CEPA

RIIO-GD2 cost assessment – econometric modelling & regional factors

June 2019

Ofgem

FINAL REPORT

IMPORTANT NOTICE

This report was prepared by Cambridge Economic Policy Associates (CEPA) Ltd for the exclusive use of the client(s) named herein.

Information furnished by others, upon which all or portions of this report are based, is believed to be reliable but has not been independently verified, unless expressly indicated. Public information, industry and statistical data are from sources we deem to be reliable; however, we make no representation as to the accuracy or completeness of such information, unless expressly indicated. The findings enclosed in this report may contain predictions based on current data and historical trends. Any such predictions are subject to inherent risks and uncertainties.

The opinions expressed in this report are valid only for the purpose stated herein and as of the date of this report. No obligation is assumed to revise this report to reflect changes, events or conditions, which occur subsequent to the date hereof.

CEPA Ltd does not accept or assume any responsibility in respect of the report to any readers of the report (third parties), other than the client(s). To the fullest extent permitted by law, CEPA Ltd will accept no liability in respect of the report to any third parties. Should any third parties choose to rely on the report, then they do so at their own risk.

CONTENTS

GLOSSARY

Term	Definition
AER	Australian Energy Regulator
ASHE	Annual Survey of Hourly Earnings
BPDT	Business Plan Data Templates
Capex	Capital expenditure
CAWG	Cost Assessment Working Group
CEPA	Cambridge Economic Policy Associates
CMA	Competition and Markets Authority
COLS	Corrected Ordinary Least Squares
CRU	Commission for Regulation of Utilities (Ireland)
CSV	Composite Scale Variables
DEA	Data Envelopment Analysis
DPCR	(Electricity) Distribution Price Control Review
DSA	Distribution Service Area
ECA	Economic Consulting Associates
ENA	Energy Network Association
FE	Fixed Effects
GB	Great Britain
GDN	Gas Distribution Network
GDPCR	Gas Distribution Price Control Review
HSE	Health and Safety Executive
IMRRP	Iron Mains Risk Reduction Programme
IPS	Intermediate Pressure System
LPS	Low Pressure System
LTS	Local Transmission System
MEAV	Modern Equivalent Asset Value
MPS	Medium Pressure System
NGN	Northern Gas Networks
NPg	Northern Power Grid
NTS	National Transmission System
Opex	Operating expenditure
POLS	Pooled Ordinary Least Squares
PR14 / PR19	Price Review 2014 / 2019 (Ofwat)
PRE	Public Reported Escape

PRI	Pressure Reduction Installations
RE	Random Effects
Repex	Replacement expenditure
RIIO	Revenue = Incentives + Innovation + Outputs
RPEs	Real Price Effects
SCF	Special Cost Factor
SFA	Stochastic Frontier Analysis
SGN	Scotia Gas Networks
Totex	Total expenditure
UR	(Northern Ireland) Utility Regulator
WWU	Wales & West Utilities

EXECUTIVE SUMMARY

Ofgem's Gas Distribution (GD) cost assessment team commissioned Cambridge Economic Policy Associates (CEPA), in association with Economic Consulting Associates (ECA), to provide advice on the cost assessment process for RIIO-GD2 (GD2) and RIIO-2 more generally.

CEPA and ECA have prepared three briefing papers for publication alongside Ofgem's consultation paper on cost assessment. The topic of each paper is as follows:

- econometrics and regional factors (prepared by CEPA).
- business support costs (prepared by ECA).
- frontier shift (prepared by CEPA).

This paper discusses econometric methods and regional factors. This includes:

- the approach Ofgem might use to reach decisions on the econometric models it selects in GD2 and the options for those models; and
- how to account for regional factors – i.e. regional differences in the external operating environment of each GDN that effect their relative cost performance but are outside of their control.

The objectives of using econometric benchmarking in GD2

A crucial element in the assessment of an appropriate level of revenue to allow network companies to cover their costs and fulfil their statutory duties and output obligations, is a judgement about the extent to which they are able to become more efficient. The use of external benchmarking – the comparison of a firm's costs to an exogenous reference level (for example the most efficient firm in the sector) – can improve the quality of this assessment and be used to strengthen the incentives facing regulated firms.

External benchmarking provides a methodology to determine a regulated company's revenue allowances independently of its own costs. This can help to reduce the problems created by the asymmetry of information between the regulator and regulated companies, increase confidence in companies' own expenditure projections at the time of the price review, and strengthen the incentives facing regulated firms by rewarding them financially for closing the gap between their actual and potential efficiency.[1]

There are a range of methods that can be used to determine the efficiency frontier for benchmarking purposes. These include linear programming methods and statistical techniques. UK regulators, including Ofgem and Ofwat, have tended to rely on econometric (parametric) techniques to estimate the efficiency frontier in recent price reviews. They have also tended to rely on deterministic approaches, the most common method being Corrected Ordinary Least Squares (COLS).[2]

[1] See Pollit & Nillesen (2003): 'The consequences of consumer welfare of the 2001-2003 Electricity Distribution Price Review in the Netherlands' for further discussion.

[2] In simple terms, the efficiency frontier under COLS is estimated using the ordinary least squares (OLS) technique. A functional form for the production / cost function is specified, and this is estimated using OLS. The calculated line of

Econometric methods require an assumption about the relationship between the firm's inputs and outputs and estimate the parameters of a function representing this. This allows for different views of industry dynamics[3] and cost drivers of the regulated companies to be considered in the benchmarking exercise. However, accounting for the complexities of real-life businesses and, in particular, the effect that differences in the external (e.g. regional) operating environment of individual firms may have on the observed costs of companies compared to their peers, can also make benchmarking challenging.

For these reasons, some stakeholders have criticised the use of benchmarking as part of regulatory proceedings, arguing that the results of benchmarking derive from arbitrary choices of the adopted technique, or the assumptions adopted within the econometric modelling process. This includes how the benchmarking accounts for the company specific differences in operating environment that may explain the differences in observed costs between companies for reasons other than (in)efficiency.

Therefore, if external benchmarking is used within regulatory proceedings such as GD2, it needs to be applied with care and a clear and logical process followed to select econometric models and account for regional and other company specific factors that may explain companies relative cost performance.

Within this context, we have considered a range of different topic areas and options for how Ofgem might practically approach its econometric benchmarking at GD2. This includes the approach / methodology to model selection, how Ofgem might go about selecting an appropriate sample size for its econometric models, options for estimation technique, choices of cost aggregation, model specification (e.g. selection of cost drivers) and the treatment of regional factors within the analysis.

Model selection methodology

The model selection methodology includes issues such as the criteria used to assess the strengths and weaknesses of different benchmarking methodologies, including statistical tests, as well as the process that is followed to ultimately identify preferred econometric models.

Within this report, we set out a process and set of proposed model selection criteria that Ofgem could use to go about developing and selecting econometric benchmarking models for GD2. We consider that our proposed approach is largely consistent with recent benchmarking exercises undertaken by Ofgem and other UK regulators, including the recommendations of the Competition and Markets Authority (CMA) during Bristol Water's appeal of its PR14 price determination in the water sector.[4]

Our key proposals are as follows:

- As part of model development, it is important to apply a **clear process that allows the evaluation of the robustness of models**. This was recognised in GD1, with Ofgem noting this helps it avoid, in particular, criticisms that it cherry-picks models or results.

---

best fit is then shifted to the defined efficient frontier for the sector (a 'corrected' form of OLS). The COLS method involves assumptions about the technological properties of the firm's production process and importantly any deviations from the frontier (the adjusted regression line) is assumed to be due to inefficiency. In contrast, other methods such as Stochastic Frontier Analysis (SFA) allow for the residual to be decomposed into inefficiency and measurement error (or 'noise'). See further discussion in CEPA (2003): 'Background to work on Assessing Efficiency for the 2005 Distribution Price Control Review'.

[3] Such as economies of scale, which is of particular importance in regulated network industries.

[4] See CMA (2015): 'Bristol Water plc – A reference under section 12(3)(a) of the Water Industry Act 1991'

- We have proposed six high-level criteria for GD2 model selection: **economic / technical rationale**, **incentive properties**, **consistency with wider GD2 policy**, **data requirements & reliability**, **transparency** and **robustness**.
- Given the relatively small data sample available for econometric benchmarking in GD2, and the CMA's findings during Bristol Water's PR14 appeal, we believe it is particularly important that Ofgem develop a **strong technical and business rationale** and **economic prior assumptions** for models selected for the GD2 cost assessment.
- As well as the ex-ante rationale for the model specification, Ofgem should also carefully consider what **insights the model coefficients themselves provide of the impact of different explanatory variables on the GDNs' costs**.

Selection of sample size

Econometric modelling uses statistical techniques and, therefore, the larger the data set used typically the more robust the estimation of the model is expected to be. The greater the volume of data and observations, for example, the more variables that can in principle be considered within the modelling.

Within the GB gas distribution sector there are four separate Gas Distribution Network (GDN) groups[5] and eight distribution service areas (DSAs), meaning there is a relatively limited cross-sectional data set available for benchmarking. However, through its annual regulatory reporting process, and the business plan data templates (BPDTs) GDNs will be required to submit as part of their price review business plans, Ofgem will have data available on the GDNs costs over a relatively long time period, including both historical data and forecast data, to help inform setting of the cost baselines in GD2.

As Ofgem has adopted at previous price reviews, this would suggest that it should use a panel dataset for its econometric modelling as this will help to increase the available sample size. We **support Ofgem continuing with this approach**, although using a panel data set raises a number of subsequent methodological issues and questions that will need to be considered.[6]

Given the available data set, we suggest that Ofgem:

- Consider econometric models that are estimated **using a range of both historical and forecast data**, as was the case at RIIO-GD1 (GD1).
- Models that use forecast (i.e. business plan) data may provide useful information to inform the GD2 determination, **particularly if the cost structures of the GDNs are expected to change** during the forthcoming regulatory period.
- There are, however, risks with *estimating* econometric models using forecast data, most importantly that it **reduces the independence** of the benchmarking from the GDNs own

[5] Cadent, Scotia Gas Networks (SGN), Wales & West Utilities (WWU) and Northern Gas Networks (NGN).

[6] For example, how to account for time effects in the specification of the model (if at all).

business plans. The longer historical data set available for benchmarking compared to GD1, may also weaken the pure statistical case for using forecast data.

Whilst other regulators have included international comparators in the sample[7] we would not propose that Ofgem include other (international) gas distribution network operators in its modelling:

- In this specific context, rather than helping increase the robustness of the econometric modelling – by increasing the available sample size – we expect this would lead to **less reliable results**, given the potential issues with data comparability.

Estimation technique and efficiency frontier

Based on our assessment of the feasible techniques and the key methodological issues for GD2, we suggest the following practical proposals for the estimation techniques that Ofgem might use for its econometric benchmarking and how to define the efficiency frontier. These follow closely the precedent set by previous Ofgem price controls. This is perhaps unsurprising given the extensive work that has been done by Ofgem and its consultants in the past.

We propose that Ofgem:

- use **COLS** as its primary estimation technique for GD2;
- **test models using Random Effects (RE)** as an alternative estimation technique;
- use the **upper quartile** as a starting point for the efficiency target / frontier for the GDNs; and
- **explores not applying a glide-path** towards the efficiency target.[8]

Given the relatively early stage of the price review process[9], Ofgem may also wish to **investigate the use of SFA as an estimation technique** in GD2. However, we consider:

- it should in the first instance be used as a **cross-check / point of comparison** to efficiency scores produced using the estimation technique adopted at GD1 (i.e. COLS)[10];
- the emphasis should be placed on demonstrating that an SFA approach leads to **considerably more reliable results** than Ofgem's existing (relatively simple and replicable) benchmarking approach, even though there are well known limitations with COLS; and
- any SFA model would need to be subject to considerable **robustness assessment and sensitivity testing**.[11]

[7] For example, the Australian Energy Regulator (AER) has considered models that use companies from Ontario and New Zealand in its econometric benchmarking of electricity distributors.

[8] For GD1, although Ofgem used the upper quartile benchmark as the efficiency frontier, it also decided that the GDNs would only need to close 75% of the gap to the upper quartile during the price control.

[9] Although we note that the GDNs' first draft of GD2 business plans is due in the next few weeks.

[10] For example, does the upper quartile efficiency frontier set a relatively challenging efficiency target when compared to the results of SFA modelling?

[11] Particularly as regards the imposed assumptions used for the distribution of the inefficiency component, given the lack of precedent of use within previous regulatory proceedings undertaken by Ofgem.

Cost aggregation

Any benchmarking model is expected to score strongly on some criteria for model selection and less strongly on others. This reflects trade-offs across the many different decisions that need to be made in the model selection process, but particularly the approach to the aggregation of costs in the benchmarking.

As was the case at GD1, we recommend that:

- Ofgem adopt **a tool kit approach to its cost assessment** in GD2. Ofgem's benchmarking should include a range of aggregated and more disaggregated cost assessment models and cost pooling methods, rather than relying on a single model or level of cost aggregation.

Building on the high-level model selection criteria set out above, we propose that Ofgem should use a **set of principles for selecting cost pools** for more aggregative benchmarking:

- This should include factors such as the complementarity, impacts on incentives and risk of biased / inaccurate regression models of different types of costs being grouped together for benchmarking purposes, as well as the complexity of drawing cost boundaries between activities.

With this in mind, we propose that Ofgem consider developing data / statistical tests to identify how complementary the different types of expenditure and their expected explanatory variables are for benchmarking in GD2. Before grouping costs, this may include:

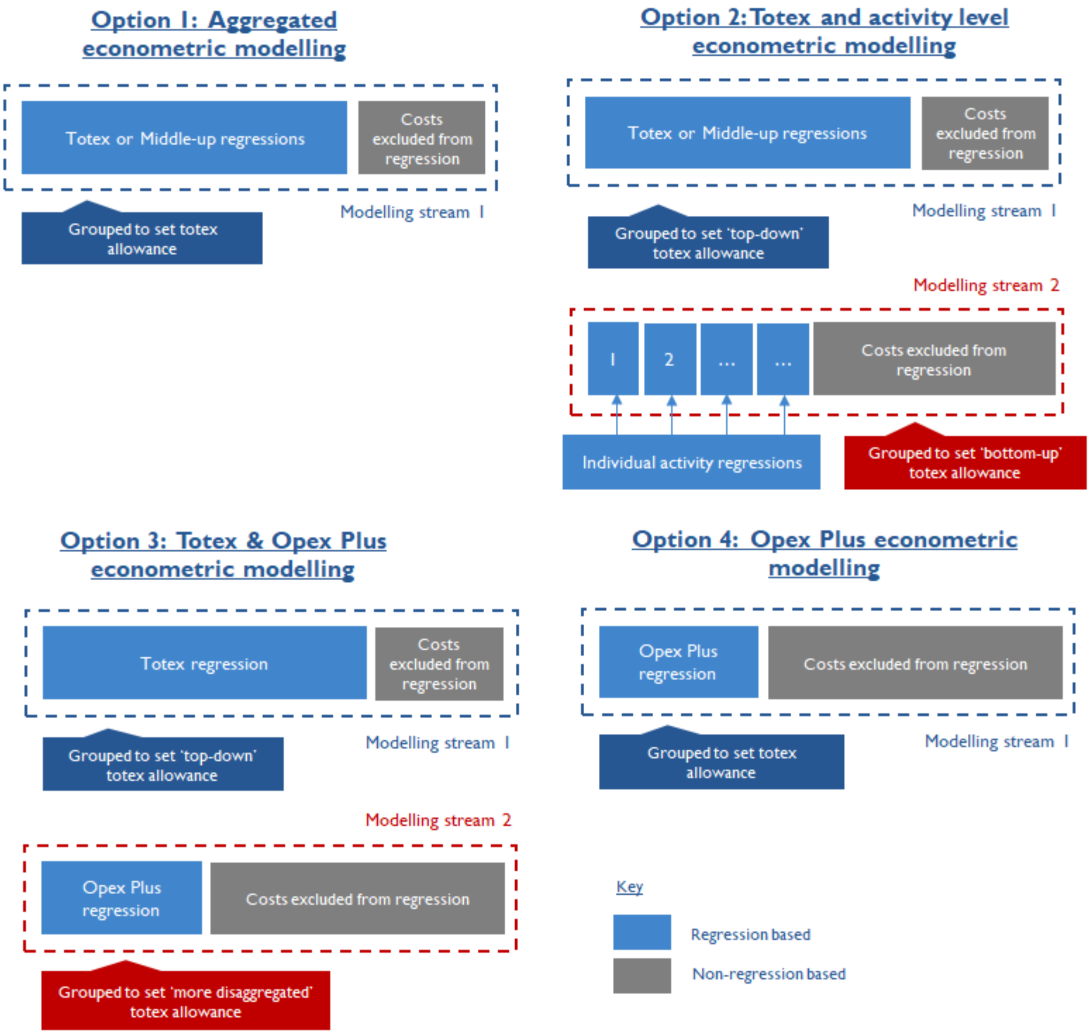
- Testing for year-on-year volatility in expenditure in particular areas / activities that appears to be unrelated and/or correlated with changes in business scale drivers.
- Testing the expected consistency of workload and other cost drivers between different types / areas of spend before costs are grouped together for benchmarking.

Within this report, we set out a number of strawman options for cost pooling, including a continuation of the cost aggregation approach used for GD1, that this decision-making framework could in principle be applied to when developing econometric models for later stages of the GD2 price review. Figure 1 below provides a summary illustration of those options with:

- different levels of cost aggregation used for benchmarking (e.g. totex vs. activity level (e.g. maintenance opex, mains reinforcement capex) regressions); and/or
- alternative approaches for combining the regression and non-regression analysis at different cost aggregation levels to set totex allowances (e.g. top-down vs. bottom-up assessment). [12]

[12] There are of course, many other options that could in principle be considered by Ofgem. We selected the illustrated four options because they cover a broad spectrum of different approaches.

Figure 1 – Illustrative cost pooling options



Source: CEPA analysis

** Activity level regression include emergency Opex, repairs Opex, mains reinforcement Capex, Repex etc.*

We believe that the final approach to cost pooling should be driven by the factors and criteria we have outlined above – i.e. **no final decision on the preferred approach should be made at this stage of the GD2 review**. However, consistent with a number of the recommendations the CMA made on benchmarking during Bristol Water's PR14 appeal, we suggest that:

- Ofgem at least investigate an approach where costs would only be benchmarked using aggregated regressions at what we term an **Opex "Plus" level** – aggregated opex *plus* other activities (e.g. certain capex items) considered complementary in the model specification.

While this approach may lead to greater weight on bottom-up/engineering analysis to set the final totex allowances than was the case at GD1, the potential benefits are:

- aggregative econometric benchmarking models that potentially have a **clearer statistical, economic and engineering logic and fit**; and

- less concern that the variations in capex and repex expenditure patterns, particularly repex, lead to potentially less reliable benchmarking results.

However, we also recognise the positive incentive properties, and benefits for regulatory consistency, in retaining a 'top-down' totex based benchmarking framework, as adopted in GD1, as part of the GD2 toolkit. Totex modelling helps reduce distortions between the treatment of different types of costs and requires less precise boundaries between costs and activities to be identified. Some of the concerns with totex modelling applied in other sectors are, to some extent, also reduced in the GDN context with the use of workload drivers within Ofgem's econometric model specifications (see discussion below).

Therefore, while we consider it sensible to investigate other aggregated specifications, such as the type of Opex Plus model discussed above, we suggest that:

- Ofgem continue to develop a **set of preferred totex models** (i.e. total controllable expenditure subject to similar adjustments that were considered for these aggregative models at GD1) that can be applied and consulted on in later stages of the price review.

Cost drivers and model specification

We expect that **network output, scale** and **operating environment** are amongst the most significant external, i.e. exogenous, drivers of the GDNs relative expenditure as well as their relative (in)efficiency.

These effects can be captured in econometric models using scale variables such as number of customers and MEAV[13] and variables that proxy the operated environment of the GDNs (e.g. density). In the case of operating environment, the effect can also be controlled for in pre or post modelling adjustments.

The age and condition of each GDN's network will also be expected to influence the expenditure of each company at any given point in time, in particular, the requirement for maintenance, enhancement and the large asset replacement investment programmes within the sector. To an extent, Ofgem has controlled for this effect in previous price reviews by including workload variables[14] in its regressions.

However, there are well-known issues with using workload variables in external benchmarking exercises. In particular, workload variables may suffer from endogeneity and managerial incentive problems.[15] However, workload drivers can help control for the effect of asset condition and differences in GDN investment cycles, and, provided forecast workloads are actually delivered during the price control period, should also allow a more accurate predication of future costs once a programme / volume of work is agreed.

We have proposed above that Ofgem should continue to apply a tool-kit approach that includes various alternative models and cost pool aggregations in the cost assessment process. In light of this recommendation, **we would expect Ofgem to also consider a range of alternative econometric model specifications for GD2**, including:

- Aggregative (e.g. totex) models that consider both scale – e.g. MEAV – and disaggregated activity variables – e.g. workload drivers.
- Models that take different approaches to control for time effects on GDN costs (either time trends or time dummy variables (or neither)).

[13] Modern Equivalent Asset Value (MEAV)

[14] For example, the km of iron mains replaced, or km of new pipeline installed in a given year.

[15] Workload drivers are inputs rather than true network outputs.

- Models that account in different ways for the impact of the external operating environment, in particular, regional factors (as discussed below).

Ofgem included Composite Scale Variables (CSVs) in its GD1 regressions when the sample was considered too small to handle multiple drivers and/or when some of the cost drivers were statistically insignificant, but engineering knowledge and industry understanding led it to believe that combining different drivers into one CSV could better account for changes in costs within the sector.

We suggest that Ofgem continue to consider models with CSVs for GD2. However, given some of the disadvantages with their use – e.g. that they place constraints on the underlying cost drivers (i.e. fixed weights) and can make the interpretation of the model coefficients more challenging particularly when combining scale and workload drivers in the composite variable – we would suggest that Ofgem also consider models that exclude CSVs. This will mean the expected sign and size of coefficient of different variables and model specifications can be logically tested during the model development process.

Regional / company specific factors

There are three main approaches that Ofgem could in principle use to account for regional and other company-specific factors in econometric modelling.

- **Pre-modelling adjustment**: adjustments are made to the cost data before estimating econometric cost models in an attempt to make data more comparable between companies. This approach was taken by Ofgem at GD1.
- **Within model adjustment**: as noted above, regional factors can be captured directly through explanatory variables within the econometric models. For example, a measure of the density of the GDNs license areas might be included as an explanatory variable in the model specification.
- **Post-modelling adjustment:** predicted costs derived from the econometric models are adjusted through a Special Cost Factor (SCF) process for regional factors considered to have not been sufficiently captured through a pre-modelling adjustment or within model explanatory variable.[16]

We have examined the advantages and disadvantages of each approach and proposed a decision procedure to help frame the choice of which to apply. Our main conclusions are:

- In the first instance, key factors should be included within the econometric modelling. This allows their effect to be statistically tested. However, a number of practical considerations mean this is not always feasible, for example, because of the relatively small data set available.
- A key advantage of pre-modelling adjustment as an alternative approach is that it provides a clear, conceptually simple estimate for the impact of the given factor. However, the choice and structure of adjustments can be argued to be somewhat arbitrary.
- A key advantage of post-modelling adjustment is that it can help reveal accurate information, as companies are required to pass a high evidential bar to justify changes, while it also recognises that econometric modelling will never perfectly capture all cost drivers.

[16] This approach has been used by Ofwat in recent price controls.

The choice of the most appropriate approach we would suggest is more of a modelling / statistical issue as opposed to a policy one. Each approach is attempting to achieve the same policy objective.

Therefore, as a general principle, **we would recommend that Ofgem explore a range of options** for the treatment of regional factors. These should be tested against prior expectations and the performance of models compared against each other before a final decision on the best approach is taken.

In particular, we believe the treatment of both **regional wages and density could in principle be explored further as within model adjustments**. However, due to regulatory precedent, and issues with including these factors within-model explored in the main report, the feasibility of using this approach will need to be demonstrated through a robust model development and selection process. While Ofgem may wish to explore alternative approaches at GD2, our initial expectation would be that a continuation of the approach that was taken at GD1 will likely remain appropriate.

1. INTRODUCTION

Ofgem's Gas Distribution (GD) cost assessment team commissioned Cambridge Economic Policy Associates (CEPA), in association with Economic Consulting Associates (ECA), to provide advice on the cost assessment process for RIIO-GD2 (GD2) and RIIO-2 more generally.

CEPA and ECA have prepared three briefing papers for publication alongside Ofgem's consultation paper on cost assessment. The topic of each paper is as follows:

- econometrics and regional factors (prepared by CEPA).
- business support costs (prepared by ECA).
- frontier shift (prepared by CEPA).

This paper discusses econometric methods and regional factors. This includes:

- the approach Ofgem might use to reach decisions on the econometric models it selects in GD2 and the options for those models; and
- how to account for regional factors – i.e. regional differences in the external operating environment of each GDN that effect their relative cost performance but are outside of their control.

The remainder of this introductory section sets out the context to the cost assessment in GD2.

1.1. OFGEM'S APPROACH IN GD1

Economic benchmarking was a major part of the toolkit of the methodologies Ofgem used at RIIO-GD1 (GD1) to assess the GDNs historical and forecast relative efficiency – i.e. the gap to the best performing company(s) within the sector (sometimes referred to as 'catch-up' efficiency).

Ofgem developed a range of econometric models to test the efficiency of the Gas Distribution Networks' (GDNs) expenditure based on different levels of cost aggregation, time periods of data (historic and forecast) and alternative model specifications (e.g. number and range of factors (explanatory variables) used to explain costs). For GD1, Ofgem used panel time-fixed effects models estimated using the Corrected Ordinary Least Squares (COLS)[17] technique and Cobb-Douglas functional form.

The Final Proposals for GD1[18] combined two cost assessment modelling streams (top-down and bottom-up) to set the GDNs allowances:

- **Top-down benchmarking approach:** The top-down modelling stream consisted of a single aggregated pool of costs at a 'totex'[19] level. Items excluded from this top-down model were assessed separately and then added to the results from the regression modelling.

[17] In simple terms, the efficiency frontier under COLS is estimated using the ordinary least squares (OLS) technique. A functional form for the production / cost function is specified, and this is estimated using OLS. The calculated line of best fit is then shifted to the defined efficient frontier for the sector (a 'corrected' form of OLS). We discuss in further detail estimation techniques for econometric benchmarking in later sections of the report.

[18] Ofgem (2012): 'RIIO-GD1: Final Proposals – Supporting document – Cost Efficiency'

[19] Controllable opex plus capex and repex.

- **Bottom-up benchmarking approach:** The bottom-up modelling stream combined seven activity level regressions, including disaggregated opex, capex and repex activity models, with thirteen non-regression disaggregated cost activities.

For both benchmarking approaches, efficient costs were defined as the upper quartile level of efficiency. In the case of the top-down modelling stream, this was defined using the COLS technique.

