RPI-X@20: The future role of benchmarking in regulatory reviews

A FINAL REPORT PREPARED FOR OFGEM

May 2010

RPI-X@20: The future role of benchmarking in regulatory reviews

RPI-X@20: The future role of benchmarking in regulatory reviews

Executive Summary

Frontier Economics was commissioned by Ofgem to produce a report on the future role of benchmarking in regulatory reviews in light of the proposals emerging from the RPI-X@20 review. In particular, Ofgem asked us to consider the potential role of total cost benchmarking and provide practical recommendations on the preferred approach for all four of the networks regulated by Ofgem (i.e. electricity transmission, gas transmission, electricity distribution and gas distribution).

Context

Over the next decade, the context for benchmarking is likely to change for a number of reasons. Most importantly, the need to decarbonise the economy has important implications for energy networks, and consequently for the use of benchmarking at price control reviews.

Decarbonisation and other changes in the energy sector are likely to change the scale of network activities, and alter the scope and cost structure of those activities in ways that are difficult to predict. Although there is uncertainty over what operators might need to do in the future, stimulating investment and innovation is likely to be important. If there are changes in underlying network cost structures, this might weaken the ability of historic cost analysis to inform on future expenditures. Ofgem's proposed output led approach to regulation, if implemented, could also alter the context for benchmarking making it likely that more potential cost drivers will need to be considered.

The threat of cost disallowance arising from ex post benchmarking – and potentially the stranding of past investment under a total cost benchmark – could have the effect of undermining incentives to invest, innovate and incur cost in transforming networks to meet the carbon challenge. Further, since the 'new' costs required in the context of decarbonisation are subject to uncertainty in both their volume and timing, there is a risk that benchmarking historic costs will not provide reliable, informative results - the sample size is unlikely to be sufficiently large to discriminate between inefficiency and the 'noise' caused by operators doing different things at different times. Finally, since a number of the networks might change their role materially over the coming years, there is potentially an argument that making use of historic benchmarking to inform directly on allowances might be less appropriate than in the past.

Consequently, our view is that the greater risk to customers is in setting inappropriate targets at regulatory reviews rather than in failing to penalise inefficient performance ex post, especially when the attempt to penalise inefficiency may undermine incentives to innovate and deliver the wider set of outputs that is now required.

These considerations imply that, in the short to medium term at least[1], there should be less emphasis on ex post benchmarking of historic costs and a greater emphasis on benchmarking future plans, assessing the extent to which they represent value for money for customers. Undertaking this analysis at the total cost level creates a more equal treatment of opex and capex in benchmarking, minimising the extent to which benchmarking might distort input choices.

However, it would be unwise to establish a regime based on the benchmarking of plans alone. It is well understood that in a regulatory context there is an incentive to inflate plans. Benchmarking of plans alone is unlikely to mitigate that incentive. Ofgem should therefore seek to supplement its assessment of future plans with an assessment of historic cost, using the results as a way to challenge operator plans, rather than to determine allowances mechanistically. The Information Quality Incentive (IQI) is an additional mechanism for addressing this incentive to inflate, although alone it is unlikely to countervail such incentives completely. The portfolio of evidence arising from this process will support Ofgem in its engagement with the operators through the price control, rather than being used to identify future allowances directly.

We recognise that the regulatory prescription described here is not readily applicable to the transmission networks. Since there are only three electricity transmission operators and a single gas transmission operator, the scope for benchmarking future plans against other future plans is likely to be limited. Further, it is unlikely that a regulator or transmission operator in another regime will have any information that is readily comparable and even if this information does exist it may not be able to be shared with Ofgem, reducing the prospects for international benchmarking of future plans in transmission. As a result, the central plank of the assessment process described above appears impractical for the transmission networks. We have therefore developed alternative proposals for transmission.

Criteria for assessing possible methodologies

Based on our assessment of the context described above we have considered a number of candidate benchmarking methodologies that Ofgem might adopt. We have assessed these methodologies against a number of criteria.

- **Robustness**: the benchmarking process and the resulting performance assessment must be regarded as robust by the operators and peer reviewers. A technique that produces results that are not sufficiently robust will be of

---

1 Over time as more data is collected and the industry heads towards a new "steady state", it will be possible to revert to greater use of ex post benchmarking, as has been the case in the past.

little use in a regulatory context and will struggle to stimulate information revelation.

- **Transparency**: if benchmarking methodologies are clear it will aid the ability of all stakeholders to understand the rationale for the selected approach. It will also be clear to the operators what conduct is being encouraged.

- **Promotion of efficiency**: benchmarking techniques should promote not just efficient cost management, but also strike an appropriate balance between low costs and desired outputs. Benchmarking methodologies should also minimise the extent to which they distort incentives to favour one cost type over another.

- **Consistency with the wider regulatory framework**: benchmarking should foster the high level objectives of the wider regulatory regime and strike an appropriate balance between different objectives. Benchmarking should also encourage operators to innovate while providing appropriate protection from unnecessary expenditure for customers.

- **Reasonableness of data requirements**: any benchmarking technique will only have merit if the necessary data exists to populate it.

- **Adaptability**: given the likelihood of material changes in the availability and relevance of certain data over time as network roles evolve, there is merit in pursuing a benchmarking technique that can adapt and remain fit for purpose.

- **Resource cost**: approaches that impose significant additional cost on Ofgem and the regulated operators should only be adopted if they deliver materially better information.

The range of possible techniques

To aid description of the wide range of approaches that might be adopted, we have identified four key dimensions in which there are material choices over how to benchmark operators.

- **Costs**: the most important choice is whether to assess performance on a "top down" basis, i.e. using total cost measures, or on a more disaggregated basis (e.g. separate assessment of opex and capex).

- **Cost drivers**: benchmarking models need to take account of the key cost drivers of the business. A comprehensive set of cost drivers should be included in order to capture as fully as possible the scale of the network task that each operator is required to undertake. Additional variables could capture valued outputs, such as quality of supply.

- **Sample**: should the sample be extended to include operators from other countries? Should historic data be supplemented with data contained in future plans?

- **Technique**: a wide range of different techniques are available based on for example well known statistical methods. Alternatives adopted by other regulators include non-parametric techniques and techniques that make use of fundamental engineering models.

Candidate methodologies

Using the dimensions specified above we have specified in more detail four approaches that have been subjected to more detailed assessment. Consistent with our remit, we have focused most attention on total cost benchmarking.

- Option 1: regression analysis of the costs, explanatory factors and outputs contained in business plans submitted by the operators at each review.

- Option 2: regression analysis of historic cost against a range of explanatory factors and outputs, grouping competing costs together. We describe this approach as DPCR5 "lite".

- Option 3: benchmarking through the use of a Total Factor Productivity (TFP) based technique (similar to the approach adopted by the Dutch energy regulator for example).

- Option 4: benchmarking through the use of Model Network Analysis (MNA) or Reference Network Analysis (RNA).

We have considered the merits of each of these approaches in detail, using the criteria outlined above and taking account of the specific context of each of the networks. On the basis of these assessments we have developed a set of recommendations.

Recommendations

Electricity distribution

Of the four network sectors, electricity distribution is the sector most likely to witness a step-change in its activities over the next decade as a result of decarbonisation objectives. Energy efficiency measures, small-scale generation plant, the electrification of heating and the likely roll-out of electric cars on a wider scale will all impact the volume and direction of flows on distribution networks in ways that appear difficult to predict. There is at present no consensus on the ways in which electricity distribution networks might need to adapt to accommodate them. This gives rise to the need to encourage innovation and proactivity in the Distribution Network Operators (DNOs) and to a potentially material uncertainty over the future activities and cost structures of distribution networks.

We are proposing a regulatory regime that is centred on the benchmarking of future plans at a total cost level. It will also be important to ensure that historical information is used effectively to check the basis on which plans have been prepared. Given the volume and value of future investments there is likely to be a continued roll for expert scrutiny of business plans and the use of historic benchmarks. We are therefore recommending the continued application of historic benchmarking, but with a modified role. In past reviews historic benchmarking of operating costs has been used more prescriptively to determine future allowances. In future, we believe that evidence gained from historic benchmarking should be used to challenge the validity of future plans, acting as a further point of traction for the regulator, rather than for setting allowances in a more mechanistic manner. Effective and targeted use of analysis of historic data should assist Ofgem in identifying any unwarranted inflation of operator plans. Given the relevance of this concern, as discussed above, we also envisage a continued role for incentives for efficient and effective forecasting, such as the existing IQI mechanism.

If Ofgem wishes to continue to undertake ex post benchmarking to support its assessment at regulatory reviews, we would advise it to focus attention on benchmarking business support costs. As Ofgem identified at DPCR5, these costs are only weakly, if at all, substitutable for more direct network related expenditures. As a consequence it is unlikely that ex post benchmarking of business support costs would reduce materially incentives to innovate and invest.

Over time, as the present uncertainty is reduced and the industry returns to a new steady state, Ofgem could consider implementing an alternative approach. Our review has identified that a TFP based approach has a number of important strengths. Under the right circumstances, a TFP regime could deliver significant benefits to customers through the creation of strong incentives while also simplifying the existing arrangements. Once the present uncertainty has played out, such a system will warrant active consideration in our view.

A summary of our recommended approach is presented in **Table 1** below.

Table 1. Summary of recommendations for Electricity Distribution

	Recommendation
Costs	Total cost, making use of two measures. Planned operating expenditure plus a measure of capital consumption. Planned operating expenditure plus planned capital expenditure.
Cost drivers	Ideally, the full set of explanatory factors presented in Section 4.3, guided by empirical analysis at each review. Include directly, where possible, outputs[2], if supported by empirical analysis.
Sample	The scenarios presented in the 14 DNO business plans. Make use of historic costs (as per Option 2 in Section 5) to increase the scope for plans to be tested.
Technique	While Stochastic Frontier Analysis (SFA) is usually preferred when undertaking efficiency analysis, data contained in operator plans will not contain statistical noise. This allows the robust use of Ordinary Least Squares (OLS) or Corrected OLS (COLS).

Gas distribution

Much of the discussion presented above applies equally to the Gas Distribution Networks (GDNs). In particular, there are a number of uncertainties over the future role of the gas industry as a result of decarbonisation. Elements of the decarbonisation programme, such as the electrification of space heating, will tend to reduce the demand for gas (and hence gas networks). On the other hand, gas distribution networks may have an important role in supplying small scale gas-fired electricity generation plant. Such plant might be needed to ensure system

2 To clarify a point of nomenclature, in this report we distinguish between 'explanatory factors', which are sometimes referred to as outputs in a DEA or regression analysis (i.e. where any potentially significant cost driver might be termed an 'output') and 'outputs' in the sense that the word is used in Ofgem's *Emerging Thinking*. To demonstrate the difference, a benchmarking study would ideally seek to take account of differences in connection density by including an appropriate variable. This would be referred to as an output in the benchmarking study, but is clearly not an 'output' in the sense implied by Ofgem. We therefore identify outputs in the sense in which Ofgem have used the word in their *Emerging Thinking* specifically as 'outputs', while outputs in a benchmarking context are referred to as 'explanatory factors'. See Footnote 6.

stability as the penetration of intermittent renewable generation increases. Further, there are potentially new uses for the gas networks resulting from decarbonisation. Whether or not biogas will become important, for example, is currently uncertain, as is the impact this might have on the gas network costs and activities.

As with electricity distribution, therefore, these uncertainties could lead to changes in the underlying scale and scope of gas distribution network activities. Consequently, our central recommendations for the gas distribution sector are very similar to those we have provided for electricity distribution.

We therefore propose a regime centred on Option 1, the benchmarking of future plans, supported by historic cost benchmarking via Option 2. We see substantial merit in focusing the regime on each operators' value for money proposition, where proposed costs are matched against the delivery of valued outputs. In particular this approach appears consistent with the philosophy set out in Ofgem's *Emerging Thinking*. It also removes the need for Ofgem to address the question of how to interpret and use total cost benchmarking based on purely historic costs. Under this approach Ofgem will not find itself in a position where it could consider writing off past investments. To do otherwise would be an important departure from past practice and would be viewed as a new risk for the regulated networks.

Notwithstanding the above arguments, there is likely to be scope to make more use of historic benchmarking in gas distribution than in electricity distribution. It appears reasonable to assume that the uncertainty over future gas networks activities appears less material than for electricity networks. A number of the envisaged changes (in particular electric vehicles and the deployment of distributed generation) are only likely impact electricity networks. It might be argued that the need to encourage innovation in gas distribution is consequently diminished. Ideally ex post benchmarking of historic costs would be undertaken on the basis of total costs. However the inherent lumpiness of capital investment coupled with some uncertainty over the future role of networks suggests that historic cost benchmarking might continue to focus on benchmarking competing operating costs in the short to medium term.

A summary of our recommendations for the gas distribution sector is presented in **Table 2** below.

Table 2. Summary of recommendations for Gas Distribution

	Recommendation
Costs	Total cost, making use of two measures. Planned operating expenditure plus a measure of capital consumption. Planned operating expenditure plus planned capital expenditure.
Cost drivers	Ideally, the full set of explanatory factors presented in Section 4.3, guided by empirical analysis at each review. Include directly, where possible, outputs, if supported by empirical analysis.
Sample	The plan scenarios presented in the 8 GDN business plans. Make use of historic costs (as per Option 2 in Section 5) to increase the scope for plans to be tested.
Technique	While Stochastic Frontier Analysis (SFA) is usually preferred when undertaking efficiency analysis, data contained in operator plans will not contain statistical noise. This allows the robust use of Ordinary Least Squares (OLS) or Corrected OLS (COLS).

Electricity and gas transmission

From our review of a range of possible approaches, it is clear that there are material obstacles to the direct implementation of several of them. The approaches that we have recommended for distribution, centred on the benchmarking of forward looking plans, are impractical at the transmission level as the relevant data will not be available.

We propose that Ofgem considers adopting a high level DEA benchmark of the recent historic costs of the transmission operators against a small number of European peers. Given limits on the data that is likely to be available we recognise that this approach is unlikely to provide definitive results. We therefore propose that this analysis is undertaken at an early stage of the regulatory review process to enable it to be used to inform discussion between the regulator and the operators, rather than to guide more mechanistically the setting of allowances. We also propose that Ofgem supports this analysis through the use of a range of other pieces of analysis, including continuing with its prevailing practice of seeking expert review of the operators' plans.

The benchmarking we envisage at the transmission level is very different from that proposed for distribution, where the richer set of data available allows for more refined and ambitious benchmarking. An iterative approach will need to be adopted for transmission, using the data that is available to the maximum extent and regarding benchmarking analysis as an important point of engagement with the operators. Given the limits of this analysis, we propose that Ofgem makes use of as many pieces of analysis as possible. For example it could compare the TFP of transmission operators against the TFP of the distribution operators. Similarly, for the three GB electricity transmission operators it might be possible to undertake some simple regression analysis using the available panel of data. Since there might be only 10-20 data points in the sample, it is likely that only very simple regression (perhaps in line with the analysis used at DPCR4) will be possible. Finally, there is clearly merit in continuing with a range of other analysis, including expert scrutiny of business plans as at present.

While none of these approaches individually is likely to provide sufficiently definitive information on the reasonableness of the transmission operators' costs, together they will provide Ofgem with the widest possible range of information. This could form the basis of a constructive dialogue with the operators throughout the price control process.

The proposed DEA approach is summarised in **Table 3** below.

Table 3. Summary of recommendations for the Transmission businesses

	Recommendation
Costs	Total cost, making use of a standardised measure of capital consumption. Given the sample, there will be a need to adjust costs to reflect a wide range of factors, including exchange rates, tax etc.
Cost drivers	Drawn from the set of explanatory factors presented in Section 4.3, but guided in practice by the data that is publically available. Include directly, where possible, outputs, but recognising that limited data is likely to be available.
Sample	The GB operator(s) supplemented by a number of operators from other countries (e.g. 4-6 others). Use historic data for the most recent year available to develop a cross section.
Technique	DEA, using a 1 input multiple output model. We would propose that Ofgem investigates both a constant and variable returns to scale frontier. From a regulatory perspective a VRS frontier is likely to be reasonable since the scale of operation of a transmission operator is fixed by the size of the country in which it operates. However, a VRS frontier can reduce the ability of the DEA model to discriminate between operators, and in small samples can often result in all operators appearing efficient.

1 Introduction

Frontier Economics[3] has been commissioned by Ofgem to produce a report on the future role of benchmarking in regulatory reviews in light of the proposals emerging from the RPI-X@20 review. In particular, Ofgem wishes to consider the potential role of total cost benchmarking. Practical recommendations have been provided for all four networks (electricity transmission, gas transmission, electricity distribution, gas distribution), reflecting the context specific to each.

The remainder of this report is comprised of the following sections.

- Section 2 provides the motivation for benchmarking within a regulatory context. We also discuss the challenges associated with undertaking benchmarking analysis, the context within which benchmarking will be undertaken at subsequent reviews by Ofgem, and consequences for the role of benchmarking.
- Section 3 describes the criteria we have used to assess the merits of different candidate benchmarking methodologies.
- Section 4 reviews the wide range of techniques that could be adopted, providing some initial insights into the merits of different approaches.
- Section 5 focuses attention on a short list of potential benchmarking approaches, providing concrete specifications, assessing the merits and identifying any application issues that arise.
- Section 6 provides our recommendations for each of the four networks.

[3] We are grateful for the advice and insights provided by Professor Tom Weyman-Jones over the course of this assignment. Professor Weyman-Jones bears no responsibility for any remaining omissions or errors in the report.

2 Motivation

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

In this section we review briefly why benchmarking can be helpful to a regulatory authority in determining the prices and/or revenues that a regulated operator might be allowed to charge/recover. We also review some of the practical difficulties that arise in identifying any benchmark. We conclude this section by discussing the context within which future benchmarking studies will be undertaken and identifying the likely consequences for future benchmarking.

