diversified**—the index should be well diversified such that an anomaly in a single component of the index does not result in an inappropriate change in the overall index value. The index should not be subject to frequent and substantial revisions, so that tariff changes do not need to be rescinded or revised. It should also not be greatly affected by compositional changes, so that movements reflect genuine price changes

Q19 - What common input and expenditure categories are appropriate for structuring RPEs?

With respect to the GDN cost categories, the current level of disaggregation into labour, materials OPEX, materials CAPEX and plant and equipment seems appropriate. Any further disaggregation would make it difficult to find suitable indices that are significantly correlated with the cost categories.

For instance, if labour is split into direct and contract labour, it would be difficult to attain appropriate indices for some specialized contract labour categories. Moreover, as it is a company's choice to employ direct or contract labour, providing an RPE allowance for contract labour may incentivise firms' employment choices, which may be contrary to regulatory intention.

Q20 - How should we identify an appropriate ongoing efficiency assumption?

Ongoing efficiencies can be estimated using direct or indirect approaches:

- **Direct approaches** involve direct comparisons of companies over time to estimate the frontier-shift improvements that they have achieved historically. Common approaches consider data envelopment analysis (DEA)[41] and stochastic frontier analysis (SFA).[42]

- **Indirect approaches** rely on the use of indirect comparators (i.e. companies outside the gas distribution sector or other sectors of the economy) to derive a benchmark. They involve the selection of sectors with characteristics comparable to the assessed companies, and assume that the rate of technological progress in the selected set of comparators is a good indicator of the rate of technological progress that the regulated sector should achieve. Existing studies typically consider sector-level data from the EU KLEMS database.[43]

Ofgem's advisors (CEPA) recommend the use of the EU KLEMS dataset for the assessment of ongoing efficiencies at RIIO-GD2, similar to the approach used at GD1.[44]

[41] DEA is a non-parametric approach that is widely used internationally when benchmarking regulated companies. For a more detailed discussion, see Thanassoulis, E. (2001), 'Introduction to the Theory and Application of Data Envelopment Analysis: A Foundation Text with Integrated Software', Springer.

[42] SFA is an econometric, parametric approach which accounts for statistical noise by making functional form and distributional assumptions. For a more detailed discussion, see Kumbhakar, S.C., Wang, H.J. and Horncastle, A.P. (2015), *A practitioner's guide to stochastic frontier analysis using Stata*, Cambridge University Press.

[43] For a review, see Oxera (2016), 'Study on ongoing efficiency for Dutch gas and electricity TSOs', January.

[44] CEPA (2019), 'RIIO-GD2 cost assessment – frontier shift', June 2019.

We consider EU KLEMS to be an appropriate basis for productivity analysis, given its widespread usage in regulated utility sectors. EU KLEMS includes data on economic growth, productivity, capital formation and technological change at the industry level for all EU member states. The most recent data release specifically contains information on value-added (VA) productivity growth measures.

However, we also note some limitations of this data source. First, EU KLEMS does not contain alternative productivity measures such as gross output (GO) based productivity for recent years (i.e. after 2007), limiting the period of analysis over which productivity can be assessed on this basis. The main advantage of using GO-based measures is that gross output includes the contribution of intermediate inputs to production. Even on a VA basis the EU KLEMS database only goes up to 2015.

Second, EUKLEMS is also based on historical data and, as discussed above, several years out of date. Its use for setting an ongoing efficiency challenge assumes that the past is a good predictor of the future. Existing evidence shows that the productivity slowdown experienced by the UK since the financial crisis is likely to persist to some extent going forward.[45] As such, (i) the timeframe over which productivity performance is measured requires careful consideration as it has to be representative of the RIIO-GD2 period; and (ii) account needs to be taken of current forecasts.[46] To derive GO productivity measures for recent years, alternative datasets may be considered to collect additional information on intermediate input volumes and prices (e.g. OECD STAN).

Third, TFP estimates obtained from the EU KLEMS productivity database are likely to encompass other effects, such as catch-up improvements and scale effects. As such, some adjustments or assumptions are required to isolate ongoing efficiency.