In the case of the bottom-up modelling workstream, Ofgem first added the costs assessed at the each of the regression cost activities and non-regression cost activities. It then benchmarked the GDNs aggregated costs at the upper quartile level consistent with the top-down modelling stream. This was to avoid cherry picking when combining the results of benchmarking at a more disaggregated level.

For GD1, Ofgem accounted for regional factors by generally adjusting the cost data *before* estimating econometric cost models in an attempt to make data more comparable between companies. Ofgem adjusted for three regional factors based on its analysis, including for regional labour, urbanity and sparsity.

Ofgem also made several company specific adjustments largely based on information provided by GDNs and supported by independent data. This included adjustments to NGN's opex for a salt cavity, Scottish Independent undertaking, and London's medium pressure undertaking.

We discuss the approach that Ofgem used to account for regional factors in GD1 in further detail, along with other elements of the GD1 modelling methodology, in later sections of the report.

1.2. CONTEXT TO GD2

Ofgem has set out in its RIIO-2 Sector Framework Decision that it intends to use the GD1 cost assessment approach as a starting point for GD2. While Ofgem has stated that it will consider areas where there may be scope for improvement, or opportunities to explore alternative approaches, the GD2 cost assessment will be an evolution of the GD1 approach, including the econometric modelling.[20]

This means that the GD1 determination provides important regulatory context to GD2.

As part of its RIIO-2 Sector Framework Decision and Consultation process, and the Cost Assessment Working Group (CAWG)[21] helping to inform Ofgem's GD2 cost assessment, Ofgem has started to consider some of the areas it may look to adapt or change during the later stages of the price review. However, a comprehensive view of the overall approach to cost assessment, including the approach to econometric modelling, will only be possible once GDNs business plans are submitted.

There are a number of high-level issues specific to the GD sector that need to be considered while engaging in cost assessment. Any approach and resulting econometric models need to appropriately reflect recent (and potential future) developments in the industry in order to produce accurate results. This is particularly important if the approach for GD2 will seek to evolve that used in GD1 and where the GDNs expect a change in the scale and scope of their operations in GD2 (e.g. due to decarbonisation).

---

[20] Ofgem (2019): 'RIIO-2 Sector Specific Methodology Decision – Gas Distribution', p. 90

[21] https://www.ofgem.gov.uk/publications-and-updates/riio-gd2-working-groups

Of particular importance is the **available data** set. The final models used in the GD1 determinations employed either four years of historical data or two years of forecasts. A number of years have passed since then and so Ofgem now has a larger set of available outturn data, as well as several years of forecast data for GD2 (once business plans are received). This may afford Ofgem a greater degree of freedom in the structure of the models used. However, it should be noted the dataset have only extended in one dimension. While there are more *years* of data, the number of *companies* has remained the same. As was the case at GD1, the available data set may restrict the econometric models Ofgem can appropriately employ in GD2, given the limited *between-group* variation. [22]

1.3. REPORT STRUCTURE

The remainder of the report is structured as follows:

- Section 2 considers how Ofgem might go about evaluating the strengths and weaknesses of different econometric benchmarking techniques and selecting models for the GD2 assessment.
- Section 3 discusses the issues associated with selecting a sample for the benchmarking analysis given the issues around data discussed above.
- Section 4 reviews different estimation techniques Ofgem might consider for GD2 and the issues they raise for the benchmarking process.
- Section 5 sets out a number of different options, including the approach adopted at GD1, that Ofgem might consider for pooling costs in the econometric benchmarking.
- Section 6 reviews various topics related to econometric model specification, in particular, possible cost drivers / explanatory variables for use in the modelling.
- Section 7 considers different options for the treatment of regional factors and issues Ofgem would need to consider if investigating the use of these options.
- Section 8 provides a brief conclusion to the report.

Three appendices provide supporting material:

- Appendix A provides a brief summary of regulatory precedent of the cost assessment process in other recent UK price reviews (including GD1, RIIO-ED1 (ED1) and PR19[23]).
- Appendix B provides a more detailed tabular presentation of the options discussed in Section 5 for aggregating costs in the econometric benchmarking.
- Appendix C summarises results of illustrative regressions developed during the process of this study to investigate some of the options discussed in the report.

The models presented in Appendix C should be considered as illustrative only as they have not been subject to a robust model selection and diagnostic process.

[22] In this case, between-group variation relates to the dispersion of the data points between different GDNs in a given year. Within-group variation relates to the differences for a single GDN over time.

[23] The price review currently being undertaken by Ofwat for water and wastewater companies in England & Wales.

In addition, we note that we have developed these illustrative models using input data provided to us by Ofgem on the GDNs costs and cost drivers. We have not sought to verify this data or investigate its accuracy or suitability for cost benchmarking purposes.

2. MODEL SELECTION METHODOLOGY

In this section we consider how Ofgem might go about evaluating the strengths and weakness of different benchmarking techniques and selecting an approach to efficiency modelling at GD2.

The model selection methodology includes key issues such as the criteria that are used to assess different benchmarking methodologies, including statistical tests, as well as the process that is followed to ultimately identify preferred econometric models for setting cost baselines.

In the subsections below, we consider the approach that was applied by Ofgem in GD1 and set out a proposed methodology for the upcoming price review based on what has been set out in the GD2 consultations and CEPA's own modelling approach guidelines.

2.1. MODEL SELECTION CRITERIA

As part of econometric model development, it is important to set out a clear set of criteria that allows the evaluation of the robustness and suitability of models and different possible benchmarking techniques for the efficiency analysis. This was recognised in the process followed at GD1, with Ofgem noting this helps it avoid, in particular, criticisms that it cherry-picks models or results. A clear set of evaluation criteria help to demonstrate that the results of the modelling can be considered suitably robust and valid for the purposes of informing or setting cost baselines at the time of the price review.

In the subsections below, we review the model evaluation criteria used by Ofgem at GD1 and discuss the model selection criteria that we propose Ofgem consider for GD2.

2.1.1. Ofgem's GD1 model selection criteria

Ofgem set out at the Initial Proposals stage of the price review a detailed step-by-step guide of the cost efficiency assessment methodology it followed to determine cost baselines for GD1.[24] This included both regression (econometric based) and qualitative benchmarking methodologies.

The process that Ofgem followed to complete the regression-based component of the assessment comprised the application of various data normalisations and adjustments before modelling was undertaken and a set of criteria to evaluate the GD1 regression models. The evaluation criteria included:

- Identifying and selecting cost drivers using engineering knowledge.
- Using cost drivers within the regression models that are outside of the control of the GDNs.
- Using reliable data.
- Including key cost drivers.
- Available sample size and how this may impact the specification of the regression model.
- Using models with sensible modelled outcomes.
- Using a rational function form of the model.

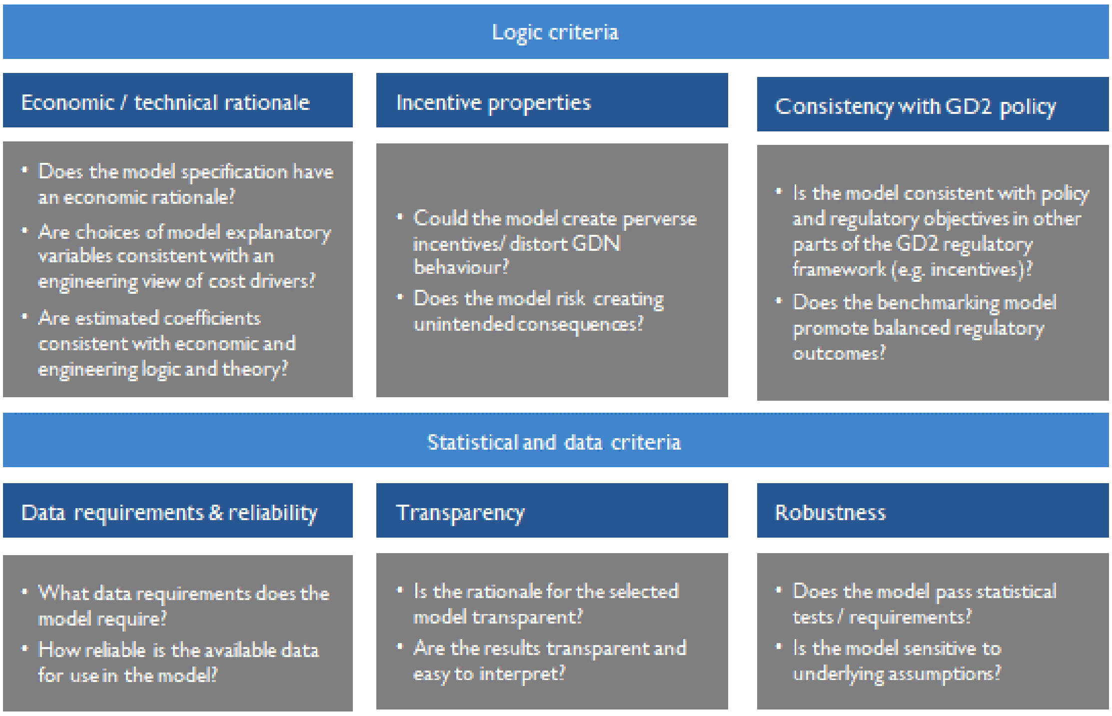
[24] Ofgem (2011): 'RIIO-GD1: Initial proposals – Step-by-step guide for the cost efficiency assessment methodology'

- Good statistical fit of the model and use of statistical tests to test for model robustness.

2.1.2. Proposed model selection criteria for GD2

We propose that Ofgem apply a set of model selection criteria broadly consistent with the approach it followed in GD1, i.e. a set of criteria that test the logic, reliability, transparency and robustness of different econometric models. The high-level criteria we propose that Ofgem consider for evaluating its econometric models in GD2 are summarised, in no particular order of priority, in Figure 2.1 below. [25]

Figure 2.1: Proposed model selection criteria



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

2.1.3. Technical and economic rationale

When developing econometric benchmarking models, a particularly important criterion is the economic and technical rationale for the models, both the specification and the results.

This includes consideration of questions such as:

- Are the selected explanatory variables in the models in line with a priori expectations of what should be important cost drivers?
- Are the estimated model coefficients consistent with a priori expectations in terms of magnitude and sign?

---

[25] A similar set of criteria are proposed in our work for Ofwat on efficiency analysis for PR19. CEPA (2018), *'PR19 Econometric Benchmarking Models'*, available here. We use an alternative graphical presentation in that case but consider the overall criteria to be the same as above (incentive properties and consistency with wider regulatory policy are for example, grouped under the high-level criteria of economic and technical rationale).

- Similar to the criteria applied at GD1, do the models lead to sensible outcomes for setting cost baselines?

Establishing a clear economic and engineering motivation for the key drivers and specification of the functional form of the benchmarking model is particularly important in the regulatory context of GD2 where Ofgem is seeking to assess the efficiency of the GDNs cost performance. A clear narrative of how selected models are consistent with the expected economic and technical drivers of operating and capital expenditure within the sector helps to build credibility in the benchmarking process and its results.

A strong technical and economic rationale for the selected models also helps supports:

- Better interpretation of the modelling results and what they imply (statistically) as regards the relationship between expenditure and cost drivers within the sector.
- Confidence in the reliability of predicted and forecast costs for the price control, particularly where only a relatively limited data set is available for benchmarking purposes.[26]

The importance of establishing the technical and economic rationale for selected econometric models, the explanatory variables they include and the assumptions that they may impose on the benchmarking process, was emphasised in the recommendations of the CMA during the Bristol Water appeal.[27] It is an approach that Ofwat has sought to apply in the development of its econometric models for PR19[28] and was stated as criterion applied by Ofgem in the cost driver selection and model specification process for GD1.

2.1.4. Incentive properties

The incentives a benchmarking methodology may create for companies and their management is also an important issue to consider in the model selection.

Certain model specifications may reward/penalise companies for behaviour that is within their control (e.g. how they choose to operate or manage their network) or may lead to perverse incentives to adopt a particular business model / behaviour which is not necessarily consistent with expected overall operational efficiency. This is one of the reasons to develop models with exogenous variables.

Certain approaches to econometric benchmarking may also have stronger overall efficiency properties when GDNs choose to incur different types of expenditure during the price control period, or when forecasting their expected costs during the GD2 business plan development process.

For example, more aggregated (e.g. totex level) benchmarking models, at least in principle, may help to minimise distortions for GDNs to prefer one type of expenditure over another and, therefore, incentivise the companies to seek out the most efficient approach to delivery. In contrast, even with the presence of a totex incentive regime, as will be the case with RIIO-2, separately benchmarking different types of controllable expenditure may lead to the networks preferring one type of expenditure over another type, because a more or less stringent benchmarking approach is applied to certain types of costs.

---

[26] See below and Vivid Economics / Arup (2017): 'Understanding the exogenous drivers of wholesale wastewater costs in England & Wales' for further discussion of the issues raised by small data sets.

[27] CMA (2015), *'Bristol Water plc. A reference under section 12(3)(a) of the Water Industry Act 1991'*, available here.

[28] See Ofwat (2019): 'PR19 – Supplementary technical appendix: econometric approach'

We discuss in further detail the advantages and disadvantages of different approaches to cost pooling in the econometric modelling in Section 4.

2.1.5. Consistency with wider GD2 regulatory policy and objectives

We would expect Ofgem to also consider whether the selected econometric models are consistent with wider policy (e.g. related to outputs incentives) and objectives in other areas of the price control.

For example, models including the degree of shrinkage or customer interruptions could provide a good statistical fit for GDN's expenditure. However, if their use results in such explanatory variables being positively associated with allowed revenues, this could provide a perverse incentive to companies. Interactions with the GD2 incentive regime must, therefore, be considered.

The selected benchmarking models and techniques should, therefore, support the wider regulatory framework (e.g. on outputs and incentives) Ofgem are developing for GD2. The benchmarking approach should seek to encourage balanced delivery and behaviour from the GDNs to:

- on the hand, explore opportunities for innovation and cost efficiency savings during the price control and business plan preparation process; and
- on the other hand, deliver against the expected outcomes and price control outputs / deliverables specified for the regulatory period.

2.1.6. Data requirements

The economic literature has proposed multiple methods for the estimation of cost assessment econometric models – commonly discussed methods include OLS, Random Effects (RE), Fixed Effects (FE) or Stochastic Frontier Analysis (SFA). Each of these methods introduces different assumptions, in particular about the composition of the error terms. This in turn has implications on the nature of company (in)efficiency.

We provide more detail and a discussion of the relative strengths and weaknesses of a number of these methods in Section 4. However, as well as the theoretical advantages and disadvantages of these models, the appropriate choice is also driven by practical concerns.

Small datasets such as the one available for GDNs will limit the degree to which advanced econometric methods and functional forms can be employed. As discussed in the introduction, while the length of the dataset has clearly increased since GD1, the number of comparators has not changed (i.e., there are still only eight licensees in four groups). This means that while the number of useable observations has increased, the overall 'between' variance (i.e., the relative performance between the GDNs) has not been enhanced to the same degree. This is important as the additional time series observations do not necessarily allow for additional explanatory variables to be included in the model.[29]

Therefore, although Ofgem will have available additional (multiple) years of data (i.e. cross-sections) compared to when it completed the GD1 price review, there are still likely to be limitations on the benchmarking framework which may reduce the confidence in certain benchmarking techniques.

[29] This issue has been recognised by Ofgem during its previous price control determinations during which pooled, or panel data has been used.

We discuss data selection issues further in Section 3 when we consider how Ofgem may go about selecting a relevant sample period for its GD2 models.

2.1.7. Transparency

Transparency is an important regulatory principle in all parts of the price control review process. It are particularly important in the context of the cost assessment process, given selected benchmarking models are used to directly set, or help inform, cost baselines and potentially introduce reductions in expenditure compared to network operators forecasts for the purpose of setting allowed revenues.

As illustrated in Figure 2.1, two key components of the transparency of the benchmarking methodology is the clarity of the rationale provided for why one model is preferred over another – e.g. based on the technical and economic criterion described above – and the transparency and clarity of understanding of the results from the selected models. Transparent models are an important part of demonstrating and testing the overall validity of the results of benchmarking analysis.

It is likely that simpler econometric models are more likely to lead to transparent models. However, it is quite possible that more complex estimation techniques and model specifications may provide a better overall fit for the data and/or have a stronger economic and technical logic. This highlights that there are likely to be natural tensions and trade-offs between criteria in the model selection process, with some models expected to perform better under some criteria, and less well under others.

In general, the transparency of the benchmarking process is also promoted by providing a clear description and rationale of the data used and any adjustments made to the data ahead of conducting the analysis and using a standard and readily available statistical package to conduct the modelling.

2.1.8. Robustness

There are many components and attributes that contribute to whether one set of econometric models might be considered more robust than an alternative set. Evaluation of econometric models against the criteria already listed above will lead to more robust models, e.g. by helping to ensure that the models have a strong technical and economic rationale and, prima facie, will be expected to lead to reliable predictions of GDNs relative efficiency and conclusions of predicted costs for the forthcoming price control.

As well as the types of logic criterion described above, the robustness of the models can also be evaluated using a series of statistical tests. Ofgem has at previous price reviews, including GD1, used a number of statistical tests to evaluate the robustness of its models. We have set these out in Table 2.1 below.

Table 2.1: Ofgem's statistical tests for GD1 and ED1 models

Test	Description	GD1	ED1
Ramsey RESET test for omitted non-linearities	Tests whether there are any omitted non-linearities in the model. It can assist in choosing between Cobb-Douglas and other function forms (see further discussion in Section 6).	✓	✓
F-test for parameter stability (Chow test)	For each model, this test helps to evaluate whether the true coefficients of a pooled OLS model are significantly different from the true coefficients of the same model run on each individual cross-section of the data. If this test fails (p-value of less than 0.01) this provides evidence that panel data analysis may not be appropriate.	✓	
White test for heteroscedasticity	Tests whether the error variance is constant across observations. To account for the fact that the variance between observations coming from one company and those coming from different companies could be different, cluster robust standard errors can be used in this test where appropriate.	✓	✓
Test for outliers	Tests whether observations are outliers.	✓	
Jarque-Bera test for normality/ Skewness and Kurtosis test (SKtest)	The test evaluates statistical significance based on the assumption that residuals follow a normal distribution. This is more important in small samples because we cannot use central limit theorem, which states that for large sample sizes the sampling distribution of the estimator converges to normality.	✓	✓
Panel robust standard errors	Clustered robust standard errors to allow for the fact that the set of observations in the panel are not independent but clustered by GDN. Not a test per se, but allows for appropriate testing of statistical significance.		✓

Source: CEPA

For GD2, while we do not consider that there are strong reasons to change the range of statistical tests Ofgem has previously used for assessing the models, we do believe it is useful to explicitly set out a full process by which models could be selected. For example, there are some additional statistical tests that should be set out explicitly:

- the statistical significance of the coefficients on individual parameters;
- correlations between cost drivers/ explanatory variable where more than one are used;
- goodness-of-fit statistics (where applicable);
- stability of the model (e.g. how robust are the models to changes in the underlying assumptions and data); and
- predictive power of the models.

2.1.9. Summary

In Table 2.2 below we set out our expectation of the relative importance of each of the core statistical tests and also the broader high-level criteria discussed in proceeding subsections.

We have grouped the model selection tests and the criteria by level of importance:

- **Very high-level importance.** Criteria / tests that when failed, would disqualify the model automatically from the benchmarking process.

- **High.** Tests and criteria that when failed would raise serious concerns about using the model, either to inform or set directly cost baselines in GD2.
- **Medium**. Tests and criteria that when failed raise some concerns about using the model in GD2 but the model could be used with caution if it passes other tests.
- **Low**. Tests and criteria that when failed would raise relatively limited concerns about using the model to inform or set GD2 cost baselines.

Table 2.2: Model selection criteria and tests summary

Level of importance	Model robustness test
Very High	• Clear economic and technical narrative / rationale for selected model • Jointly statistically significant (F-test) • Overall goodness of fit (e.g. R^2) • Consistency with policy in other parts of the price control • Expected to generate appropriate incentives for company and management behaviour • Approach is consistent / feasible given available data and quality of data
High	• Consistency with a priori expectations of magnitude and signs of estimated coefficients • Stability of efficiency rankings • Stability of inefficiency range • Transparency of results and ease of interpretation • Omitted non-linearities (e.g. RESET test)
Medium	• Sensitivity to: ○ removal or addition of a year ○ the removal of the most or least efficient company ○ introduction of quadratic terms • Statistical significance of individual parameters (t-test) • Pooling test • Within-sample forecasting power
Low	• Multicollinearity tests • Homoscedasticity • Normality • Test of pooled OLS versus RE (Breusch-Pagan test) • Hausman test for fixed effects

Source: CEPA

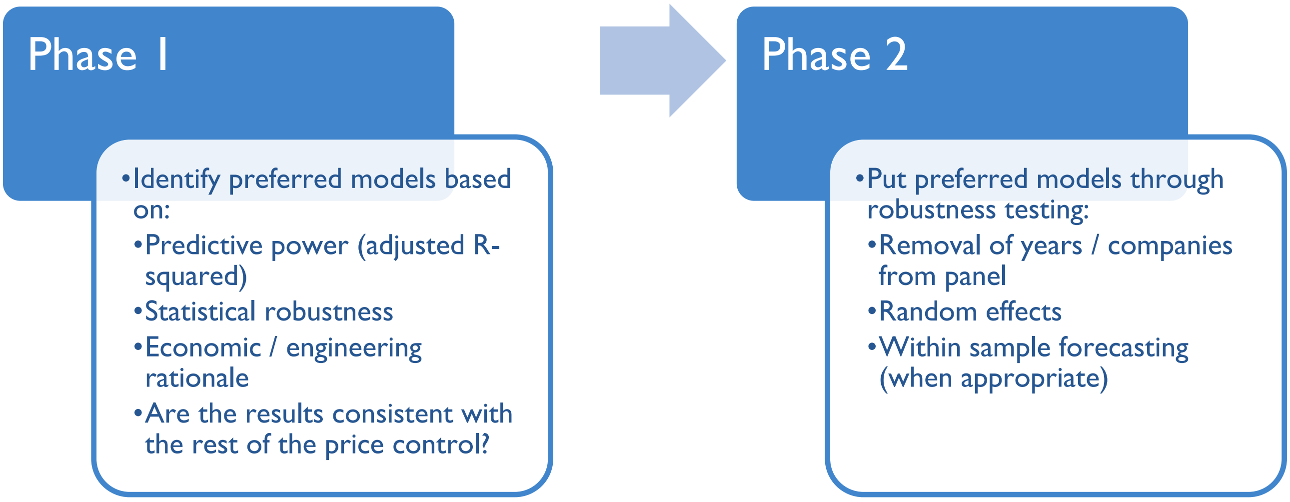
Ideally, the final models selected would pass all model evaluation criteria and tests they are submitted to. However, setting such a high standard could make it very difficult to develop any models at all. Passing all statistical and logic tests is very challenging in applied work.

As a result, it is important to understand what a model failing a test means for its potential use in GD2. As discussed above, trade-offs between test results and evaluation criteria are an inherent part of model development, meaning that a failure of one test may not necessarily result in the rejection of the model. Nevertheless, if we consider there are significant concerns which mean a particular model is not robust, we would go back through our iterative process and consider model alterations.

2.2. Process of model development

In order to keep the modelling process manageable, we typically recommend carrying it out in two phases, with the stringency of assessment increasing as models pass through various criteria. This process is set out in Figure 2.2. below and was the approach CEPA followed to develop models for Ofwat for PR19.[30] We suggest Ofgem investigate following a similar approach to identify a preferred set of models for GD2.

Figure 2.2: Model development phases



Source: CEPA

Following this process, in the first phase, models would be selected from a comprehensive set that meet the minimum characteristics required for a model to be considered further. For example, in CEPA's recent work for Ofwat as part of PR19, the models selected at the end of Phase 1 were those for which:

- all variables were individually significant at a 10% confidence level;
- no two variables included in a model were correlated by more than 90%;
- all coefficients were consistent with prior expectations based on economic and engineering rationale;
- the adjusted R^2 was higher than 80%[31]; and
- the coefficients were consistent with Ofwat's incentives for PR19 (e.g. models where leakage would grant higher allowance to companies would be excluded).

In the second phase, those models that were selected in Phase 1 would be evaluated further by running the types of robustness (statistical) tests discussed above.

---

[30] CEPA (2018), *'PR19 Econometric Benchmarking Models'*

[31] A target R^2 of 80% may not be appropriate in the context of GD2 due to the relative lack of available cost driver data compared to Ofwat.

2.3. Conclusions

In this section, we have set out a process and set of proposed model selection criteria that Ofgem could use to go about developing and selecting econometric benchmarking models for GD2.

Our key proposals are as follows:

- As part of model development, it is important to apply a **clear process that allows the evaluation of the robustness of models**. This was recognised in GD1, with Ofgem noting this helps it avoid, in particular, criticisms that it cherry-picks models or results.
- We have proposed six high-level criteria for GD2 model selection: **economic / technical rationale**, **incentive properties**, **consistency with wider GD2 policy**, **data requirements & reliability**, **transparency** and **robustness**.
- Given the relatively small data sample available for benchmarking in GD2, and the CMA's findings during Bristol Water's PR14 appeal, we believe it is particularly important that Ofgem develop a **strong technical and business rationale** and **economic prior assumptions** for models selected for the GD2 cost assessment.
- As well as the ex-ante rationale for the model specification, Ofgem should also carefully consider what **insights the model coefficients themselves provide of the impact of different explanatory variables on the GDNs' costs**.

The model selection criteria set out and discussed above have informed our own assessment of possible candidate benchmarking methodologies and models that Ofgem could explore for GD2, as discussed in subsequent sections of the report.

3. SELECTION OF THE SAMPLE

In this section we consider the selection of the sample / data that is used for the econometric modelling and some of the methodological issues this raises for benchmarking in GD2. The subsections below:

- Review some of the high-level conceptual questions and issues typically considered in selecting a sample for econometric benchmarking (Section 3.1).
- Discuss certain specific methodological issues in sample selection that arise from the high-level questions identified previously (Section 3.2, 3.3 and 3.4).
- Present our conclusions on sample selection for GD2, recognising the relatively early stage of the GD2 review process, based on discussion in proceeding subsections (Section 3.5).

3.1. SAMPLE SELECTION ISSUES

Econometric modelling uses statistical techniques and, therefore, the larger the data set used, typically the more robust the estimation of the model is expected to be. The greater the volume of data and observations available the more variables that can in principle be considered within the modelling.

As discussed in previous sections of the report, within the GB gas distribution sector there are four separate GDN groups[32] and eight distribution service areas (DSAs), meaning there is a relatively limited cross-sectional data set available for benchmarking. However, through its annual regulatory reporting process, and Business Plan Data Templates (BPDTs) that GDNs will be required to submit as part of their GD2 business plan submission, Ofgem will have data available on GDNs costs over a relatively significant time period, including historical data and forecast data, to help inform setting the cost baselines in GD2.

As Ofgem has adopted at previous price reviews, this would suggest that it should use a panel dataset for its econometric modelling as this will help to increase the available sample size. We support Ofgem continuing with this approach, although as discussed in later parts of the report, using a panel raises a number of subsequent methodological issues that then need to be considered (see Section 4 and 6).