2.1 Why benchmark?

This section is intended to provide a brief background in certain elements of the theory of regulation. Within the academic literature there is a strong theoretical case for making use of benchmarking in an incentive based regulatory regime. We provide a high level overview of these insights here. Readers familiar with this material can proceed to later sections of this report.

To summarise the argument, benchmarking, if applied appropriately, is an effective tool for overcoming the asymmetry of information between the regulator and the regulated operators. In the absence of benchmarking, the regulator must judge whether or not a regulated operator is performing at the maximum possible level of efficiency on the basis of information revealed by the operator itself. If there are penalties for inefficient behaviour, the operator is unlikely to reveal information that makes it appear inefficient. Therefore, in order to provide an incentive for the operator to reveal this information, the regulator must offer some sort of reward. Ultimately, the cost of that reward is borne by customers.

If benchmarking is possible, the regulator will be able to compare operators to retrieve information on the possible level of performance 'independently' of the information revealed by the operator itself. Benchmarking can therefore be used to reduce the size of the reward offered to operators, without necessarily reducing the strength of incentives.

2.1.1 The managerial firm

Economic theories of the "managerial firm" recognise that the interests of managers and shareholders are not perfectly aligned. Shareholders are usually concerned solely with the objective of maximising profits. Managers will also wish to ensure that the firm makes a profit, but will have other objectives too, such as their remuneration, the effort they need to put into their job, the desire for personal progression etc. If managers are not focused solely on delivering profits, suboptimal performance (from the perspective of shareholders) might result.

However, it can be difficult for shareholders to understand the sources of operator performance. If, for example, profits are falling, shareholders do not necessarily know if this is the result of poor performance by managers, or a consequence of unforeseen and uncontrollable events. In short, there is an asymmetry of information between shareholders and managers.

In response to this information asymmetry, shareholders can choose to either:

- meticulously scrutinise every detail of the business in order to monitor managerial activity and performance better; or
- seek to design employment contracts and incentive schemes that align the interests of managers and employees far more closely with those of shareholders.

In most circumstances the former approach is impractical and would be very costly to implement. The latter approach is typically preferred and is often the only practical alternative. Indeed, performance contracts, in which managers and employees benefit directly from increasing profits, are now commonplace.

This gives rise to a variety of questions over how such contracts should best be designed.

2.1.2 Parallels with regulation

There are clear parallels between the shareholder-manager relationship described above and the relationship between regulator and regulated operator. The regulator wishes to ensure that the operators it regulates deliver a range of valued outputs in an efficient manner, but is unable to monitor perfectly the conduct of the regulated operators.

Incentive regulation seeks to overcome these challenges by aligning the interests of the operators with those of the regulator (where the regulator is required to act in the best interests of customers). It centres on the design of regulatory arrangements that allow operators to earn additional profits if they deliver a range of outputs and manage their costs effectively. This type of regulation has been implemented in the past by Ofgem across the four networks it regulates.

Incentive regulation can be described by considering the 'value' created by improvements in efficiency. In principle, customers should benefit from efficiency improvements. However, in order to provide incentives to make efficiency gains, the regulated operator must also benefit from making such improvements. Thus, the value created by efficiency improvements must be shared somehow between customers and the regulated operator.

Regulatory incentive mechanisms can be thought of as determining the share of improvements that the regulated operator is permitted to retain. The remaining value will accrue to customers. This creates a trade off between strong incentives for improvement (and hence more material improvements) and passing benefits to customers (so that they can benefit from those improvements).

If the share of any incremental improvement retained by the operator is very high then the reward for making improvements is also high, and operators can be expected to deliver more significant improvements. However it follows that the share of those benefits that subsequently accrues to customers is low. Conversely, if the share of any incremental improvement to be retained by the operator is low, then the operator has a smaller incentive to make savings and can be expected to deliver less significant improvements. However, a greater share of these improvements will flow to customers.

It is not clear from this characterisation whether stronger or weaker efficiency incentives are most beneficial for customers. With the former, customers gain a small share of large efficiency improvements, whilst with the latter customers gain a large share of small improvements.

While this discussion is highly abstract, it reveals a key tension in regulatory design between strong incentives and the speed with which improvements are passed through to customers[4]. Regulators will seek to strike an appropriate balance between these concerns. Although achieving this balance might be straightforward in principle, in practice it is challenging.

The type of incentive regulation described above can give rise to an additional problem. Under most regulatory regimes, the regulator collects information from operators on the costs they expect to incur in the future. These forecasts help the regulator form its view of the level of allowances it should put in place. At the subsequent review, if an operator has spent less than its allowance, that underspend would be deemed an 'efficiency improvement', and the operator would retain a share of that improvement. In other words, regulated operators can increase their profits by spending less than they are allowed, and allowances

4 To provide an extreme example, the strongest possible incentives for improvement would arise if the regulator promised the operator it could retain all of the future benefit arising. While this would put in place incentives to stimulate the most material possible improvements in performance, it would be of no absolutely no benefit to consumers.

depend (at least to some extent) on the forecasts provided by the operators. Operators therefore have an incentive to inflate their forecasts to try and convince the regulator that future costs will be "high". If successful, the operator would be able to underspend its "high" allowances and increase its profits.

Ofgem's Information Quality Incentive (IQI) mechanism, first applied at DPCR4 and subsequently at the GDPCR and DPCR5, is an example of a mechanism that attempts to counteract this incentive.

2.1.3 Making use of outside information

If there is more than one regulated operator, the framework described above can be modified. For example, if there were two essentially identical firms undertaking the same task, a comparison of the performance of both operators would yield useful information to the regulator for the purposes of setting allowances. Suppose Operator A is able to deliver the required service to the same standard as Operator B, but for a lower price. Since the operators are identical, the regulator knows that Operator B can achieve the same level of performance as Operator A. Given this, the regulator could simply require Operator B to match the performance of Operator A.

Importantly, the information revealed by a comparative efficiency analysis of this kind is retrieved 'independently' of the operators themselves. The regulator is able to determine allowances without the need to reward operators for bringing forward information. As such, if comparative analysis is possible, customers will not need to pay a price through an incentive scheme.

The availability of additional information beyond that revealed by the regulated operator, therefore, allows the regulator to secure benefits for customers without incurring all of the costs described above.

2.1.4 Comparative regulation

We can extend the example above to a case where there are many similar firms. In this case, suppose the regulator committed to set future prices for all of these operators at the same level for all firms, based on the industry average level of costs incurred (i.e. imposing an average benchmark drawn from industry wide data). Operators that are able to beat the average, i.e. lower their costs below the industry average, will therefore earn additional profits, while less well run firms that are unable to match the industry average will see their economic profits eroded. Unlike the single firm case, where once the operator has revealed the ability to do the same job for less the regulator will capture (at least some of) the benefit for customers, the operator's own cost reductions will reduce the industry average only slightly, allowing that operator to retain most of the benefit for itself.

A model of this kind, based on a benchmark, can therefore provide very strong incentives to operators to improve performance. However, what is true for one operator is true for all, implying that it is reasonable to expect all operators to improve performance to a similar degree. Hence, while strong incentives are provided to operators, these benefits can (and will) be rapidly transferred to customers.

The high level regulatory framework described above fits with the view that regulation can act as a proxy for competition. Indeed the regulatory rules described above mimic closely the sorts of competitive dynamics that can be found in unregulated markets where direct competition is possible.

2.2 What are the challenges associated with benchmarking?

The highly stylised descriptions provided above capture the theoretical benefits of undertaking benchmarking. However, real network businesses are complex, making use of a wide range of inputs and delivering a diverse basket of outputs. This gives rise to a range of practical difficulties associated with benchmarking. We discuss a number of these briefly below.

2.2.1 Minimising distortions between cost types

Operators incur many different types of cost in undertaking a range of different activities, all of which contribute to the delivery of network services.

Where those different cost types are exposed to different incentive arrangements, and where there is scope to substitute those costs for one another, there is a danger that operators respond to those incentives by reclassifying costs from one type to another, in order to increase profits. Where the difference in incentive strength is particularly material, there is even a risk that operators might physically substitute one cost for another, i.e. might change the specification of a project, because even though it results in a higher cost in aggregate, it increases the profits earned through incentive arrangements. The equalisation of incentives across competing costs is now a well established principle in incentive design.

Ideally such equalisation should extend to benchmarking. For example, suppose we have a regulatory process that determines a future allowance for one cost type based on an aggressive benchmarking methodology, while a second competing cost has an allowance set on the basis of operator plans with minimal challenge. Operators would rationally seek to classify more costs to the cost type subject to weaker scrutiny. If instead the operator classified those costs as the type subject to aggressive benchmarking it would give rise to a material risk of those costs being disallowed.

To avoid these distortionary input incentives, competing costs should not only face the same incentive arrangements, but should also be subjected to as similar a benchmarking process as possible (ideally identical). This would more perfectly align the interests of the operators with those of customers, leaving an incentive to seek the lowest cost solution rather than favouring one type of cost over another.

2.2.2 Total cost benchmarking

The above argues in favour of treating all costs (or at least all competing costs) equally, leading to the conclusion that total cost benchmarking might be part of the preferred approach to benchmarking[5]. While there are many benefits from adopting such an approach, principally around reducing potential distortions described above, it does give rise to a number of technical and regulatory issues. In particular, the treatment of capital costs can prove difficult. There are two obvious approaches to the measurement and benchmarking of capital costs.

First, the flow of capital expenditure could be added to other costs and this total expenditure (totex) subject to a single benchmarking process. However, capital costs are sometimes "lumpy" in nature. Consequently, the benchmarking results for a totex model might be volatile from year to year and the results for any single year might be an unreliable guide to prevailing performance. Similarly, operators might be at different points in their investment cycle. Some operators might be undertaking significant renewal programmes, while others might have recently completed large programmes and will therefore incur lower levels of spend. Again, it is not clear that a totex assessment will provide reliable results if such differences are material. Some of these problems could be solved by, for example, benchmarking a rolling average of totex (or equivalently aggregate expenditure over a price control period). However, if efficiency performance is changing rapidly this could result in certain operators appearing inefficient as a consequence of past inefficiencies that have already been addressed.

Alternatively, operators could be benchmarked on the basis of their ongoing operating costs together with a measure of their capital consumption (i.e. depreciation plus return on regulated asset value). This has the effect of smoothing capital costs since no single year of capex has a disproportionate impact on measured capital consumption. However, there are issues surrounding the interpretation and application of such a benchmark. The measurement of capital consumption is potentially controversial. The most obvious basis for such analysis is to make use of the prevailing regulatory accounting arrangements, i.e. to use the existing RAV and consistent depreciation and return allowances.

5 It should be noted that benchmarking is not necessarily the only tool that can be used to assess total costs. Ofgem is likely to employ a suite of tools over time, of which benchmarking is one, that will be used to inform final decisions on allowances.

However, the RAV is a regulatory construct that serves a number of purposes and is not necessary an ideal economic measure of the capital resources consumed in any given period of time. As such the use of existing RAV as a basis for measuring capital consumption could give rise to benchmarking results that require very careful interpretation. For example, where an operator is found to be inefficient on total cost, is the implication that some of their past capital investments be written off? Adopting such a policy would represent a material change in regulatory approach, exposing operators to a new risk that previous investment decisions might be subjected to ex post review and then, at least partially, written off.

At DPCR5 Ofgem put in place a set of incentives under which there was an equal treatment of all competing costs. Costs were split into two pots, network operating costs and closely related indirect costs and business support costs, based on the view that business support costs could not be substituted to any material degree with network operating costs and hence no material distortion would arise from separate treatment. When implementing total cost analysis in future there is potentially merit in undertaking analysis where business support costs are benchmarked separately from all other costs. This would align Ofgem's benchmarking with its existing incentives framework, ensuring the most equal possible treatment of competing costs. This provides a different approach to minimising potential incentives to distort input choices that remains consistent with the main arguments in favour of total cost benchmarking.

An alternative approach is to benchmark only ongoing operations, taking the existing network configuration and past investment levels as a given, largely removing capital costs. This model does not hold current network managers to account for past decisions, only for ensuring that the existing network is maintained in a least cost manner. While this removes the issue of capital measurement, the obvious drawback is that it provides only very weak incentives to plan the ongoing expansion of the network optimally. Given the very material investments that are likely to be required over the coming years, this approach therefore appears undesirable.

2.2.3 Capturing important differences between operators

No two network businesses are exactly same. Since no two operating regions are exactly the same, some observed differences in costs might be justified by these differences. Ideally, the benchmarking methodology employed should seek to take appropriate account of such factors, allowing any differences in performance as measured by that technique to be ascribed to differences in relative efficiency. In practice, this can be difficult to achieve.

Differences in perceived performance can arise from a number of potential sources including underlying differences in:

- input costs (e.g. labour rates, local taxes);

- operating environment (e.g. geography, topography, soil properties, and the urban/rural nature of certain areas);
- past configuration decisions and planning constraints; and
- current managerial and operating efficiency.

Some of these factors are straightforward to correct for, such as local taxes. Others are far more complex. For example, what allowance should be made for operators in urban areas, to account for time lost due to congestion? Or should serving an urban area, which will be densely populated, be regarded as giving rise to an advantageous environment, since many customers can be served with fewer network assets than would be required to serve a rural area that is likely to be sparsely populated?

Further complications arise when attempting to benchmark operators in different countries, as might be necessary for transmission where there are few operators within GB. There could be material differences in a number of additional areas to those identified above, including:

- legislative framework (e.g. employment, environmental, planning, tax and health and safety law etc);
- regulatory arrangements (e.g. data collection processes, incentive frameworks, scope of licensed activities, boundary/interface with other businesses etc);
- cost of capital and other financing arrangements; and
- exchange rates.

Designing a benchmarking methodology that accounts perfectly for all of these factors is unlikely to be possible. As a consequence it will be necessary to interpret the results of efficiency analysis with care. It might be necessary to accept that benchmarking results will need to be used, along with a range of other evidence, to inform on efficiency rather than providing a definitive basis for comparison.

2.2.4 Outputs

Ofgem has indicated its intention to put in place arrangements that focus on providing strong incentives for operators to deliver a basket of outputs[6], defined

[6] The use of the word "outputs" in this context gives rise to some scope for confusion in nomenclature. In a DEA benchmarking context any potentially significant driver of cost might be referred to as an output. These outputs, in the context of a DEA study, are typically the same set of independent variables that might be included on the right hand side of a regression equation in a regression based benchmarking study, and the description of these right hand side variables as outputs in regression studies is also relatively common. As a consequence the set of variables that might be regarded as outputs in a benchmarking study will include some measures that would not be

at a high level along six dimensions[7]. Ideally benchmarking should promote this objective, ensuring that operators are rewarded for providing valued outputs.

The ideal approach that might be adopted is to include a set of additional variables, which measure the desired outputs directly, within the final benchmarking model. Operators that deliver higher quantities of outputs would then be rewarded for this activity automatically through the benchmarking methodology, since this high level of output delivery will be factored into the model prediction of the underlying efficient level of costs. In principle this allows operators to choose the level of output delivery they think is right for the customers attached to their networks, supported by direct engagement with those customers and taking account of the incremental impact on costs.

While such an approach clearly has a number of highly desirable features, there is considerable doubt over whether it is practical. The high level principles that will underpin a set of outputs are now taking shape, but there is uncertainty over exactly how those outputs might be parameterised. It is not clear that all of these outputs will necessarily lend themselves to being captured by a sparse number of numeric variables that are capable of being measured in an accurate and consistent manner across all operators. If a less prescriptive approach is adopted, in which operators can measure their performance on certain outputs on an operator specific basis, it is not clear that those variables will be suitable for direct inclusion in a benchmarking model. Similarly, it is unlikely that there will be historic data available on these output measures, limiting the extent to which ex post panel data analysis might be possible.

If highly robust data on outputs becomes available in the future, the inclusion of many output variables within a benchmarking model will quickly consume degrees of freedom and could weaken the statistical validity of the results. Furthermore, if it is possible to overcome the difficulties associated with including international data in a benchmarking sample it is very unlikely that comparable data on outputs will be available for international operators.

The extent to which a complete set of explanatory factors and output variables might be included robustly within a formal benchmarking study will depend on the properties of the variables selected and volume and quality of data available.

---

regarded as an output in the sense in which the word is used in Ofgem's *Emerging Thinking*. For example, as described above, a benchmarking study would ideally seek to take account of differences in connection density by including an appropriate variable. This would be referred to as an output in a benchmarking study, but is clearly not an output in the sense implied by Ofgem. In the interests of clarity, in this report we will identify outputs in the sense in which Ofgem have used the word in their *Emerging Thinking* as outputs, while variables that might be described as outputs in a benchmarking sense will be referred to as "explanatory factors".

7 Ofgem has identified 6 areas in which it will seek to develop and implement outputs. These are reliability, safety, the environment, conditions for connection, customer satisfaction and social obligations.

To an extent, this is an empirical matter that should be investigated by the relevant price control teams during work on each specific network. It is possible that, in the short term, a practical solution might be to consider outputs separately, alongside less ambitious benchmarking that excludes some or all output measures. As future data on costs and outputs is collected over time more sophisticated analysis might become feasible. In section 4 we discuss in more detail the potential for inclusion of outputs in a benchmarking regime

2.2.5 Making use of the results

The outcome of a benchmarking exercise is typically some measure of the distance between each operator's observed cost and the model's estimate of efficient cost. It is then necessary to decide how these "efficiency" scores might be translated into cost allowances. Numerous approaches have been adopted by different regulators around the world. Some examples include:

- deriving efficiency scores based on the absolute frontier of performance but allowing operators off the frontier some time to achieve that level of performance (i.e. a glide path);
- applying a less demanding frontier (e.g. based on average performance or the upper quartile) but perhaps requiring that level of efficiency immediately;
- putting in place an allowance based partly on the operator's own costs and partly on the efficient cost level identified by the benchmarking model; and
- creating a number of groups of operators that are regarded as having similar efficiency and requiring the same improvements of each.