Alternatively, more direct approaches, such as SFA and DEA, can be considered to decompose productivity growth achieved by the industry into its constituent parts and therefore isolate the impact of frontier shift. These approaches could be based on the same company level data that Ofgem will use in its assessment of relative efficiency and, as such, would ensure consistent between the catch-up and frontier shift assessment. The direct approaches have been used extensively by European energy regulators to determine ongoing efficiency and has extensive regulatory precedents and academic evidence.[47]

Q21 - How should we determine frontier shift?

In line with the terminology used in the CEPA report, we are using the term ‘ongoing efficiency’ to refer to improvements that can be made from the adoption of new technologies or innovative processes and is considered in combination with changes in the prices of the inputs net of inflation (i.e. real price effects, or RPEs). While, the difference between ongoing efficiencies and RPEs would provide an estimate of ‘frontier shift’.

[45] Productivity estimates at the economy-level are also provided by the Office for Budget Responsibility and the Bank of England on a regular basis.

[46] See Oxera (2019), ‘Establishing an appropriate efficiency challenge’, prepared for Wales & West Utilities Limited, June 2019.

[47] For example, the German energy regulator considers a combination of DEA and SFA in estimating the Xgen factor (productivity factor) for network operators in the gas and electricity market.

Two general approaches to frontier shift can be used – indirect or direct approaches (see our response to question 20).

If an indirect approach is used, ongoing efficiencies are based on comparisons outside the sector in question. The frontier shift approach is often based on total factor productivity (TFP) estimates and this approach has a lot of regulatory precedent in the UK and continental Europe.

If a direct approach is used then the historical performance is directly estimated. Approaches such as SFA and DEA can be used, and frontier shift can be directly estimated using company-level data.

Although the latter approach is in principle a more robust and accurate way to assess frontier shift, it relies on the quality of the company-level data used. For example, a recent DEA application using operator-level data provided by Ofgem showed quite volatile results.[48] The indirect approach is a possible alternative in such cases but, given the 'aggregate' nature of the assessment, a range of estimates and sensitivities would need to be considered.

Combining the elements of our cost assessment

Q22 - Should we set the efficiency benchmark at the upper quartile level?

(Please also refer to our response to question 1).

The upper quartile benchmark is currently used as an ad-hoc adjustment to account for noise, or errors, in companies' predicted costs.

The assumption that any difference between a firm's actual and predicted costs resulting from the regression analysis is entirely due to inefficiency may indeed result in overestimating inefficiency, as a company can be estimated to be inefficient simply because its inefficiency is confounded with noise and/or company heterogeneity.

The upper quartile adjustment is intended to make such an adjustment; however, it has no academic basis and requires judgement as to where to move the cost function to and assumes the same degree of noise is present for each company. Despite requiring empirical evidence on the level of data/modelling errors, which is contextual, Ofwat and Ofgem have been applying regulatory discretion in using a UQ benchmark without rigorous testing in various price reviews. In contrast, in PR13, ORR tested for the level of noise by comparing to SFA results,[49] and, in Royal Mail's price and service quality control review 2006-10, Ofcom's advisers similarly used SFA and made adjustment to its COLS and DEA results to achieve consistent outcomes.[50]

[48] Ajayi. V., Anaya. K. and Pollitt. M (2018), 'Productivity growth in electricity and gas networks since 1990', 21 December.

[49] ORR (2013), 'PR13 Efficiency Benchmarking of Network Rail using LICB', August

[50] LECG's (2005), 'Future Efficient Costs of Royal Mail's Regulated Mail Activities', June.

In contrast to a UQ benchmark, where the adjustment is the same for all companies, it is well known in econometrics that the accuracy of the model prediction decreases as you move farther out from the central data. It is also well known in econometrics that the accuracy of the model prediction decreases as the sample size decreases. In GD there are only 8 cross-sectional observations (networks), however, these are not truly independent observations as there are only 4 ownership groups. This is the smallest cross section where top-down benchmarking is used in a GB regulatory context. As such, examining the accuracy of the modelling for RIIO-GD2 will be critical.