In addition, there may be good reasons why all the GDNs' cost data available for benchmarking review may not be used in its econometric modelling. For example, there may be inconsistencies in the reporting of data across years, or other reasons why only a part of the available data set is preferred or used. We discuss these issues in further detail in Section 3.2 below.

At GD1, Ofgem also used a consistent panel / sample period for all its econometric models – i.e. across top-down, middle-up and disaggregated activity level regressions. At least in principle, there is the option of considering using different time periods for different models. We consider this issue in Section 3.3.

Finally, in addition to the selection of the time period of the sample, in any benchmarking study there is also the possibility of expanding the sample to include international data (expand the company cross-section). Although for distribution networks, Ofgem has typically not sought to used international comparators in its econometric modelling (e.g. other European gas distribution operators), regulators in some other jurisdictions have attempted this. We briefly consider this issue in Section 3.4.

---

[32] Cadent, SGN, Wales & West Utilities and Northern Gas Networks.

3.2. Selection of sample period

As discussed above, for GD2, Ofgem, once the GDNs' business plan templates are submitted, will have access to a much longer dataset than it had for previous price reviews in the sector. This includes

- five years of historical GDPCR1 data;
- five years of historical GD1;
- three years of GD1 forecasts; and
- five years of GD2 forecasts.

Historical data provides information on how GDNs *actually* performed while the forecast information provides information on *GDNs' views* on how costs and workloads will develop over GD2. However, as the latter data is forecast it is subject to uncertainty and is based on GDNs' managements' view of its future expenditure requirements.

At the GD1 price review, Ofgem estimated its models on two different periods: (1) four-years of historical data; and (2) two-years of forecast data. Ofgem stated that it did not use the eight years of forecast data as "*the models based on the full eight years forecast period fail[ed] [its] model specification tests.*"[33] We are not aware of Ofgem testing a combination of the historical and forecast data. We note that for ED1, Ofgem used models that combined the historical and the forecast data.

As discussed above, typically, a statistical / econometric modeller will seek to include as much accurate data in their models as is available. However, the use of an extended dataset needs to be carefully considered as we are dealing with expenditure data that can include:

- Allocations of costs that may change over time. For example, in GDPCR repex was not split into the different tiers.
- A *structural break(s)* that means there is a change within the time series of historical data or that historical data is no longer a good indicator of future performance.[34]
- The GDNs' own assumptions of how their costs will develop over GD2.

We discuss each of these issues in the subsections below.

[33] Ofgem (2012), page 7. Ofgem also noted that it used 2-years of forecast data because the immediate forecasts for the first two years of GD1 were considered more robust than the latter years because of the compounding effect across the sector from a wide range of asset health and work volume assumptions.

[34] The latter for example, could include specific (material) costs that may exist in the historical data that do not exist in the forecast data and/or vice-versa, and changes in GDNs' expenditure relative to their workload.

3.2.1. Allocation of costs

In selecting a time period for the sample in GD2 models, this should only impact disaggregated models.[35] Unless the change in cost allocation methodology can be applied to the historical data (or the forecast data), there is not a lot that can be done about it.

In developing its models for GD2, Ofgem should check whether there have been any changes in allocations[36]. We are not aware of any issues aside from a lack of repex tier. If there has then the modelling should be done using the cost allocation that best fits the forecast costs, or the expenditure should be aggregated up to a point where it is consistent across time.

3.2.2. Structural break

A structural break occurs when the parameters in a time series change abruptly. This could be a sudden change in the fixed element of a particular activity (e.g., the fixed cost captured in the 'constant')[37] and/ or it could be from a change in the relationship between expenditure and the cost drivers/ explanatory variables (i.e., a change in the coefficient of the cost drivers/ explanatory variables). An example of the former could be a similar fall in a common (non-size related) cost across the GDNs. An example of the latter could be, a unit cost over the first five-year period that averaged £20 across all the GDNs, but in the following year it drops to £10 and averages at this level over the next five years.[38] Structural breaks can often be identified by visual inspection of the data, e.g., using line graphs to identify sudden changes in the data. However, in some cases structural breaks may not be easy to identify visually (e.g. changes in relationship between costs and cost drivers) but they can be statistically tested.[39] We note that, even when a visually structural break is identified it should be verified using the statistical tests. If a structural break is identified it does not necessarily mean that the dataset should be truncated (i.e., selecting the latest period). If the structural break is a shift in constant rather than the coefficient on the cost driver(s)/ explanatory variable(s) then the historical data can continue to be used with a dummy variable used to pick up the differential in the constant across the periods.[40] Similarly, if there is a change in the relationship then another term can be added to the model to pick up the change in the relationship.[41]

[35] Given that by definition, aggregated models at a totex level will include all controllable costs. Of course, even a totex model may still exclude certain cost items and so there will always remain some allocation issues.

[36] For example, different pipe diameter bands reported by individual GDNs across business plan templates for RIIO-GD2 and RRPs.

[37] The constant term in a regression is also commonly referred to as the 'intercept'.

[38] e.g. due to a significant reduction in input prices and/or new working practice.

[39] Statistical structural break tests are well established, and all good econometric programmes will have standard tests included. For example, Stata has a test the identifies structural breaks (sbsingle) and a test for when structural breaks are known, or believed to be known (sbknown).

[40] For example, if appropriately identified as a structural break of this kind, a 'RIIO-1' dummy variable could simply have '0s' prior to the RIIO-1 period and '1s' afterwards.

[41] This is done in the model by adding an additional variable which is the product of the dummy variable and the cost driver(s)/ explanatory variables.

This means that, should Ofgem identify a structural break, so long as there are sufficient degrees of freedom (i.e., relatively few cost driver(s) and explanatory variables in the model) then Ofgem can use the full dataset available in the model. When there are insufficient degrees of freedom, Ofgem should run separate models using historical/ forecast data post any structural break and data pre any structural break. If the models pass the selection criteria[42], then differing results from the models, and time periods of the models, should be considered on the basis set out in the next sub-section.

3.2.3. Using the GDNs' forecast data

At previous RIIO price reviews, both for gas distribution and electricity distribution, Ofgem has relied on forecast data in its econometric modelling.[43]

On the one hand, using GDNs' forecast costs has the benefit of helping to promote innovation from the GDNs by incentivising them to put forward more challenging proposals that may improve their positioning in the efficiency rankings. Including forecast data also helps extend the benchmarking sample.

Forecast costs may also be more informative for estimating regression models in the context of a changing energy system, where the scale and scope of the gas network may change in response to the decarbonisation of UK economy (we discuss this issue in further detail in Section 6 as part of the review of cost drivers). Stated alternatively, these issues may mean that models estimated on historical data do not provide a good benchmark if the GDNs underlying cost structures are expected to change in future.

On the other hand, there are certain risks in using forecast data within the econometric model *estimation* process and reasons why Ofgem might be cautious in using forecast data in its model estimation:

- If Ofgem's models are estimated using forecast data, then the benchmarking process is less independent of the companies' plans.
- Given the increase in the time period of historical data now available compared to GD1, the case for using forecasts for statistical reasons arguably is weakened.[44]
- Although decarbonisation may in future impact on the gas grid[45] it is unclear whether this will significantly change the scale and scope of GDNs operations *in GD2*.

If dealing with results based on forecast data, as compared to historical data, careful consideration needs to be given to questions of whether there is a legitimate reason for structural break/ shift in costs if any are present, and, given the GD2 incentive regime, should this be allowed?

- If the GDNs' forecasts lead to lower overall allowances (i.e. lower cost baselines), then there may be good reasons why this data should be used for setting final cost allowances, as it represents new information the GDNs are providing.

[42] As proposed in Phase 2 of our proposed modelling process set out in Section 2.

[43] In contrast, for PR19 Ofwat has used historical data to estimate the regression model and then uses this to identify predicted costs for the forthcoming price control.

[44] Particularly given some of the other problems with using forecasts as identified above.

[45] See for example Frontier Economics (2016): 'Future Regulation of the Gas Grid – a report for the CCC'

- In contrast, if the GDNs' forecasts indicate that they are seeking higher allowances for the associated workloads, then, unless they can justify the change, Ofgem might wish to rely more on historical as opposed to forecast data.

While it might be argued that taking a different approach to setting cost allowances across different cost pools and circumstances may lead to cherry picking by the regulator, it might also be argued that this approach applies a more consistent methodology where benchmarking is driven by the data and the strength of GDNs' justification for not placing weight on historical (outturn) evidence of GDN costs. We provide further discussion of using different time periods in different models in Section 3.3 below.

3.2.4. Summary

We suggest that Ofgem should consider econometric models **estimated using historical and forecast data** as part of its benchmarking for GD2. However, the issues we have highlighted above in relation to use of GDNs forecast data will need careful consideration.

One of the ways that different sample time periods could sensibly be used within Ofgem's broader cost assessment is to help support its assessment of the confidence it has in GDNs submitted costs, both at an aggregated and disaggregated level. For example, the application of forecasts from models based on historical data, relative to GDNs own forecasts in their GDN business plans, could be used to help establish both the stability of the model estimates themselves, whether or not there appears to be any evidence of significant structural change, or whether Ofgem's models appear a good predictor of costs.

Alternatively, Ofgem could also test the confidence of its models in predicted costs, by splitting historical samples and using estimates on this earlier sample period to predict later historical periods.

This process of considering evidence from models estimated using a range of different sample time periods, could, therefore, be a useful input to the new business plan incentive that Ofgem proposes to apply for RIIO-2 where the sharing factor applied to the totex incentive will be a function of Ofgem's confidence in the costs across different elements of GDNs (and other network operators) plans.[46] The different sample periods used may help with testing the stability and predictive power of Ofgem's econometric models[47] as well as its confidence in the GDNs submitted business plan costs for GD2.

3.3. Using different sample periods

As highlighted above, an additional question to consider is whether the time period chosen for different levels of aggregation of cost modelling necessarily needs to be consistent in all cases.

Would, for example, a scenario where Ofgem chose to benchmark GDNs totex using 15 years of data (historic and forecast data), but developed more disaggregated cost models (e.g. for repex) based only on historical or forecast information (e.g. 5-years), lead to cherry picking or unbalanced results?

Ofgem applied a consistent 4-year historical and 2-year forecast time period for both its aggregated (totex) and disaggregated benchmark models in GD1. Clearly there is a risk that using different time periods for different models could lead to unbalanced efficiency scores and cost allowances.

---

[46] Ofgem (2019): 'RIIO-2 Sector Specific Methodology – Core document'

[47] As proposed under our model selection criteria and tests.

However, using different time periods may not necessarily be cherry picking provided an objective justification – absent of the actual results of its benchmarking – for using different approaches is provided. For example, if a structural break is identified in the data at a more aggregated (e.g. totex) level, this does not necessarily imply the same structural break applies to all levels of GDN costs. There may also be good objective reasons to consider that more aggregated cost models should be based on a combination of historical and/or forecast data, while individual cost pools or activities within the GDNs businesses, should only be modelled using historical or forecast cost data.

In general, we would suggest that Ofgem seeks to maintain as consistent sample periods as possible across its benchmarking, as was the case at GD1. However, the discussion above highlights that there may circumstances where there is justification for adopting different approaches.

3.4. International comparators

In principle including international comparators in the benchmark sample can help to improve the benchmarking process and, in particular, allow more ambitious econometric estimation techniques by increasing the sample size for the analysis.

Regulators in other jurisdictions, such as the Australian Energy Regulator (AER) and the CRU in the Republic of Ireland have either adopted or previously considered using international comparators in their benchmarking. The AER for example, included data from Ontario and New Zealand in addition to Australian electricity distribution network operators in its benchmarking.[48] The CRU regularly uses information on the costs of UK based utility companies to benchmark companies in Ireland. However, particularly in the case of Ireland, this is largely driven by the lack of relevant domestic comparators.

We would **not propose** that Ofgem include gas distribution network operators in the econometric modelling other than the GDNs for GD2. Given the potential risks of the comparability of the GDNs data with international comparators, we consider including international comparators is unlikely to improve the robustness of the benchmarking in this context, particularly given that Ofgem already has a relatively large company cross-section compared to regulators in some other jurisdictions and sectors.

3.5. Conclusions

In this section we have considered various methodological issues and choices relevant to the selection of the sample period for econometric modelling in GD2. Our key conclusions are as follows:

- Consider econometric models that are estimated **using a range of both historical and forecast data**, as was the case at GD1.
- Models that use forecast (i.e. business plan) data may provide useful information to inform the GD2 determination, **particularly if the cost structures of the GDNs are expected to change** during the forthcoming regulatory period.
- There are, however, risks with *estimating* econometric models using forecast data, most importantly that it **reduces the independence** of the benchmarking from the GDNs own

[48] See for example, Economic Insights (2014), Economic benchmarking assessment of operating expenditure for NSW and ACT DNSPs

business plans. The longer historical data set available for benchmarking compared to GD1, may also weaken the pure statistical case for using forecast data.

Whilst some other regulators have included international comparators in their benchmarking analysis, we would not propose that Ofgem include them in the GD2 modelling:

- In this specific context, rather than helping increase the robustness of the econometric modelling – by increasing the available sample size – we expect this would lead to **less reliable results**, given the potential issues with data comparability.

4. ESTIMATION TECHNIQUES

This section focuses on the estimation techniques that Ofgem could use for its advanced cost assessment modelling in GD2. We consider the approach to cost assessment applied in GD1 and ED1, and potential strengths and weaknesses of different approaches for use in the current GD2 price review. The remainder of the section is structured as follows:

- Section 4.1 provides a brief summary of what econometric techniques are available for GD2 and Ofgem's approach in recent price control reviews.
- Section 4.2 considers a number of key methodological issues related to estimation techniques and how these could be approached in GD2.
- Section 4.3 concludes.

4.1. ESTIMATION TECHNIQUES

Estimation techniques that are available to Ofgem can be categorised as *parametric* or *non-parametric* approaches.[49] Ofgem, aside from simple unit cost models, has not previously relied on more sophisticated non-parametric approaches (e.g., Data Envelopment Analysis (DEA)). We consider that Ofgem should continue to focus on parametric approaches for GD2 as:

- **Parametric approaches** allow for a more comprehensive view of industry dynamics such as economies of scale and density. This is of particular importance in regulated network industries where these factors may vary greatly.
- **Non-parametric approaches**, unlike parametric approaches, do not allow for statistical testing of the models and between alternative model specifications i.e., non-parametric models do not provide statistical confidence in the robustness of the model.

The most commonly used or tested techniques for benchmarking of regulated infrastructure are:

- Simple unit cost/ volume models.
- Econometric **OLS**. These are often conducted on 'pooled' (i.e., multiple years of data) and is referred to as pooled OLS (**POLS**).

  As noted earlier in the report, OLS identifies the average expenditure levels for the comparators based on their cost drivers/ explanatory variables, which can be adjusted to a chosen benchmark if deemed appropriate (e.g., upper quartile company). The latter is referred to as corrected OLS or **COLS**, which we discuss in further detail below.

[49] This section provides only a brief summary of the most common estimation techniques used in UK regulation. There is an extensive academic literature of the theoretical merits of different techniques for efficiency benchmarking and their application to different data sets internationally. This section does not cover this extensive literature but instead focuses on the practical application of estimation techniques in the context of recent Ofgem price reviews.

- Econometric **RE** models. POLS does not specifically identify comparators' inefficiency, rather the error term thus captures the company effect and white noise. RE allows for a company effect to be identified which can be interpreted as inefficiency.
- **SFA**. Like RE, SFA allows for the separate identification of inefficiency. However, it requires additional assumptions and a significant amount of data (ideally lots of comparators) for the estimation process to successfully run.

The choice of modelling approach for assessing the efficiency of networks has been considered extensively by Ofgem since its inception. In addition to its own in-depth research, Ofgem has commissioned a number of reports which include our 2003 report on DPCR4 benchmarking approaches[50] and Frontier Economics' 2013 report on ED1 benchmarking approaches.[51] Various independent or company submissions have also debated the relative strengths and weaknesses of alternative estimation techniques.[52]

Given the extensive work that has been done by Ofgem and its consultants in the past, we believe that an assessment of possible estimation techniques should consider what Ofgem has ruled out in the past and why, and whether anything (e.g., data availability and detail) has changed such that the new techniques could be applied. We consider these issues in further detail in the subsections below.

4.1.1. Use of POLS in Ofgem's benchmarking

Over the years Ofgem's approach has been to use consistently:

- POLS; and
- simple unit cost/ volume models.[53]

The reasons for this are largely down to data limitations. However, we also note that these techniques are simple, transparent, and replicable.

A key concern that has been raised by stakeholders with POLS is that it does not control for systematic differences in the GDNs that are not captured in the cost drivers and explanatory variables. Ofgem has itself noted part of the observed differences in predicted vs. actual costs of the energy network companies can relate to factors other than their relative efficiency (i.e. "noise" in the data).[54]

---

[50] CEPA (2003), '[Background to work on assessing efficiency for the 2005 distribution price control review]', a report prepared for Ofgem, November.

[51] Frontier Economics (2013), '[Total cost benchmarking at RIIO-ED1 – Phase 2 report – Volume 1]', a report prepared for Ofgem, April.

[52] See for example Oxera (2014): 'Recommendations on cost assessment approaches for RIIO-ED1'

[53] We believe that this has generally been the case since DPCR3 determination in 2000.

[54] In its Initial Proposals step-by-step guide on the GD1 cost efficiency assessment methodology, in commenting on its use of an upper quartile efficiency target (discussed in more detail below) Ofgem stated "*We are defining efficient costs from our benchmarking at the upper quartile (UQ) level of efficiency rather than the frontier to acknowledge that a part of the difference in costs across the GDNs relates to factors other GDNs' relative efficiency (i.e. there are statistical errors).* Ofgem (2012): 'RIIO-GD1 Initial Proposals – Step-by-step guide for the cost efficiency assessment methodology', p. 13

This has led certain practitioners to recommend:

- use of company FE models when using panel data sets, to allow for company specific factors that cannot be suitably observed, measured or controlled for; and/or
- estimation techniques, such as RE and SFA, that can in principle distinguish between firm effects, inefficiency and statistical "noise".

Some of the advantages and disadvantages of using time invariant FE modelling in panel method-based studies, is set out in the text box below.

Text Box 1: Company FE models

Company (or firm) effects reflect factors that explain differences (i.e. the heterogeneity) in companies relative cost performance other than inefficiency and noise in the data that are due to company characteristics that are not observable or controlled for in the specified relationship between costs and cost drivers in the model. If these effects are company specific but time invariant their effect is captured in the intercept term of the regression after including company dummy variables.[55]

Company FE models can be used to take account of these cost variations that are considered to be driven by exogenous factors but are not otherwise accounted for in the regression model. FE models have well known theoretical strengths and limitations. They take account of the panel structure of the data and in principle produce unbiased and consistent parameter estimates in the presence of correlation between company effects and cost drivers.

However, they also have a number of practical limitations. With a relatively short time series there may be a risk that the company FE to some degree captures differences in efficiency / inefficiency rather than just time invariant factors that cause differences in costs between companies. As a result, estimates of efficiency based on model residuals may not be accurate. FE is also a relatively data intensive approach as it effectively estimates a dummy variable for each network company included in the sample. This can lead to imprecise model estimation when using relatively small datasets as is the case with the GB gas distribution sector.

Company-specific effects are also partly accounted for in company-specific and normalisation adjustments that regulators, including Ofgem, apply ahead of their cost modelling. This does not mean that all company specific effects may be removed, but it does mean that part of the considerable variation in company characteristics is already adjusted for ahead of the econometric modelling – e.g. for urbanity and industrial composition, etc. (see further discussion on regional factors in Section 7).

As discussed in Section 3, while the time dimension (number of years data) of Ofgem's datasets have increased since GD1 price review, the number of comparators has not, and this is the most crucial element in getting robust estimates of the networks' relative efficiencies. This is particularly pertinent when key cost drivers, such as customer numbers and MEAV do not have material variations year-on-year to align with variations in opex, repex and capex and is another reason why a FE model is unlikely to be appropriate for the GDN dataset. [56] This is one of the main reasons that RE has been proposed in recent regulatory settings rather than company FE models. As discussed above, the structure imposed on the error term allows company efficiency to be differentiated from "noise". Ofwat has used RE to estimate its econometric models for PR19, although RE also has certain methodological limitations.[57]

---

[55] See Oxera (2014) for a more extensive discussion of company-specific effects.

[56] This point was also raised in Gibbens and Zackery (2013) in response to Frontier Economics' proposed RE models for ED1. Frontier Economics (2013): 'Total cost benchmarking at RIIO-ED1, Phase 2 report. Vol 1'

[57] See CEPA (2018): 'PR19 Econometric Benchmarking Models'

As compared to POLS, in principle SFA allows for direct interpretation of the econometric residuals so that white noise / modelling errors can be separated from inefficiency. However, there are, as with all estimation techniques, known limitations with SFA:

- SFA uses a maximum likelihood estimation method. This means SFA models are relatively data intensive and have proven to be difficult to implement and function appropriately.

  For example, the AER in Australia relies on four econometric models, two of which use SFA, that are intended to separate the efficiency estimate from the error[58]. While AER has been able to get its SFA models to function it has needed to use data from New Zealand and Ontario (Canada).[59] In addition, even though it uses SFA techniques for the purpose of separating the inefficiency and noise, the AER also made a 'correction' to its SFA efficiency results to reflect "*specification of outputs and inputs, data imperfections and other uncertainties*".[60]

- SFA also requires distributional assumptions to be made on both the error term and efficiency components of the residuals.[61]

  While different assumptions can be investigated through the modelling exercise, this does mean that the results of the benchmarking may be dependent on a relatively arbitrary / strong assumption of both these components.[62]

In summary, while SFA can help to address some of the concerns with the use of POLS in a cost efficiency assessment context, there are also known practical challenges and issues. We note that whilst these issues exist, and have been well rehearsed in both academic and regulatory literatures, given the known advantages of SFA, a number of regulators have either sought to develop SFA models, or in practice apply them in regulatory contexts to set allowed revenues (e.g. BNetzA in Germany, AER in Australia. The Office of Rail Regulation (now the Office of Rail and Road (ORR)) has also previously considered SFA for benchmarking Network Rail, as has Ofcom, so SFA has some UK regulatory precedent).

While we would not suggest Ofgem necessarily rule out SFA as an approach for GD2, we believe considerable work would be necessary to test the robustness of this estimation approach and to justify the assumptions that would need to underpin any such analysis. In context of the current regulatory proceeding for GD2, we consider there may be a number of risks, even though there are intrinsic advantages from SFA models, from introducing an alternative estimation into the benchmarking process. There is also a risk that SFA reduces transparency while not necessarily improving model robustness, both of which have a significant role in our model selection criteria.

[58] AER (2019), 'DRAFT DECISION: Ausgrid Distribution determination 2019−24, Attachment 6 – Operating expenditure', November.

[59] As discussed in Section 3, we do not consider that the use of international data is appropriate for Ofgem. Our concerns about the AER's use of international data are set out in our report for ActewAGL (CEPA, 2015, 'Benchmarking and setting efficiency targets for the Australian DNSPs', a report prepared for ActewAGL, January).

[60] Economic Insights (2014), 'Economic Benchmarking Assessment of Operating Expenditure for NSW and ACT Electricity DNSPs', a report prepared for Australian Energy Regulator, November, page 47.

[61] Popular distributional forms are the half-normal distribution, the truncated normal distribution, and the exponential distribution. See Oxera (2014).

[62] Although the same issue may also be argued applies to the way that POLS is adjusted to account for the impact of potential statistical error or noise in the data using the COLS method (see 4.1.2 below).

We would suggest, therefore, that should Ofgem wish to investigate using an SFA model:

- It should in the first instance be used as a cross-check / point of comparison to efficiency scores produced using the estimation technique adopted at GD1 (i.e. POLS). For example, it might be used to test how challenging an upper quartile COLS based efficiency frontier may be in GD2.
- Given POLS with an adjustment (i.e., COLS, discussed in the following subsection) is the existing regulatory approach in gas distribution price controls, for regulatory consistency purposes, emphasis should be placed on demonstrating that an SFA approach leads to considerably more reliable results that than the existing approach, before applied in GD2.
- Any SFA model should be subject to considerable robustness assessment and sensitivity analysis, particularly as regards the assumptions used for the distribution of the inefficiency component.

4.1.2. Use of COLS to address concerns with pooled models

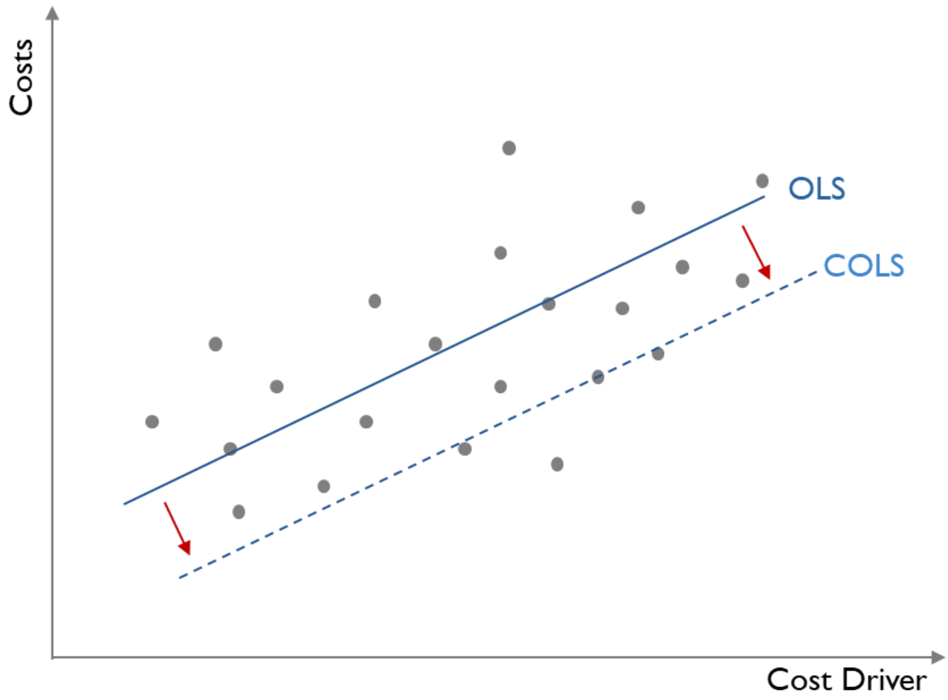
As discussed, above Ofgem has identified in previous price reviews, that one of the limitations of POLS is that it does not allow separation between the "noise", company heterogeneity, and inefficiency in the errors term. In recognition of the risk of this 'measurement error'[63], Ofgem has in the past not used the frontier performer to set the efficiency targets at previous price control reviews. Instead, it has aimed off the frontier by using an upper quartile, third, or average in the past.

This 'correction' to estimated costs can be applied to a single cross-sectional OLS and pooled OLS (multiple years of cross-sectional data). The correction is done by shifting the 'intercept' point rather than altering the slope of the line, i.e., the relationship between expenditure and cost drivers is maintained but the 'fixed' element is shifted (see Figure 4.1 below[64]).