There is clearly an element of judgement and regulatory discretion in which of these broad approaches to adopt. Such judgements might be based on, for example, the wider regulatory framework (which might make an apparently "tough" application of benchmarking results more reasonable) and also on an assessment of the robustness and accuracy of the benchmarking model.

One potential concern over the usefulness of benchmarking results is whether the activities carried out by networks are expected to remain stable over time or otherwise. If network activities are expected to change very materially then it is possible that past cost levels and structures might be a relatively poor guide to the costs that might be incurred in the future. If that were the case benchmarking of past performance would need to be used very carefully if the results were to be used to determine future allowances.

On the other hand, although the drive to decarbonise the economy is likely to alter the volume of network flows, the fundamental activities carried out by the networks might remain broadly stable. Thus, although it might be necessary to anticipate a different volume of demand, there may still be merit in benchmarking past performance in order to infer what level of cost is reasonable in the future.

It is clear from the discussion above that the context within which a benchmarking analysis is undertaken is highly relevant for how useful the results of that analysis will be. In order to identify a benchmarking methodology that is appropriate for use in price control reviews, it will therefore be necessary to have a full understanding of these contextual issues. We discuss the context for benchmarking in the next section.

2.3 Context for benchmarking

Traditionally, regulation and benchmarking have been undertaken in the context of a relatively stable industry, where the operators have experienced modest incremental growth for their services and there has been no underlying change in the characteristics of the production process. When combined with a legacy of inefficiency dating back to pre-privatisation business structures, this background created a strong expectation of falling costs.

Benchmarking was used to identify what progress had been made since the previous review, and was an input to the calculation of the "P0" cut. Further analysis identified what additional progress it was safe to anticipate during the course of the forthcoming price control period, to identify a common "X factor" to apply to all businesses[8]. Benchmarking focused on driving down operating expenditure, while funding capex plans subjected to expert review and accommodating relatively small and steady increases in demand.

We would argue that the context described above is changing in a way that has important implications for which benchmarking methodology is most appropriate. There are a number of material differences.

- **Decarbonisation**: the need to decarbonise the economy in coming years is likely to give rise to material changes in the scale of network activities, and could alter the scope and cost structure of their activities in ways that are difficult to predict at this time.
    - There is considerable **uncertainty** over what will be required of the network operators to facilitate decarbonisation[9]. It is not, at present,

---

8 Similar analysis was conducted at transmission reviews, and until recently at the combined gas transmission and distribution review, in order to identify an appropriate P0 and X.

9 It might be argued that this uncertainty is far more material for the electricity networks. There is unlikely to be a step change in the level of activity observed on the gas networks.

clear what actions will be required or what costs might be incurred in the process.

- Given this uncertainty, operators will need to be provided with an incentive framework that **encourages innovation and action**. Operators might seek to reduce their risk exposure by waiting for more information to emerge. However, policy makers might be seeking more rapid progress in order to meet wider environmental targets. Operators might also choose to undertake decarbonisation programmes only where the associated costs have been signed off by the regulator, rather than risk incurring costs that might be deemed inefficient/unnecessary in the future. This could have the effect of placing Ofgem in a position where it vets which projects should proceed and which should not, even if the operators might be better placed to make those judgements.
- Notwithstanding the uncertainty over what activity might be required, there is likely to be a need for a **substantial increase in expenditure** over the coming years, at least on the part of the electricity networks. Given the likely scale of that expenditure, **strong incentives** to seek efficiencies are likely to be a key element of future arrangements in order to ensure that money is well spent. While much of this spend is likely to be capex (supporting the view that capex assessment will be increasingly important) there is likely to be scope for some capex to be replaced by other types of expenditure (e.g. interruptible contracts to charge electric cars). If this is the case, then it is also increasingly important that the chosen benchmarking technique **does not distort the input choice** of operators.
- Not only is the anticipated expansion in network activities substantial, but there is also a significant degree of **uncertainty over exactly when expenditure increases will be required**. Ofgem's Social and Environmental Guidance, provided by the Secretary of State, outlines the key need for networks to invest ahead of time. However, this must be balanced by the need to avoid inefficient 'gold-plating' of the network. In this context, it is entirely possible that different **networks will incur "new" costs at different times and with different rates of ramp up**.
- Finally, if underlying activities and cost structures do change fundamentally, it could follow that past performance might be only a poor guide for what might be possible in the future. Using **benchmarks based on historic cost** to determine future allowances **might be less reliable** than has been the case in the past, when the industry was in more of a steady state.

- **Decreased risk of systematic inefficiency**: we would argue that after approximately 20 years of strong incentive regulation, coupled with an effective market for corporate control, it is likely that there is no longer a material, systematic legacy of inefficiency in the network businesses. This is particularly likely to be true of the electricity distribution operators, who have been subjected to comparative regulation throughout that period. If this is the case, the previously held view that costs will inevitably continue to fall in the future does not necessarily hold any longer. It is important to note that in making this assertion we do not wish to suggest that, as a consequence, there is no longer a role for incentive based regulation and benchmarking. All businesses require continued and active management in order to manage costs and ensure that their processes and operations are carried out as efficiently as possible, including adopting new innovations as they emerge. Given the context of decarbonisation described above, differences in efficiency can be expected to open up, but these will cycle as best practice spreads and different operators take the lead at different points in time. Incentives will therefore need to be provided to operators to uncover continued incremental improvements, and appropriate benchmarking will still do this effectively.

- **Ageing assets**: many of the network assets originally installed during the 1960s are reaching the end of their technical lifetime and need to be replaced. Recent increases in capex at both DPCR4 and DPCR5 reflect this increased level of replacement activity and create the context in which assessment of capex, and incentives for efficient execution of capex programmes, are even more important than in the past.

- **Outputs**: future benchmarking will take place in the context of a set of outputs. These will measure what each operator has delivered in the past and plans to deliver in the future. In the past operators might have had the ability to delay needed activity, reducing the volume of outputs delivered, in order to meet or beat allowances, then claim funding a second time for the delivery of those delayed outputs at the next review. In granting an allowance to an operator based on specified outputs, Ofgem will have more information on what the operator will deliver for that money than it has in the past. The strategy of delaying delivery of outputs is, therefore, no longer available. As a result, Ofgem might have more confidence that its framework of outputs will allow it to increase the strength of its incentive arrangements, encouraging incremental improvements in efficiency without encouraging operators to delay needed activity.

2.4 Consequences

The context described above gives rise to a number of important themes that together guide how benchmarking might be used in future reviews:

- The threat of cost disallowance – and potentially the stranding of past investment under a total cost benchmark – could undermine incentives to invest, innovate and incur cost in transforming networks to meet the carbon challenge.

- Since the 'new' costs required in the context of decarbonisation are subject to uncertainty in both their volume and timing, there is a risk that benchmarking historic costs will not provide reliable, informative results. For example, operators may incur different costs at different times in order to achieve the same decarbonisation objectives. The sample size is unlikely to be sufficiently large to discriminate between inefficiency and the 'noise' caused by operators doing different things at different times.

- Finally, since a number of the networks might change their role materially over the coming years, there is potentially an argument that making use of historic benchmarking to inform directly on allowances might be less appropriate than in the past.

Consequently, our view is that the greater risk to customers is in setting inappropriate targets at regulatory reviews rather than in failing to penalise inefficient performance ex post, especially when the attempt to penalise inefficiency may undermine incentives to innovate and deliver the wider set of outputs that is now required. These considerations may imply that, in the short to medium term at least[10], there should be less emphasis on ex post benchmarking and a greater emphasis on benchmarking future plans, assessing the extent to which they represent value for money for customers.

2.4.1 The merits of benchmarking future plans

Analysis of future plans could examine the total spend anticipated during the coming period against the present and anticipated volume of network activity. Undertaking this analysis at the total cost level creates a more equal treatment of opex and capex in benchmarking, minimising the extent to which benchmarking might distort input choices.

Assessment of planned total cost against explanatory factors and future increases in the outputs of the network focuses benchmarking on a high level assessment

10 Over time as more data is collected and the industry heads towards a new steady state, it will be possible to revert to greater use of ex post benchmarking, as has been the case in the past.

of value for money for customers. Under this approach, the risk of regulatory stranding of assets as a result of ex post benchmarking is avoided, but a further layer of discipline is imposed in the setting of targets, so that customers' interests in efficient prices are served. Such a regime could supplement and support longer regulatory periods which should provide a further natural incentive to improve cost efficiency.

2.4.2 Risks associated with benchmarking future plans

As discussed above, it is well understood that in a regulatory context there is an incentive to inflate plans. Benchmarking of those plans is on its own unlikely to mitigate that incentive entirely. While it might be able to detect inflation if this strategy was followed by a subset of companies., it would not be able to identify such inflation if this strategy was followed by all companies. The IQI is one mechanism for addressing this incentive to inflate, although it is unlikely to remove such incentives completely. In order to address this concern, there is likely to be a continued need for the targeting use of historic cost benchmarking, as we discuss below.

Linking treatment at future regulatory reviews to the ex post credibility of previous forecasts is another potential approach. Under this approach the regulator would revisit plans submitted at previous reviews to assess whether, ex post, there is reason to believe they had been unnecessarily inflated at the time.

We believe that implementing this approach might prove impractical. It might be difficult to prove that significant variation between forecast and outturn was a consequence of a misleading forecast rather than unforeseen events. It is likely that the regulated operators would also regard this approach as an unwelcome increase in regulatory risk. Also relevant is the uncertainty over exactly what activity might be required over the coming years. There is material uncertainty over what additional demands might be met and how those needs might best be served. Given this uncertainty, it is possible that an operator could submit a plan and then discover that some element of the anticipated spend was in fact not needed (for example as a result of a highly valuable innovation), allowing them to under spend their allowance very substantially. If operators believe that there is a prospect they might be "punished" for substantial under spend at the next review, it will create an incentive for them to simply spend the money anyway, i.e. the risk reducing behaviour will simply be to do whatever was in the plan. We believe, therefore, that this approach might have the effect of increasing expenditure and reducing innovation at a time when capital expenditures will be large and innovation highly valued.

2.4.3 Making use of analysis of historic costs

We propose that the solution to this is likely to be an additional assessment of forward looking plans based on an analysis of historical costs. As we pointed out earlier, although material changes in activity might well be anticipated in the future it is also reasonable to suppose that many elements of the underlying network activity will remain the same. For example, there is no reason to suppose that the cost of maintaining a kilometre of network will change very materially between now and the future. While the volumes of activity required might be highly uncertain, the costs that might be incurred *given* those volumes will be less so. This suggests that historic costs and historic cost benchmarking will remain useful in the future, but rather than being used to set explicit allowances could be used instead to challenge forecasts. This would make it clear that poorly evidenced claims for substantial increases in cost can and would be identified and removed from plans during regulatory review. The regulator could also make clear that operators found to have submitted unreliable information in one area would be prone to more active scrutiny, reflecting an increased belief that other elements of the plans might also be poorly supported. An approach of this kind coupled with a "reward" for bringing forward well evidenced, transparent and reasonably costed plans (i.e. something like the IQI mechanism) should mitigate against the underlying incentive to inflate forecasts.

2.4.4 Summary

The discussion above argues for the collection of a portfolio of evidence on efficiency using a suite of benchmarking analysis that:

- is focused on a high level assessment of the value for money offered by each operator's plan for the next 5 years[11]; and
- takes account of past costs to ensure that the core assumptions underpinning future costs are reasonable.

These ideas inform the assessments we document in subsequent sections of this report.

2.4.5 Benchmarking transmission operators

We recognise that the proposals described above are not readily applicable to the transmission sector. Since there are only three electricity transmission operators and a single gas transmission operator, the scope for benchmarking future plans is likely to be limited. Further, it is unlikely that a regulator or transmission operator in another regime will have any information that is readily comparable. Even if this information does exist, it may not be able to be shared with Ofgem, reducing the prospects for international benchmarking of future plans in

11 Additionally, it seems reasonable to argue that data on future plans will not be subject to noise in a statistical sense (i.e. the unforeseen events that cause random, rather than systematic fluctuations in performance). Given this there is no need to use SFA when analysing data in this way, placing less strain on the available degrees of freedom.

transmission. As a result, the central plank of the assessment process described above appears impractical for the transmission networks.

While it will be possible to directly benchmark elements of proposed transmission business plans (e.g. projected costs based on anticipated volumes, head office costs against similar sized organisations) there will be a continued need for the regulator to depend on:

- expert scrutiny of plans in order to reduce the asymmetry of information and identify areas where activity is obviously unnecessary or costs are obviously inflated; and
- strong incentives for the transmission operators to reveal through their actions during price controls what can be achieved.

Notwithstanding the material difficulties that arise when attempting to benchmark transmission operators, we discuss below how some high level benchmarking might support the regulators efforts at price controls.

3 Criteria for assessment

In this section we discuss the criteria we have adopted to assess the strengths and weaknesses of different benchmarking techniques. It will be immediately apparent that a number of these criteria are in conflict with one another, creating cases where it will be necessary to make trade offs. One approach to identifying the preferred balance in such cases is to create ranking of criteria. We have not formally adopted this process. However, in Section 5, where we review the merits of a shortlist of candidate methodologies, we have made certain judgements which have guided the recommendations we make in Section 6. Others might make different judgements and hence prefer a different approach.

3.1 Overview

Figure 1 provides an overview of the criteria we have used to assess the merits of different benchmarking methodologies.

Figure 1. Criteria against which to assess the merits of different benchmarking methodologies

We provide further details on each criterion in the following subsections.

3.2 Robustness

A critical property of any benchmarking process and the resulting performance assessment is that it must be regarded as robust by the operators and peer reviewers. Ultimately a technique that produces results that are not sufficiently robust will be of little use in a regulatory context. The results would not be credible to the sector and the basis of the regulatory settlement would be weakened, opening up the prospect of the decision needing to be adjusted or being overturned on appeal.

There are a number of dimensions in which robustness should ideally be demonstrated. The technique should ideally be robust to "noisy" data with results that are not inappropriately volatile, nor driven in apparently inappropriate ways by variation in data. Results should be also verifiable, in the sense that the ranking of operator performance should be consistent with other forms of assessment and other performance metrics.

Given the inevitable limitations of benchmarking, the ideal of a model that perfectly captures and balances all relevant factors is unattainable in any practical context. In this regard it should be understood that robustness is a relative concept. We might classify a number of points along a spectrum of robustness, where results might be regarded as:

- **definitive**: the results of applying the technique can be demonstrated to be highly robust along all relevant dimensions and can therefore be regarded as providing evidence on which allowances could be set with a high degree of confidence;
- **informative**: applying the technique produces results that capture most aspects of performance, but imperfectly. For example, proxy variables might be used to capture certain exogenous environmental differences between firms, implying that care should be taken when drawing conclusions on relative efficiency. The results are likely to be useful as part of a wider body of evidence with which to challenge operator forecasts and arrive at final cost allowances; and
- **unreliable**: in extreme circumstances there might be insufficient data with which to capture the salient features of the production process with confidence. While benchmarking results might provide a very broad indication of relative performance, important drivers of performance might be weakly captured making inferences difficult to draw from these results alone.

Of course, there are many intermediate points along this spectrum and the descriptions above should be regarded as illustrations of how outcomes might vary. We note that even comparatively unreliable benchmarking results might still be of use to the regulator. Suppose an operator is shown as being highly inefficient on the basis of an unreliable technique, yet after taking account of all of the factors missing from the model it was still not possible to close the gap between some operator's performance and the level predicted by the model. Even unreliable results, appropriately interpreted and supported by additional analysis in this way, might be regarded as useful evidence of inefficiency and a helpful ingredient to setting cost allowances.

3.3 Transparency

Transparent benchmarking models and processes should be preferred over those that are less transparent. If benchmarking methodologies are clear it will aid the ability of all stakeholders to understand the rationale for the selected approach, e.g. the why certain data has been used in the model and why other data has not. Similarly, it will be clear to the operators what conduct is being encouraged, how they have been rewarded for cost reductions, for enhanced quality etc, providing stronger signals on what the regulator (acting on behalf of customers) values.

Although it is not the only dimension of transparency, simplicity is an important element. More complex techniques are likely to be more difficult for stakeholders to replicate, which might limit understanding and hence the extent to which operators and others are willing to engage in debate on performance. Stakeholders will be better able to replicate a simple benchmarking method, further increasing their ability to understand the key drivers of their proposed cost allowances. For example, while Ofgem published extensive details of the approach it adopted to benchmark operators at DPCR5, the process was highly detailed, based on very many different regression models. It is likely that few of the interested stakeholders will have been able to replicate the approach and many will therefore lack an intuitive understanding of the final results.

3.4 Promotion of efficiency

In its Emerging Thinking from the RPI-X @20 review Ofgem has indicated a desire to move towards output based regulation. Under this kind of approach, performance will be monitored across a number of dimensions. This suggests that benchmarking techniques should, ideally, promote not just efficient cost management, but also striking the appropriate balance between low costs and desired outputs.

Consistent with this, building on the discussion in Section 2, benchmarking methodologies should ideally minimise the extent to which they distort incentives to favour one cost type over another. Ideally all competing costs should be exposed to benchmarking of a similar "strength".

Finally, it is likely to become increasingly important that benchmarking does not unduly encourage operators to avoid early action, innovation and investment that might be required to foster a transition to a low carbon economy. Benchmarking has traditionally involved taking "snapshots" of performance at points in time and it is possible that these techniques discourage operators from acting early, as there is a risk that those costs will be assessed as inefficient in comparison with other operators yet to act. Benchmarking processes might therefore be adopted that give rise to some institutional memory of past conduct, in order to ensure that appropriate and efficient early action is rewarded appropriately.

3.5 Consistency with other elements of regulation

Any benchmarking undertaken should support and enhance the set of incentives created by the nexus of other regulatory arrangements and should foster the high level objectives of the wider regulatory regime. For example, it would be inappropriate for the benchmarking to be designed in a way that delivery of a certain output could produce a disproportionate reward for operators. This might have the effect of encouraging operators to place a disproportionate focus on the delivery of that output. One element of this consistency should be encouraging operators to innovate while providing appropriate protection from unnecessary expenditure for customers.