As shown in Oxera (2013),[51] such ad hoc adjustments may overcompensate or undercompensate for specific companies, even when the adjustment is broadly correct across the industry as a whole. As a result, the quality of the models used is key if the same benchmarking approach were to be considered at GD2. Besides being statistically robust, the models need to control for the appropriate sets of industry cost drivers and to account for heterogeneity in the industry. Furthermore, given likely biased outcomes for individual companies, the approach, if used, should at least be cross-checked using SFA.

SFA is a more robust approach enabling a direct estimate of inefficiency and noise to be undertaken, with more limited assumptions.[52] That is, SFA will determine the appropriate split between noise and inefficiency based on the model specification and the data. Indeed, it is possible that SFA may determine that it is not possible to estimate inefficiency. This would suggest a need to review and extend the data and review the model specification.

Q23 - Are there types of expenditure that we should model using only historical or forecast data?

Historical data is robust for activities which are consistent that do not have any known cost pressures/efficiencies.

Forecast data allows GDNs to reflect workload changes and cyclical investment decisions into future business plans costs.

We have some specific costs that would need to be considered specifically for historical or future are below;

Repex contractual benefit – as we have benefitted from contractual savings in RIIO-GD1 of a material nature, if using historic data, we would need to consider 'trueing up' the

[51] Oxera (2013), 'Recommendations on cost assessment approaches for RIIO-ED1', February.

[52] While we consider that SFA should be used to directly estimate inefficiency, it could also be considered to inform the choice of an appropriate benchmark. As SFA can separate noise from inefficiency at company level, the estimated efficiency scores can be used to inform the choice of benchmark in an OLS or RE context for a particular company. For example, if an SFA model predicts a company to have a lower efficiency score than upper-quartile-corrected OLS or RE, it may indicate that an upper-quartile benchmark is lenient, possibly overcompensating for noise or advantageous company effects. The converse may hold if the SFA model predicts a higher efficiency score than OLS or RE. This would make the choice of upper-quartile (or another benchmark) for a particular company less dependent on ad hoc adjustments and judgements.

saving in the historical cost base to reflect the expectation that it will not reoccur due to contract end.

Atypical costs – these need to be normalised in both historic and future data when considering use in regressions.

Loss of metering – when carrying out emergency regressions we need to consider the impact of the metering contracts during RIIO-GD1, which means historical costs will not be reflective of future costs, for certain GDNs and can impact the regression results using historic data.

Any costs of a cyclical nature that we do not hold panel data for the duration should be reflected in a mix of historical and future data. Items like vehicles and asset investment costs should reflect a mix of the data available.

Q24 - If we use a combination of aggregated and disaggregated modelling approaches, how should we determine the weight we apply to each?

As mentioned in the response to question 22, with respect to different estimation approaches. A similar argument can be made with respect to alternative model specifications. No one knows what the true underlying relationship is and reliance on only one model risks that the outcome may be due to the idiosyncrasies of the particular model used. As such, multiple models and multiple levels of aggregation should be consider.

As mentioned in the response to question 22, different cost drivers provide proxies for the key operational factors to different extents for different companies. That is some drivers are more important than others for some companies.

As such, the results should then be compared and, based on an understanding of the models and their relative pros and cons of each, differences between the outcomes for individual companies reduced as far as possible, in order to reach achieve greater consensus.

One approach, for example, is to use a general industry wide model, but use other models to capture more company-specific factors. The difference between the two might provide estimates for company specific cost adjustment claims.

Another approach is to use some form of triangulation of results across appropriate model specification. This can provide a more balanced view of companies' cost performance. For example, in the German energy market companies are given the benefit of the doubt in that the best outcome from four models is used. [53]

Model quality could be one of the key criteria to consider when deciding on the weights to give to the different modelling approaches. This primarily refers to the operational relevance of the cost drivers included in the model as well as their statistical significance and other statistical robustness tests.

[53] Section 12 of the Incentive Regulation Ordinance, ARegV. http://www.gesetze-im-internet.de/aregv/__12.html