[63] We use the term 'measurement error' to refer to issues with the both the underlying data, noise, and instances where there may be omitted variables (i.e. systematic differences that are not captured by the explanatory variables included in the model).

[64] In this example, the COLS line reflects an upper quartile benchmark.

Figure 4.1: Illustration of OLS and COLS



Source: CEPA

While the COLS adjustment is relatively arbitrary, it is a generally simple, transparent, and replicable approach to addressing some of the concerns identified from use of POLS for efficiency analysis and, therefore, is consistent with a number of the model selection criteria we set out in Section 2.

4.1.3. Summary

We propose that Ofgem use COLS as its primary estimation technique for GD2. However, we suggest that Ofgem also test RE models given that this alternative estimation technique has been explored in previous price reviews, has been used by other regulators (e.g. Ofwat) and has certain advantages compared to use of OLS. Ofgem may also wish to **investigate the use of SFA as an estimation technique** in GD2, although we have set out above a number of factors that we believe Ofgem should take into consideration before applying this technique in practice.

We believe that there are also a number of variations within Ofgem's application of COLS technique for GD2 that can be considered:

- The choice of the 'correction' / benchmark.
- The time required to 'catch-up' to the frontier.

We discuss these in the next section.

4.2. Considerations in applying COLS

4.2.1. Choice of correction / benchmark (offset v frontier)

As discussed above, COLS uses the same statistical estimation technique as POLS, however, the 'average' line is shifted towards a 'frontier' point to reflect the efficiency target / benchmark for the GDNs. The 'correction' distance / benchmark is at Ofgem's discretion and, as discussed above, Ofgem has typically chosen the upper (top) quartile performing company[65].

Ofgem has aimed off the frontier being defined as the most efficient unit in previous price controls as an acknowledgement that it does not have perfect information, and that there is likely to still be measurement error in the data. This includes systematic differences, besides efficiency, between GDNs that is not explained by the cost drivers and explanatory variables.

Ofgem's choice of correction / benchmark has been driven by:

- **The confidence Ofgem has in the data**. If Ofgem considers that the data is accurate, and is likely to provide robust results, then it may be more inclined to set a tougher target.
- **The variability in the modelling results**. If there is a relatively large distribution in the networks' expenditure around the line of best fit this can indicate that there may be a greater degree of measurement error and Ofgem may need to be more cautious.

  For example, in DPCR5 Ofgem used the upper third for network operating costs "*due to greater variability in the data*".[66] On the other hand, if the model has good statistical results then Ofgem may have more confidence in choosing a tougher target.

There is no statistical guidance or rule around what the appropriate correction / benchmark should be. In its most recent price controls, Ofgem has relied on the upper quartile. This sets the efficiency target at a point that is being achieved (or forecast to be achieved) by a quarter of the networks. **We recommend that Ofgem use the upper quartile as a starting point again.**

We consider that the upper quartile provides a reasonable allowance for the measurement error in the modelling. We recommend that Ofgem moves away from this only if there is greater variability in the results it has observed at previous price controls.

4.2.2. Closing the gap

In addition to using the upper quartile efficiency target, Ofgem also decided that GDNs' would only need to close 75% of the gap to the upper quartile during GD1. This concession was to further recognise that the models, results, and the target were still affected by measurement error.

In other price controls including in other sectors, the use of a more gradual shift to the 'frontier' has been adopted. This includes Ofgem's DPCR3, Ofwat's PR09, and ORR's 2008 decision for Network Rail. The glide-paths have typically been used when the regulator considers that it is not feasible for the networks to achieve the full scale of the efficiency savings over one price control period.

By the start of GD2, GDNs will have had two price controls (GDPCR and GD1) to catch-up to the efficiency target, with a relatively long current price control period (eight years). We would, therefore, expect the companies to be converging towards the frontier and **suggest that Ofgem explore the case for not applying a glide-path towards the efficiency target in GD2.**

---

[65] Or in between two companies.

[66] Ofgem (2009), 'Electricity Distribution Price Control Review: Final Proposals - Allowed revenue - Cost assessment', December, page 4.

In line with ED1 and our recommendations above, we consider that use of the upper quartile is likely to be sufficient in dealing with measurement error in the models.

4.3. Conclusions

Based on our assessment of the feasible techniques and the key methodological issues for GD2, we suggest the following practical proposals for the estimation techniques that Ofgem might use for its econometric benchmarking and how to define the efficiency frontier. These follow closely the precedent set by previous Ofgem price controls. This is perhaps unsurprising given the extensive work that has been done by Ofgem and its consultants in the past. We propose that Ofgem:

- use **COLS** as its primary estimation technique for GD2;
- **test models using RE** as an alternative estimation technique;
- use the **upper quartile** as a starting point for the efficiency target / benchmark for the GDNs; and
- **explore not applying a glide-path** towards the efficiency target.

Ofgem may also wish to **investigate the use of SFA as an estimation technique** in the GD2 proceedings. However, we consider:

- it should in the first instance be used as a **cross-check / point of comparison** to efficiency scores produced using the estimation technique adopted at GD1 (i.e. COLS)[67];
- the emphasis should be placed on demonstrating that an SFA approach leads to **considerably more reliable results** than Ofgem's existing (relatively simple and replicable) benchmarking approach, even though there are well known limitations with COLS; and
- any SFA model would need to be subject to considerable **robustness assessment and sensitivity testing**.[68]

[67] For example, does the upper quartile efficiency frontier set a relatively challenging efficiency target when compared to the results of SFA modelling?

[68] Particularly as regards the imposed assumptions used for the distribution of the inefficiency component, given the lack of precedent of use within previous regulatory proceedings undertaken by Ofgem.

5. COST AGGREGATION

In this section we consider options for selecting the cost aggregation pools for benchmarking analysis in GD2. The subsections below:

- Provide a brief summary of the cost aggregation pools used for econometric benchmarking in the Final Proposals for GD1 (Section 5.1).
- Set out a series of relevant considerations for deciding on the level of aggregation and types of cost to include in the econometric modelling (Section 5.2).
- Consider options for the benchmarking cost pools in the GD2 cost assessment and their relative merits (Section 5.3).

5.1. Approach to cost aggregation in GD1

As discussed in the introduction, the Final Proposals for GD1 combined top-down and bottom-up modelling streams to set the GDNs final totex allowances. The GD1 benchmarking included models that grouped costs according to:

- **Expenditure areas:** either at a broadly total business level ('totex'), or individual expenditure area (i.e. opex, capex and repex – 'middle-up'); or
- **Activity level areas:** such as repairs, emergency call-out, connections, mains and services replacement programme etc.

In its Initial Proposals document, Ofgem noted that:

> "*We consider the different modelling approaches provide useful information in assessing GDNs' comparative efficiency. For example, totex models ensure that we consider GDNs' opex-capex trade-offs in our comparative efficiency assessment, ie that we can identify those GDNs that have minimised total costs.*
>
> *Activity level analysis enables a richer model specification, ie we can take into account a greater number of potential factors that explain costs. Our models based on the principal expenditure lines, opex, capex, and repex, strike a balance between ensuring that we consider trade-offs between cost areas but allow a richer model specification than the high-level totex model.*"[69]

Although the results of the top-down and bottom-up workstreams were used separately in the GD1 Final Proposals, it is important to note that the selection of the groups of costs for the bottom-up assessment also influenced the top-down assessment, given that the explanatory variables adopted in the bottom-up models, were also largely used in the top-down regression for totex:

- The bottom-up regressions used a series of work / activity driver and scale variables to model difference activity level costs.
- The top-down level totex regression used a Composite Scale Variable (CSV) which combined network scale[70] and workload driver variables based on the bottom-up regressions.

[69] Ofgem (2010): 'RIIO-GD1: Initial Proposals – Supporting document – Cost efficiency', p. 6

[70] The principal scale variable used was MEAV.

Our interpretation of the approach Ofgem applied for benchmarking cost pools for the GD1 Final Proposals is summarised in further detail in Appendix A.

5.2. Issues and approach in selecting a cost pool

5.2.1. Issues in selecting a benchmarking cost pool

There is an extensive literature on the different ways in which costs can be aggregated and treated for cost benchmarking purposes.[71] This literature explains how choices of the cost pools (types and aggregation of costs) used for benchmarking requires careful consideration of a series of high-level trade-offs. These are discussed in the subsections below.

Cost allocation across businesses

As benchmarking becomes more granular there is an increasing risk that different company approaches in cost allocation methodologies are the cause of differences in GDN costs, rather than (in)efficiency.[72] Therefore, the quality of data and consistency of reporting becomes increasingly important where more disaggregate benchmarking modelling is used within the price review.

While we note that Ofgem has over a period of years developed relatively detailed regulatory reporting guidelines for the GDNs through its BPDTs and RRPs, differences in the corporate policies and reporting of costs by each GDN group may still be one of the explanations for differences in observed cost performance, both across activities or individual areas of expenditure.

In contrast, one of the advantages of more aggregated – e.g. totex or middle-up – analysis is that the risks from these cost allocation and boundary issues can be reduced.

Aggregation impact of cost efficiency assessment

Where more disaggregated cost pools are adopted for benchmarking, the interactions between the benchmarking of each cost pool requires careful consideration:

- **First,** to avoid 'cherry picking' by the regulator when the results of the efficiency assessment across each cost pool are combined.
- **Second,** to avoid creating unintended incentives for the network operator when making expenditure choices.[73]

---

[71] See for example CEPA (2018): 'PR19 Econometric Benchmarking Models' and Frontier Economics (2010): 'RPI-X@20: The future role of benchmarking in regulatory reviews'

[72] Of course, this can partly be mitigated by imposing strict regulatory reporting guidelines for the purposes of reporting to the regulator during price reviews.

[73] Introducing boundaries between the benchmarking of different costs may distort managerial incentives if, for example, different cost pools are benchmarked / evaluated using different cost assessment tools.

Differences in cost performance in more disaggregated cost pool models may also reflect differences in business practices and models between companies, rather than compelling evidence of overall business (in)efficiency by the GDNs. For example:

- GDN 1 may have materially higher costs in cost pool 1 than GDN 2, slightly lower costs in cost pool 2, but overall lower costs than GDN 2.
- Benchmarking at an individual cost pool, rather than aggregative level, may mean that the benchmarking incorrectly identifies one GDN as less efficient than another.[74]

This is the trade-off effect that Ofgem identified in its GD1 Initial Proposals. More disaggregated benchmarking models may be less effective at picking up these effects. For at the margin expenditure decisions at least, disaggregated modelling may create unintended incentives for GDNs to adopt a particular business model or expenditure solution, simply to perform better under the benchmarking, rather than because it is more or less efficient from a total resource cost perspective to do so.

Choice of explanatory variables

There are also advantages and disadvantages of more or less disaggregated benchmarking for selecting the explanatory variables in the benchmarking model:

- On the one hand, as models become more granular it may be possible to better identify explanatory variables that reflect the specific costs and drivers under consideration. More disaggregated benchmarking may help to explain the causes of differences in GDN cost performance / efficiency which more aggregated models fail to achieve.
- On the other hand, it may be argued that at more disaggregated levels it is more difficult to establish explanatory variables that meaningfully reflect all of the drivers of the costs of particular activities. More aggregative cost pools are more likely to reflect the more aggregative narrative of drivers of gas distribution costs (see Section 6).

During the PR14 Bristol Water appeal, the CMA discussed similar trade-offs in water and wastewater sector benchmarking. It noted that models that focus on specific parts of the value chain / activities:

> "*may allow the set of explanatory factors to be tailored to each model, reducing risks of inaccuracy in estimated coefficients from the small sample size (and limited variation within the sample) relative to the number of cost drivers that are material for wholesale water supply.*"[75]

Timing of investment requirements

One of the known issues with more aggregated cost modelling that involves comparisons of companies totex over time, is that the differences in the level of reported cash expenditure between the companies, both within a given year or over a price control period (or longer), may be due to differences in investment requirements at a given point in time, rather than differences in companies' relative efficiency. Again, this issue was highlighted by the CMA during the Bristol Water PR14 appeal.

---

[74] To the extent trade-offs can be made between cost pool 1 and cost pool 2 in this example.

[75] CMA (2015): 'Bristol Water plc: a reference under section 12 (3) a of the Water Industry Act 1991', p. A4(1)-32

The CMA identified two risks to the accuracy of aggregated regression models:

- The estimates derived from totex models for a particular company may provide an inaccurate guide to its expenditure requirements because they may take insufficient account of the extent to which investment requirements in the price review period in question, differ from those that applied (or will apply) on average across all companies in the data period used.
- The CMA also identified a risk that investment cycles may lead to less accurate models: "*Variations over time and between companies in capex, which are driven by variations in investment needs, may give* [sic] *add noise to the data and distort the estimated coefficients from the econometric model and worsen the accuracy of the estimated expenditure for each company that is derived from the model.*"[76] [77]

Within the gas distribution sector, capex and the higher value spend repex programme, may raise similar issues for the GD2 benchmarking.

Forecast repex spend varies between GDNs depending on the network's condition and, therefore, the expected programme of work needed to reduce risk on each network. As with other sectors such as water, wastewater and electricity distribution, capex in gas distribution networks can be lumpy. This means that more aggregative totex models that primarily rely on scale variables may not be able to control for all the factors that influence GDNs costs over the price control.

A number of potential solutions have been proposed to these issues:

- In GD1, Ofgem used workload drivers within the aggregated regressions, to help account for differences in the volume of activity / work that the repex programmes of each GDN, in particular, that may drive variations in costs.
- For the PR14 modelling, Ofwat smoothed capex in its econometric models. Ofgem adopted a similar approach at GD1 by using the seven-year moving average workloads for top-down and middle-up capex regressions.
- Some stakeholders have also proposed including measures of the condition or quality of companies' capital stock as explanatory variables in the regression, to help account for the impact of differences in network asset condition.

With regards inclusion of measures of asset condition / health within regression models, this could create perverse incentives for the GDNs by rewarding companies which maintain poorer condition assets. Within the water sector, the CMA has also noted that while including measures of companies' levels of activity in the regressions (as adopted in GD1) may help address some of the issues discussed above, levels of activity "*are not necessarily indicative of their investment requirements; the activity levels reflect management choices and working practices and do not cover all aspects of water companies' investment requirements.*"[78]

We discuss capex smoothing and the use of workload drivers in further detail in Section 6.

---

[76] Ibid, p A4(1)-29

[77] NERA (2014): 'Ofgem's Slow Track (Draft Determination) Benchmarking – Memo for Scottish Power Energy Networks', also suggest that there may be economies of scale and scope in replacement refurbishment programmes that make lumpy investment programmes efficient.

[78] Ibid, p. A4(1)-28

An alternative approach to all of the above, would be to simply exclude all, of elements of, the GDNs capex and repex programmes from the regressions. This has:

- The advantage that regression analysis can focus on costs that can be explained by a more consistent set of explanatory variables, with potentially more accurate estimates.
- The potential disadvantage that the benefits of benchmarking at a more aggregative level, to capture trade-offs between activities and costs, are reduced.

Disaggregated to aggregated, or aggregated to disaggregated?

As noted above, although the GD1 cost assessment involved separate top-down and bottom-up modelling workstreams, the bottom-up regressions had a significant influence on the top-down modelling given the totex model specification aggregated many of the explanatory variables used in the disaggregated (activity level) regressions. In this respect, the approach Ofgem took at GD1 might be described as a 'disaggregated to aggregated' approach to benchmarking the GDNs:

- An advantage of this approach is that the aggregative totex model specification can reflect the more granular analysis of expected cost and workload drivers within benchmarking of different respective activities of the GDNs businesses.

  The totex modelling workstream[79] can then be used to help reduce the risk (see 5.1.2 above) that the relative efficiency and trade-offs of different GDN business models used to deliver a range of activities and functions, are not lost within the efficiency assessment.
- The disadvantage of this approach, in contrast to a modelling exercise that potentially seeks to identify a set of aggregative explanatory variables of GDNs costs for aggregative modelling, is that all the models are based on a similar set of explanatory variables.

5.2.2. Proposed principles for selecting a cost pool

Given the trade-offs that exist in selecting a cost pool for benchmarking purposes, **we propose that Ofgem continue to retain a toolkit approach** to its cost assessment that considers various alternative approaches to cost pooling, rather than relying on a single model or modelling aggregation workstream.

We would recommend that Ofgem **use a set of principles** to determine the aggregation of its cost assessment models, the expenditure which it includes / excludes from its toolkit of regression models and, importantly, the weight it places on benchmarking results at different levels of cost aggregations. These should build on the high-level model selection criteria set out in Section 2.

While we would expect Ofgem to refine its views of the relevant principles and criteria based on the information it receives in the submitted GD2 business plans, our initial thoughts on the key factors we would expect the GD2 assessment to consider, include the following:

- **Complementary:** Is there a strong technical / economic reason to believe that activities or groups of expenditure are complementary and should be benchmarked together and a consistent set of cost drivers can be identified?

[79] Based on a similar aggregative view of cost drivers across GDN business activities.

- **Cost trade-offs:** Can GDNs make trade-offs in expenditure between the different activities / areas included in the cost pool, and so benchmarking those activities / costs together will help avoid biased relative efficiency results or unintended managerial incentives for the GDNs?
- **Cost boundary complexity:** How complex is the boundary of cost reporting data that needs to be defined to benchmark the identified cost pool / activity (e.g. how well defined is the group of costs within Ofgem's regulatory reporting templates)?
- **Risk of inaccurate / biased models:** Is there too much "noise" in the data to be confident that including certain types of expenditure within aggregated regressions could lead to inaccurate model results, or coefficient estimates that are difficult to interpret using engineering/economic logic?

We propose that Ofgem consider various data / statistical tests to identify how complementary the different types of expenditure and their expected explanatory variables are for benchmarking purposes in GD2. Before grouping costs, this may include:

- Testing for year-on-year volatility in expenditure in particular areas / activities that appears to be unrelated and/or correlated with changes in business scale drivers.
- Testing[80] the expected consistency of workload drivers between different types / areas of expenditure before costs are grouped together for benchmarking.

5.3. Options for RIIO-GD2 cost pooling

There are of a variety of different approaches that Ofgem could in principle consider for cost pooling in its benchmarking. We have set out below four options that reflect a range of approaches we would view as practical and feasible approaches for GD2 and capture a broad spectrum of different approaches that Ofgem could consider when developing its methodology for GD2.

The four options can be described as follows:

- **1: Aggregated modelling:** This option would involve only aggregated, high-level, forms of econometric benchmarking. We assume Ofgem would consider under this pooling approach a range of totex models as part of a tool kit assessment[81] and the results then combined with residual costs evaluated using technical and qualitative assessment.

  This means there would not be a separate bottom-up cost assessment workstream.
- **2: Totex and disaggregated (activity based) modelling:** This option would be similar to the approach Ofgem followed for GD1. The cost assessment would combine a "top-down" (totex regression based) and "bottom-up" workstream (including disaggregated cost activity level regression modelling) with residual costs evaluated using technical and qualitative assessment.

---

[80] Through engineering judgement and explanation within the GDNs plans of the expected key workload drivers of different cost activity levels and programmes.

[81] For example, Ofgem developed two totex models at ED1, one largely scale based, the other activity based.

- **3: Totex and "Opex Plus" modelling:** This option would combine a "top-down" totex modelling workstream and a more disaggregated workstream where total opex is benchmarked using regression analysis on a pooled basis that includes other costs (e.g. elements of capex and/or repex) where complementarities (e.g. in cost drivers) and trade-offs for pooling exist.[82]

  The disaggregated regression modelling in this case is described as "Opex Plus" as it combines total opex and other costs where the types of criteria listed in 5.2.2 above are met.

- **4: "Opex Plus" modelling:** Under this option there would be a single cost assessment workstream that combines an Opex Plus regression and residual expenditure evaluated under separate technical and qualitative activity level assessments.

  Costs would only be benchmarked using regression analysis at an aggregative level to the extent that the pooled costs are considered to be complementary, have trade-offs and can be robustly explained by a consistent set of cost drivers / explanatory variables (e.g. scale variables).

For all four options, the expectation is that within the residual expenditure pool, costs would be grouped and assessed according to activities that are considered appropriate for assessment together[83] as was the case in the bottom up GD1 cost assessment.

It is important to note that different options are not necessarily mutually exclusive.

For example, Ofgem could consider an approach where it develops a set of preferred totex models – common to Options 1 – 3 – and a series of more disaggregated modelling workstreams that group disaggregated cost pools together to the extent the types of criteria described in the previous section are met. The more disaggregated modelling could also in principle be used to inform the specification of the aggregative models or, as was the case at GD1, Ofgem could consider developing a set of middle-up models to accompany the totex and disaggregated activity modelling in Option 2.

A variant of either Option 1 or Option 2 could also be where the types of middle-up (opex, capex and repex) regressions developed at GD1 are used in replacement, or as evidence alongside, either the totex modelling (in the case of Option 1) or the disaggregated modelling (in the case of Option 2).[84]

We provide a summary illustration of each option in Figure 5.1 below. In each case, we seek to illustrate the cost aggregation level used in the econometric modelling and how under each option different modelling workstreams could be combined to set totex allowances for the price control. Appendix B provides a more detailed tabular illustration of each option using the cost pooling presentation that was adopted by Ofgem for its GD1 Initial Proposals for cost assessment.

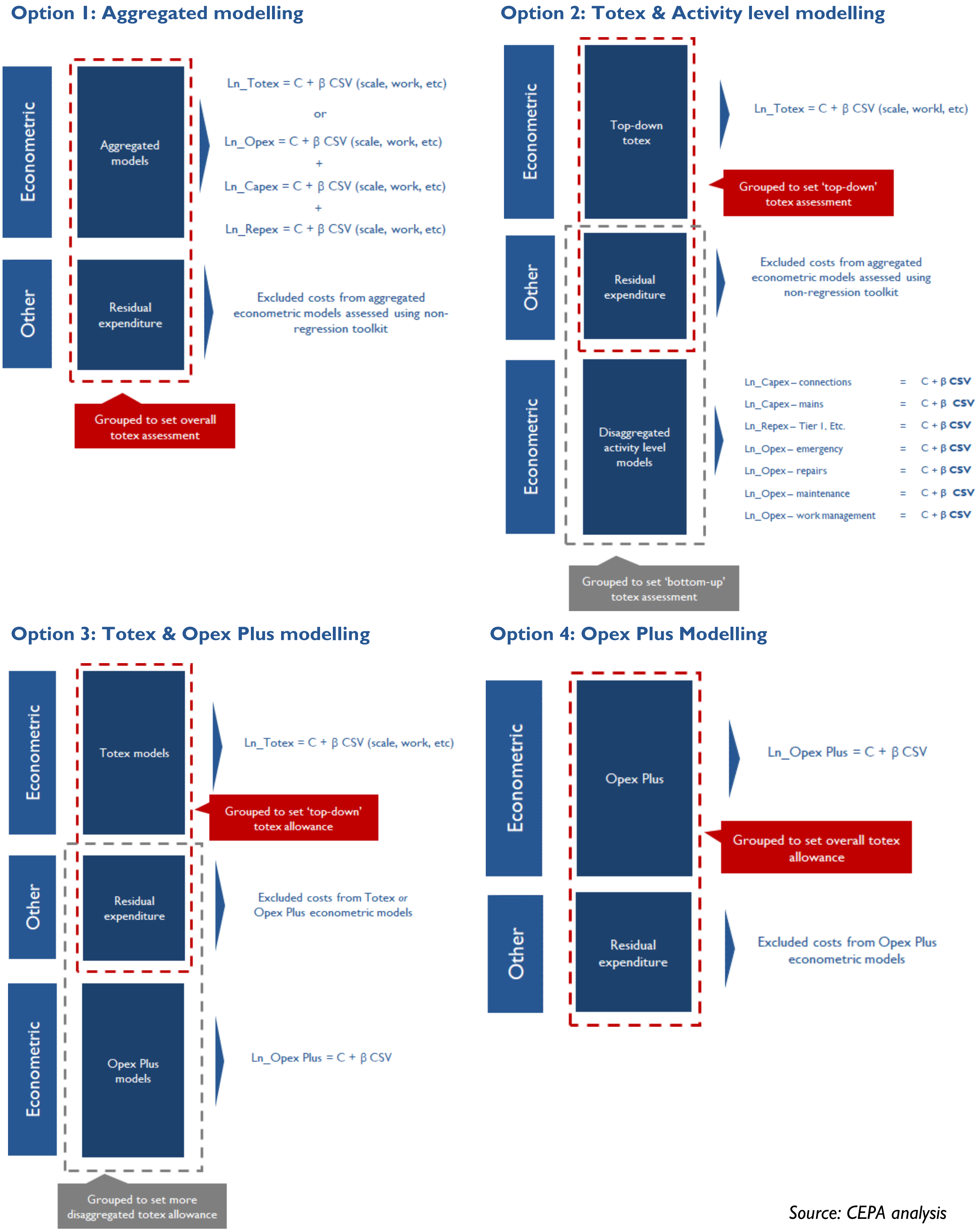
---

[82] As with Option 1 and 2, any residual costs would be evaluated using separate technical and qualitative assessment and added to the results of the top-down and more disaggregated modelling workstreams.

[83] For example, repex may be grouped for cost assessment purposes together as a single activity or broken into separate activities (Tier 1, Tier 2 etc.) as deemed appropriate.

[84] In the case of Option 2, the middle-up regressions could, for example, be used to help inform specification of the totex model similar to the approach at GD1.

Figure 5.1: Cost aggregation modelling options



Source: CEPA analysis

** Regression specifications in figures show CSVs. This is for illustration purposes.*

5.4. Assessment of options

We expect the preferred approach will depend on how well the models perform against model selection criteria and how confident Ofgem is with placing weight on more aggregated totex benchmarking models as opposed to bottom-up modelling in setting final totex allowances. Building on the discussion in Section 5.3, some of the relative strengths and weaknesses of the different options described above are as follows:

1: Aggregative modelling

- The main advantage of this approach is it balances trade-offs between costs and seeks to mitigate problems from cost allocation differences between companies, to the extent these exist for regulatory reporting purposes. A totex / more aggregative modelling approach has strong incentive properties if based on exogenous cost drivers / explanatory variables.
- In principle, the results of the regression should also be relatively easy to interpret if based on scale variables (arguably less so if the regression is a blended activity / scale CSV as was adopted by Ofgem in the GD1 Final Proposals).
- However, as discussed above, there is a risk that more aggregative models do not capture exogenous cost drivers that may explain variations in GDN costs appropriately (see discussion of investment requirements above). This may mean that more adjustments are necessary outside of the models (e.g. ex-post to the modelling) which make assessment complex and non-transparent.
- The efficiency assessment is likely to be sensitive to model choices (e.g. explanatory variables). The final aggregative model specification is also more likely to need to reflect prior economic / engineering expectations of the drivers of GDN costs than more disaggregated cost assessment toolkits, where the findings from disaggregated (e.g. activity level) models can in principle be used to inform the specification of the aggregative totex model.