3.6 Reasonable data requirements

It is possible to develop highly sophisticated approaches to benchmarking, which address directly many of the key concerns set out in Section 2. However, these techniques will only have merit if data exists with which to populate them. As we have discussed above an ideal benchmarking model might include numerous explanatory factors, outputs and variables capturing regional differences, together with squared and interaction variables, leading to a rich description of the activity of each business. Such a rich specification is likely to be impossible given limitations on the data that is available. In distribution the available cross section is relatively small and there are still few years of data (although this will increase over time). The availability (or unavailability) of data will inevitably limit the extent to which ideal models might be implemented and will rule out certain proposed models that would be impossible to make operational.

3.7 Adaptability

It is becoming clear that the activities we will ask networks to undertake in the future might be different from those undertaken now. For example, it is anticipated that there will be the need for distribution networks to serve an increasing fleet of electric vehicles in future. It is unclear how this additional network activity might be best encouraged and delivered. Similarly, it is likely that the focus of certain outputs might change over time. The definition of some output measures can change, making some outputs more or less measureable as a consequence, and inevitably leading to breaks and/or gaps in the available data.

Over time it is reasonable to suppose that the volume of data available will increase, as more years of cost data are collected on a broadly comparable basis.

The availability of more data might allow for more ambitious benchmarking models in the future, capturing a richer set of outputs.

Given the likelihood of significant changes in the available data, there may be merit in pursuing a benchmarking technique that could evolve over time.

3.8 Proportionate resource cost

Finally, it is important to consider the resource cost of implementing a benchmarking methodology. All relevant resource costs should be considered, including the cost of time spent by Ofgem, the operators and advisors in gathering and processing the data. If a technique requires only modest resource input and yet is found to be fit for purpose, this is clearly to be preferred to other techniques that might require a larger resource commitment. Similarly, it might be prudent to not pursue modest or uncertain benefits that could arise only following a significant resource investment. The counter-argument to this line is that, in comparison with the aggregate cost allowances of the sector, the resource cost of benchmarking is likely to be small. Therefore even small improvements in the accuracy of results might be worth paying for.

4 Approaches to benchmarking

A wealth of benchmarking techniques exists with which to compare the performance of operators. This section is not intended to provide a definitive overview of all of the techniques that might be adopted, nor is it intended to provide an academic review of their properties. Instead, we provide a framework for categorising the techniques to make identifying and describing candidate methodologies more tractable.

4.1 Overview

In order to aid description of the range of techniques, we have identified four key dimensions in which there are material choices over how to benchmark operators. This framework is illustrated in **Figure 2**.

Figure 2. Benchmarking options

Costs

- Top down, bottom up?
  - Total cost?
  - Competing cost?
  - Opex and capex?
  - By activity/ process?
- If total cost, then which measure?
  - Totex?
  - Opex + capital consumption?
- Group or licence (distribution only)?
- Regional/company specific adjustments?

Cost drivers

- High level or detailed?
- Should "outputs" be included? If so how?
  - Direct inclusion of full set?
  - Single composite of all outputs?
  - Excluded?
- Exogenous/ environmental variables?
  - Complex measure?
  - Proxy?
  - Excluded?

Sample

- International sample or national?
- Cross section or panel?
- Historic data only?
- Future plans?
  - E.g. total cost outcome if company plans are achieved?

Technique

- OLS (COLS/MOLS)
- SFA
- DEA
  - Stochastic DEA
- TFP (and other productivity based approaches)
  - Further choice in specification, e.g. returns to scale etc.
- Reference/model network techniques

The following sections provide details of the options in each dimension.

4.2 Which costs?

There are many different ways in which costs could be structured, aggregated and treated for inclusion in a benchmarking study. We identify the key dimensions of choice below.

4.2.1 Top-down or bottom-up

Top-down benchmarking is based on seeking a single, high level measure of all financial inputs used in a business. Total resource use can then be compared to the basket of explanatory factors and outputs delivered, to derive an overall assessment of the relative value for money delivered by each operator.

Top-down benchmarking of this kind places the focus firmly on this aggregate assessment of value for money for the customer. It is "blind" to the more detailed input choices made by the operator that ultimately lead to the recorded total resource use. In principle, top-down benchmarking removes any residual incentive distortions of the kind discussed in Section 2.2.1. For example, it is irrelevant whether operators choose to replace or maintain assets, to contract out or keep work in-house. Such benchmarking simply seeks the business that incurs the lowest level of total cost. Assessment of this kind is consistent with a light-touch approach to regulation, where business decisions are left entirely in the hands of the operators. However, as we discuss below, total cost benchmarking gives rise to a range of challenging issues regarding the quantification and treatment of historic capital costs.

The alternative to top-down benchmarking is to undertake a bottom-up analysis. Here different types of cost might be subjected to different types of benchmarking analysis. Costs might be split according to different types of activity (e.g. wages and salaries, IT, repair costs, reinforcement costs etc), with each cost type entering a different model and being compared to different cost drivers. The benefit of such an approach is that has the potential to yield more information to the regulator on *why* different operators might be efficient or otherwise. Whereas top-down benchmarking might reveal aggregate efficiency, bottom-up allows the regulator to offer an explanation for why this judgement has been reached.

However, bottom-up benchmarking creates a number of implementation issues:

- First, the greater the degree of disaggregation, the greater the burden created in monitoring the quality of the reported cost data. With more boundaries between costs there is an increased prospect of differences in accounting treatments leading to differences in apparent performance that are actually differences in reporting methodologies[12].

- Second, as discussed in Section 2.2.1, introducing boundaries between costs and benchmarking those costs in different ways can create material distortions of managerial incentives.

- Third, much care is needed when applying the results of several bottom-up analyses to determine allowances in order to avoid "cherry picking" efficient performance in each activity, thereby creating an aggregate efficiency standard that cannot be met by any real operator.

Between the polar extremes of top-down and bottom-up benchmarking there exists a middle ground. It might be possible to identify costs that compete with one another and ensure that all such groups of costs are included within a single benchmarking model. For example, all direct and indirect network expenditures could be benchmarked together, with business support costs benchmarked separately. Such an approach would reduce many of the difficulties that go alongside more detailed disaggregated benchmarking, by removing the sorts of artificial boundaries between costs that create material scope for distortion.

4.2.2 Treatment of capital costs

Capital costs give rise to a number of benchmarking challenges. Capital costs tend to be lumpy in nature and inherently less likely to be stable from year to year. They also give rise to long lived assets that require maintenance over time but provide services over the course of their lifetime. How should such costs be treated in a top-down benchmark?

One approach is to adopt a project finance model of the operator with a measure of "capital consumption" entering into a total cost benchmark. One would benchmark opex plus depreciation plus return, which has the effect of creating a smoother measure of capital, reducing the consequences of short run fluctuations. This gives rise to further questions, e.g. how should the depreciation term be derived (e.g. accounting, regulatory or economic) and how should the return element be calculated? Such questions are often more material in principle than in practice, since operators often have similar stocks of assets (i.e. asset bases with broadly similar age profiles), but this is ultimately an empirical matter and different approaches to valuing and incorporating capital could give rise to material differences in measured performance.

---

12 Ofgem has first hand experience of these difficulties. While the provision of clear reporting guidelines and close monitoring of returns can help alleviate some of these concerns, the process is time consuming and costly and some differences in reporting approaches are likely to remain.

An alternative is to benchmark total expenditure (totex), comprised of the flow of opex and capex. However, given the lumpiness of capex and potential fluctuations from year-to-year, the results of this approach might be volatile. Given this volatility, a third approach that could be adopted is to make use of aggregate (or equivalently average) expenditure over a period.

Under any benchmarking model it is also necessary for the regulator to consider how it intends to apply the results of the model. In particular, if a firm is found to be inefficient on a project finance style benchmark of total cost, is it then appropriate to disallow some capital consumption? In effect this would be equivalent to writing off some portion of past capex and would represent a significant departure from Ofgem's past conduct and from long standing regulatory commitments.

4.2.3 By licensee or by ownership group:

There are 14 separate electricity distribution licence areas and 8 gas distribution licence areas. However, a number of these licences are under common ownership. It is therefore necessary to consider whether to benchmark at the level of the licence or the ownership group. There is an obvious benefit in conducting analysis at the licence level, as this considerably increases the available sample size. Moreover, it seems reasonable to suppose that such analysis will be robust, given the licence conditions that exist to ensure that each licensee must not cross subsidise. However, given common ownership there is the risk that, at least in principle, operators could choose to allocate costs between their licence areas strategically. Conducting some analysis at the group level might therefore provide a useful cross check.

4.2.4 Prior adjustments for regional differences in input prices

Costs could be adjusted for regional factors prior to benchmarking. For example, adjustments might reflect differences in underlying labour costs between regions. The alternative to prior adjustment (in a regression based assessment at least) is to include within the model an additional variable that captures differences in regional input prices. This would allow empirical testing of the extent to which differences in input prices might explain differences in costs.

In our view regression with input price variables is the superior approach, as analysis with prior adjustment may lead to specification error and omitted variable bias. However, we note that in cases where there is little data, such an approach might not provide conclusive results.

4.3 Which cost drivers?

As we discussed in Section 2.2, benchmarking models need to take account of the key cost drivers of the business. Ideally a comprehensive set of cost drivers should be included to best capture the scale of the network task that each operator undertakes. In addition to measuring the task at hand, it might also be necessary to include additional variables that capture valued outputs, such as quality of supply. Taking account of Ofgem's desire to move to an output based regulatory regime, it might also be desirable to include output measures directly within a benchmarking model. This would provide operators with an increased incentive to deliver those outputs and would, potentially, allow different operators to choose to deliver different levels of output service at different levels of cost.

In the case of energy networks the key cost drivers are often also the key services provided to customers. Relevant cost drivers include the number of connection points (e.g. in distribution the number of customers connected), load served (e.g. coincident or non-coincident peak load) and the volume of energy transported. Customers will also be impacted by the quality of service provided by their network operators, particularly the reliability and continuity of service. Since the provision of higher quality is not costless, it is appropriate to include quality variables within a benchmark. Similar arguments could also be made for the inclusion of losses within a formal benchmark, although we understand that the present approach to the measurement of losses using settlements data could make the available data unreliable.

A further important cost driver at the distribution level is the connection density of customers, which will help to determine the volume of network assets needed to provide network services to customers. All other things being equal, the less densely populated a region the greater the volume of network assets required to serve a given number of customers. Potentially, then, serving a densely populated area gives rise to a lower cost base than serving sparsely populated area. However, any potential benefit might be offset (in part or in full) by other factors, such as the need to bury all network facilities in urban areas and the potential cost of congestion in highly urban areas.

This suggests that it might be helpful to include within the model a large number of variables, including for example:

- the number of connections;
- peak load;

- volumes distributed/transmitted[13];
- quality variables;
- losses;
- a proxy for network density at the distribution level, e.g. network length[14]; and
- other output variables.

As noted earlier, it is also desirable to include input price series as explanatory factors in a benchmarking model, to reduce the risk of variable bias and specification error[15]. Additionally, it is often argued that it is appropriate to split a number of these cost drivers into different voltage levels (e.g. high, medium and low voltage), to capture better the underlying load mix served by different operators and account accurately for differences in cost that arise at these different levels.

At this stage we note that such a richly specified model is unlikely to be implementable given the size of sample that is likely to be available (even in the case of electricity distribution where a panel already exists and there are 14 DNOs). One solution to this challenge, as has been used by Ofgem in the past, is to aggregate several cost drivers into a single composite driver. In effect, this is equivalent to imposing a restriction on the estimation of the benchmarking model. Weights for a composite might be informed by engineering analysis or an initial regression analysis, although it is likely that a composite driver might only be used when data is limited, implying that a multistage stage regression analysis of this kind might not be statistically robust.

4.3.1 Cost drivers used at DPCR5

During the course of DPCR5 Ofgem made use of a wide range of cost drivers in its opex benchmarking. These variables are summarised in **Table 4** below.

---

13 It might be argued that the volume distributed/transmitted is not a significant driver of cost since the flow of energy through the network does not give rise to a significant direct cost. Most network infrastructure is scaled to serve expected peak demands.

14 The use of actual network length as a cost driver in a regulatory context can give rise to perverse incentives. Operators are able to control their network length and could install more network assets. In principle therefore, operators might see the installation of unnecessary network length as a way in which they could appear more efficient. A proxy variable that that captures these spatial characteristics but is beyond the control of the operator, for example a variable derived from a Model Network Analysis, has better regulatory properties since it eliminates this incentive.

15 The inclusion of input price variables also allows the researcher to separately identify allocative and technical inefficiency

Table 4. Cost drivers used at DPCR5

	High level	Low level
Scale variables	Number of customers Network length Units distributed Number of streetlights Number of services Modern Equivalent Asset Value (MEAV) Assets causing man hours of work	Assets requiring man hours of work for inspection and maintenance Length of cable replaced
"Work" variables	Network spend Volume/unit cost Direct costs (less non operational capex) Network spend plus non operational capex	Total overhead faults Total underground faults Total spans of trees cut Spans of trees affected

Source: Ofgem

Scale variables

A number of the variables we identified as suitable cost drivers were also used by Ofgem and were classified as high level scale variables (i.e. number of customers, network length and units distributed). Ofgem used these variables primarily in benchmarking opex at a more aggregated level. We would have concerns about the repeated use of the other variables in this classification for regulatory purposes, in particular the use of MEAV and assets causing man hours of work. Both of these variables might be subject to the same critique as was identified above for network length. Similarly, neither of these variables seems to provide operators with an incentive to expand their network optimally, since higher volumes of installed assets will result in better measured efficiency (all other things being equal).

Ofgem also made use of a number of low level variables in its analysis. Such variables have merit if a disaggregated approach is to be adopted. As set out in Section 2.2.1, a disaggregated approach might have the effect of creating arbitrary boundaries between expenditures, increasing the scope for behaviour to be distorted. In the following section we expand further on this and set out our preference for more aggregated benchmarking in order to minimise these distortions.

Work variables

We have a number of concerns with the repeated use of what Ofgem has classified as work variables at DPCR5 in regulatory benchmarking. There is an obvious circularity in using a subset of costs to explain other costs. The potential incentive effects are clear, i.e. there is a marginal incentive not to reduce one set of costs in order to help explain/justify another. This could give rise to significant distortions.

The low level variables used by Ofgem, while clearly drivers of cost, will have the effect of creating a check on the unit cost of the relevant activities (e.g. benchmarking the cost of fixing overhead faults against the number of overhead faults is equivalent to measuring the unit cost of an overhead fault). While this might reveal differences in the underlying efficiency with which faults are repaired, the mix of fault types is also likely to be a significant driver of the unit cost of fault repair (i.e. not all faults will give rise to the same volume of work). More generally, a process that involves assessing unit costs, absent any additional incentive arrangements, provides little incentive for operators to consider carefully the volumes of work they undertake, only that the work be undertaken at a low unit cost (see Section 4.5.1 for further discussion of the weaknesses of unit cost based approaches to benchmarking).

4.4 Which sample?

The greater the volume of data that is available the more variables it is possible to include within a model and the more robust the estimation of that model is likely to be. As a consequence, it is likely to be unambiguously desirable to include as much data as possible. In the case of the distribution networks, it is clearly better to include all operators in the survey and to make use of data for as many years as is available (i.e. to run analysis with a panel).

The key decision to be taken here is whether to augment that sample with international data, given the difficulties that international benchmarking gives rise to as described in Section 2.2.3. For the distribution networks, where there is a reasonable body of data, the work involved in extending the sample to include international data is unlikely to be merited. For transmission however, substantive benchmarking of any kind is unlikely to be possible without making use of international data.

It is also helpful to consider whether benchmarking should be restricted to historic outturn costs alone, or whether benchmarking of future plans could prove informative. For example, it would be possible to undertake total cost benchmarking of the operators' Forward Business Plan Questionnaire (FBPQ) returns in order to identify whether the implied future charges of a given operator are broadly reasonable or appear out of line with the rest of the industry.

4.5 Which technique?

There are many different techniques that could be adopted to bring together the cost and cost driver data in order to estimate relative efficiency. We provide a very brief overview of those techniques here, which is not intended to be a full exposition of the relative properties of each approach. A large body of academic literature exists on each technique, including numerous practical applications of these techniques to a range of data sets.

- **Ordinary Least Squares (OLS)**: this technique makes use of regression based techniques in order to identify the relationship between the left hand side (LHS) dependent variable (e.g. total cost) and the right hand side (RHS) independent variables (i.e. cost drivers). A considerable body of academic work sets out how this framework can be used to support hypothesis testing, including testing whether a given cost driver is indeed a significant driver of cost in a statistical sense. Loosely speaking, the standard OLS model will return a predicted cost consistent with the average level of performance in the sample. Given this it is traditional to "shift" the frontier to reflect a better than average performance. For example the frontier might be shifted to match the level of the best performing business in the sample, i.e. Corrected OLS (COLS).

- **Stochastic Frontier Analysis (SFA)**: SFA is also a regression technique that shares many of the properties of OLS. However it adopts a more sophisticated approach to determine the location of the efficient frontier. SFA is a technique that allows the researcher to seek to identify whether observed differences in performance should be regarded as systematic evidence of inefficiency, or whether they arise as a result of "noise" in the data. In order to make use of SFA it is necessary to have more data, since the technique consumes additional degrees of freedom.

- **Data Envelopment Analysis (DEA)**: DEA is an extension of simple ratio analysis to cases where there are potentially many inputs and many outputs (i.e. explanatory factors). DEA techniques can be used to establish which firms produce the most of each output(s) for the least input(s) and by how much inefficient firms need to decrease their input (or increase their outputs) in order to also be regarded as efficient.