2: Totex and disaggregated modelling

- This would be the least change approach relative to GD1. The approach helps to avoid cherry-picking if Ofgem triangulates across different model findings.
- More disaggregated models may be able to more accurately capture relevant cost drivers than Option 1 and, therefore, more accurately predict future costs. As discussed above, findings from disaggregated models can also be used to inform the more aggregative modelling workstream.
- The disaggregated analysis can also help to explain where sources of efficiency / inefficiency may arise from in combination with the aggregative totex models and may be a useful approach in the context of the proposed Business Plan Incentive for GD2.
- However, as indicated above, there is also the risk that differences in the reported costs / boundaries between individual activities across GDN groups lead to misleading conclusions on the relative efficiency of the individual GDNs if using a bottom-up assessment.
- While in principle more disaggregated modelling may allow a richer model specification, e.g. in the choice of explanatory variables / cost drivers used, retaining the activity level regression specifications at GD1 raises a number of issues:

- In general, they relied on workload volume drivers – which are not a fully exogenous driver of GDNs costs and may still not fully capture the variation in factors that drive differences in investment programmes of the GDNs within a regression model.
- At GD1, Ofgem also aggregated up the disaggregated workload drivers into the totex regression specification. In applying this approach, Ofgem potentially lost some of the benefits of totex vs. disaggregated analysis.[85]

3: Totex & Opex Plus modelling

- This approach captures the trade-offs and complementary nature of different opex activities as part of the bottom-up as well as top-down modelling workstream. There is less risk from drawing boundaries between different opex activities in the more disaggregated workstream.
- The treatment of repex and capex is potentially more understandable within the overall cost assessment, as relative to Option 1 and 2, it is less bundled into the econometric assessment where there is a risk of omitted variable bias.
- However, to the extent that aggregated totex regressions remain a part of the modelling toolkit, as is envisaged under this approach, the types of criticisms that apply to Option 1 and 2, would also apply to Option 3.

4: Opex Plus modelling

- Opex may be better fit for more aggregated regression modelling than capex and repex given the commonality of cost drivers (e.g. scale, work drivers) between opex activities and the nature of the expenditure. Consistent with the principles set out above, it may be sensible to only group other costs with Opex (the 'Plus') in econometric benchmarking where there is clear justification.
- Arguably this is the most consistent approach with the CMA recommendations from the Bristol Water PR14 appeal. More aggregative regression analysis, where possible, captures the trade-offs between different opex and capex activities to the extent that these can be considered to be accurately captured in regression analysis using a consistent set of explanatory variables.
- This option has the potential disadvantage that some of the benefits of benchmarking at a more aggregative level, that capture trade-offs between activities, are reduced. Although this approach potentially leads to more accurate models, and a clearer and more consistent set of explanatory variables across costs grouped together for benchmarking purposes.
- Could potentially be described as a more disciplined application of the 'disaggregated to aggregated' approach to cost modelling. The emphasis is on justifying why expenditure should be *included* in an aggregative benchmarking model, as opposed to a methodology like Option 1 or 2 where the emphasis is on why expenditure should be *excluded* from totex regressions.[86]

---

[85] Essentially doing the same thing albeit on different levels of aggregation of costs.

[86] While in principle, both approaches should produce similar answers, in practice, we expect that Option 4 will produce less aggregative econometric modelling (i.e. less included in aggregated / totex models) and a more bottom-up based cost assessment at GD2.

5.5. CONCLUSIONS

Any benchmarking model is expected to score strongly on some criteria for model selection and less strongly on others. This reflects trade-offs across the many different decisions that need to be made in the model selection process, but particularly the approach to the aggregation of costs in the benchmarking.

As was the case at GD1, we recommend that:

- Ofgem adopt **a tool kit approach to its cost assessment** in GD2. Ofgem's benchmarking should include a range of aggregated and more disaggregated cost assessment models and cost pooling methods, rather than relying on a single model or level of cost aggregation.

Building on the model selection criteria set out in section 2, we propose that Ofgem should use a **set of principles for selecting cost pools** for more aggregative benchmarking:

- This should include factors such as the complementarity, impacts on incentives and risk of biased / inaccurate regression models of different types of costs being grouped together for benchmarking purposes, as well as the complexity of drawing cost boundaries between activities.

With this in mind, we propose that Ofgem consider developing data / statistical tests to identify how complementary the different types of expenditure and their expected explanatory variables are for benchmarking in GD2. Before grouping costs, this may include:

- Testing for year-on-year volatility in expenditure in particular areas / activities that appears to be unrelated and/or correlated with changes in business scale drivers.
- Testing the expected consistency of workload drivers between different types / areas of expenditure before costs are grouped together for benchmarking.

We believe that the final approach to cost pooling should be driven by the factors and criteria we have outlined above – i.e. **no final decision on the preferred approach should be made at this stage of the GD2 review**. However, consistent with a number of the recommendations the CMA made on benchmarking during Bristol Water's PR14 appeal, we suggest that:

- Ofgem investigate an approach where costs would only be benchmarked using aggregated regressions at what we term an **Opex "Plus" level –** aggregated opex *plus* other activities (e.g. certain capex elements) considered complementary in the model specification.

While this approach may lead to greater weight on bottom-up/engineering analysis to set the final totex allowances than was the case at GD1, the potential benefits are:

- aggregative econometric benchmarking models that potentially have a **clearer statistical, economic and engineering logic and fit**; and
- less concern that the variations in capex and repex expenditure patterns, particularly repex, lead to potentially less reliable benchmarking results.

However, we also recognise the positive incentive properties, and benefits for regulatory consistency, in retaining a 'top-down' totex based benchmarking framework, as adopted in GD1, as part of the GD2 toolkit. Totex modelling helps reduce distortions between the treatment of different types of costs and requires less precise boundaries between costs and activities to be identified. Some of the concerns with totex modelling applied in other sectors are, to some extent, also reduced in the GDN context with the use of workload drivers within Ofgem's econometric model specifications.

Therefore, while we consider it sensible to investigate other aggregated specifications, such as the type of Opex Plus model discussed above, we suggest that**:**

- Ofgem continue to develop a **set of preferred totex models** that can be applied and consulted on in later stages of the price review.

6. MODEL SPECIFICATIONS AND COST DRIVERS

In this section, we focus on options for the specification of Ofgem's econometric cost assessment models in GD2, in particular, the cost drivers and proxy explanatory variables that could be used within the models. This section is structured as follows:

- Section 6.1 considers possible cost drivers and explanatory variables that Ofgem could consider for the GD2 econometric modelling.
- Section 6.2 discusses in further detail, the issues with using certain explanatory variables in the regression models, including use of workload drivers, CSVs and time trends/dummies, and the functional form of the regression.
- Section 6.3 provides some high-level conclusions.

6.1. COST DRIVERS

Economic theory and engineering logic would suggest there are a number of different drivers of the costs of gas distribution networks, including scale factors, such as the numbers of customers served by the GDNs, external operating environment and the quality of services provided. External – i.e. non-controllable – differences in input prices (e.g. regional wages or capital (materials) prices) of the firm may also drive differences in GDNs relative capital and operating expenditure performance over time.

What makes a good cost driver / explanatory variable for the purposes of econometric benchmarking and the GD2 cost assessment process more generally? Building on the model evaluation criteria in Section 2 and principles in Ofgem's recent GD2 sector methodology consultation paper, we would expect cost drivers to be assessed against whether they:

- make economic (or engineering) sense;
- are accurately and consistently measurable;
- have a stable relationship with costs over time; and
- are (as far as possible) beyond the control (or influence) of network companies.

In this sub-section we discuss a range of cost drivers and the explanatory variables that could be used as a proxy for those drivers within Ofgem's econometric modelling.

We start by setting out a plausible causal narrative of what might be expected to be the *exogenous* (i.e. external) drivers of total expenditure within the GB gas distribution sector, followed by a discussion of more endogenous drivers, e.g. workload and volumes, that might be used as explanatory variables within the models, but are in part influenced by the choices of the GDNs themselves and technically are better described as inputs to the delivery of network services and outputs.

We then look to identify possible explanatory variables[87] for use in the econometric models that could be used to reflect / proxy these cost drivers.

[87] An explanatory variable will have data available but may only constitute a partial proxy for the effect expected from a cost driver.

6.1.1. Exogenous cost drivers

We expect the **number and type of consumers and the scale of the network** to deliver services to these consumers would have a strong external influence on GDNs costs.

There is strong reason to believe that the scale of activities that need to be undertaken across the GDNs' businesses will have a strong external influence on their total expenditure, in particular, the costs of operating and maintaining the gas networks for a given fixed stock of capital assets. GDNs' operations and maintenance expenditure, like network operators in other sectors, includes operational control, emergency call out, routine and non-routine maintenance work and various organisational overheads[88] of looking after the existing fixed asset base. The aggregate size of these tasks and activities can be expected to depend on the size of the gas distribution network operated by the GDN. [89]

As well as high-level company scale, as captured through variables such as total number of customers or connected properties to the network, scale in particular activities or variables may also be an important driver of the GDNs relative cost performance. For example, the number of particular types of customers and connected properties, and the pressure tier of their connection.[90]

As well as number of customers, demand metrics such as the peak demand served by each GDN, or proxies such as the volume of annual network throughput, will also influence the scale of GDNs' activities. They can be considered **measures of output** from GDNs.

Quality of service and other output requirements imposed via the GDN licence and/or the price control may also influence relativities in network costs.

From a cost benchmarking perspective, the standards and outputs (e.g. related to safety), and indeed any external environmental conditions, that are common to all the sampled businesses, can be omitted from the efficiency assessment.[91] However, if there are good reasons to believe that output service standards and quality expected and/or delivered in one GDN area are fundamentally different to other GDN areas, then this may need to be controlled for in the benchmarking.

In principle differences in quality could be picked up pre-regression through normalisation adjustments (similar to how Ofgem applied normalisation adjustments for regional wage input prices at GD1) or including explanatory variables for service quality directly in the regressions, similar to the options available

[88] In practice, organisational overheads may not necessarily be directly related to network scale in that we understand many of the maintenance activities are outsourced by GDNs to third party organisation. Therefore, direct labour, for example, may not always be increase proportionally to network scale / size of asset base.

[89] See Turvey (2004): 'On benchmarking and TFP comparisons'

[90] Our expectation is that domestic customers connected to the GDNs low-pressure distribution systems will be one of the main drivers of the companies' costs and, as discussed below, that this effect would tend to be higher in densely populated urban environments. In contrast, I&C users connected to the higher-pressure tiers of pipes may be a less significant cost driver, although still important.

[91] For example, if the quality standard is constant across GDNs then it should be the volume of repairs and emergency events that are a driver of relative efficiency which is more likely to be dependent on the number of size and characteristics of the network – so, number of customers, asset base etc.

for the treatment of sparsity / urbanity effects. However, one reason not to include quality within the assessment is that similar to asset age and condition (see below), quality is not a fully external cost driver for the GDNs. Ofgem highlighted this issue in the GD2 sector specific methodology annex noting:

> "*we have reservations over the incorporation of quality. Quality is an output which is in GDNs' control, undermining its use as a cost driver. Also, incorporating quality in regression analysis does not inform the value that consumers place on the level of quality delivered.*"[92]

Another reason not to include quality of service in the analysis is it may be challenging to do so in more disaggregated benchmarking models, where the link between quality of service and the costs of specific activities is more difficult to establish through a set of relevant explanatory variables.[93]

In contrast, at a very aggregated cost modelling level, service quality and customer satisfaction, may not be a significant driver of GDNs costs (although this is proposition that can be tested empirically through the model development process).

We would expect that the **age and condition of the gas distribution system** that a GDN operates will impact the costs it incurs.

Although the composition and characteristics of gas distribution networks across GB are similar[94], the **age and condition** of the network will influence the operational efficiency of the network and the need for maintenance, enhancement and asset replacement investment programmes. Differences in the age and condition of GDNs networks may one of the causes of differences in the observed costs of the company.

However, an important point to consider with these causal factors of expenditure, is whether they can be considered *exogenous* drivers of GDNs' costs (e.g. related to the historical development of each distribution network) or instead the result of endogenous decisions made by the GDNs themselves. There is a risk, from a competitive benchmarking perspective, that while capturing information on asset age, health and composition as explanatory variables of GDNs' costs within Ofgem's models provides a more accurate representation of the expected *input* activity and the expenditure of each of the GDNs, the benchmarking framework could also create perverse incentives for the GDNs.

For example, the MEAV of the gas distribution network – or a subset of the network related to a particular cost pool or activity – could be considered as an explanatory variable to reflect the scale, characteristics and composition of the network asset base.[95] Monitored metrics of asset health or condition risk could be used as a measure to understand how age and condition of network assets impacts GDN costs. However, MEAV or measures of network asset health condition in particular, are in principle under the control of the

[92] Ofgem, December 2018. RIIO-2 Sector Specific Methodology Annex: Gas Distribution. Paragraph 6.24.

[93] In part this is an argument for more aggregated (e.g. totex or middle-up) based benchmarking.

[94] e.g. similar pipe materials, operating pressures, diameter ranges, technology / techniques employed to repair / replace the pipe networks.

[95] Using MEAV as a cost driver is effectively a weighted asset scale / volume index, where the weights are determined by the assumed unit costs of the asset base.

GDNs.[96] Using these as explanatory variables within benchmarking models may incentivise the companies to increase MEAV or run a relatively poor-quality network if this allows short term outperformance of regulatory assumptions today.[97]

Of course, the GDNs may not be able to materially impact on MEAV in the short run as compared to the long run. Participants in the GD2 cost assessment working group, for example, noted that although the GDNs can influence MEAV they would need to spend a significant amount to influence regression results, relative to the fixed historical value of the network asset base. It may also be argued that gas distribution networks involve very long-lived assets, and their current age and condition, is therefore, largely the result of investment decisions that predated GDNs ownership and stewardship of the networks.

The **operating environment** in the GDN licence area will also impact on the total expenditure of the gas distribution operators.

We have set out above how both the size and characteristics of the gas network can be expected to impact network costs. The types as well as number of customers (non-domestic vs. domestic), mix of population, demographics etc. may also impact the activities of the GDN and use of the distribution system.

For example, the types of connections and consumers in one area may have different preferences for how they consume energy compared to another network, influencing the profile of gas load that needs to be served by each GDN. Societal regional factors, such as use of different types of space heating fuels, local input prices etc., may also have a heterogenous impact on GDN costs.

The network may also be more complex to build, maintain and replace, depending on the geography of the GDN's licensed service area – e.g. coastal or mountain area, urban vs. rural operating environment.

Differences in the density / sparsity of population in each GDN licence areas, in particular, influence the companies observed costs. The effect on the costs of each company, however, could be ambiguous:

- GDNs could face lower costs in densely populated areas (e.g. reducing travelling distances for maintenance or duplication of depots and spare parts).
- GDNs could face higher costs in densely populated areas from working in more urban areas (e.g. increase expenditure for traffic management, service diversion, etc.).

Networks and assets within densely and sparsely populated urban / rural areas may, therefore, give rise to different operating environments and associated operating costs. Some specific locations may also impact the expected costs from certain activities[98].

---

[96] GDNs choose the assets they invest in and, to some extent, the maintenance philosophy / asset integrity standard that they target for their respective gas distribution networks.

[97] For example, reduce maintenance expenditure and future regulatory allowances will provide for the relatively poor health of the network at future price reviews, through increased future expenditure allowances.

[98] For example, GDNs may incur higher and more costly non-routine maintenance and repair work resulting from third party activities in heavily built up urban areas due to civil construction work damaging network infrastructure.

Ofgem included various adjustments in its GD1 assessment for urbanity and sparsity. Regarding sparsity, in its Initial Proposals document, Ofgem stated:

> *"The productivity impact of sparsity relates to the productive time lost during the additional time spend on travelling in a sparse area when attending emergency and repairs, ie the extra non-productive time spent on the journey instead of attending to the job … We accept that more resources are required to meet the emergency and repairs requirements in a more sparse area given limited travel patterns and the consequent increase in travel time required to cover operations."* [99]

Regarding urbanity, Ofgem stated:

> *"Some of the GDNs have suggested that there are additional costs associated with working in urban areas. These costs include street works issues such as additional requirements to close roads or put in place traffic controls, premium time working, requirements for full reinstatement of roads and congestion of underground assets. The additional costs can be split into higher than average salaries and other costs that rise from working in an urban environment … We do not consider that there is a need for additional urbanity adjustment for regional labour rates …*
>
> *However we accept arguments that in practice there are lower levels of productivity in London associated with more congested infrastructure, depth of infrastructure and reduced access."*

The **input prices of the GDNs**, both changes over time and differences in prices between regions of GB, may impact GDNs expenditure.

Relativities in regional wages are an example of how differences in *regional* input prices, that are not under the control of the GDNs, might impact the relative cost performance of the GDNs. There are other input prices to the GDNs production of network services, e.g. the cost of materials and price of transport infrastructure, that may in principle also differ depending on the part of the country an GDN serves.

From an econometric modelling perspective, the time effect of changes input prices on GDNs costs – and unit costs more generally (i.e. frontier shift) – can be accounted for within the modelling through the inclusion of a time trend in the model's specification (e.g. a time trend was an explanatory variable included in Ofgem's econometric models used for the final ED1 determination[100]). An alternative is to include a measure of input prices directly in the specified production / cost function.[101]

As we discuss in Section 7 there are a number of different ways that regional differences in input prices (e.g. wages) can also be accounted for in the econometric modelling process. This includes directly introducing explanatory variables into the models or via pre or post modelling adjustments.

---

[99] Ofgem (2012): 'RIIO: Initial Proposals – Cost assessment'

[100] Ofgem (2014): 'RIIO-ED1: Final determinations for the slowtrack electricity distribution companies – Business Plan Assessment'

[101] This approach was investigated by Frontier Economics during the ED1 price control and by CEPA during the model development for PR14. See Frontier Economics (2013): 'Total cost benchmarking at RIIO-ED1 – Phase 2 report – Volume 1' and CEPA (2014): 'Ofwat – Cost assessment – Advanced Econometric Assessment'

The **transition to a low carbon economy** may impact on the operational and capital expenditure each of the GDNs needs to incur in future price controls.

There is considerable uncertainty of the future role of natural gas within the GB energy system as a result of current recommendations and targets for decarbonisation.[102] This may in future have an impact on the expected scale and scope of gas distribution network activities. Although the uncertainties over the future role of the gas network in the energy system were starting to be considered during the GD1 review process, it was not an issue addressed explicitly within the econometric benchmarking.

A report by KPMG for the Energy Network Association (ENA) (2016)[103] considered a range of scenarios for the decarbonisation of space heating within the UK alongside decarbonisation of power and transport (as part of a whole system approach). The four scenarios considered indicated different practical obstacles and incremental whole system costs depending on the pathway for decarbonisation.

These different pathways may be expected to have different external impacts on the incremental costs that GDNs may need to incur at future price reviews. For example:

- Pathways that involve a fall in gas demand / flows may lead to greater decommissioning costs, whether that is in repurposing the gas network for alternative use, or permanent abandonment of assets. They may also require less network reinforcement costs if rising demand on the network is not triggered and scale of network is reduced. [104]
- In contrast, in pathways that involve more extensive role out of hydrogen as a fuel, GDNs may be required to incur some network reinforcement costs. In this case, there may be no or limited reduction in the overall scale of their network businesses, but the scope of some activities involved in providing distribution services may need to adapt.

SGN in its report on the future of gas networks, states that:

> "*We believe a key objective during the GD2 price control period will be to further develop the blend of lower carbon gas within our networks which will allow customers to be supplied with low carbon energy without the need for new appliances or additional gas network investment. In GD2 we would envisage BioSNG will be injected at scale along with greater quantities of biomethane, subject to the required government support being available.*
>
> *We would expect to see hydrogen added to that blend provided there is the evidence to show it can be distributed and utilised safely.*"[105]

While this indicates that changes from decarbonisation are likely to become an increasingly important issue for the GDNs, there is still considerable uncertainty of the impacts on the gas network and, in particular, whether the factors identified above will have a material impact on GDNs costs in GD2.

[102] See CCC (2019): 'Net Zero – The UK's contribution to stopping global warming'

[103] KPMG (2016): '2050 Energy Scenarios: The UK Gas Networks role in a 2050 whole energy system.

[104] See discussion in Frontier Economics (2016): 'Future Regulation of the Gas Grid: Impacts and institutional implications of UK gas grid future scenarios a report for the CCC'

[105] Scotia (2018): 'The future of gas networks'

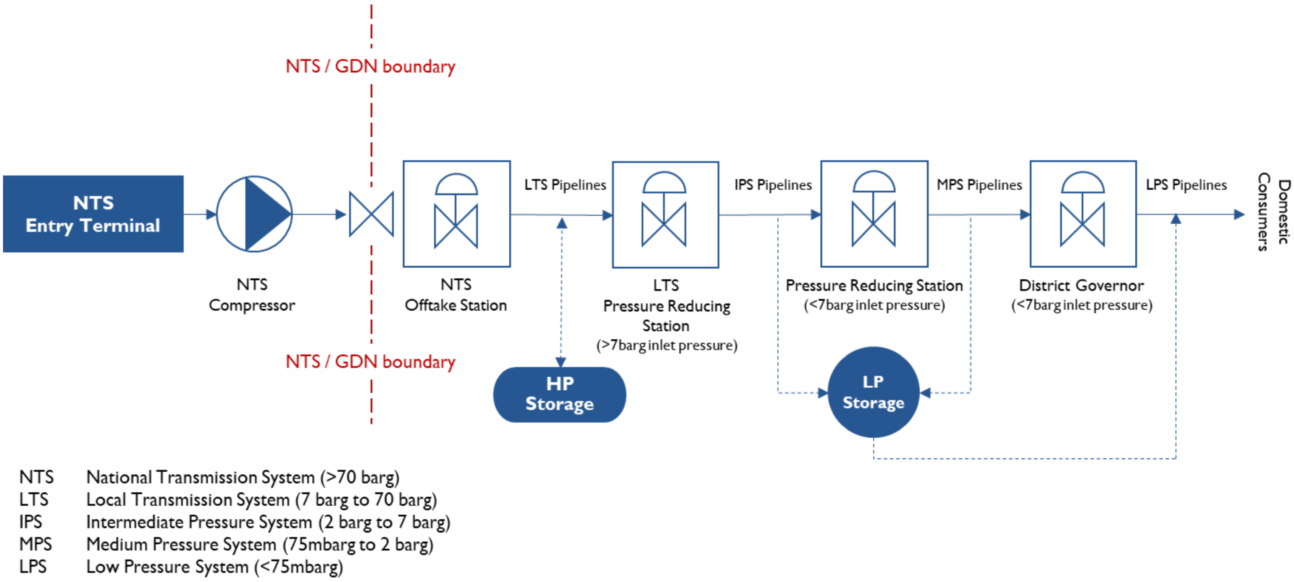
6.1.2. Workload (endogenous input) cost drivers

Ofgem has at previous GD price controls used workload drivers as proxy cost drivers and explanatory variables in its econometric models. We discuss some of the methodological and regulatory issues with this approach in section 6.2 below (including endogeneity issues). However, given workload drivers have been a significant part of previous GD cost assessments, we have considered, from a purely engineering perspective, drivers Ofgem could consider if using workload drivers to benchmark GDNs efficiency.

We have used an asset management approach to identify relevant cost / workload drivers by asset grouping, on the basis that the network assets of all the GDN's in GB were designed, constructed and are currently operated under a common set of industry standards and recommendations developed by the former British Gas (subsequently Transco) and the Institution of Gas Engineers and Managers. Furthermore, all GDN's must comply with the same legislative requirements with respect to safety.

Although currently under different ownership and operation, there has been very little, if any, divergence in the way in which these assets are designed, built and operated. – i.e. all the GDN's are carrying out the same activities on these assets but in slightly different ways that result in greater or lesser efficiencies. The various network asset groupings are also common across all GDN's (as shown by Figure 6.1 below[106]) and, we understand, are the cost pools that all GDN's base their financial plans around.

Figure 6.1: GDN network asset groupings



We understand that each GDN currently undertakes this process under its asset management planning activities (ISO 55000 Asset Management) and that efficiencies fall out of the various asset management strategies that each GDN chooses to adopt for each asset grouping (including the opex v. capex trade-offs based on evidential asset condition data and assessments of asset health and criticality). On this basis, a simplified illustration of asset groupings (cost pools) and the associated opex, capex and repex cost drivers that might be considered for these groupings, is provided in the table below.

[106] The asset groupings comprise: LTS Pipelines; IPS Pipelines; MPS Pipelines; LPS Pipelines; Pressure Reducing Installations – 3 broad categories; Services; Meters; and Storage Facilities – high pressure and low pressure (currently being decommissioned).

Table 6.1: Asset grouping illustration

Asset Grouping	Cost Drivers		
	Opex	Capex	Repex
LTS Pipelines	Number of OLI surveys Number of completed pipeline repairs Annual costs of: • CP monitoring • aerial survey • 3rd party damage prevention • valve maintenance • special crossings	Kms of new LTS pipeline Kms of LTS reinforcement pipeline	
IPS Pipelines	Number of completed pipeline repairs Annual costs of: • CP monitoring • 3rd party damage prevention • valve maintenance • special crossings	Kms of new IPS pipeline Kms of IPS reinforcement pipeline	Kms of replaced or diverted IPS pipeline (non-rechargeable)
MPS Pipelines	Number of completed pipeline repairs Annual costs of: • CP monitoring • leakage survey • 3rd party damage prevention • valve maintenance • special crossings	Kms of new MPS pipeline Kms of MPS reinforcement pipeline	Kms of replaced or diverted MPS pipeline (non-rechargeable)
LPS Pipelines	Number of completed pipeline repairs Annual costs of: • CP monitoring • leakage survey • 3rd party damage prevention • valve maintenance • special crossings	Kms of new LPS pipeline Kms of LPS reinforcement pipeline	Kms of non-discretionary mains replacement Kms of discretionary mains replacement
Pressure Reduction Installations (PRIs)	Number of annual inspections Number of annual functional checks	Number of new build PRI's	Number of PRI rebuilds/refurbs
Services	Number of completed service repairs	Number of new service connections • statutory • competitive	Number of service replacements under IMRRP Number of service replacement non-IMRRP
Meters	Annual costs of meter maintenance: • domestic • commercial • industrial	Number of new meter installations • domestic • commercial • industrial	Number of replacement meter installations • domestic • commercial • industrial
Storage Facilities	Annual costs of maintenance for: • LP holders • HP bullets		

Source: CEPA

6.1.3. Identifying explanatory variables

Having identified plausible external and workload drivers of GDN costs, for econometric modelling purposes, it is necessary to determine observable explanatory variables that can be used as proxies.

Potential variables that could be used to reflect each of the exogenous cost drivers discussed in the previous subsection are discussed in the table below. In each case we provide:

- **Explanation / commentary** on how each explanatory variable might be expected to influence the GDNs costs.
- Building on the discussion of model selection criteria in Section 2, the **expected sign of the explanatory variable** if included in a regression model.
- A short **discussion of some of the issues** that may be associated with the use of each of the listed explanatory variables.