- **Total Factor Productivity (TFP)**: TFP is an approach to measuring productivity where all factors of production are considered. A range of approaches could be adopted to identify TFP, including methods based on index numbers (as in the Dutch regime), DEA and OLS/SFA. All of these techniques focus on measuring the rate of change of explanatory factors and outputs relative to the rate of change of all input factors[16]. A TFP approach might not be regarded as a benchmarking methodology at all. TFP estimates typically focus on the rate of progress of an entire industry, although it is possible to identify firm specific TFP changes when using DEA or SFA based techniques. Where an industry wide rate of progress is estimated, it could then be imposed on all operators, implying that firms able to improve performance at a faster rate than the average will earn excess returns at the expense of operators performing worse than average, providing strong incentives for such improvements. Clearly such an approach is only reasonable if there is a view that all operators have similar opportunities to achieve performance improvements, otherwise unmerited windfall profits/losses will be earned. As noted above, a TFP methodology based on the derivation of an index measuring all explanatory factors and outputs and an index measuring all inputs underpins the regulatory regime imposed on the Dutch distribution operators. If a traditional index based methodology is to be employed then it is necessary to determine the price of all explanatory factors and outputs, which is not always straightforward.

- **Reference/Model Network Analysis (RNA/MNA)**: These techniques are based on constructing an engineering model of each regulated operator. RNA aims to construct a detailed model making use of a considerable body of data on the layout of the network (i.e. creating a detailed spatial model). A MNA approach makes a number of simplifying assumptions in order to reduce the complexity of the model and the volume of data that is required to undertake the work. Both approaches produce an optimal network specification, setting out the network that could be constructed in order to undertake the network activities of the operator in question. These stylised models can then be used in a number of ways. For example, the modelled network can be "costed" in order to derive an estimate of the efficient cost of constructing/operating a network to serve the region in question. Actual cost can be compared directly to this estimate of efficient costs, or operators can be assessed on how close they are able to get to this benchmark relative to other operators in the sample. Alternatively, this approach can be adopted to derive structural variables such as modelled network length and modelled transformer capacity. These variables could be used as additional explanatory factors in one of the other techniques described above (e.g. OLS, SFA or DEA). Modelled variables of this kind are arguably better

16 It is also possible to derive partial measures of productivity, of which the most commonly encountered example are labour productivity indices.

proxies for the underlying operating environment of different regions than, for example, actual network length as they are beyond the control of the operators and therefore give rise to no perverse incentives.

4.5.1 Other approaches to benchmarking

In addition to the techniques described above, a range of other approaches to benchmarking might also be adopted. Many of these might be regarded as process benchmarking methods, including subjecting plans to detailed scrutiny by suitably qualified experts.

For example, benchmarking can be conducted by examining individual unit costs. A number of specific "units of work" can be identified, allowing outturn cost of undertaking one unit of such work to be assessed. Where future/planned unit costs vary from current or observed unit costs, such analysis can provide a useful challenge.

Practical difficulties can arise where tasks are too complex or heterogeneous to be readily mapped to a set of unit activities. Where judgement is required to make such mappings, differences in data reporting or treatment can give rise to apparently material differences in unit costs. Analysis conducted on this basis can also be criticised for "cherry picking" the best level of performance observed on each unit of work, potentially creating an aggregate target operator that is not achievable in practice. Finally, unit cost studies provide no insights on the volumes of work that should be anticipated. For example, a firm might be identified as an excellent performer on a unit cost analysis, but this provides no insight on their skill in identifying what work should be undertaken.

More generally, expert scrutiny of plans can provide important insights into the merits of what has been proposed, identifying obvious weaknesses and/or inconsistencies. However, the appointed experts will never have the same set of information available to them as the planning team of the regulated network. Inevitably, therefore, there will be limits on the extent to which they will be able to assess plans effectively.

While techniques of this kind can play a potentially useful role in a holistic review of efficiency, it is unlikely that any can be relied upon alone.

5 Candidate methodologies

In this section we review a short list of candidates that have been assessed in some detail, based on the fuller range of options set out in Section 4. We provide a more detailed specification of each approach together with an assessment of their properties. In Section 6 we set out our recommendations for each sector based on the discussion in this section, but homing in on our preferred option in light of the context for each network.

5.1 Option 1: total cost benchmarking of future plans

Currently, Ofgem uses historic cost data to assess the relative performance of network operators on operating expenditure. However, as discussed in Section 2, the forthcoming period is likely to require more proactivity and innovation from the networks than has been the case in the past, particularly in electricity distribution. There is also the possibility of a trend-break in the types of activities that network operators will be undertaking over the next decade, and a consequent change in underlying cost structures. In this context, and particularly with regard to the need to encourage innovation, benchmarking of historic costs may no longer be a relevant tool with which to inform future revenue allowances. As discussed in Section 2.4, undertaking strong ex post benchmarking of historic costs to inform future allowances might reduce incentives for the network operators to invest in a timely manner and pursue potentially risky innovation.

We therefore propose, as a first candidate benchmarking methodology, a technique in which we benchmark future business plans. Below we set out this approach in further detail and discuss how it might work in practice.

As we discuss in Section 6, we regard it as highly unlikely that it will be possible to benchmark forward looking transmission plans as data on a sufficient sample of international peers will not be available on a timely or comparable basis. Consequently, the discussion of this option focuses on distribution.

5.1.1 Cost definition

We envisage that benchmarking could be carried out using two separate sets of forward-looking cost data:

- planned total expenditure (operating expenditure + capital expenditure)[17]; and

17 If capital expenditures were particularly volatile, an potential helpful approach might be to benchmark aggregate expenditure over the forthcoming price control period in order to minimise the consequences of this variation for the robustness of the analysis. However, this would have the

- planned future total cost (opex + depreciation + return based on planned capex).

The relative merits of these cost definitions were discussed in Section 2.2.2. From this discussion, it is clear that there are both merits and drawbacks associated with using either measure. In particular analysis that makes use of prevailing regulatory accounting arrangements and existing RAVs might require careful interpretation. We would argue that benchmarking operator plans using both definitions of cost will improve the richness of information provided to Ofgem. This would provide a wider body of evidence with which to challenge the operators and increase Ofgem's leverage in negotiations with the network operators.

With regard to the second of these measures (planned future total cost), we would propose that when calculating its estimate of capital consumption Ofgem makes use of the prevailing regulatory accounting rules (i.e. it uses the prevailing depreciation methodology and proposed cost of capital). We believe that this approach is pragmatic and helps to reduce the administrative burden of undertaking the analysis. In principle, using alternative accounting methodologies could yield materially different results. However, in practice results are often reasonably robust to modest changes in methodology. Operators in the same country and under the same regulatory regime are subject to similar demands and pressures. Profiles of capital expenditure over time are often reasonably correlated and it is only material differences in expenditure profiles that might make the results sensitive to the accounting methodology chosen.

5.1.2 Cost drivers

In Section 4.3 we identified a set of cost drivers that can be used as explanatory variables in the benchmarking analysis. These include:

- the number of connections (split between load and generation);
- peak load;
- volumes distributed/transmitted;
- quality variables;
- losses;
- a proxy for network density at the distribution level, e.g. network length; and

---

effect of reducing the proposed panel analysis to a cross section, reducing the sample size and potentially limiting the extent to which all relevant explanatory factors and outputs can be captured.

- other output variables.

In principle, the benchmarking model should attempt to incorporate a complete set of cost drivers. However, the set identified above represent a relatively large number of RHS variables. It is likely that the inclusion of all of these cost drivers in a general form specification (that includes levels, squares and interaction terms) will be impractical, given the limited sample size. To address this, it might be possible to adopt alternative ways to include certain explanatory variables. For example, the cost of anticipated Customer Interruptions (CIs), Customer Minutes Lost (CMLs) or Energy Not Served (ENS) could be monetised and included as a direct network cost, as could the estimated cost of losses. This would provide appropriate incentives for networks to try and minimise these costs, indirectly incentivising the desired outcome. An approach of this kind would preserve degrees of freedom while maintaining the incentives to which Ofgem wishes to expose the operators.

Given the forward looking nature of the proposed benchmarking, projections of these cost drivers will be required. It is likely that these projections would need to be obtained from the operators own business plans, which gives rise to a perverse incentive for the operators in submitting those forecasts. Ofgem would therefore need to test the validity of operator forecasts rigorously. Operators would need to explain the basis for their forecasts, given prevailing levels and historic trends, the assumptions made by others in the sector, the context of the details of their business plans and the outcome of engagement with the relevant stakeholders.

It will be necessary to take account of the final set of outputs. In parallel to this study Frontier has also undertaken an assignment to identify how Ofgem's vision of a more outputs focused regulatory regime might be put into practice. That report, published alongside this one, has identified a core set of outputs that might feature directly in a benchmark.

- **Reliability**: we propose that Ofgem makes use of CI and CML/ENS[18] (ENS at the transmission level) as key outputs to provide incentives for the delivery of a suitably reliable network. To preserve degrees of freedom we would propose that the expected level of these quality variables be monetised at "prices" to be determined by Ofgem and the cost of customer interruptions added to planned expenditures. Our report on outputs also proposes that the networks prepare an annual "reliability report" to monitor the operators' efforts to anticipate future changes in demands on their

18 The discussion here is focused primarily on measures relevant to the electricity industry. Given the binding nature of safety concerns, gas networks are typically designed such that interruptions to supply are very rare indeed. In the light of this, we anticipate that the gas price review team might choose not to include reliability measures directly in their benchmarking analysis.

networks. However, it would not appear appropriate to take account of this directly in the proposed benchmarking methodology.

- **Connections**: we are proposing a regime under which Ofgem will ask operators to ensure that new connections are made within a prescribed timetable, with financial penalties imposed for failing to meet this standard (or alternatively rewards for meeting/beating that standard). The planned volume of connections to be made could be included as an additional output variable, providing greater consistency between proposed budgets and agreed output volumes.

- **Environment**: as we explain in the our outputs report, we are proposing that the wider role of the electricity networks in bringing forward new generation will be incentivised through the reliability and connections incentives. As a consequence it is appropriate to include only a narrow measure of environmental performance in the benchmarking model. For example, the business carbon footprint would be a key part of this narrow measure of environmental performance. We would propose that this narrow measure be monetised and added to planned cost in the same manner as CI/CMLs, e.g. making use of prevailing carbon prices to cost the planned carbon footprint.

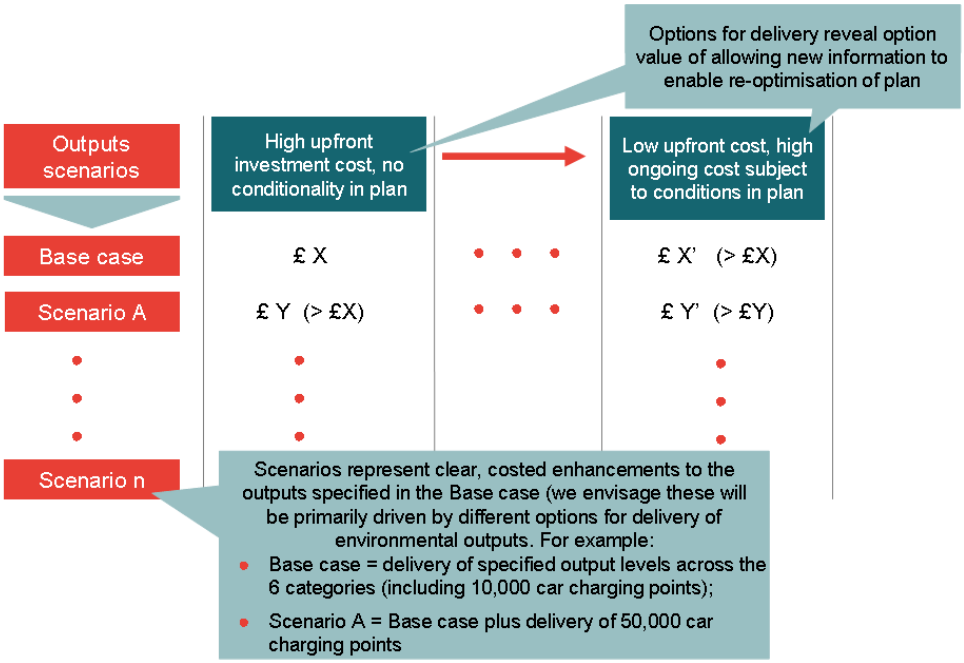
With regard to a proxy variable to capture exogenous differences in operating environment, there are two obvious choices available. Ofgem could make use of actual network length as a proxy variable, although this has the scope to create a well understood perverse incentive to install more pipes, wires and cables. Alternatively Ofgem could undertake a MNA (as discussed below) to derive a proxy variable that is beyond the control of the operators. We would recommend that this second option is adopted, although we recognise that this would have important resource implications. The decision on how to proceed in this area is likely to be a subject that the specific price control teams will wish to revisit on a sector by sector basis.

5.1.3 Sample

Currently, network operators submit a Forecast Business Plan Questionnaire (FBPQ) as part of the existing price control process which contains similar information to that which would be required to carry out benchmarking. The sample gained from FBPQ submissions will present a panel of data over the length of time specified in the questionnaire.

At this stage we believe it is useful for us to present an outline of the information that we anticipate might be included in future business plans. This is summarised in **Figure 3**, which illustrates what might be the structure of operator business plans in future.

Figure 3. Illustration of the possible structure of business plans submitted at future regulatory reviews



As shown in **Figure 3**, we expect that future business plans will contain several scenarios. Each scenario might be premised on the delivery of different volumes of the six outputs (in particular the three key output categories identified above) and would indicate what underlying activities would be needed to support that delivery.

We also anticipate the operators will bring forward a range of scenarios in their business plans that might vary in terms of the timing with which expenditures are made. The operator might submit a scenario in which expenditures to support the delivery of a certain output are front loaded, and a second in which expenditures are phased over time, with short term needs potentially met by more costly short run remedies. In total, the phased delivery plan might be expected to be more costly than the plan that rolls out fresh capex sooner. However, the plan with extensive upfront expenditure has less option value. Should new information come to light that reveals that the specified output is no longer as valued as was previously envisaged, there is the option to "cancel" the plan with phased expenditure and save potentially substantial sums of money. In contrast, the plan in which costs are sunk early would have resulted in now unwanted assets and stranded investments, which customers would need to fund.

The conclusions of the above are that we envisage a regulatory regime in future in which plans contain a richer set of information with more scenarios and greater optionality. As a consequence, it is reasonable to expect Ofgem to have the ability to benchmark a set of plans from each operator, providing the scope for a richer understanding of the potential incremental costs associated with different network expansion opportunities.

There is, however, a related concern regarding the submission of several plans by each operator. In a regulatory process of this kind, operators could seek to manipulate the analysis through the submission of some scenarios in which costs are particularly inflated, alongside others that are only slightly inflated. Such a strategy might be used to convince Ofgem that the more reasonable (but still inflated) plans should be accepted in full, as direct comparison with highly inflated plans reveals that they are apparently good value for money. This strategy could be effective if Ofgem simply benchmarks all plans at the outset, as a high cost plan will have the effect of "dragging up" the location of the regression line. Conduct of this kind is essentially a further example of the well known incentive to inflate discussed in Section 2.4.2, which could be addressed through targeted use of historic cost benchmarking.

In the Emerging Thinking document, Ofgem outlined proposals for partially lengthening the price control period. Ofgem may therefore be interested in collecting forecast data for different outputs over different timeframes. However, for the purposes of benchmarking, only cost and output data over the same timeframe is relevant. Given that forecast data becomes less reliable the further into the future the projection, the data request will need to be limited to ensure forecasts can be made accurately. Requesting data over a five year period would provide a sufficient sample to carry out a robust panel analysis. We would therefore propose that collecting forecast data for a five year period would be appropriate for the purposes of benchmarking.

While data from other countries might create a larger sample and allow more ambitious econometrics, we regard it as impractical to seek similar data from operators in other countries. It is highly unlikely that the relevant information will be produced to the required timetable and on a sufficiently comparable basis. In any event, for electricity and gas distribution, the sample is likely to be sufficiently large (14 DNOs, 8 GDNs for five years or more) to allow robust statistical estimation.

5.1.4 Technique

Given Ofgem's past preference for regression techniques, and hence the familiarity of the GB operators with them, we propose that Ofgem continues to adopt an OLS based approach to benchmarking. While SFA is typically the preferred technique, the absence of statistical noise in data based on future operator plans suggests that there is no requirement for an SFA approach to be implemented. Making use of OLS/COLS will consume fewer degrees of freedom, providing scope for more ambitious statistical analysis.

We would propose that Ofgem begins with the usual translog specification and tests down to a reduced model. The detail of how to proceed with specification testing is, again, an empirical matter for the price control teams.

We are proposing that Ofgem conducts analysis using two cost measures and potentially several sets of alternative operator plans. This will produce a body of evidence on which Ofgem will need to engage with the operators to challenge the operators' plans. This approach to benchmarking will not yield an unambiguous point estimate of efficiency therefore, but a range of evidence that will require assessment and discussion with the operators.