Note that we have not considered explanatory variables related to the transition to a low carbon economy given it is unclear to what extent this will be a material cost driver in GD2, or at least a statistically significant enough driver of GDNs costs to be included within the relatively aggregated benchmarking analysis that is envisaged under the econometric workstream of the cost assessment.[107]

[107] We might expect costs driven by biogas or hydrogen conversation of the grid, for example, to be captured through a separate technical / qualitative assessment outside of the modelling.

Table 6.2: Possible explanatory variables for cost benchmarking – exogenous drivers

Explanatory variables	Explanation / comments	Expected sign of explanatory variable?	Issues
High-level output / demand metrics			
Customer numbers	The expected expenditure of GDNs, operational and maintenance expenditure in particular, would be expected to increase with scale, as proxied by the number of customers connected.	Positive	As discussed above, the type of customers served by a GDN may also be important to consider. For example: • Domestic vs. Non-Domestic • Mix of rural and urban customers
Throughput (GWh)	Expected demands on the gas distribution network provide a proxy for the scale of the network. The higher the throughput of the network, the larger the expected scale of the network.	Positive	Network infrastructure is sized to meet expected peak demand, so the volume of gas distributed does not give rise to direct cost. However, would expect a positive correlation (albeit imperfect) between throughput and peak demand (i.e. higher throughput, higher peak demand).
Measure of peak demand / capacity	Provides a proxy for maximum system capacity. It also acts as an output variable as it is a measure of yearly peak demand.	Positive	Has been used in some international benchmarking studies.[108]
Climate (heating degree days)	Provides an output measure. The higher the number of heating degree days the higher the level of gas consumption, which may lead to higher costs incurred by GDNs.	Positive	Can be proxied by a simpler variable such as throughput?

[108] Journal of Regulatory Economics (2002): 'International Cost Benchmarking for Monopoly Price Regulation: The Case of Australian Gas Distribution', Carrington. R, Coelli. T, Groom, E. Volume 21, Issue 2, pp 191 - 216

Explanatory variables	Explanation / comments	Expected sign of explanatory variable?	Issues
Service quality metrics			
Shrinkage	Shrinkage represents the volume of gas lost from the network. Assuming a negative coefficient, this may provide an indication that as companies spend more on maintenance and capex, they provide a better quality of service. Including shrinkage in the model specification may help account for the trade-offs the GDNs face in reducing shrinkage (costs vs. value this provides to customers / society (e.g. avoided carbon emissions)).	Negative	A key driver of reduction in shrinkage has been iron mains replacement. If the sign is in fact positive, the inclusion of this explanatory variable could provide a perverse incentive to increase the level of shrinkage. Shrinkage should be expected to reduce as the iron mains replacement programme is delivered. Level of shrinkage (along with other service quality metrics discussed below) expected to reflect past expenditure and incentives under price control output measures – interaction would need to be considered before including shrinkage and other quality measures below. As discussed above, shrinkage (and other metrics of service quality discussed below) are an output which to an extent is under the control of the GDNs.
Interruptions (duration, volume, etc.)	Assuming a negative coefficient, this may provide an indication that as companies spend more on maintenance and capex, they provide a better quality of service through reduced number of interruptions.	Negative	If the sign is in fact positive, the inclusion of this explanatory variable could provide a perverse incentive. The inclusion of such an explanatory variable must be carefully considered given an interaction with any financial incentive for interruptions. As above with shrinkage, interactions with price control output measures and historical network spend would need to be considered.
Customer satisfaction survey scores	Assuming a positive coefficient, this may provide an indication that as companies spend more on opex (e.g. business support costs or emergency call out services), they provide a better quality of service through increased customer satisfaction scores.	Positive	If the sign is in fact negative, the inclusion of this explanatory variable could provide a perverse incentive. The inclusion of such an explanatory variable must be carefully considered given an interaction with any satisfaction financial incentive.

Explanatory variables	Explanation / comments	Expected sign of explanatory variable?	Issues
Operating environment and system characteristics			
Network length	While network length is a proxy for the scale of network activities, it is a relatively aggregative scale variable.	Positive	Use of actual network length in a regulatory context may give rise to perverse incentives to install more network assets to appear more efficient.
MEAV	At a relatively high-level, MEAV captures the composition and complexity of the gas distribution system's asset base – which drives tasks and activities related to operation and maintenance – as well as scale. At GD1, Ofgem used MEAV as a scale variable in its econometric modelling.	Positive	See critique above on network length. Although GDNs may need to commit significant spend to materially impact on MEAV in the short run. In an environment where network operators are being expected to identify smarter / operational solutions to network issues[109], is MEAV a good benchmarking metric given that GDNs will improve their efficiency scores by increasing the volume of installed assets in their networks? Can different MEAVs be used for different cost pools / activities? [110]
% of network assets expired	Provides a proxy for the age and condition of the network. The higher the % of network assets expired the higher the opex and maintenance costs as GDNs respond to a greater frequency of faults.	Positive	This variable could be considered quite endogenous. Would need to ensure that the inclusion of this variable does not perversely incentivise GDNs not to replace their network. May not capture the condition of all network assets.
% of iron mains replaced	Older iron mains would be more likely to leak due to the effects of corrosion than the more modern polyethylene pipes, which are now being installed.	Negative	As with % of network assets expired, this variable could be considered quite endogenous (indeed, likely more so given some GDN discretion over how iron mains are replaced).

[109] See for example, SGN 'Real time networks project'

[110] This approach was used at GD1, where MEAV for different activities was used in individual cost activity level regressions (e.g. maintenance).

Explanatory variables	Explanation / comments	Expected sign of explanatory variable?	Issues
	Companies who have replaced a high percentage of iron mains will have relatively lower workload levels (emergencies, repairs and maintenance). Similarly, all else being equal, companies who have replaced a high proportion of iron mains in the past will have less iron mains to replace in the future, leading to lower repex.		Better to consider broader measures of network asset health / risk (as collected by GDNs) than this simple metric?
Network asset health and risk reporting	Ofgem and the GDNs have been working to develop various measures of asset condition and health as part of the Network Asset Risk Metric (NARM). To the extent these metrics, developed as part of the Networks Output Methodology (NOMs), provide an indicator of network performance and condition, they could be used to control for the effect of asset condition within the econometric modelling.	Ambiguous	Like other variables under this category could be considered relatively endogenous compared to other categories such as high-level output metrics. We understand that a common asset health and risk reporting methodology in the gas distribution sector has only been developed relatively recently.[111] May raise issues for models that include long time periods of historical data. If including this as a variable in the benchmarking models, any interactions with other elements of the price control regime would need to be considered.
Total connections / total length of mains	A higher number of connections per a km of mains suggests a denser network. There are reasons why high density may drive increased costs (e.g. access issues, higher disruption), but low density (i.e. sparsity) can similarly increase costs through longer travel time and emergency response requirements. Therefore, density is a priori considered to have an ambiguous effect. Other sectors have considered including quadradic terms for this type of variable to reflect a possible "U-shape"	Ambiguous	Can be treated as a single variable which helps with degrees of freedom.

[111] See for example ENA (2017): 'Network Output Measures: Health and Risk Reporting methodology and framework consultation'

Explanatory variables	Explanation / comments	Expected sign of explanatory variable?	Issues
	relationship between the impact of sparsity and urbanity on network utility costs.		
Percentage of "urban assets"	In the water sector, Arup and Vivid Economics have previously suggested the use of variables that reflect the rural-urban split of the distribution of populations or assets.[112] They argued assets in urban areas cost more to operate due to, for example, harder access, traffic permissions and restricted land footprints. High population density urban areas may also impact level of third-party damage and consequential maintenance, repair, etc. costs.	Ambiguous	Limited regulatory precedent. As discussed above, while there may be cost impacts driven by sparsity/urbanity factors, there are costs and benefits associated with both. As we discuss later in the report, impacts of urbanity can also be captured through pre or post modelling adjustments (as was the case in GD1).
Ofwat high density style variable	This variable reflects the percentage of the population living in densely populated areas. It is constructed using ONS population and population density data at a local authority district (LAD) level. Each LAD is defined as highly dense if people per square kilometre (measure of density) is greater than a specified threshold (e.g. 2000). The variable is then created by dividing total population in areas that are defined as dense by the total population for each company. Hence, the variable is between 0% and 100%.[113]	Ambiguous	Relatively complex to calculate compared to alternative variables.

[112] Vivid Economics and Arup (2017), *'Understanding the exogenous drivers of wholesale wastewater costs in England & Wales'*, available here.

[113] See Ofwat (2019): 'PR19 – Supplementary technical appendix: econometric approach'

Explanatory variables	Explanation / comments	Expected sign of explanatory variable?	Issues
Input prices			
Input prices, including regional differences in input cost	Changes in the input prices, for example, the cost of labour and materials (e.g. steel or polyurethane (PE) pipes) required to deliver network outputs over time or between different regions of GB, may account for some of the observed differences in GDNs costs both at a particular point in time, or over a period of time. Including input price indices (e.g. labour or materials) in the regression specification can help to avoid the risk of omitted variable bias.	Positive – as input prices increase, costs increase	Requires a robust set of input price indices to be identified. The impact of regional wages on GDNs costs can also be captured through a pre or post modelling adjustment (the former approach was adopted at GD1, as discussed in later sections of the report).
Other			
Time trend variable	A time trend variable can be used to control for effect of GDNs real expenditure changing over time relative to the other cost drivers and explanatory variables, including input prices.	Ambiguous – given that a time trend will account for a number of factors changing over time including ongoing productivity	A time trend may capture a number of drivers of changes in unit costs over time, e.g. changes in quality of service or the impact of ongoing efficiency as well as changes in input prices (i.e. frontier shift), to the extent these effects are not captured by other explanatory variables.

Source: CEPA analysis

Potential variables that could be used to reflect the types of endogenous workload / input drivers discussed in the previous subsection are discussed in Table 6.3 below.

As with the more exogenous explanatory variables reviewed in Table 6.2, in each case we provide:

- **Explanation / commentary** on how each explanatory variable might be expected to influence the GDNs costs.
- Building on the discussion of model selection criteria in Section 2, the **expected sign of the explanatory variable** if included in a regression model.
- A short **discussion of some of the issues** that may be associated with the use of each of the listed explanatory variables.

Table 6.3: Possible explanatory variables for cost benchmarking – endogenous drivers

Explanatory variables	Explanation / comments	Expected sign of explanatory variable?	Issues
Public reported Escapes			
Total number of PREs	External + Internal PREs. Each escape requires the GDN to send an operative to investigate the report. PREs as a whole are not necessarily caused through the fault of the GDN, as this includes third-party damage and escapes from outside the controlled network.	Positive	May potentially reward companies where the network is in relatively poor condition (except in the case of internal PREs). Risk of incentivising a 'narrow' repair strategy if the variable does not recognise the potential for multiple repairs following a single report. May not capture the complexity and differences of repair costs incurred between GDNs.
External PREs	External PREs are those found to originate on the GDN's network.	Positive	
Internal PREs	Internal PREs are those found not to originate on the GDN's network (e.g. from a customer's appliance). There are therefore largely outside of company control.	Positive	
External condition reports	External condition reports are external PREs that require repair by the GDN.	Positive	
Capex workload drivers			
Capex connection workload drivers	The number of new connections will drive the number of new statutory and competitive service connections, new meter installations etc.	Positive	Used in GD1. Potential endogeneity issues, as discussed in Section 6.1 above.
Capex mains workload drivers	As indicated in Table 6.1, capex mains workload likely to be proxied / driven by km of new pipelines installed by the GDNs.	Positive	
Repex workload drivers			
Various factors have been identified as drivers of GDN repex workload and costs:	Km of mains replaced or diverted	Positive	Used in GD1 cost assessment
	Km services replaced	Positive	
	Technique – e.g. open cut trench, live mains insertion, dead mains insertion Broader pipe risk management options, e.g. remediation, was identified by GDNs in GD2 cost working group	Unclear	Very endogenous variable – adopted technique rather than volume of work based.

Explanatory variables	Explanation / comments	Expected sign of explanatory variable?	Issues
	Number of service re-laid and individual length of renewed services	Positive	These factors may be a driver of repex but can they be accommodated in aggregated econometric models, particularly given the relatively small data set available?
	Diameter of pipe replaced	Unclear	
	Number of tie-in connections per km required	Positive	
	Location – e.g. carriageway, modular footpath or grass verge, was noted as a relevant driver in the GD2 cost assessment work group. Possible explanatory variables that were suggested included: the proportion of mains in the footpath, carriageway and verge.	Unclear - dependent on location variable chosen?	

Source: CEPA analysis

6.2. Other model specification topics

In this section we provide thoughts on a number of broader issues for the specification of econometric benchmarking models. This includes:

- Selection of functional form.
- Use of time dummies and time trends in the models.
- Use of workload drivers.
- Use of capex smoothing to address 'lumpiness' issues.
- Use of CSVs.

6.2.1. Selection of functional form

The specification of the functional form is an important aspect of the econometric methodology. Different functional forms introduce different assumptions on the relationship between the dependent and explanatory variables. In particular, the choice is framed by assumptions on the nature of economies of scale.

Among the multiple options presented in the academic literature, the models used in GD1 employed a **Cobb-Douglas** form. This is a standard approach used in cost assessment literature as it allows for economies of scale. In the case of a single explanatory variable, this can take the form:

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

Where β_1 represent the coefficient to be estimated and ε reflects the component of costs not explained by the cost driver. β_1 can be interpreted as the elasticity of costs with respect to the driver – if the cost driver increases by 1%, cost can be expected to increase by β_1%. Therefore, if β_1 is less than one, an activity can be said to have increasing returns to scale (with respect to the given driver).

Cobb-Douglas models are relatively easy to replicate and interpret but suffer from the imposition of a single degree of economies of scale being assumed across the industry (i.e. all companies are assumed to have the same level of economies of scale). Therefore, the use of this form could require the introduction of other variables that can reflect variations in economies of scale across companies.

In PR14, Ofwat implemented a version of the 'translog' functional form, which introduces squared and cross-product terms. This is a highly flexible functional form and used routinely in academic literature.[114]

However, the use of these models makes it more difficult to identify the specific effect of each variable on costs. They also require the introduction of a larger number of explanatory variables in each one of the models to account for these variations. In the CMA determination for Bristol Water following the company's PR14 appeal, the CMA noted that Ofwat's models were difficult to interpret and, given the small sample size and the data requirements of translog, its use seemed overly ambitious.[115]

---

[114] It can be noted that Cobb-Douglas is nested within the more general translog form.

[115] CMA (2015), *'Bristol Water plc. A reference under section 12(3)(a) of the Water Industry Act 1991'*, available here.

When deciding on a functional form, there are a number of principles to consider drawing from the high-level model selection criteria discussed in Section 2:

- **Transparency**. Ideally, models should be simple and easy to interpret. As noted above, the CMA has found translog models to be overly complex. In GD1, all Ofgem's models used a single explanatory variable – in part to reduce the degrees of freedom lost in the small dataset – but this also avoided the interdependencies that occur with the multiple interactions in the translog form.
- **Technical justification**. During initial development of econometric models, the functional form should also be based, in part, on an underlying economic and engineering understanding of a GD business. For example, if there is a strong rationale for believing an explanatory variable has a U-shaped relationship with costs, this may justify the use of squared terms in the functional form.
- **Specification tests**. While initial functional specifications can be informed by a priori rationale, the final form employed should perform well against specification tests. The Ramsey RESET test (see Section 2) is a general test for omitted nonlinearities (e.g. quadratic and/or interaction terms). If a model fails to satisfy this criterion, a different functional form may be justified.

We would recommend that Ofgem **adopts the Cobb-Douglas functional form** for its GD2 modelling given the relatively small sample size and the fact that this is likely to allow a greater number of explanatory variables to be included in the model specification compared to other alternatives.

6.2.2. Modelling "time effects"

An issue for Ofgem to consider for GD2 is whether to use a time trend variable or year dummy variables within the econometric analysis.

A year time trend variable or yearly dummy variables are important where it is considered that the GDNs' real expenditure (i.e., expenditure adjusted for RPI or CPI) change over time relative to the costs drivers and explanatory variables. In other words, if the values of the explanatory variables stayed the same over time, is there any expectation that the real expenditure will change over time?

As indicated in Table 6.2, real expenditure is expected to change over time due to ongoing efficiency and real price effects (RPEs) i.e., frontier shift, and potentially other exogenous factors such as changes in quality of service (which may not be picked up in other measures). Therefore, we consider it appropriate to test whether a time trend variable or yearly dummy variables are required in Ofgem's models.

The key difference between a time trend and dummy variables is that a time trend assumes that the annual changes are the same and does not capture any one-off annual impacts,[116] while the yearly dummy variables allow for each year to have a different impact on the average performance of the GDNs. Specific year dummy variables can be used with a time trend if any outlier years are identified (e.g., due to one-off expenditure that affects all GDNs' equally).[117]

[116] Aside from capturing these in the average trend.

[117] In contrast, if the one-off expenditure affects only one or a few GDNs then this should be dealt with within ex-ante ('normalisation') adjustments.

For GD1, Ofgem included year dummy variables within its benchmarking model. For ED1, Ofgem used a time trend. We also note that Ofwat used a time trend for PR14 and the AER in Australia also uses a time trend. Ofwat considered including a time trend or time dummy variables for individual years in its PR19 models, however, concluded that "*such variables did not have a stable or significant effect on the model.*"[118]

While we consider there are reasons to support the use of a time trend variable over annual dummy variables – given Ofgem is seeking to capture the average change in frontier shift, for which the time trend is more appropriate – as identified by the CMA during the PR14 Bristol Water appeal[119], there may be circumstances where year-to-year fluctuations in industry costs do not fit well with a time trend effect, but impact on all GDNs simultaneously (time dummy specifications will control for such effects).

We therefore **recommend that both time trend and dummy variable approaches are tested as part of GD2 model development process**. While both approaches should lead to similar results, cross-checking between the two approaches may help identify any issues that need further investigation.

However, it is important to note that it may not be necessary to include a time trend or time dummies in the models if it proves statistically insignificant and/or produces results that are not deemed sensible. As discussed above, Ofwat at PR19 decided not to include a time trend or year dummies in its wholesale econometric cost models as they were not statistically significant in its models.

6.2.3. Workload drivers

As noted above, Ofgem has made extensive use of workload drivers in previous price control benchmarking exercises. This was the case during its GD1 cost assessment, both in the aggregated totex and disaggregated models Ofgem developed, and at ED1.

The repex programmes undertaken by the GDNs are perhaps the best example of this in the gas distribution sector. The level of activity that has historically being undertaken has in part been driven by *external policy* – the HSE requirements for the IMRP – as well as decisions by the GDNs of how they choose to maintain and replace their ageing network assets.

The scale and complexity of the repex programme for each GDN is (in part) a consequence of the condition of the asset base each management group inherited and imposed HSE regulatory policy, as well as the choices they made (and will in future make) on how to remove remaining risk from the network. Ofgem in a number of its regressions for GD1 used workload drivers to capture these effects.

The text box below reviews the approach that was adopted in the final proposals for benchmarking repex as part of the series of bottom-up regressions Ofgem used to derive final allowances. A similar type approach was also used for the modelling of capex activities and some of the disaggregated opex activity models that Ofgem developed for GD1.

[118] Ofwat (2019): 'PR19 – Supplementary technical appendix: econometric approach', p. 16

[119] CMA (2015): "'Bristol Water plc – A reference under section 12(3)(a) of the Water Industry Act 1991: Report", p. A4(2)-22

Text Box 2: Treatment of repex programme in RIIO-GD1

The GD1 final proposal split repex workload into discretionary workload and non-discretionary workload:

Non-discretionary repex:

- Tier 1
- Tier 2A (above risk threshold)
- Other non-standard mains
- Services

Discretionary repex

- Mains and associated services (tier 2B below risk threshold, tier 3, iron mains > 30 metres from a property, other mains), multi occupancy buildings (MOBs)

The efficiency of all mains and services repex were assessed in the Final Proposals using regression modelling techniques. This involved taking synthetic unit costs (£/m for mains, £/service for services) for different types / categories of mains and services, multiplying the synthetic unit costs by GDN submitted workload (km mains, no. of services for services) to derive a synthetic value / cost of workload, and regressing the synthetic value / cost of workload explanatory variable against actual repex spend.

Effectively costs are regressed on a manufactured variable of activity (the explanatory variable) and so may be described as a weighted unit cost assessment. While this approach is likely to in part reveal underlying differences in inefficiency between the GDNs, the costs of delivering a safe, reliable etc. gas network will also depend on the mix of work undertaken which could be done more or less efficiently depending on how each of the GDNs choose to structure their maintenance and asset replacement programmes. [120]

Source: CEPA

The advantages of using workload drivers in the modelling include:

- Used within both disaggregated and aggregated models, they can be used to help control for the effects of different workloads by the GDNs that may be due to factors outside of the companies' control, e.g. variation in asset condition that drive variations in GDNs year-on-year costs, e.g. maintenance and repair costs, or repex work due to the inherited state of the network.
- It is a potentially useful approach when attempting to accurately forecast costs using regression-based models. To the extent that GDNs are expected to have very different year-on-year fluctuations in costs that do not reflect changes in aggregated scale variables (e.g. MEAV or no. of customers) the regression models can be used to specify an expected relationship between costs and workload (whether based on forecast or historical cost data, smoothed vs. unsmoothed data, etc.) and then used to produce a forecast cost allowance based on either accepted company forecasts or a detailed technical / engineering review of forecast workload volumes.

---

[120] SGN in their GD2 cost assessment working group presentation note: "*Repex efficiency scores are very volatile and companies switch positions over time – suggesting that cost driver is not properly explaining Repex well. Alternative cost drivers could be considered:*

- *Using mains laid does not recognise companies that optimise design and avoid higher mains workload*
- *Need to account for of abandonment ratios*
- *Consider use of standardised industry abandonment ratio which is set for GD2*
- *Consider impact of higher insertion*
- *Recognise broader pipe risk management options, e.g. remediation"*

As highlighted in Table 6.2, other repex drivers noted in the GD2 cost assessment working group were: abandonment to lay ratios, the proportion of mains in the footpath, carriageway and verge, and number of connections. SGN also commented on the use of synthetic unit costs stating: *"Concerns over synthetic unit costs: Outdated and may no longer be a good reflection of the relativities between costs of different types of work; they do not test if workload levels are efficient."*

- In contrast, one of the criticisms of more aggregate benchmarking at a totex level using primarily scale variables (i.e. excluding measures of asset condition / health or workload drivers) is that they may be biased by unobserved heterogeneity.

Some of the well-known disadvantages of using workload drivers are:

- Incentives problems are typically identified with using explanatory variables within the cost modelling that are within the control of the company and its management.

  Under certain specifications, workload drivers could lead to perverse incentives where companies are rewarded for running relatively poorer condition networks provided price control outputs and deliverables under the broader regulatory settlement can still be met.

  Using workload drivers incentivises the GDNs to put forward high workload forecasts in business plans, even if in practice these are not delivered.

- As discussed in the text box above, using synthetic workload explanatory variables, particularly within disaggregated activity cost models, may mean that the econometric models are really only benchmarking weighted average unit costs of the GDNs during the selected sample period for the regression analysis.[121]

There are also a number of more practical issues to consider with the use of workload drivers as explanatory variables within the context of the GB gas distribution sector:

- While at previous price reviews, GDNs may have had less control over the workload in their repex programmes, the change in HSE policy since the start of GD1 has meant that the GDNs have more freedom to select mains and services replaced than previous price reviews. Arguably this may extenuate the endogeneity / incentive problems from using workload drivers than was the case when the approach was originally applied in GB gas distribution sector benchmarking.

- The model coefficients in the *disaggregated* models are relatively easy to interpret from a technical / economic logic point of view, as indicating a benchmark of unit costs for different work activities within the industry (although arguably the approach is less transparent for the repex programme where a synthetic workload driver is used). However, this is not the case for the more aggregated totex models when combined in a CSV.[122]

- Interactions with other areas of the price control – it may not be appropriate to reflect some aspects of activity within the cost allowance modelling if they have already been accounted for elsewhere in the price control.

---

[121] This is not necessarily a problem if the only alternative is following a more "bottom-up" cost assessment methodology. We note that a unit cost x work volume approach is normally how a more technical / engineering-based cost assessment is undertaken. Therefore, the same issues with use of workload volumes apply in both cases. The issue in this case, is perhaps more whether a detailed engineering review of benchmark unit costs for a programme of work is considered to be a more or less effective approach for setting disaggregated cost activity allowances, or whether the type of regression based analysis Ofgem has used at previous GD price controls is a more effective, aggregative, benchmark of the weighted unit cost of delivering a weighted volume/package of work.

[122] For example, the CSV used in the totex regression for GD1 Final Proposals, includes a combination of workload and scale variables. This makes it difficult to test the engineering / economic logic of the modelled relationship between the adopted cost drivers in the modelling and changes in totex. We discuss this issue in further detail in the subsection below on use of CSVs in the benchmarking framework.

For example, GDNs are provided a pass through for most shrinkage, with an incentive to further reduce it below a baseline; the introduction of financial incentives on interruptions would cause similar interaction issues (see discussion above in Table 6.2).

The GDNs repex programme, in particular, raises a number of challenges of Ofgem not using workload drivers within its econometric modelling. However, as discussed above their use – particularly in more aggregated totex models – does raise a number of issues.

We consider that Ofgem should continue to develop models in GD2 that include workloads as explanatory variables. However, it may wish to test more aggregated scale (e.g. Totex or Opex plus) models, similar to one of two totex models used for the final ED1 determination, that rely on explanatory variables that are more exogenous than work variables. Arguably this may be a more feasible approach with the greater flexibility GDNs have been provided on how they deliver their repex programmes, following changes in HSE policy, and the relative maturity of the repex programmes within GB.

However, the predicative power of benchmarking models that exclude workload variables would need to be carefully considered given the risk that they may suffer from omitted variable bias, in particular, for factors that cannot be easily quantified such as condition of the network and typology. While in principle the specific external operating environment of each GDN can be accounted for in the model (see explanatory variables discussion in Table 6.2) there is still a risk of unobserved heterogeneity which the introduction of workload drivers, albeit relatively endogenous drivers, may help to better account for.

In addition, we expect that models that place greater weight on scale variables in the model specification, may require significantly greater use of ex post adjustments to the modelling[123], which may lead to questions over the original model specifications in the first place.

6.2.4. Capex smoothing

As discussed in Section 5, capex in network companies is generally 'lumpy' over time, which is either due to the need to replace existing assets as and when needed or because expansion of a network is on a stepped basis rather than continuous. This means that capex does not generally move 'smoothly' in line with the cost drivers, which causes difficulties with the modelling estimation in more aggregated (e.g. totex) models.

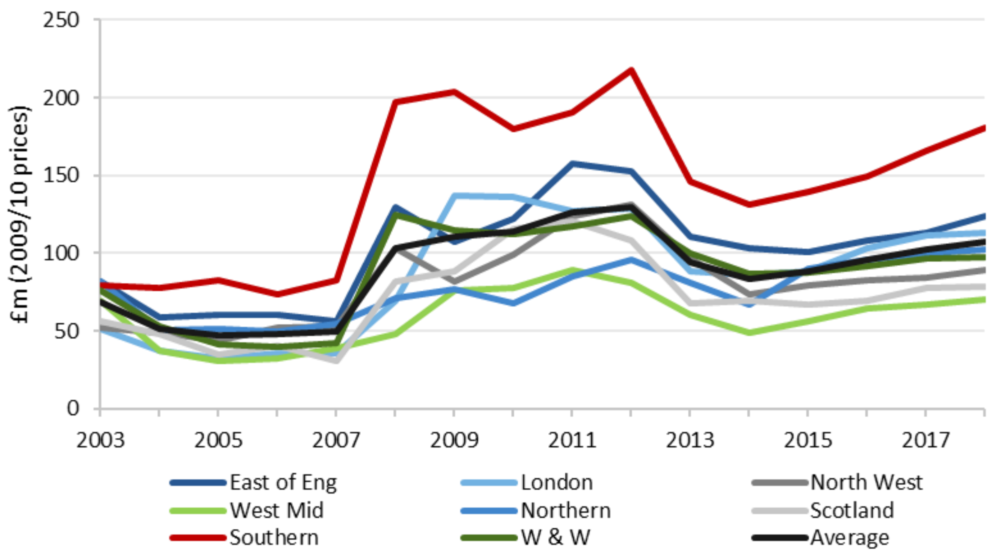
One solution that has been adopted to address this issue is to use smoothed capex, which was adopted in GD1, ED1 and Ofwat's PR14 modelling.[124]

There are a number of different ways that the lumpiness and potential differences in investment cycles driving capex and repex within the sector could be investigated and accounted for within the modelling. Figures 6. 2 to 6.4 below illustrates RAV additions and capex and repex reported in the gas distribution price control financial model and input files provided by Ofgem of GDNs' costs.

[123] Sometimes referred to as Special Cost Factor (SCF) adjustments.

[124] For GD1, capex was smoothed using a seven-year moving average. Historic data for repex at the tier 1-3 disaggregation level was not available and so smoothing was not applied.

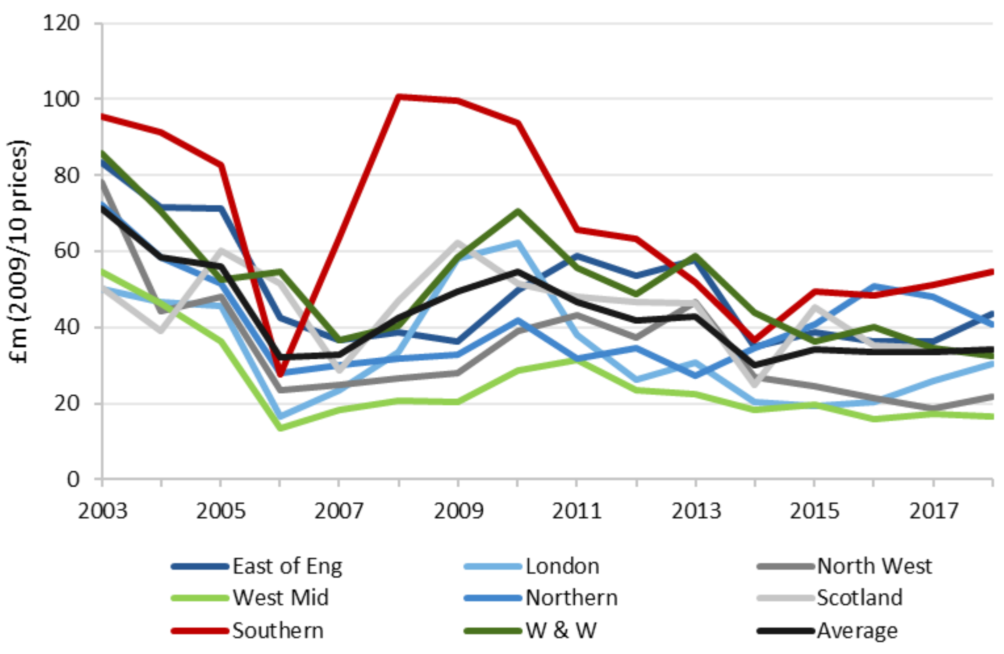
Figure 6.2: PCFM RAV additions



Source: CEPA analysis of GD1 PCFM

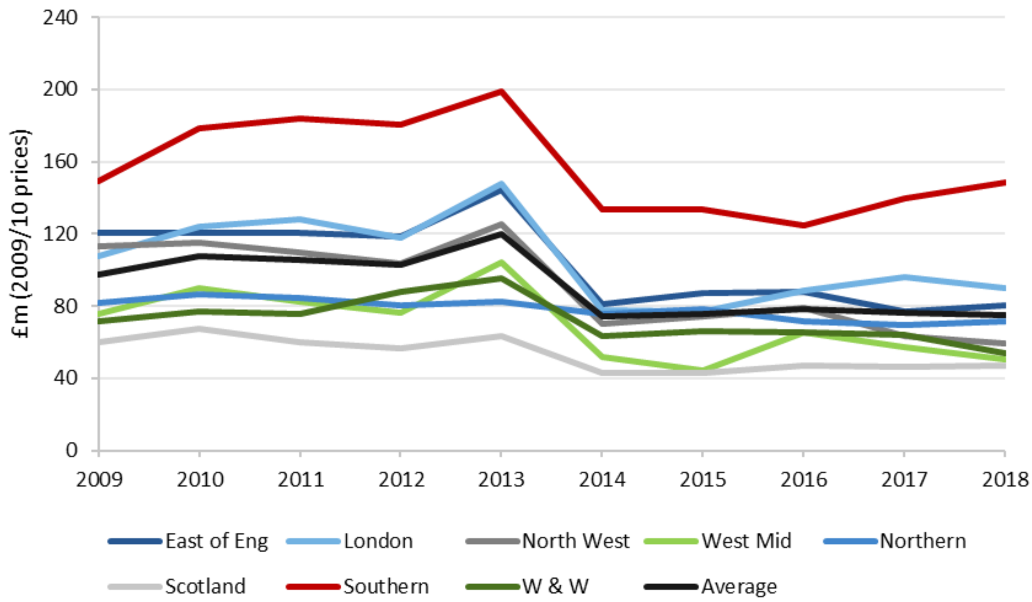
Note: the observed trend in RAV additions in GD1is in part caused by the increasing capitalisation rate during the price control

Figure 6.3: Capex (including RPEs – 2009/10 prices)



Source: CEPA analysis of Ofgem data

Figure 6.4: Repex (including RPEs – 2009/10 prices)



Source: CEPA analysis of Ofgem data

The three charts illustrate the year-on-year volatility in RAV additions (and capex and repex separately), both at an industry and company level, although some evidence of overall consistency in the investment cycle and expenditure trend across the industry (there is a noticeable trend in both reported capex and repex across GDNs).[125] Another point to note is that year-on-year volatility in capex has been greater than in repex, although repex has been a higher area of spend for the GDNs on average.

We suggest that smoothing vs. not smoothing is a statistical issue that will need to be investigated during the process of the GD2 econometric model development. Smoothing may be adopted where it can be demonstrated to improve the performance of the models. There are two key questions in principle to consider: (i) under what circumstances should costs be smoothed by Ofgem; and (ii) having decided to smooth, over what time period should the smoothing / moving average be calculated?

We briefly consider both issues below.

Under what circumstances should costs be smoothed?

As discussed above, the choice whether to smooth or not is largely a statistical issue, with smoothing used to the extent that it helps to improve the performance of Ofgem's econometric models, in terms of estimating an elasticity relationship between the GDNs costs and selection of cost drivers.

[125] As highlighted in the footnote to Figure 6.2 upwards trend in RAV additions, as compared to the observed trend in capex and repex, is due to the increased rate of capitalisation of the repex programme during GD1.

At a more aggregated cost level, the objective of the totex and middle-up regressions Ofgem developed for GD1 was to estimate a relatively stable relationship between total controllable costs and selected explanatory variables using both forecast and historical data. As the GD1 regressions smoothed capex and capex workloads using a moving average (seven-years), the regressions will have, therefore, tested a longer-term elasticity relationship between capex and selected workload drivers.

On this basis, smoothing is likely to have contributed to the original objective of Ofgem adopting a more aggregative benchmarking methodology in the first place. While smoothing in principle weakens the purpose of the totex approach[126], we consider that it is a still a useful approach to test ways to improve the performance of Ofgem's GD2 models. We would, therefore, recommend Ofgem **continues to test options for capex smoothing in its aggregated cost model selection process for GD2**.

We note that while Ofgem smoothed capex in its aggregated regressions at GD1, it did not smooth repex. While Figure 6.4 would indicate that there is less year-on-year volatility in repex we suggest that Ofgem at least investigate models that also smooth repex. This may support an assessment of a stable long-term relationship between total costs and workload drivers and may be particularly relevant where there are significant differences in the assumptions that GDNs have made on workloads and network condition in their business plans (if forecast costs are used by Ofgem to estimate its models).

As far as we are aware, Ofgem did not apply smoothing in the disaggregated (cost activity level) regressions developed at GD1. Given for capex and repex cost activities, these regressions effectively benchmark weighted unit costs, it is unclear to us why Ofgem would not also investigate the impact / benefits of smoothing in this case as well. Averaging costs and cost drivers over a period of time is a fairly standard approach in bottom-up unit cost assessment and applying smoothing in this context would be consistent with this approach. While the implication of this is the regressions are likely to be closer to a longer-term cross-sectional analysis of the relationship between capex/repex and workload drivers, this may be a useful alternative approach for Ofgem to consider alongside the methodology adopted at GD1.

Smoothing / averaging period

From a historical perspective, possible options for smoothing include using the full 10-year data set available to Ofgem at the time of the determination, or simply using current price control years.[127] Figures 6.2-6.4 suggest that there are trend differences in both capex and repex between GDPR and GD1 that will need to be investigated in further detail before a preferred moving average assumption is adopted.

Using forecast data raises further complications when choosing the smoothing period.

Although we would expect the GDNs as part of their business plans to consider in detail the expected profile of work and associated spend, this will by definition be somewhat assumptions driven. The wide range of assumptions for network condition and workload that might be adopted by the GDNs in their business plans, could in principle influence the results of the efficiency analysis.[128]

---

[126] Totex benchmarking is intended by design to capture the trade-offs that GDNs can make on a year by year basis between different capital and operational solutions. By smoothing capex and/or repex within the assessment, but not opex, costs are treated differently within the efficiency analysis.

[127] Clearly there are many other variants using the sample data that sits across these different time periods.

[128] To the extent that forecast expenditure and cost drivers are used in the model specification (i.e. sample period). This was an issue Ofgem identified at the GD1 review if using long term expenditure and cost driver forecasts.

On the one hand, this issue may support using longer-term moving averages when smoothing. Smoothing may help to reduce the risk that different GDN assumptions of relatively lumpy forecast expenditure and workloads inappropriately impacts the efficiency analysis. On the other hand, these differences in assumptions may in practice be viewed as beneficial in the spirt of benchmarking GDNs forecast costs and business plans (e.g. allows business plan proposals to influence the model specification).

We suggest that Ofgem model alternative variants of moving averages applied to smooth the capex and/or repex costs and test the impact on econometric model results. As with other aspects of the model development and selection process, smoothing different options should be considered to the extent they are considered to help improve the performance of the models against the selection criteria.

6.2.5. Use of CSVs

Several of Ofgem's GD1 models relied on CSVs, which combined multiple cost drivers into one composite variable. Ofgem stated that the criteria for constructing a CSV were when:

- the sample was considered too small to handle multiple drivers; and/or
- some of the cost drivers were statistically insignificant, but engineering knowledge and industry understanding led it to believe that combining different drivers into one CSV could better account for changes in costs within the sector.[129]

One of the disadvantages of using CSVs are that they place constraints on the underlying cost drivers (i.e., fixed weights on multiple drivers). In addition:

- CSVs, particularly when scale and workload drivers are combined into a single variable, make the interpretation of the economic / technical logic of the relationship between costs and cost drivers within econometrics more difficult to test.
- As we have discussed in previous sections of the report, the ability to understand and interpret the results from econometric modelling was one of key issues that the CMA raised with the modelling that Ofwat adopted for the PR14 determination (CSVs are less supportive of this approach).

However, CMA also criticised the modelling framework used at PR14 where a large number of explanatory variables were included in the model specification. In part this was because this made the models challenging to understand but also because it considered this may lead to less precise coefficient estimates, particularly where there may be a high degree of correlation between variables (e.g. scale). Using CSVs where there is only a small data sample for benchmarking in GD2 helps to address these issues.

We suggest that Ofgem compare and contrast the results of models that include and exclude CSVs as the former may help to develop and test the economic / technical rationale and expected coefficients of explanatory variables included in the more aggregative cost models. Developing models without CSVs may also help to determine the weights used within CSVs.

[129] Ofgem (2012): 'RIIO-GD1: Initial Proposals – Step-by-step guide for the cost efficiency assessment methodology'

6.3. CONCLUSIONS

This section has set out a causal narrative of the cost drivers of the GDNs and possible explanatory variables that could be considered to proxy these drivers in Ofgem's GD2 cost modelling.

Our assessment suggests that network output, scale and operating environment are amongst the most significant external, i.e. exogenous, drivers of the GDNs relative expenditure as well as their relative (in)efficiency. These effects can be captured in econometric models using scale variables such as number of customers and MEAV and variables that proxy the operated environment of the GDNs.

The age and condition of each GDN's network will also be expected to influence the expenditure of each company at any given point in time, in particular, the requirement for maintenance, enhancement and the large asset replacement investment programmes within the sector. To an extent, Ofgem has controlled for this effect in previous price reviews by including workload variables in its regressions.

However, there are well-known issues with using workload variables in external benchmarking exercises. In particular, workload variables may suffer from endogeneity and managerial incentive problems.[130] However, workload drivers help control for the effect of asset condition and different investment cycles, and, provided those forecast workloads are actually delivered during the price control period, should also allow a more accurate predication of future costs once a programme / volume of work is agreed.

We have proposed in previous sections of the report that Ofgem should continue to apply a tool-kit approach that includes various alternative models and cost pool aggregations in the cost assessment process. In light of this recommendation, **we would expect Ofgem to also consider a range of alternative econometric model specifications for GD2**, including:

- Aggregative (e.g. totex) models that consider both scale – e.g. MEAV – and disaggregated activity variables – e.g. workload drivers.
- Models that take different approaches to control for time effects on GDN costs (either time trends or time dummy variables (or neither)).
- Models that account in different ways for the impact of the external operating environment, in particular, regional factors.
- Models that adopt a Cobb-Douglas functional form.
- Models that include and exclude CSVs so that the expected sign and size of coefficient of different explanator models and specifications can be logically tested.

In Appendix C, we provide some illustrative regressions to demonstrate how some of the options discussed in this section might compare to the model specifications used at GD1.

In the next section of the report we discuss in further detail, how Ofgem might look to account for regional factors within its econometric modelling framework.

[130] As discussed above, workload drivers are inputs rather than true network outputs.

7. REGIONAL FACTORS

This section focuses on the possible approaches to account for regional factors in the econometric benchmarking. At a very high level, the objective of cost benchmarking is to compare companies against each other to help determine the efficient level of expenditure required to achieve a given output. However, as discussed in section 6, there are a number of reasons why companies may not be directly comparable, even when they are part of the same sector. Some costs may be driven by region- or company-specific factors outside of company control.

In regard to the efficiency modelling, regional differences between GDNs must be properly considered to ensure efficiency assessments are accurate. Factors outside of company control could lead to higher or lower costs that are not the result of efficient / inefficient behaviour.

Within regulatory benchmarking studies, these factors are typically captured within the efficiency modelling at one of three points in the modelling process:

- **Pre-modelling adjustment**: adjustments are made to the cost data before estimating econometric cost models in an attempt to make data more comparable between companies.
- **Within model adjustment**: regional factors can be captured through explanatory variables within the econometric models (where possible).
- **Post-modelling adjustment:** predicted costs from the econometric models are adjusted for regional and/or special cost factor (SCF) that are considering to not been sufficiently captured through a pre-modelling adjustment or within model adjustment.

In the subsections below we:

- Provide a summary of the approach taken to regional and special cost factors at RIIO-1 (with a focus on GD1).
- Consider the theoretical advantages and disadvantages of the three approaches to adjusting for such factors.
- Consider the practical issues in applying regional adjustments within the RIIO-2 cost assessment framework.

7.1. Approach at RIIO-1

Ofgem applied a pre-modelling adjustment approach in GD1. Ofgem adjusted for three regional factors based on its analysis (shown in Figure 7.1 below).

Ofgem also made several company specific adjustments largely based on information provided by GDNs and supported by independent data. This included adjustments to NGN's opex for a salt cavity, Scottish Independent undertaking, and London's medium pressure undertaking.

We review Ofgem's approach to regional factors in GD1 in more detail below.

Figure 7.1: Regional adjustments

Regional labour cost adjustment	• The difference in efficient labour costs between GDNs experienced due to geographical location.
Urbanity	• Estimate of efficient additional cost experienced by GDNs working in highly urban environments (inside the M25).
Sparsity	• Estimate of efficient additional cost experienced by GDNs working in highly sparse environments (namely, in Scotland).

Regional labour cost adjustment

Ofgem has a well-established approach to applying regional labour cost adjustments (RLCAs):

- differentials for London and the South East relative to the rest of GB;
- using Annual Survey of Hourly Earnings (ASHE) hourly data for the differentials; and
- using local authority population data to map ASHE data onto GDN areas.

Ofgem's GD1 and ED1 approaches were very similar aside from the Standard Occupational Classification (SOC) level that was chosen for the ASHE data. GD1 used 2- and 3-digit SOC level data while ED1 used only 2-digit SOC level data. Ofgem's ED1 approach passed scrutiny by the CMA during Northern Power Grid's (NPg) appeal of this component of its price review.[131] However, the CMA did not specify that 2-digit SOC level ASHE data needed to be used in future regulatory determinations. Rather, the CMA said that Ofgem should continue to review the data and choose the most appropriate SOC level.

Sparsity and urbanity

As discussed in Section 6, Ofgem has recognised at previous price reviews the impact the operating environment may have on energy networks' costs. This includes the impact of sparsity and urbanity.

Ofgem made an adjustment for **sparsity** in both its GD1 and ED1 determinations. Both decisions focused on additional costs of servicing customers in remote areas, namely in Scotland. The initial sparsity indices Ofgem used for GD1 were criticised by GDNs for the wide variations across the GDNs. In its Initial Proposals Ofgem reduced the impact of the indices on less sparse networks, and further reduced it in its Final Proposals. For ED1, Ofgem made an adjustment to SSEPD for the higher costs of working in a remote location. Relative to ED1, Ofgem seemed to provide much greater allowances for sparsity in GD1.

Ofgem's GD1 **urbanity** adjustment focused on the higher costs for networks operating within the M25. Its adjustment relied on analysis provided by both NGGD and SGN, which indicated that productivity was 15-25% lower in London than other areas. Ofgem placed more weight on SGN's arguments and made a 15% labour productivity adjustment. An adjustment for reinstatement and transport activities was also made by

[131] CMA (2015): 'Northern Powergrid (Northeast) Limited and Northern Powergrid (Yorkshire) plc v the Gas and Electricity Markets Authority – Final determination'

treating reinstatement transport costs as 100 per cent contract labour and applying contract labour indices to repairs and maintenance reinstatement.

7.2. Decision making framework

Section 6 illustrates that there are many different factors in the external operating environment of network companies that may account for observed differences in companies' costs. How to specifically account for these regional factors can be divided into two decision making stages:

- Determining what regional factors should be considered.
- Determining when each regional factor should be considered in the modelling process.

7.2.1. What regional factors should be considered?

Before a regional / company-specific factor adjustment approach is chosen, it is necessary to determine what costs should be appropriately adjusted. For this we can consider a number of general principles. For example, companies should be able to sufficiently justify that:

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

7.2.2. When should regional factors be considered in the modelling process?

Once it has been determined that a given regional factor should be considered for an adjustment, the theoretical merits of the three adjustment approaches discussed above to account for these factors can be considered. Which is most appropriate may differ for each of the factors identified:

- **Pre-modelling adjustment:** data is adjusted ahead of modelling. This is appropriate when regional/ company specific costs affect the accuracy of the modelling.[132] Pre-modelling adjustments can then be reversed out after the efficiency analysis (i.e. added back into modelled cost allowances).
- **Within model adjustment:** the regional factor is controlled for through the explanatory variables included in the cost assessment model. This would be subject to such a model specification satisfying the appropriate specification and robustness tests.
- **Post-modelling adjustment:** prior to the allowance or efficiency assessment taking place companies are permitted to submit claims to the regulator if they consider certain regional or SCFs are not sufficiently captured through pre-modelling or within model adjustments.

[132] i.e., systematic differences across the GDNs or one-off/ outlier costs affect the coefficients and efficiency scores produced by the modelling.

Table 7.1 provides a summary of the advantages and disadvantages of the three general approaches to regional company-specific factors:

Table 7.1: Pros and Cons of regional adjustment approaches

Adjustment	Pros	Cons
Pre-modelling	• Can ensure that the resulting adjustment index is consistent with technical and economic rationale before being applied • Conceptually simple approach that arrives at a tangible (monetary) impact for each regional factor • Clear monetary effect can be related back to the business plan or specific company activities	• Imposing an a priori rationale may lead to outcomes inconsistent with what the data suggests • The structure of the adjustments chosen is to some extent arbitrary • Removal of costs from econometric modelling could disincentivise the development of viable mitigation methods
Within model	• Allows the data to 'speak for itself' rather than imposing own judgements • The effect of regional factors is captured in the same way as other cost drivers in the model, which allows their relative effects to be more easily compared	• Greater data requirement – potentially difficult if applied in context of small data sets • Model may become difficult to interpret • Limited control over the effect of inclusion – explanatory variable could be a poor proxy, or the adjustment applied to inappropriate costs. This risks perverse results
Post-modelling (SCF)	• Recognises that econometric modelling may not be able to fully capture all idiosyncratic cost drivers • If the SCF process is suitably constructed, it will incentivise companies to reveal accurate cost information • Potentially lower burden on regulator – companies are expected to provide the evidence of unique circumstances to justify efficient costs different to that implied by the model when submitting SCF claims to the regulator	• If the SCF process is not suitably constructed, it could lead to companies considering it a 'one-way bet' • A heavy reliance on SCFs may indicate that the previous modelling process is not fit for purpose • Approach may lead to companies focusing on SCF claims and distract from the business planning process • Risk of double-counting effects if variables included in model already capture the effect • Can lead to an increasingly complex regime, e.g. the concept of 'implicit allowances' – to what extent do the models already account for the effect of regional factors on costs? • Can be difficult for companies to provide evidence for SCFs unless the regulator publish their minded-to model specifications ahead of SCF submission being made

Source: CEPA analysis

While a judgement on what is the most appropriate approach should be taken case-by-case, it can still be possible to identify some key determinants on the basis of Table 7.1. For example:

- **Pre-modelling** adjustments are perhaps most appropriate when the specific costs relating to the factor can be very clearly identified. In this case, econometric modelling, or an ex post adjustment / SCF process, may not be necessary.

  An example may be atypical costs, which can be removed ahead of the econometric modelling and then evaluated elsewhere within the cost assessment in order to ensure the companies are being assessed on a comparative basis.

Even where there is not a direct monetary value available for the pre-modelling adjustment, constructed indices from published data on regional wage differentials (as an example) can offer a transparent and easily replicable approach to account for regional factors, even though the consequence is the structure of the adjustment can be argued to be somewhat arbitrary.[133]

- **Within model** adjustment should arguably be the default position from a 'pure' econometrics viewpoint but will only feasible if there is suitable data (e.g. a large enough dataset to ensure sufficient degrees of freedom are available).

  A practical advantage of this approach is that model specification tests can be used to test the inclusion of explanatory variables to account for the regional factor. For example, are the coefficients statistically significant and in line with technical / economic rationale? A potential disadvantage is the increase in complexity of the existing models from including regional factors as variables in the models – even where data is available the inclusion of additional explanatory variables may make the overall specification harder to interpret in some cases.

  This may mean that directly accounting for regional factors within the econometric models may create a tension amongst the different model selection criteria identified in Section 2 and a judgement in the round needs to be made about whether this approach is appropriate.

- **Post-modelling** adjustment can be used for costs that are not sufficiently captured by either pre-modelling and/or within model adjustments. An adjustment is therefore used to correct for omitted variable bias within the model where there is sufficient evidence this may exist. Hence, this approach can complement pre-modelling and within modelling approaches.

  The main practical advantage of this approach, from Ofgem's perspective, is that the emphasis is placed on the network companies to justify adjustments for regional factors, rather than adjustments needing to be accounted for ahead of any benchmarking. From one perspective, this may lead to a more transparent regulatory process. However, as Table 7.1 identifies, there may also be the risk the process is a one-way-bet for the companies, unless a high evidential bar is set for any ex post adjustment and the regulator considers symmetrical adjustments.[134]

There is regulatory precedent of all three adjustment approaches being used or considered as part of benchmark analysis. Using the example of regional wage adjustments:

- As discussed above, Ofgem has previously used a pre-modelling approach, adjusting for expected differentials in labour costs in the London and the South East using ASHE data.
- For PR14, Ofwat included a regional wage variable in their econometric cost models (i.e. a within model adjustment).

[133] As opposed to if the adjustment was derived from observed differences in the network companies' actual expenditure (e.g. regional labour costs).

[134] In PR19, Ofwat has stated that it will apply a high bar when assessing SCF claims for each company. In cases where factors are expected to affect more than one company, it will also consider symmetrical adjustments.

- The Northern Ireland Utility Regulator (UR) applied a regional wage factor adjustment within Northern Ireland Water's most recent price control review (PC15) as a *post modelling adjustment*.
- Ofwat at PR19 have so far chosen not to apply a regional wage adjustment at all, following an analysis of the performance of both pre- and within-modelling adjustments.[135]

While we note the within- / post-model approach is quite different conceptually from the pre-model adjustment approach, there is little practical reason why both cannot be tested.

Where there is not a strong theoretical rationale for the choice of approach, the one that best improves modelling performance should be applied. An example of this approach was taken as part of Ofwat's model development in PR19 when considering regional wages.[136] CEPA's final proposed models did not include any regional wage adjustment – meaning any adjustment would have to be justified by companies as part of the SCF process – but this was concluded following extensive sensitivity analysis. For example:

- a regional wage variable was included within the model, but was not robustly significant or of the expected sign;
- a pre-modelling adjustment was tested, but did not significantly improve the capacity of the models to explain costs; and finally
- staff costs were considered to represent a small cost pool.[137]

If Ofgem were to continue to seek to account for differences in regional labour costs and other regional factors such as urbanity or sparsity though a *pre-modelling methodology*, there does not appear to be an obvious reason why it would, as a starting point, move away from the approach that it has used for the GD1 and ED1 reviews. However, given the importance of these adjustments, for thoroughness, Ofgem may want to consider alternative methodologies as part of RIIO-2.