5.1.5 Discussion

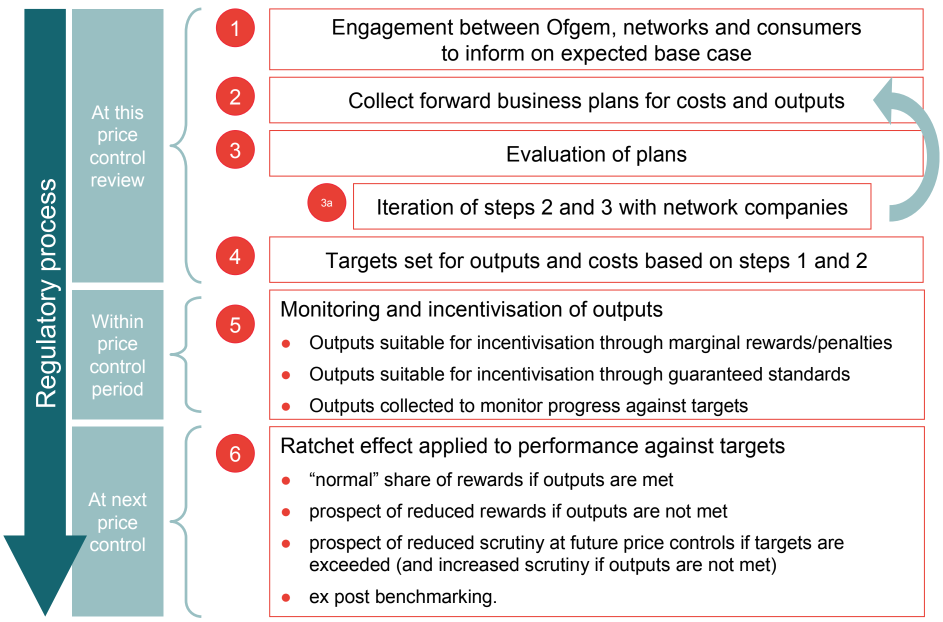
In making future assessments, Ofgem will need to trade off two competing objectives. On the one hand, Ofgem will wish to ensure that operators facilitate the timely roll out of the technologies required to decarbonise the economy, including investigating all options to do so and innovating where necessary. At the same time, Ofgem will seek to ensure that networks are efficient in their spending and provide customers with value for money. We believe that benchmarking future plans strikes an appropriate trade off between these objectives. As we discussed in Section 2.4, using ex post benchmarking of expenditures to set future allowances could have the effect of increasing regulatory risk and stifling operator innovation. Benchmarking future plans reduces this risk (in particular the risk that operators are provided with an incentive to simply do whatever was in the plan), but maintains an important element of competition between firms to seek value for customers. By establishing a regime based on benchmarking future plans, Ofgem would signal clearly to operators that in future they will be assessed on what they are charging customers compared to the basket of explanatory factors and outputs they provide.

In the context of decarbonisation and uncertainty over the next decade, placing emphasis on providing value for money is an important outcome of the benchmarking analysis. Benchmarking plans also eliminates concerns that arise in other candidate methodologies over the need to disallow past capital investment. The focus here is on future plans not ex post analysis of historic capex.

A further advantage of benchmarking future plans is the additional flexibility it provides. If it becomes apparent that certain outputs are no longer relevant over the course of the price control period, it is possible to instruct network operators to adjust their forecasts appropriately at the start of the next period. This opportunity is restricted with ex post benchmarking, where data needs to be collected over a sufficient period to allow robust assessment, and required changes to the data cannot be made at short notice[19]. As time progresses and stakeholders develop their understanding of new outputs and potential changes in network activities and cost structures, the data requirements for future plans will be able to adapt and develop accordingly.

In this regard, it is helpful to discuss how we envisage benchmarking fitting into the regulatory process, which is illustrated in **Figure 4**.

Figure 4. The role of business plans in the regulatory process



Following the submission of plans by the operators, Ofgem would benchmark these plans using the approach described above. This benchmark of plans, supported by targeted use of historic benchmarking, would lead to detailed discussions with the operators and the refinement of those plans. Ultimately, the operators and the regulators would agree a set of allowances, alongside a set of targets for the delivery of the relevant outputs. It is likely that the final agreement on allowed costs and the target level of outputs would be a conflation of a number of the scenarios originally submitted by the DNO, amended to

19 For example, if a particular cost driver were deemed no longer relevant in the third year of a price control, the data collected for that driver for the purposes of ex post benchmarking at the next price control would be redundant. However, with benchmarking forward plans, the data required for inclusion in that plan can be altered at relatively short notice. In other words, ex ante benchmarking potentially locks Ofgem and network operators into a pre-defined set of outputs.

reflect information revealed through benchmarking and through dialogue over the course of the regulatory review.

During the subsequent price control review, the operators would monitor and submit to Ofgem progress according to that plan (i.e. on expenditures and output delivery relative to plan). It will be particularly important for operators to bring forward to Ofgem updated information that might provide evidence to suggest that the assumptions that underpinned the final plan and allowances were proving to be inaccurate. This would provide the operator with the opportunity to explain any departure from plan at the next review and agree with Ofgem the appropriate regulatory treatment of departures from plan.

At the next review, Ofgem would inevitably wish to assess how the operator had performed relative to plan. Where particularly significant differences between planned and actual spend have materialised, there will be a strong temptation to put in place an ex post adjustment. On this point, we would urge Ofgem to be cautious. As we have discussed, if the operators believe that a significant departure from plan will be "punished", then there will be a strong temptation for them to simply stick to the plan whether this represents best value or otherwise. Ofgem must seek to reassure operators that, while inefficiency and waste will be punished, adjusting plans in the light of new information will be rewarded. We anticipate more extensive dialogue between Ofgem and the operators during review periods than has traditionally been the case.

In principle Ofgem could undertake analysis that makes use of a panel that is comprised of both future plans and historic expenditure/output delivery, in order to create a large sample with which to work. On this we would suggest that Ofgem adopts a cautious approach in the short to medium term. As discussed above, the inclusion of historic data directly within the sample could discourage investment and innovation. In any event, if the activities of the operators do change materially over the coming years then the direct inclusion of historic data might be less revealing. We would advise that Ofgem considers extending the sample to include historic data only once it is confident that the prevailing uncertainty over network activities has reduced, so that the incentive problems identified above are no longer material.

As we have discussed in Section 2.4.2, one of the risks of this approach is that it leaves unaddressed the scope for operators to respond to the well understood incentive to inflate their plans. This suggests that Option 1, if adopted will need to be buttressed with some targeted use of historic cost benchmarking, in order to increase the likelihood of those following such a strategy being identified.

As discussed above, we believe that Ofgem should conduct analysis using both totex and total cost measures. These two sets of information will provide Ofgem with greater insights on differences in expenditure that are more related to the timing with which capital programmes are implemented, rather than underlying differences in efficiency. Firms planning to undertake an investment programme (for example an electricity DNO putting in place reinforcement required to support electrification of space and water heating) in the forthcoming regulatory period might appear relatively less efficient on the totex benchmark, but the difference should be less pronounced on the total cost assessment (where capital expenditures are more smoothed). Similarly, an operator that begins to roll out a capex programme after its peers have completed that investment will perform worse on the totex assessment, but should perform relatively better on the total cost assessment. The envisaged analysis will provide Ofgem with the fullest possible range of information on which to understand the drivers of such differences between operators.

Table 5 below provides a summary assessment of Option 1 against the criteria identified in Section 3. Our summary view is that we see many merits associated with benchmarking forward looking plans. There are technical and regulatory challenges to be addressed in adopting this approach, but we believe that they can be overcome. As a consequence, Option 1 represents an important element of our proposals at the distribution level. As discussed in Section 6, however, we do not regard it as a practical approach for the transmission networks.

Table 5. Assessment of Option 1

Criteria	Option 1: Total cost benchmarking of future plans
Robustness	Likely to achieve a sufficient sample to carry out robust panel data analysis, at least for the distribution networks. For transmission networks, issues with sample size are compounded by additional difficulty associated with obtaining comparable forward-looking international data. Given uncertainty about future and potential difficulties with fully capturing exogenous differences between operators, results are likely to be informative rather than definitive. Operators may not be able to assess each other's plans effectively, either because of confidentiality issues, or because of fundamental difficulties with understanding peers' business models and planning processes. However, the ability to make such assessments is likely to be essential if operators are to accept the results of benchmarking.
Transparency	Use of OLS technique with two top-down cost categories is simple and replicable. However, since benchmarking future plans represents a significant departure from traditional regulatory practice, there is likely to be a period of learning. The benchmarking will not be used to directly determine allowances. It would form a body of evidence on which Ofgem would engage with the operators, supporting a process through which the individual operator plans are agreed.
Promote efficiency	Strong focus will be placed on providing value for money for delivered outputs. The proposed cost measure should ensure equal treatment of all costs, minimising distortions over input choices. However, risk of gaming of plans must be reduced to avoid efficient plan but inefficient action. This could be achieved through a stronger IQI and appropriate ex post assessment (although this must not promote 'spending to plan' at the expense of innovation).
Consistency with wider regulatory framework	Incorporation of appropriate outputs in cost drivers will ensure that wider regulatory objectives are achieved. However, there are likely to be a set of desired outputs that are not suitable for inclusion in benchmarking. Wider regulatory framework must ensure that these desired outputs are appropriately incentivised.
Reasonable data requirements	Strong emphasis will need to be placed on accuracy of business plans. This will require networks to undertake appropriate forecasting if data is to be reliable. There are likely to be significant difficulties with forecasting some of the core explanatory factors and outputs. However, data requests at a sufficiently high level should be relatively easy to provide (much of the information is already contained in the FBPQ and future plans will request greater volumes of data).
Adaptability	Flexible incorporation of outputs will ensure adaptability. Sample size will only increase if longer plans are requested (i.e. forecast data does not grow over time). However, as underlying cost structures and industry best practice for new activities is identified and understood, benchmarking will be flexible enough to adapt.
Proportionate resource cost	Fewer regressions and high-level data requests imply lower resource cost than, for example, DPCR5 benchmarking. Ensuring that forecast data contained within business plans is accurate will increase resource requirements. However, it is likely that extra focus would have been placed on this anyway, regardless of the benchmarking framework.

5.2 Option 2: benchmarking of historic competing costs (DPCR5 "lite")

Benchmarking future plans as described in Option 1 represents a significant departure from current regulatory practice. An alternative approach is to retain the benchmarking of historic costs, using performance in the previous price control period to inform revenue allowances for the next one. This is the approach described below.

Option 2 retains the principle of aggregating competing costs for the purposes of benchmarking, wherever this is sensible. In this sense, Option 2 represents a departure from some of the developments made at DPCR5, where certain elements of the benchmarking analysis were carried out at a highly disaggregated level. As discussed in Section 3.3, there is a danger that the level of detail and complexity of the benchmarking analysis at DPCR5 made it difficult for interested stakeholders to replicate the analysis and, arguably, reduced the transparency of the results. Further, the reintroduction of cost boundaries between competing costs in the benchmarking has the potential to dilute the wider principle of equalising incentives introduced at DPCR5. Aggregating competing costs for benchmarking improves the transparency of the analysis and avoids distorting incentives. Option 2 might therefore be described as a DPCR5 "lite" benchmarking analysis.

5.2.1 Cost definition

We begin by discussing the appropriate cost definition in the context of the distribution networks.

In principle, it would be preferable to benchmark total historic costs, using the two cost definitions already discussed above. However, in practice there may be difficulties in implementing total cost benchmarking immediately. In particular, it may prove difficult to include capital expenditure in total cost benchmarking in the short run. Given the inherent 'lumpiness' of capex and the uncertainty arising from decarbonisation, it is not clear that a benchmark of past capital expenditure will be highly informative in determining the appropriate level of future allowances at present. This gives rise to new and uncertain challenges for the networks and the likelihood of heterogeneous approaches to delivering new objectives across the networks, reducing further the merit in benchmarking historic capex.

Once the necessary network activities become more well-established and industry best-practice is identified, capex might return to a 'business-as-usual' trend. Historic cost benchmarking would then become more relevant for comparing efficiency across networks. The exact timeframe within which capex will return to a BAU trend, and whether capex is then sufficiently stable for historic benchmarks to be informative over future spend, must be determined empirically.

In defining the costs to be used for historic cost benchmarking, we have retained a high level, top-down approach. This is designed to minimise potential distortion of input incentives, as well as ensuring transparency.

We would propose that Ofgem focus on the following cost categories in the short run:

- Business Support Costs (BSCs)
- "Network-related" costs (i.e. Network Operating Costs (NOCs) + Closely Associated Indirect costs).

These two categories are delineated on the grounds that there is only weak scope for substitution between these cost types. As Ofgem concluded at DPCR5, placing a boundary between BSCs and network-related costs does not therefore distort input incentives.

In the longer run, as the nature of future network activities and cost structures becomes better understood we would envisage Ofgem being able to undertake total benchmarking of historic costs, adding a measure of historic capex (i.e. depreciation + return on capital) to be network related costs.

At the transmission level, benchmarking of the kind described here will only be possible if data is collected from other countries. Since the operators on which data might be collected will be subject to different regulatory regimes, it is unlikely that data on, for example, business support costs, will be available on an appropriately comparable basis. It will instead be necessary to work with whatever high level data is publically available. In light of this, a different approach is likely to be necessary for the transmission networks, focused on a high level assessment of total cost. We return to this in Section 6.

5.2.2 Cost drivers

At the distribution level, the cost drivers we believe would be relevant for benchmarking historic cost are as set out in Section 5.1.2 above. Similar considerations apply with regard to the ambitiousness of the benchmarking and preservation of degrees of freedom. However, with historic cost benchmarking the issues discussed above with regard to gathering projections of these cost drivers are no longer relevant.

Once the outputs framework is properly embedded and tested, and as more data becomes available over time, appropriate outputs (i.e. those that are unambiguously measurable and comparable across operators) may also be included as a direct cost. We anticipate that there is likely to be limited scope for this at the next GDPCR and, potentially, at DPCR6, but the scope for their inclusion will increase over time.

At the transmission level, it will be necessary to accept that a full set of potential output factors is unlikely to be available. It will be necessary to work with the data that is readily available. Again, we expand on this in Section 6.

5.2.3 Sample

Ideally, historic cost benchmarking would utilise a large panel of historic data to ensure robustness. As time progresses, the panel size will increase and the model estimations will become more robust. However, care must be taken to ensure that historic data is a relevant yardstick by which to assess future expenditure. If the coming decade(s) represent a period of uncertainty, and we are likely to witness a trend-break in terms of the activities and related costs of a network, then it might be inappropriate to use historic cost benchmarking to influence future allowances. The extent to which such a trend break occurs will be an empirical matter for investigation at the relevant regulatory review.

A similar discussion to that contained in Section 5.1.3 regarding the inclusion of international comparators is relevant here. There is unlikely to be a sufficient need to warrant the resource implications at the distribution level. However, it will be necessary to collect data for international comparators to deliver robust estimations for the transmission networks. In this regard, the list of explanatory factors and outputs set out above becomes more of a "wish list". It might not be possible to collect the necessary data on a consistent and timely basis.

5.2.4 Technique

As for Option 1, at the distribution level, we propose the use of OLS/SFA, as deemed appropriate by the researchers at the relevant regulatory review, testing down from a general translog specification.

At the transmission level however, we anticipate that it will not be possible to gather a large panel of data. This scarcity of data suggests that it might not be possible to undertake regression analysis with sufficient rigour. With this in mind, we are proposing the use of DEA as the preferred technique for the transmission networks. Arguably, DEA is more robust to small sample analysis. DEA also has the benefit of identifying explicitly the frontier firms that determine the estimated scores of firms off the frontier. This creates scope for an iterative or multistage research agenda, which we describe in Section 6.

5.2.5 Discussion

The framework described above for historic cost benchmarking is simple, specified at a high level, and practical to implement. Benchmarking historic costs is a well developed practice for energy network regulation, and the network operators will be familiar with the techniques used and the incentives and implications of the analysis. The cost categories defined above are designed to minimise potential distortions of input choice. Conducting the analysis on this basis will therefore ensure consistency with the DPCR5 equalised incentives framework, which we presume Ofgem intends to roll out where possible to the other networks in due course.

The OLS/SFA technique identifies the frontier of performance for each cost category. For BSCs, the results could be used to directly inform allowed revenue. However, results for network related costs would require careful interpretation. It is not obvious that historic cost benchmarking could be used to inform directly on allowances if network activities and underlying cost structures change materially in the future relative to the past.

For transmission networks, familiar issues of sample size are still important (see for example Section 2.4.5). The change in technique is designed to address this concern, but Ofgem will need to be pragmatic in implementing this approach at the transmission level.

In general, it is important to re-iterate that there are significant limitations and weaknesses associated with the use of historic benchmarking at present, as we have discussed in Section 2. Ex post benchmarking to inform future allowances has the potential to weaken incentives to innovate and could encourage operators to simply follow the plan. There is also the possibility that future activities and cost structures might be materially different from those observed in the past. Related to this is a concern that operators might (rationally and efficiently) roll out new services at different times and that historic cost benchmarking, given the likely sample size, might struggle to distinguish between early action and managerial inefficiency.

Table 6 below provides a summary assessment of Option 2 against the criteria identified in Section 3.

Table 6. Assessment of Option 2

Criteria	Option 2: DPCR 5 "Lite"
Robustness	Likely to obtain a sufficient sample to carry out robust panel data analysis, at least for the distribution networks. The sample will grow over time, improving robustness in the future. However, in the short term, data on new outputs will not be available for ex post benchmarking. Furthermore, given the issues surrounding the use of ex post benchmarking over the next decade, Option 2 may not be appropriate to inform future revenue allowances. Results might therefore be considered informative rather than definitive.
Transparency	Use of OLS with top-down cost categories is simple and replicable. Option 2 is a natural development of the analysis used at DPCR4 and DPCR5. Benchmarking historic costs is therefore an analysis that is familiar to network operators.
Promote efficiency	In the long-run, the proposed cost measure should ensure equal treatment of all costs, minimising distortions over input choices. In the short run incentives to distort input choice will be minimised by benchmarking as far as possible all competing costs. There is a risk that ex post benchmarking might not incentivise delivery of required outputs and could stifle innovation. Thus, although Option 2 provides strong incentives to reduce costs, this might come at the expense of delivering new outputs.
Consistency with wider regulatory framework	Aside from the scope for ex post benchmarking to stifle innovation, in the short run and given limitations on data, it might be difficult to include all relevant explanatory factors and outputs directly in Option 2. If Ofgem wishes to focus on provision of outputs, it will be necessary to provide these incentives outside benchmarking analysis. This could potentially add to the complexity of the regulatory package, and would require complex balancing of incentives.
Reasonable data requirements	High level data requests should be easy to complete and analyse. In contrast to Option 1, business plans will be less relevant. Option 2 is therefore less open to the possibility of gaming. For transmission operators, cross-sectional data may be used in order to reduce the burden of collecting data from international comparators.
Adaptability	Historic cost benchmarking may not allow the same level of flexibility and adaptability provided by benchmarking of future plans given the new requirements of networks. However, the volume of available data will increase over time which could allow more ambitious benchmarking.
Proportionate resource cost	High level specification of data requirements and fewer regressions imply lower resource cost than, for example, the DPCR5 analysis. Resource cost is likely to be relatively low.

5.3 Option 3: yardstick competition based on an estimate of total factor productivity

The third option we have considered is, arguably, not a benchmarking technique at all. Here we envisage a yardstick regime built on empirical analysis to determine industry wide TFP estimate. The X factor derived from this industry wide analysis would be applied to all operators. Operators able to outperform the industry average would earn returns above the regulated rate, while the opposite would be true of operators unable to progress at the same rate. Below we provide further description of the methodology that could be adopted and its properties.

Cost definition

To produce an estimate of TFP, by definition, all factors of production should be included. A total cost concept is therefore the appropriate measure.

For the purposes of discussion, we propose to define the total cost measure to be used as operating cost plus annual capital consumption (i.e. depreciation plus regulated return). This is the cost measure that underpins the present Dutch distribution regime. Defining the cost measure in this way allows us to use the Dutch regime as a basis for drawing conclusions regarding the likely performance of TFP analysis. It also demonstrates that a regime of this kind is practical, since it has already been implemented.

In the Dutch regime, the cost estimate at the industry level is the numerator in an index based method used to derive TFP. We discuss how this cost measure might be used in more detail below.

5.3.1 Cost drivers

Just as the cost definition should cover all factors of production, the model used to estimate TFP should capture the relevant explanatory factors and outputs of the business. The set of explanatory factors that we would propose to include are discussed in Section 4.3 and form the basis of the model we consider here. These explanatory factors and outputs could be included as right hand side variables in a regression, or as outputs in a DEA model[20].

We note that the explanatory factors and output measures proposed here are not entirely consistent with the Dutch regulatory model. Since the Dutch regulatory model derives an estimate of TFP using an index methodology, it is necessary to assign prices to all explanatory factors and outputs that are to be included in the

20 More accurately, if DEA is the preferred technique then certain of these variables, including specifically quality variables that are typically "bads" rather than "good", should be classified as inputs rather than outputs.

model, while such price data is not necessary under either a regression or DEA approach. The Dutch regime therefore uses an output set that includes all elements of the distribution tariff basket, with each output, so defined, valued at the weighted average industry level. As a consequence the outputs of the operators can be readily measured and converted into a monetary value. At present the Dutch regime takes no account of differences in the operating environment between the different distribution operators, although this is an area where EK, the Dutch energy regulator, has previously sought to make adjustments and might seek to do so in future.

The Dutch electricity distribution regime also takes direct account of quality/reliability measures (electricity distribution only). Effectively every disconnection/CML attracts a "fine" that the operator is required to pay. Incentives to provide incremental quality improvements are therefore directly proportional to the size of the fine, where these are parameters of the regime that the regulator may change. The aggregate cost of poor quality across the sector is measured and the industry compensated on the basis of average quality performance. It therefore follows that operators with better than average quality performance are compensated for average level quality, thereby earning additional profits – a yardstick dimension is therefore brought to quality performance. In principle, other outputs (in both the traditional benchmarking sense and the wider Ofgem sense of the word) could receive similar treatment, with the regulator-determined price used to calibrate incentives.

5.3.2 Sample

At the distribution level there is unlikely to be a need to extend the sample to include operators in different countries. The panel of data that is available from GB countries is likely to suffice. Limiting the sample to include the entire GB industry but no other operators will also allow the regime to be calibrated to ensure that there is, at the industry level, recovery of all costs. Only historic data would be used, with future prices rolled forward in line with past TFP. As with the Dutch regime, at each review it would be necessary to correct ex post for realised progress, where that departed from the estimate established at the preceding review.

Without extending the sample to include international comparators, there is no clear way in which this approach might be applied to the transmission networks. Clearly, using information on an estimate of TFP drawn from a sample of international transmission operators will only be appropriate if it is reasonable to anticipate that the rates of progress anticipated for the GB operators are similar to those observed elsewhere.

5.3.3 Technique

As discussed briefly in Section 4.5, there are three broad ways in which TFP might be estimated. These are:

- non-parametric approaches, typically using DEA;
- parametric methods, using either OLS or SFA; and
- index number methods.

Regardless of the method adopted, each will yield an estimate of TFP that can be applied to all operators. Here we would propose that both DEA and OLS/SFA should be used in order to identify the likely range within which TFP might fall. One would not anticipate material differences arising between the two techniques, but in the event one is found it would need to be investigated in detail to understand the technical reasons for any difference. Where the two techniques yielded similar results, this would increase the perceived robustness of the methodology.

For the parametric estimation of TFP, we would recommend that a similar process should be used to determine an appropriate model specification, i.e. beginning with a general translog specification and testing down. The extent to which this procedure is possible is ultimately an empirical matter that will need to be tested at the relevant network review.

At the each review, this forward looking estimate will be updated to reflect actual outturn productivity improvements over the period. As a consequence all operators can be sure that they will be exposed to an X factor that accurately reflects the outturn rate of change of TFP for the whole sector (albeit with a lag of several years).

5.3.4 Discussion

There are a number of important strengths to a regime based on a common estimate of TFP.

As noted above, a regime of this kind provides a strong incentive for operators to improve efficiency. As the process for estimating TFP becomes more mechanistic there is also a reduced need for Ofgem to regulate the detail of operator activity. The task is reduced to collecting headline data and publishing results of data manipulation. Consequently, the approach can be very light touch, potentially requiring few resources on the part of either Ofgem and/or the operators. Similarly, the technique removes all material distortions to input choice, since it focuses on total cost.

We understand that questions have been raised over the ability of regimes based on yardstick competition to deliver industry change in support of, for example, decarbonisation. Similar concerns have arisen in the past with regard to the ability of the regime to ensure that reliability is maintained in the future. We believe that these concerns could be addressed by modifying the approach to ensure that operators are "paid" for providing the outputs Ofgem is targeting in its *Emerging Thinking*. We note that the Dutch regulator has recently concluded a review of its regulatory arrangements for the distribution operators and intends to continue with its present arrangements, although it will seek to investigate how the arrangements can be made more flexible to account for important differences between operators and the potential for exceptional expenditures to arise.

In principle the yardstick approach guarantees that the regulator has a long "institutional memory". Investment now that produces savings later will result in poor performance in one review, but outperformance in subsequent periods. In principle then, the investment can be assessed by operators on their merits. In practice network managers might (quite rationally) have short term horizons, e.g. failure to deliver at least the industry average rate of TFP progress might have consequences for an individual's career, encouraging them to delay some investments. Arguably, this could be overcome by including "outputs" within the TFP formula so operators benefit directly from delivery. The "price" associated with these outputs could be flexed to match policy objectives.

Notwithstanding these strengths, there are important concerns about implementing a TFP based regime, a number of which relate not to the regime itself but to the timing of its introduction.

A regulatory regime based on an estimate of TFP would clearly represent a radical departure from the regime that Ofgem has developed across the four network industries. While the "building blocks" of regulation can still be observed, they would play a rather different role in a TFP based regime and there are a number of important consequences of adopting such an approach.

Adopting a regime centred on TFP would probably require Ofgem to move towards a price cap, while at the distribution level, the Ofgem approach has evolved into what is now close to a revenue control, where operators are allowed to recover anticipated costs and revenues are increasingly invariant to changes in demand. A TFP based regime is a far more natural fit with a price cap based regime (since it has the effect of granting operators an amount of input for each unit produced). Indeed, the Dutch distribution regime is a pure price cap, where prices (based on the usual tariff basket calculations) are allowed to flex in line with this index. Price cap regimes can give rise to an incentive for the regulated operator to encourage increased demand. While this incentive might be regarded as helpful with regard to certain outputs (e.g. quality, connecting distributed generation) Ofgem is unlikely to regard incentives to expand all explanatory factors and outputs as an unambiguous improvement.

Ofgem would no longer be able to guarantee that all operators would earn a reasonable rate of return on past investments. If operators are unable to match industry average performance a potentially substantial difference could open up between costs and revenues over time. A yardstick regime of this kind requires strong commitment on the part of the regulator in the event of a lagging operator experiencing financial distress, otherwise the incentive properties of the regime are materially weakened.

In this context, the credibility of such a regime depends on there being a consensus that the opportunities for future progress are broadly similar for all operators. If this is not the case then it is possible that the operators that are currently most efficient (having already made great progress) will be financially disadvantaged by a yardstick regime, while current laggards will experience windfall profits from simply catching up to the prevailing frontier. The variation in efficiency, as measured using the techniques deployed at DPCR5 for the electricity distribution sector, suggests that a move to TFP could give rise to windfall gains/losses of this kind. If this approach is to be introduced, therefore, there will be a need to test that there has been an appropriate level of convergence in operator performance.

As we discussed in Section 2.3, there is likely to be uncertainty over the extent to which historical cost trends can be expected to be informative about future costs. This might arise either because fundamental cost structures change in the light of new services provided by the networks, or as a consequence of the timing with which different networks roll out those new services. As a result there is a risk that a methodology of this kind might produce an estimate of TFP that bears little relation to future movements in TFP. While it is possible to put in place a regime that includes an ex post true up that will ensure that operators are exposed to outturn TFP, there is likely to be a considerable delay before revenues adjust to reflect industry costs, which might give rise to concerns over regulatory credibility and short term funding. Coupled with the strong incentives to reduce costs that a yardstick regime is known to exhibit, the net effect might be to discourage needed investments.

Finally, it is worth considering whether there is scope for the results of a TFP analysis of the distribution operators to be used to inform on the performance of transmission operators. It might be argued that the underlying technologies are sufficiently similar that long term trends in productive efficiency in one should be broadly replicated in the other. While it would appear inappropriate to base a price control on the strong application of this principle, it might provide a further piece of evidence with which to challenge the past and proposed expenditures of the transmission operators.

Our assessment is summarised in **Table 7** below. On balance, given the important role that early action and risky innovation might be expected to play in achieving a variety of environmental goals, we do not consider it appropriate to introduce a TFP-based regime at this stage, although it may be suitable to apply in future, due to its highly desirable incentive and risk-sharing properties.

Table 7. Assessment of Option 3

Criteria	Option 3: TFP
Robustness	A range of techniques exist to estimate industry TFP. A suite of techniques could be employed to develop a highly robust range of estimates of industry TFP. Since TFP focuses on rates of change at the industry level it is arguably more robust to models that capture only imperfectly exogenous differences between operators and other factors that complicate estimation of operator specific inefficiency.
Transparency	Communication of the approach and results should be relatively straightforward. We anticipate the techniques being well understood by a wide range of qualified experts allowing ready replication of the analysis by interested parties.
Promote efficiency	The technique could only practically be applied to a relatively formal yardstick competition regime. Hence it would support a regime in which there are very strong incentives to achieve efficiency improvements. The proposed cost measure should ensure equal treatment of all costs, minimising distortions over input choices. Potential concerns might arise over the extent to which it might discourage early action and more risky innovation as a consequence, both of which are likely to be highly valued. If the energy network industry is about to experience a trend break, as it undertakes a range of new activities to support decarbonisation, then past trends in TFP might be a very poor guide to future TFP.
Consistency with wider framework	Implementation of a regime based on yardstick competition would require some significant regulatory reforms. Operators would need to accept the danger of potential sustained periods of excess returns/losses. Capital might be funded at the industry level, but not for all operators individually. TFP would fit more naturally into a price cap system. It is not clear that there is a consensus that all operators are yet equally efficient. It might therefore be argued that it is currently inappropriate to implement a yardstick regime. The regime would need to be supported by additional strong incentives to deliver outputs, given the strong incentives to avoid expenditure where possible. These arrangements would require careful calibration.
Reasonable data requirements	Given the focus on measuring industry wide change, rather than operator specific inefficiency, the data requirements for Option 3 are comparatively modest. However, since a panel is required, the scope for application in transmission (based on international data) might be more limited.
Adaptability	The approach can be seen as highly formulaic, which might limit the scope for new developments (such as the emergence of new outputs) to be captured by the technique. The recent review of the Dutch regime suggests that within these limits the system can adapt to reflect changes over time. Given the potential lag before operators are exposed to actual, rather than estimated, industry TFP, strong regulatory commitment to the regime would be required.
Proportionate resource cost	The focus on high level measurement of industry change arguably reduces the need to measure operator specific factors in great detail. There is a reduced need for multiple model runs and disaggregation of data. The measurement of TFP, even if several techniques are applied, would involve relatively modest resource input.

5.4 Option 4: model network analysis

The final methodology we have considered makes use of model network analysis[21]. This approach uses a stylised model of each network as a basis of comparative regulation, with operator performance measured by comparing the ratio of actual cost to modelled cost for each operator.

5.4.1 Cost definition

The model we consider here makes use of an annual, total cost measure equal to opex plus annual capital consumption. As a consequence operators are provided with an incentive to manage all costs and the potential to distort input choices are minimised. Since capital costs are included in the cost measure, operators are provided with incentives to plan network expansion efficiently. This is in contrast to models where only operating costs are included, which provides operators with an incentive to operate existing assets efficiently.

5.4.2 Cost drivers

Unlike the models described above, MNA based methodologies do not make use of the usual benchmarking cost drivers. Undertaking MNA would require the collection of data on the operating environment of each operator. Data on connection density and land use is typically collected at relatively high level of granularity (e.g. at the level of the local council ward). This information is then passed through an engineering algorithm to determine the volume of assets required to serve each detailed sub-region. The derived model networks for each sub-region can then be aggregated across all sub-regions to produce an estimate of the total inventory of assets that would be required to serve the specified region.

Operators might be required to provide the data they have available, for example a database of all connection points that includes a post code. Otherwise, data drawn from government sources might be used.

5.4.3 Sample

At the distribution level, a sufficiently large sample could be produced by including only GB operators. Limiting the sample to GB operators would help to reduce the resource cost of implementing this approach by removing the need to collect additional data from other countries and ensure its comparability.

---

21 Model network analysis makes use of a range of simplifying assumptions, as discussed in Section 4.5. More detailed Reference Network Analysis has not been considered in detail at the distribution level, as we anticipate a very substantial resource cost arising from the detailed data collection and analysis that would be required. However, RNA is likely to be required at the transmission level, as the simplifying assumptions used in MNA (principally the homogenous distribution of connection points within subsets of the operator's service area) are unlikely to be valid.

As we have observed above for other techniques, this approach has limited scope for application at the transmission level without collecting data from other countries.

5.4.4 Technique

As discussed above, an engineering model is used to derive an optimal network for each regulated operator, making use of a range of detailed data about each network.

Once the optimal network has been derived, Ofgem can:

- use standardised costs and the inventory of assets from the model network to determine optimal network costs;
- compare actual total cost with the optimal to estimate an efficiency ratio for each network; and
- use the relative "distance" between optimal and actual to inform on relative performance.

It is important to note that operators are not compared to their stylised ideal directly. Instead they are benchmarked on the basis of the distance of their actual costs from the ideal, relative to the rest of the sector.

While we envisage using MNA for the distribution level, RNA is likely to be required at the transmission level for the results to be meaningful. This follows because there are far fewer connection points in a transmission network than with a distribution network. The "detail" of the specific assets, routes etc. is therefore likely to be far more important at the transmission level than at the distribution level where connection patterns are likely to be more homogenous.

5.4.5 Discussion

Ofgem has not made use of MNA or RNA in any of its previous reviews, although the fact that use of reference networks is a core element of the Swedish regulatory model demonstrates that such a regime is practical.

It is clear that a move to benchmarking through MNA would be a significant change in approach. The operators are likely to be sceptical of such models, creating the need for the work to be highly robust, clearly communicated to the sector and open to testing. This process is likely to require a large time investment by Ofgem and significant engagement with the operators. The technique is therefore likely to be resource intensive.

It is also likely that operators that perform poorly against an optimal network benchmark would invest significant effort in identifying reasons why the underlying (and necessarily stylised) engineering modelling was not accurately reflecting important points of detail. There is scope for Ofgem (and their consultants) to be tied up with company engineers in lengthy debates about the merits of the modelling and its application to their specific network. Operators are likely to invest significant resource in critiquing the regulator's models. The adoption of this approach also seems to move detailed network planning into the heart of the regulatory interaction, an area in which it is likely that the operators will always have great expertise. This arguably increases the likelihood of the results of this approach being subject to successful challenge by the operators.

A regime based around MNA, given the proposed total cost measure, would provide the operators with many desirable incentives, in particular to manage costs and plan network reinforcement/expansion as efficiently as is possible. However, the benchmarking methodology alone does not incentivise the delivery of other increased levels of all explanatory factors (including outputs), creating a need for other incentive mechanisms to be added to the regime. Any adaption of the regime to take account of other outputs would therefore need to be effected through specific incentive mechanisms.

The results of an MNA based assessment of efficiency could not be used to determine allowances directly. Rather, the regulator would be provided with a metric for determining which operator's total cost was closest to the modelled optimum. The model also lacks a forward looking dimension, since it is based on engineering analysis of prevailing network configurations. If the demands met by the network will evolve very materially over the coming price control period it is not clear that the MNA will necessarily reflect this well, limiting the extent to which it might be able to inform on the value for money offered by different operator plans.

It is worth noting that the transmission modelling would be complex and it is unlikely that any outsider could successfully argue that their modelling was more robust than that undertaken by the transmission operators of their own networks. It also follows that the modelling of international peers will also need to be undertaken in great detail, which might be time consuming and costly. There would be a need to gain a thorough understanding of the operation of any operator that the transmission operators might be compared to in order to ensure that drawing that comparison was reasonable. Overall, therefore, it seems that the "battle of the models" critique of this approach is particularly significant at the transmission level.

Notwithstanding the discussion above, we believe that MNA can play an important role in improving the robustness of other benchmarking approaches. As we discussed in Section 4, MNA can be used to gain a deeper understanding of how the connection density of a given operator's service area might be expected to impact on network configuration and hence on its costs. For example, MNA has been used successfully in this narrower role in Austria, where stylised models of the networks were developed in order to provide proxy variables (modelled network length, by voltage level) that captured connection density. These proxy variables were then used in a regression analysis to assess company efficiency. An approach of this kind ensures that important structural effects can be captured robustly using variables that are not under the direct control of the operators, where the alternative is to make use of actual network length which is. While we are not convinced that Ofgem should consider adopting a benchmarking approach in which MNA plays a central role, we believe that MNA warrants further investigation as a way to improve the robustness and incentive properties of other benchmarking approaches.

Our summary assessment of this technique is set out in **Table 8** below.

Table 8. Assessment of Option 4

Criteria	Option 4: MNA
Robustness	Assessments based on MNA are only as robust as the engineering models that underpin the work. Work would be required to develop and validate the models, detailed data and the results. There is the scope for an assessment of this kind to reduce to a "battle of the models", particularly at the transmission level where the "optimal" network would be sure to be subjected to a very detailed critique.
Transparency	MNA can only be implemented through the development of specific engineering tools. There would be a need to communicate clearly how Ofgem's MNA models operated. It might be difficult for operators to replicate this modelling without investing in their own shadow analysis. Overall the technique might be regarded as less transparent than more traditional benchmarking approaches.
Promote efficiency	Given the proposed total cost measure, the methodology should promote efficient input choice and should encourage efficient planning. The technique alone does not provide specific incentives to focus on the delivery of outputs.
Consistency with wider framework	Significant further analysis might be required to convert the results of MNA analysis into allowances for operators. As noted above, it is not obvious that this approach provides a direct incentive for operators to deliver valued outputs. Indeed there is a danger that the stylised model discourages expenditure that might enhance output delivery/quality. There would be a need to put in place other incentive arrangements.
Reasonable data requirements	At the distribution level detailed data on operating regions would be required. This detailed data would need to be thoroughly tested and checked. Detailed GIS data would be required at the transmission level, for GB and international operators.
Adaptability	The inability of this technique to provide direct incentives for operators to deliver outputs reveals limits to the extent to which the technique alone might be able to adapt to fit evolving regulatory needs. However, the technique would be robust to changing network configurations, although this robustness only arises as a consequence of the need to repeat the underlying engineering analysis at each review.
Proportionate resource cost	The development, testing and communication of MNA/RNA models would require significant time commitment on the part of both the operators and Ofgem. It is likely that this effort would need to be repeated each price control, unless the networks were thought to be in steady state.

6 Recommendations

This section sets out our recommended approach to benchmarking for each of the network operators. We provide a discussion of the context for benchmarking for each network and the rationale for the chosen technique in each case. We have also attempted to highlight any regulatory issues that are likely to arise if the recommendation is adopted and provide some initial thoughts on how they might be addressed.

6.1 Electricity distribution

6.1.1 Discussion

Of the 4 network sectors, electricity distribution is the sector most likely to witness a step-change in its activities over the next decade as a result of decarbonisation objectives. Energy efficiency measures, small-scale generation plant, the electrification of heating and the likely roll-out of electric cars on a wider scale will all impact the volume and direction of flows on distribution networks in ways that appear difficult to predict. While all of these developments are expected to arise imminently, there is no consensus on the ways in which electricity distribution networks might need to adopt to accommodate them. This gives rise to a compelling need to encourage innovation and proactivity in the DNOs and to a potentially material uncertainty over the future activities and cost structures of distribution networks.

6.1.2 Recommended approach

Benchmarking in a regulatory context has traditionally focused on comparing the historic costs incurred by the regulated operators in order to try to identify best practice. This best practice can then be used as a basis for determining future cost allowances, in the light of what operators have revealed to be possible by their past actions. For the reasons discussed above (particularly the risk that ex post benchmarking might provide incentives for operators to simply stick to the plan and discourage innovation), this approach appears unsuitable for electricity distribution going forward. For this reason, among others, we are proposing a regulatory regime that is centred on Option 1, the benchmarking of future plans at a total cost level. While we see many merits in Option 3, again, the present context for benchmarking makes a regime based on TFP analysis less desirable at this time. TFP is likely to be a more useful technique in the longer term, as a new steady state emerges for the industry.

It will also be important to ensure that historical information is used effectively to check the basis on which plans have been prepared. Given the volume and value of future investments there is likely to be a continued role for expert scrutiny of business plans and the use of historic benchmarks. We are therefore recommending the continued application of historic benchmarking, but with a modified role. In past reviews historic benchmarking of operating costs has been used more prescriptively to determine future allowances. In future, we believe that evidence gained from historic benchmarking should be used to challenge the validity of future plans, acting as a further point of traction for the regulator, rather than for setting allowances in a more mechanistic manner. Effective and targeted use of analysis of historic data should assist Ofgem in identifying any unwarranted inflation of operator plans. Given the relevance of this concern, as discussed above, we also envisage a continued role for incentives for efficient and effective forecasting, such as the existing IQI mechanism.

If Ofgem wishes to continue to undertake more formal ex post benchmarking to support its assessment at regulatory reviews, we would advise it to focus attention on benchmarking business support costs. As Ofgem identified at DPCR5, these costs are only weakly, if at all, substitutable for more direct network related expenditures. As a consequence it is unlikely that ex post benchmarking of business support costs would reduce materially incentives to innovate and invest.

Adopting the approach we recommend will provide Ofgem with a range of benchmarking analysis with which to assess operator plans. It will have analysis of a set of operator plans under two benchmarking approaches (totex and total cost) in addition to analysis of the consistency of these plans with historic expenditure. It is unlikely that there will be a single preferred model that can be applied mechanistically to determine allowances. Ofgem will instead need to adopt a holistic approach to using this portfolio of information to determine final allowances and output targets.

Over time, as the present uncertainty is reduced and the industry returns to a steady state, Ofgem could consider implementing an alternative approach. Our review has identified that a TFP based approach has a number of important strengths. Under the right circumstances, a TFP regime could deliver significant benefits to customers through the creation of strong incentives while also simplifying the existing arrangements. Once the present uncertainty has played out, such a system will warrant active consideration in our view.

6.1.3 Summary of recommended approach

A summary of our recommended approach is presented in **Table 9** below. Benchmarking future plans will form the central comparative efficiency analysis, with revenue allowances set on the basis of the results and subsequent dialogue with the operators. This will be supplemented by high-level benchmarking of historic costs to provide comfort that plans are not unnecessarily inflated.

Table 9. Summary of recommendations for Electricity Distribution

	Recommendation
Costs	Total cost, making use of two measures. Planned operating expenditure plus a measure of capital consumption. Planned operating expenditure plus planned capital expenditure.
Cost drivers	Ideally, the full set of explanatory factors presented in Section 4.3, guided by empirical analysis at each review. Include directly, where possible, outputs, if supported by empirical analysis.
Sample	The scenarios presented in the 14 DNO business plans. Make use of historic costs (as per Option 2 in Section 5) to increase the scope for plans to be tested.
Technique	While Stochastic Frontier Analysis (SFA) is usually preferred when undertaking efficiency analysis, data contained in operator plans will not contain statistical noise. This allows the robust use of Ordinary Least Squares (OLS) or Corrected OLS (COLS).

6.2 Gas distribution

6.2.1 Discussion

Much of the discussion presented above applies equally to the Gas Distribution Networks (GDNs). In particular, there are a number of uncertainties over the future role of the gas industry as a result of decarbonisation. Elements of the decarbonisation programme, such as the electrification of space heating, will tend to reduce the demand for gas (and hence gas networks). On the other hand, gas distribution networks may have an important role in supplying small scale gas-fired electricity generation plant. Such plant might be needed to ensure system stability as the penetration of intermittent renewable generation increases. Further, there are potentially new uses for the gas networks resulting from decarbonisation. Whether or not biogas will become important, for example, is currently uncertain, as is the impact this might have on the gas network costs and activities.

As with electricity distribution, therefore, these uncertainties could lead to changes in the underlying scale and scope of gas distribution network activities. Consequently, our central recommendations for the gas distribution sector are very similar to those we have provided for electricity distribution.

We therefore propose a regime centred on Option 1, the benchmarking of future plans, supported by historic cost benchmarking via Option 2. We see substantial merit in focusing the regime on each operators' value for money proposition, where proposed costs are matched against the delivery of valued outputs. In particular this approach appears consistent with the philosophy set out in Ofgem's *Emerging Thinking*. It also removes the need for Ofgem to address the question of how to interpret and use total cost benchmarking based on purely historic costs. Under this approach Ofgem will not find itself in a position where it could consider writing off past investments. To do otherwise would be an important departure from past practice and would be viewed as a new risk for the regulated networks.

Notwithstanding the above arguments, there is likely to be scope to make more use of historic benchmarking in gas distribution than in electricity distribution. It appears reasonable to assume that the uncertainty over future gas networks activities appears less material than for electricity networks. A number of the envisaged changes (in particular electric vehicles and the deployment of distributed generation) are only likely impact electricity networks. It might be argued that the need to encourage innovation in gas distribution is consequently diminished. Ideally ex post benchmarking of historic costs would be undertaken on the basis of total costs. However the inherent lumpiness of capital investment coupled with some uncertainty over the future role of networks suggests that historic cost benchmarking might continue to focus on benchmarking competing operating costs in the short to medium term.

6.2.2 Summary of recommended approach

A summary of our recommendations for the gas distribution sector is presented in **Table 10** below.

Table 10. Summary of recommendations for Gas Distribution

	Recommendation
Costs	Total cost, making use of two measures. Planned operating expenditure plus a measure of capital consumption. Planned operating expenditure plus planned capital expenditure.
Cost drivers	Ideally, the full set of explanatory factors presented in Section 4.3, guided by empirical analysis at each review. Include directly, where possible, outputs, if supported by empirical analysis.
Sample	The plan scenarios presented in the 8 GDN business plans. Make use of historic costs (as per Option 2 in Section 5) to increase the scope for plans to be tested.
Technique	While Stochastic Frontier Analysis (SFA) is usually preferred when undertaking efficiency analysis, data contained in operator plans will not contain statistical noise. This allows the robust use of Ordinary Least Squares (OLS) or Corrected OLS (COLS).

6.3 Electricity and gas transmission

6.3.1 Discussion

Benchmarking transmission operators is challenging. There are typically few transmission operators within the same country, often only one. As a consequence, benchmarking of transmission operators is typically only possible through the collection of international data, exacerbating the usual problems of ensuring comparability. Problems of comparability between operators are also more material for transmission operators than distribution operators. While there are important differences between distribution regions that need to be captured, the provision of network access to very many smaller consumers is a far more homogenous activity than transmission, where there are fewer connection points and consequently far more specificity in network design. The exact location of load and demand, the key determinant of transmission network configuration, will vary very substantially between different operators. There is also more scope for network design to be inhibited by planning restrictions, which again might be more operator and country specific.

From the review of techniques presented in Section 5, it is clear that there are material obstacles to the direct implementation of several of the techniques. The technique(s) that we have recommended for distribution, centred on the benchmarking of forward looking plans, is impractical at the transmission level. The necessary data on international peers will simply be unavailable as no other transmission operator will collect the required data on a comparable basis. Even if such data were collected, the data may not be available publically, so Ofgem may have to enter into negotiations with other regulators. In the case of electricity transmission, there is perhaps limited scope for the use of econometrics if the three GB operators submit multiple scenarios. There is no similar scope in gas transmission where there is only one operator. Given this, we expect that the direct comparison of plans will therefore need to be conducted on a more ad hoc basis. While this can still be expected to yield some helpful insights, it is likely to be informative at best.

The collection of data is also likely to limit the direct applicability of Option 3 (TFP). In order to implement TFP it is necessary to collect a panel of broadly comparable data. Collecting such a panel for a number of other TSOs might be difficult and time consuming. Even if the data could be collected, arguments over the suitability of a TFP based technique at this point in time (where the industry is potentially facing a material change in its activity) are likely to be compelling. It is not clear that recent movements in TFP across Europe will be a reliable guide to future spend by the GB transmission operators. In Section 5.3 we raised the prospect of using an estimate of TFP derived from the sample of distribution operators to inform on the transmission operators. Further analysis of the extent to which this proposal has merit is probably warranted.

Option 4, the RNA approach, appears to be a potentially fruitful way to proceed. However, to implement RNA successfully it would be necessary to develop robust underlying engineering models for at least the GB operators. This is likely to require a significant resource input by both Ofgem and the transmission operators[22]. As discussed in Section 5.4, there would remain the danger of the modelling being challenged by the regulated operators and of the interaction between regulator and operator becoming unnecessarily focused on stylised modelling. Ofgem might wish to seek an engineering assessment of the likelihood of a RNA approach yielding information that is regarded as sufficiently robust for regulatory use.

Given this discussion, it is our view that Option 2 (benchmarking historic expenditure) represents the best way forward. However, given the specific

---

22 Since there are three GB Operators, there is some limited scope for application of the proposed RNA technique to the three alone. Extending the sample to include operators elsewhere in Europe would greatly increase the volume of work required. There would be a need to gather a range of detailed data on the location and shape of transmission connected load and demand in a region where Ofgem has no jurisdiction to require the provision of that information.

challenges of benchmarking transmission operators, there will be a need to tailor the methodology, as we discuss below. Further, we propose that Ofgem seeks to supplement this analysis with a range of other approaches. Since no single approach is likely to provide definitive results, there is merit in gathering together as wide a portfolio of information as is possible,

6.3.2 Recommended approach

We propose that Ofgem considers adopting a high level DEA benchmark of the recent historic costs of the transmission operators against a small number of European peers. Given limits on the data that is likely to be available we recognise that this approach is unlikely to provide definitive results. We therefore propose that this analysis is undertaken at an early stage of the regulatory review process to ensure that it is used to inform discussion between the regulator and the operators, rather than to guide more mechanistically the setting of allowances. We also propose that Ofgem supports this analysis through the use of a range of other pieces of analysis, including continuing with its prevailing practice of seeking expert review of the operators' plans.

The sample against which the GB operators should be compared using DEA should include data drawn from other European countries. Data on recent expenditures is likely to be available from the relevant operator accounts and or published regulatory accounts. High level data on relevant outputs and explanatory factors could also be collected, using the candidate variables described in Section 4.3 as a guide. However, the exact choice of which explanatory factors and output measures to use is likely to be strongly guided by the data that is available in the public domain. Given the nature of the data that is available, it is likely that not all outputs will be included directly in the benchmarking model, again as comparable data will not be available for other operators. There will also be a need to adjust the cost data for each operator for a wide range of potential differences. The relevant potential differences for which to adjust are set out in Section 2.2.3. We anticipate that it will be difficult to develop a panel of data on which to undertake analysis and therefore propose that Ofgem focuses on undertaking cross sectional analysis.

As a consequence of the limited data that is likely to be available, benchmarking will almost certainly be undertaken on the basis of an imperfect set of cost drivers. It is unlikely that there will be sufficient data to allow the robust estimation of an SFA model, or a process of "testing down" from a general translog specification. Given this, we consider it more appropriate to undertake benchmarking using DEA, which is arguably more robust in cases where there is less data. However, when undertaking DEA with such data, it will be essential to interpret the results of the analysis with care. We would propose an iterative approach.

Ofgem would begin with a first cut DEA run making use of the data collected. This first pass would yield a set of initial "efficiency scores" that should be used to check the integrity of the data further. Operators that were found to be very highly efficient, or very highly inefficient, should have their data subjected to further and more thorough investigation. The refined data would then be used to produce what might be called a "draft" set of efficiency scores. The draft results could be used to assess whether there was *prima facie* evidence of inefficiency. This analysis could form the basis of a dialogue between the operator and the regulator on why the costs of the GB firm might be higher (i.e. for reasons other than managerial inefficiency). It should then be possible to assess the potential materiality of any differences identified and account for it in the study. If the identified gap could not be explained entirely, then the regulator would be able to identify that the costs of the GB firm are higher than appears necessary given the costs incurred by other transmission operators.

Given the limits of this analysis, we propose that Ofgem makes use of as many pieces of analysis as possible. For example it could compare the TFP of transmission operators against the TFP of the distribution operators. Similarly, for the three GB electricity transmission operators it might be possible to undertake some simple regression analysis using the available panel of data. Since there might be only 10-20 data points in the sample, it is likely that only very simple regression (perhaps in line with the analysis used at DPCR4) will be possible. Finally, there is clearly merit in continuing with a range of other analysis, including expert scrutiny of business plans as at present.

While none of these approaches individually is likely to provide sufficiently definitive information on the reasonableness of the transmission operators' costs, together they will provide Ofgem with the widest possible range of information. This could form the basis of a constructive dialogue with the operators throughout the price control process.

The proposed DEA approach is summarised in **Table 11** below.

Table 11. Summary of recommendations for the Transmission businesses

	Recommendation
Costs	Total cost, making use of a standardised measure of capital consumption. Given the sample, there will be a need to adjust costs to reflect a wide range of factors, including exchange rates, tax etc.
Cost drivers	Drawn from the set of explanatory factors presented in Section 4.3, but guided by the data that is publically available in practice. Include directly, where possible, outputs, but recognising that limited data is likely to be available.
Sample	The GB operator(s) supplemented by a number of operators from other countries (e.g. 4-6 others). Use historic data for the most recent year available to develop a cross section.
Technique	DEA, using a 1 input multiple output model. We would propose that Ofgem investigates both a constant and variable returns to scale frontier. From a regulatory perspective a VRS frontier is likely to be reasonable since the scale of operation of a transmission operator is typically invariant. However, a VRS frontier can reduce the ability of the DEA model to discriminate between operators, and in small samples can often result in all operators appearing efficient.

FRONTIER ECONOMICS EUROPE
BRUSSELS | COLOGNE | LONDON | MADRID

Frontier Economics Ltd 71 High Holborn London WC1V 6DA
Tel. +44 (0)20 7031 7000 Fax. +44 (0)20 7031 7001 www.frontier-economics.com