Therefore, in the next subsection, we provide an initial assessment of how two key regional factors could in principle be considered through a within model adjustment:

- density; and
- labour price inputs (regional labour).

We note that we have only provided a brief discussion of the issues that would need to be considered if Ofgem were to consider either option as an alternative to the pre-modelling adjustments used at GD1. In particular, the inclusion of explanatory variables within the models would need to be subject to the model selection and evaluation process we have set out in Section 2.

---

[135] See Ofwat (2019): 'PR19 – Supplementary technical appendix: econometric approach'

[136] CEPA (2018), *'PR19 Econometric Benchmarking Models'*, available here, Box 1

[137] This final factor is an example of a 'materiality threshold' that could be applied as part of the decision procedure. For costs that do not immediately appear to represent a material divergence from the rest of the industry, companies should be expected to justify why their efficient costs are different through the SCF process.

7.3. ACCOUNTING FOR REGIONAL FACTORS WITHIN THE REGRESSION

7.3.1. Density

As discussed in Section 6, the impact of density on network company costs is ambiguous.

- **The costs of the company may increase with density.**

  Working in highly dense urban areas is more complex as it is likely to require, among other things, a more detailed consideration of the deployment of other utilities (e.g. the company would need to consider whether there are water mains before opening the road) and local authorities (e.g. requirement to request special local permits to stop the traffic or parking in areas where the work needs to be undertaken). The congestion of highly dense urban areas may also lead to lower productivity of workers. All of these factors may lead to higher costs.

- **However, the costs of the company may also increase with sparsity.**

  As discussed in Section 6, when serving rural areas, a network company could need higher staff numbers and/or assets to provide the same level of services as a more densely populated license area. A company operating in a sparse area will face the challenge that their staff are going to spend longer periods travelling, which may lead to higher costs. To mitigate the travelling time and to minimise customer disruption the company may decide to increase the number of depots, which may also lead to a relatively higher number of employees.[138]

The ambiguity of the relationship between density and costs has caused some stakeholders to hypothesise that there is a 'u-shaped' relationship between costs and density reflecting the arguments made above (i.e. costs are higher for companies operating in very dense and very sparse areas).[139]

Some regulators have attempted to include explanatory variables to capture the ambiguous effect of density on costs in a number of price reviews. A sample of proxies for density that have already been considered in the context of cost assessment are summarised in Table 7.2.[140]

[138] In addition to the direct cost increase from higher employee numbers, the network company may also experience an increase in company idle time, and in turn a decrease in employee productivity relative to a company operating in less rural areas.

[139] See for example Ofwat (2018), *'Cost assessment for PR19: a consultation on econometric cost modelling'*, available here.

[140] This tables includes analysis performed in section 6.

Table 7.2: Sample density explanatory variables

Explanatory variable	Discussion
Total connections / total length of mains	This group of variables reflect network activity or use per unit of network size and are a 'simple' way to capture the density of a network. For example, a higher number of connections or households per km of mains suggests a denser network. A quadratic version of this variable could also be tested, which could capture the u-shape relationship between density and costs.
Number of customers / service area[141]	
Ofwat-style weighted average density variable	This variable reflects the percentage of the population living in densely populated areas. Ofwat include a quadratic version of this variable in some models to capture the u-shape relationship between density and costs. It is constructed using ONS population and population density data at a LAD level. Each LAD is defined as highly dense if people per square kilometre (measure of density) is greater than 2000. The variable is then created by dividing total population in areas that are defined as dense by the total population for each company. Hence, the variable is between 0% and 100%.
Percentage of "urban assets"	In the water sector, Arup and Vivid Economics have previously suggested the use of variables that reflect the rural-urban split of the distribution of populations or assets.[142] They argued assets in urban areas cost more to operate due to, for example, harder access, traffic permissions and restricted land footprints.
Density standard deviation	This group of variables attempt to capture the variation of density within companies. They have been previously considered as part of initial scoping of the ED1 model development.[143] One way in which to do this is to consider the distribution of a 'basic' measure (e.g. customers per service area), weighted by the area of each sub-region. The use of these explanatory variables would be to recognise that GDNs serve a wide variety of different types of terrain, including relatively sparsely populated rural regions, moderately dense suburban regions and (possibly highly dense) urban regions.
Density skewness	
Density kurtosis	
Density Gini coefficient	
Total / share of surface area below / above a given density threshold	

Source: CEPA analysis

The cost assessment approach used at PR19 by Ofwat in the water sector is a good case study of some of the advantages of including density within econometric regression.

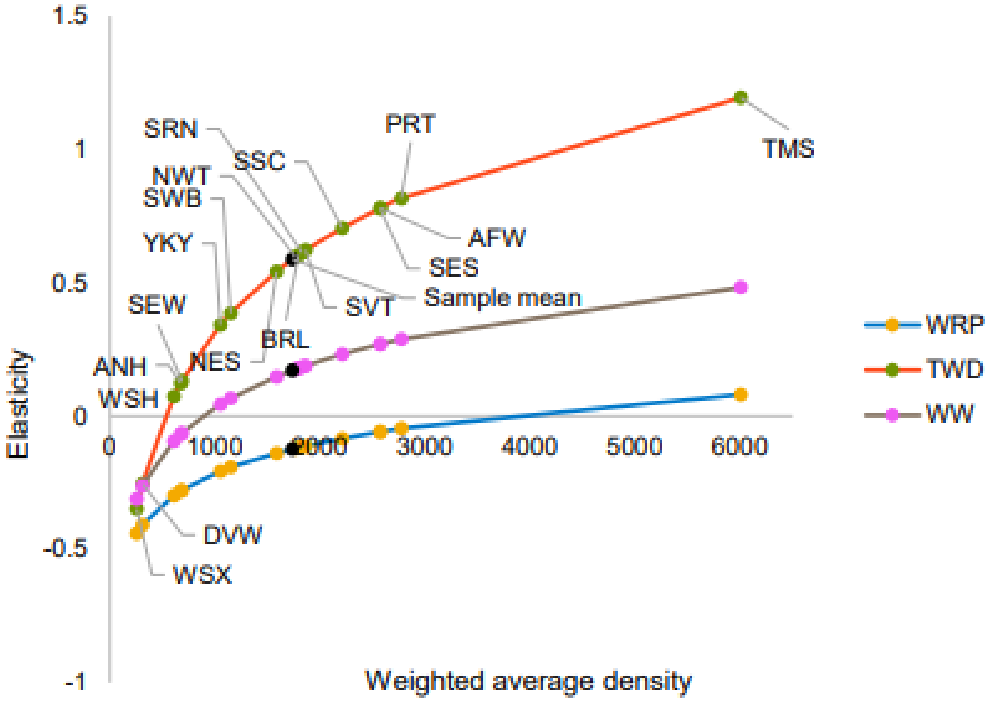
As opposed to making pre-modelling adjustments based on indices of high-level comparative statistics, this allows the data to speak for itself. In Ofwat's Cobb-Douglas functional form, the coefficients of explanatory variables can be interpreted as elasticities. The inclusion of a quadratic density term in their models allows Ofwat to conduct a discussion on how the elasticity of costs varies with respect to density across companies, as shown in Figure 7.2, and relate this to the economic and technical rationale behind their a priori expectations.

[141] This was a variable considered during ED1 modelling development – see for example, Ofgem (2014), *'RIIO-ED1 Draft Determination – business plan expenditure* assessment', pg. 173 available here.

[142] Vivid Economics and Arup (2017), *'Understanding the exogenous drivers of wholesale wastewater costs in England & Wales'*, available here.

[143] Frontier Economics (2013), *'Total cost benchmarking at RIIO-ED1 – Phase 2 report – Volume 1'*, prepared on behalf of Ofgem, available here.

Figure 7.2: The distribution of weighted average density elasticities in different Ofwat models



Source: Ofwat

Note: Water resources plus (WRP) reflects a mid-level aggregation of company costs; treated water distribution (TWD) is a granular level of cost activity; wholesale water (WW) can be consider analogous to a totex-level aggregation.

To illustrate some of the issues discussed above we have run two illustrative regression models and compared the results to the totex model – including sparsity and urbanity adjustments – used at GD1:

- Totex on the GD1 totex CSV[144] with no pre- or within model adjustments for sparsity or urbanity included in the model.
- Totex on the GD1 totex CSV after controlling for density through a within modelling density variable.[145]

We summarise the results, in the text box below.

[144] The RIIO-GD1 totex CSV is specified as: $CSV = meav^{0.37} repex_{driver}{}^{0.42} connections_{driver}{}^{0.02} mains_{driver}{}^{0.02}$
$$external\_conditions\_reports^{0.06} maintenance\_meav^{0.06} emergency\_CSV^{0.06}$$

[145] In our illustrative model, density is defined as the number of connections divided by network length.

Text Box 3: Illustrative density regressions

We have estimated four variations of the GD1 top-down totex model: (i) no sparsity / urbanity adjustment; (ii) pre-modelling sparsity / urbanity adjustment applied; (iii) linear density term included in model; and (iv) linear and quadratic density terms included to capture u-shape relationship between density and costs.

The cost drivers and the dependent variable are regressed in logarithmic form and cluster robust standard errors are used. The econometric output from each model specification is presented below.

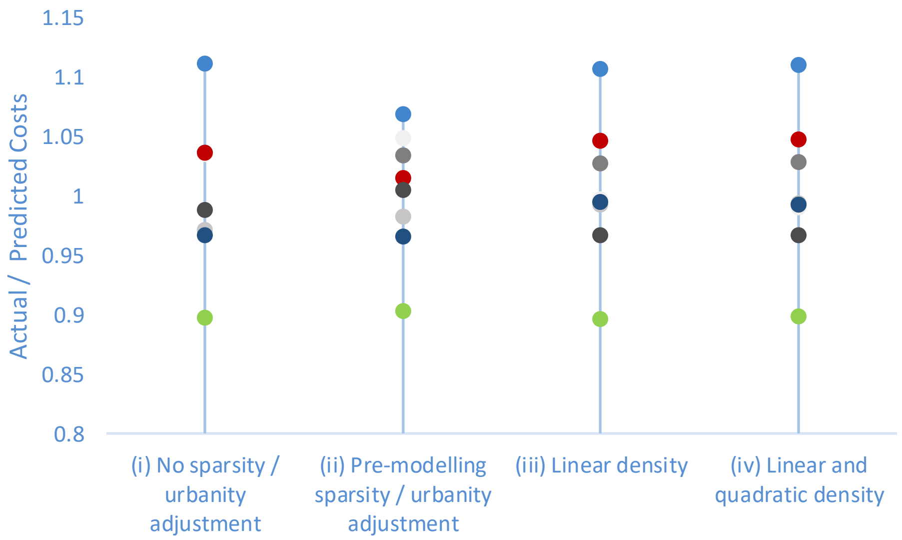
	(i) No sparsity / urbanity adjustment	(ii) Pre-modelling adjustment (GD1 approach)	Within model density controls	
			(iii) Linear	(iv) Linear and Quadratic
Totex CSV	0.739***	0.758***	0.739***	0.743***
Density			-0.049**	0.211
Density Squared				0.016
Time trend	-0.017***	-0.019***	-0.019***	-0.019***
Constant	34.952***	37.549***	38.161***	38.398***
Observations	80	80	80	80
R-squared	0.891	0.924	0.898	0.897

*Source: CEPA analysis. *** p<0.01, ** p<0.05, * p<0.1.*

By comparing the output presented above, the inclusion of the density variable increases the adjusted R-squared compared to when no sparsity / urbanity pre-modelling adjustment is made (0.891 to 0.898). The density variable is negative and statistically significant at a 10% level.

However, the predictive power of the power is greater when the urbanity / sparsity pre-modelling adjustment is applied, which implies that our choice of density explanatory variable may not be capturing the full effect of urbanity / sparsity on costs. Ofgem may want to investigate this result further by exploring alternative measures of density in the econometric models as well as testing the inclusion of density in other models (e.g. opex).

For illustration purposes, the efficiency scores (actual costs divided by predicted costs) implied by each model specification are presented below. Relative to not making any regional factor adjustments, the range of efficiency scores narrow slightly when within-model density controls are included in the model. But the efficiency range is even narrower when the pre-modelling sparsity / urbanity adjustment is made. Interestingly, the inclusion of the quadratic density term is not statistically significant, which implies that the u-shape relationship between density and costs may not be as significant as first thought.



Another practical issue to consider regarding how account is made for density in the modelling is whether it is possible to fully control for the costs any adjustment applies to.

One of the advantages of a pre or post-modelling adjustment is that it can be focused only in cost areas with a clear technical or business justification for these factors to drive costs. At GD1 for example, the sparsity adjustment was only applied to emergency and repair costs, as it can be argued it is justifiably more costly to maintain low response times where customers are sparsely distributed.

In contrast, in the simple regression example provided above, the impact of density when included within the econometric model is on *all* totex. The justification of the elasticity effect of sparsity or urbanity on totex is perhaps more challenging to explain. However, we recognise that the estimated magnitude of the coefficient should reflect this to some extent.

7.3.2. Labour price inputs / regional wage differentials

As with density/sparsity, it is possible to consider how input prices – in particular wages – might vary across different regions, and lead to different costs outside of company control. In GD1, Ofgem applied a regional labour adjustment to costs pre-modelling. However, this could be achieved within the models.

In their initial model development on behalf of Ofgem, Frontier Economics placed input prices as a key category of cost driver to be included in totex models.[146] Frontier recommended Ofgem consider a range of explanatory variables to capture the impact of input prices on costs. Of the four specifications identified, all included a capital price index (BEAMA index for Basic Electrical Equipment). The Frontier models then either applied a national or regional labour price adjustment, based off of indices primarily sourced from the ASHE database provided by the ONS:[147]

- SIC_35 (regional)
- SIC_35 (national)
- SIC_3513 (national)
- BEAMA_electrical_labour (national)

A similar approach to that suggested by Frontier for ED1 was applied by Ofwat in PR14. Ofwat constructed a regional wage variable using a weighted average of ASHE data on regional wages. However, during the Bristol Water appeal, the CMA identified a number of concerns with this approach and said the 'econometric models did not seem to work particularly well'.[148] The CMA noted that a lack of sufficiently granular ASHE data meant it was not possible to investigate specific issues as it does not provide data for the specific geographic areas that individual water companies serve.

[146] Frontier Economics (2013), *'Total cost benchmarking at RIIO-ED1 Phase 2 report – volume 1'*, available here.

[147] The UK Standard Industrial Classification (SIC) codes reflect different classifications of labour (here related to ED).

[148] CMA (2015), *'Bristol Water plc: A reference under section 12(3)(a) of the Water Industry Act 1991'*, appendices 4.2 and 4.3, available here.

7.4. Conclusions

This section has considered the issues surrounding the three main approaches to regional and company-specific adjustments. We have examined the advantages and disadvantages of each approach and proposed a decision procedure to help frame the choice of which to apply. Our main conclusions are:

- In the first instance, key factors should be included within the econometric modelling. This allows their effect to be statistically tested. However, a number of practical considerations mean this is not always feasible.
- A key advantage of pre-modelling adjustment as an alternative approach is that it provides a clear, conceptually simple estimate for the impact of the given factor. However, the choice and structure of adjustments can be argued to be somewhat arbitrary.
- A key advantage of post-modelling adjustment is that it can help reveal accurate information as companies are required to pass a high evidential bar to justify changes, while it also recognises that econometric modelling will never perfectly capture all cost drivers.

The choice of the most appropriate approach we would suggest is more of a modelling / statistical issue as opposed to a policy one. Each approach is attempting to achieve the same policy objective.

Therefore, as a general principle, **we would recommend that Ofgem explore a range of options** for the treatment of regional factors. These should be tested against prior expectations and the performance of models compared against each other before a final decision on the best approach is taken.

In particular, we believe the treatment of both **regional wages and density could in principle be explored further as within model adjustments**. However, due to regulatory precedent, and issues with including these factors within-model explored in the main report, the feasibility of using this approach will need to be demonstrated through a robust model development and selection process. While Ofgem may wish to explore alternative approaches at GD2, our initial expectation would be that a continuation of the approach that was taken at GD1 will likely remain appropriate.

8. CONCLUDING REMARKS

This report has considered various topics related to the use of econometric benchmarking and treatment of regional factors within Ofgem's GD2 cost assessment framework. We have set out in previous sections of the report specific conclusions and proposals for further consultation by topic area.

There are, however, a number of important themes that apply across all these topics and much of the discussion throughout the report of how Ofgem might look to build upon, and where appropriate change or adapt, the approach that it adopted for benchmarking in GD1.

One common theme is that benchmarking is a challenging process in regulated network industries, particularly in the context of GB gas distribution where only a small panel data set is available for the analysis. Therefore, benchmarking needs to be applied with care within the regulatory proceedings and a clear and logical process set out and followed to select econometric models, and account for the effect of regional, and other company specific, factors that may explain differences in GDNs costs.

While this includes the need to undertake rigorous statistical and robustness testing of the econometric models, as we have discussed throughout the report, UK regulatory precedent, including recent price review appeals to the CMA, demonstrates that the replicability (transparency), economic and technical justification, and the incentive properties of the econometric benchmarking models and techniques employed, is equally important in the context of the proceedings.

A second common theme is the importance of retaining a tool-kit approach.

In light of the known limitations of econometric benchmarking with small datasets, there is no single model that will capture all of the various effects of company heterogeneity and time on the observed differences between the GDNs reported costs. For this reason, we believe it is important that Ofgem consider a range of different approaches, including both disaggregated and aggregated econometric benchmarking models, in forming its views of the appropriate efficiency frontier and cost baselines for GD2.

Finally, we have highlighted how there are a range approaches that can be considered to account for differences in the outputs, input prices and external operating environment of the GDNs and how they in turn account for the observed differences in GDNs costs for reasons other than (in)efficiency.

Accounting for these factors using pre, within or post modelling adjustments, all have their theoretical and practical advantages and disadvantages. As with other parts of the benchmarking toolkit, we would, therefore, encourage Ofgem to investigate a range of different approaches. This is consistent, in our view, with regulatory best practice for external benchmarking. However, we also believe that investigating different approaches will help Ofgem to build a better understanding of how a range of external factors impact GDNs costs, and may also help Ofgem assess the confidence it has in GDNs forecast expenditure in GD2, once the final price control business plans have been submitted.

APPENDIX A BENCHMARKING REGULATORY PRECEDENT

In this appendix, we consider recent cost assessment approaches across UK network regulation that are relevant to the issues considered in this report.

A.1. APPROACH FOLLOWED IN RIIO-GD1

The general modelling approach followed for RIIO-GD1 is described in the step-by-step methodology published as a part of Ofgem's initial proposals.[149]

Around 60% of company costs were assessed using regressions techniques. Bottom-up modelling of seven disaggregated cost categories and a top-down modelling of all included totex were undertaken.[150] These were performed on four years of historical data and two years of forecast data. The final view of efficient costs was derived by taking an unweighted average of the results from the four approaches.

Ofgem applied similar statistical robustness tests developed for DPCR5 when identifying their preferred models (see Table 2.1 in the main report).

Ofgem's econometric modelling approach at GD1 was criticised at the time by a number of industry stakeholders. We briefly summarise some of these below:

- **Omitted / alternative cost drivers / explanatory variables.** Several of Ofgem's models relied on CSVs, which combined multiple cost drivers into one.

  Ofgem's use of CSVs was criticised as they place constraints on the underlying cost drivers (i.e., fixed weights). Some companies also had concerns that the use of workload drivers may mean that efficiencies in workload volumes are not adequately captured.

- **Investment cycle and lumpy capex.** Ofgem's totex models assumed that companies have a common, synchronous investment cycle but in reality, this may not be the case. This may be reflected by the fact that companies with relatively large capex programmes appeared inefficient in the totex model but relatively more efficient in the disaggregated benchmarking.

- **Time period.** A number of companies questioned Ofgem's decision to not place any weight on the 8-year forecast models, and instead rely on the historical and two-year forecast model.

  Ofgem rejected the use of the 8-year forecast models as it did not consider these to be robust due to differences in the GDN's forward-looking assumptions. This included the range of assumptions of deterioration in asset health and work volumes across GDNs.

- **Cherry picking.** At GD1, Ofgem provided clear guidance on how it planned to approach model development and selection, which helped to avoid criticisms that it cherry-picked models / results. However, some companies argued that the summation of bottom-up regression activities does not avoid cherry picking between regressed and non-regressed activities, as a large percentage of costs were non-regressed (35%).

---

149 Ofgem (2012), *'RIIO-GD1: Initial Proposals – Step-by-step guide for the cost efficiency assessment methodology'*, available here.

150 'Middle-up' models (based on opex, capex, and repex categories) were also estimated but not used as the results were broadly similar to the totex modelling.

As part of the CAWG, Ofgem and stakeholders performed a lessons learned exercise looking back at GD1. Some of the main takeaways were as follows:

- The mixing of top-down and bottom-up models was appropriate, but the weighting may need to be reviewed and cherry picking should be avoided.
- Separate assessments or alternative techniques could be considered where benchmarking via regressions is in inappropriate or not viable, and to account for network specificities.
- Some cost categories may need to be combined (e.g. emergency and repair) and some cost drivers may need to be reviewed (e.g. MEAV).
- Quality could potentially be included in models.
- Repex should be a particular area of focus.

Table A.1 overleaf summarises the cost pools / benchmarking aggregation that Ofgem adopted for its GD1 Final Proposals on cost assessment.

Table A.1: Cost pooling in RIIO-GD1 determination

Cost area	Bottom-up	Middle-up	Top-down
Opex			
Work management	Activity regression	Pooled opex activity regression	Pooled totex activity regression
Emergency	Activity regression		
Repairs	Activity regression		
Maintenance	Activity regression [1]		
Other direct	Non-regression		
Business support	Non-regression		
Xoserve	Non-regression	Non-regression	Non-regression
SIUs	Non-regression	Non-regression	Non-regression
Holder decommissioning	Non-regression	Non-regression	Non-regression
Holder site remediation	Non-regression	Non-regression	Non-regression
Loss of metering	Non-regression	Non-regression	Non-regression
Tier 2/3 survey costs	Non-regression	Non-regression	Non-regression
MOBs surveys	Non-regression	Non-regression	Non-regression
Interruptible contracts	Non-regression	Non-regression	Non-regression
Smart metering set up	Non-regression	Non-regression	Non-regression
Capex			
Connections	Activity regression	Pooled capex activity regression	Pooled totex activity regression
Mains reinforcement	Activity regression		
LTS & storage	Non-regression		
Governors	Non-regression		
Other operational capex	Non-regression		
Holder decommissioning	Non-regression		
Fuel poor extensions	Non-regression	Non-regression	Non-regression
Repex			
Tier 1 – mains & services	Pooled repex activity regression	Pooled repex activity regression	Pooled totex activity regression
Tier 2 – above threshold			
Tier 2 – below threshold			
Tier 3 – mains & services			
Other repex			
Other costs			
Street-works	Non-regression	Non-regression	Non-regression
RPEs	Non-regression	Non-regression	Non-regression
Ongoing productivity	Non-regression	Non-regression	Non-regression

Source: CEPA analysis of Ofgem GD1 determination

Note 1 – Final Proposals document states that Ofgem set a final maintenance opex allowance as "the maintenance allowance on the basis of the combined maintenance plus LTS pipelines regression, net of our technical assessment of efficient LTS pipelines costs."

A.2. APPROACH FOLLOWED IN OTHER REGULATORY CONTEXTS

Price control
RIIO-ED1 (2015-2023), GB electricity distribution, Ofgem
Cost assessment in ED1 followed a similar approach to that used in GD1. Comparative analysis was done at a totex level and on a more disaggregated basis (i.e. cost activity level). In reaching decisions on the fast-tracking of companies, Ofgem placed greater weight on the disaggregated models as it enabled a richer model specification. A Cobb-Douglas functional form was employed, using Corrected OLS, as with GD1. Advanced techniques such as RE or SFA were not considered appropriate for separating out inefficiency due to the limited time-series variation compared to cross-sectional variation. For the fast-tracked companies, three years of historical data was used for the model estimation; using the full set of available historical and forecast data was considered but not employed as the model diagnostics were poor. In contrast, the slow-track determination used 13 years of data (5 years of DPCR5 and 8 years of ED1) as Ofgem considered this to 'better take account of the scope of efficiency savings and likely patterns of costs', and was consistent with the use of historic and forecast data in the disaggregated analysis. In terms of weightings, bottom-up and top-down totex models accounted for 25% each while the disaggregated model accounted for 50% of the final assessment.
PR19 (upcoming), England & Wales water and sewage, Ofwat
Econometric models will be used to assess 'base' costs, which include opex and maintenance capex and excludes enhancement capex (i.e. 'lumpy' capex). Ofwat have noted that they believe there is scope for econometric models to include some enhancement and expect to revisit this in future. Models across a range of business activities and disaggregation was used to 'triangulate' the cost allowances. To date in PR19 models have been developed using historical data only. Ofwat made a number of adjustments relative to their PR14 approach following Bristol Water's appeal to the CMA. In particular, the 'translog' functional form used in PR14 was criticised by the CMA and Ofwat have consequently simplified their approach – Cobb-Douglas is used a starting point, with additional non-linear or cross-product terms only added when there is a clear economic or technical rationale. Two panel data estimation techniques were considered, POLS and RE. The statistical testing supported the use of RE and so this method was used to estimate all of Ofwat's models and the initial assessment of business plans.
Firmus Energy appeal (GD17), N. Ireland gas distribution, CMA (following UR determination)[151]
The GD17 price control runs from 2017 to 2022. The Utility Regulator (UR) applied a top-down econometric benchmark for the first time when determining opex allowances. The top-down analysis was not used to directly determine Firmus Energy's (FE's) allowance, but instead 'informed' it. The analysis was used as a sense-check of the bottom-up analysis directly used. The two preferred models used a CSV comprised of customer numbers (50%), gas volumes (25%) and network length (25%). One model also included a time trend while the other included iron mains as proportion of the network (as a proxy for network quality). POLS was used as the estimation technique and the data was from the eight GB GDNs across seven years. FE appealed elements of GD17 to the CMA, including criticisms on a number of grounds of the UR's cost assessment approach, the majority of which were rejected. One area of particular interest was the CMA's response to FE's challenge that the UR failed to take proper account of the impact of sparsity. The CMA concluded that the UR was not wrong when setting maintenance allowances that did not allow for the relative sparsity of FE's network compared to the gas distribution network in the Greater Belfast, Larne and East Down areas in Northern Ireland operated by Phoenix Natural Gas (PNG). While FE demonstrated it had a sparse network, it did not provide persuasive evidence of whether and to what extent these differences affect FE's maintenance costs.

[151] https://www.gov.uk/cma-cases/energy-licence-modification-appeal-firmus-energy

APPENDIX B COST AGGREGATION OPTIONS

Table B.1: Illustration of cost pooling under ***Options 1 & 2***

Cost activity	Option 1	Option 2	
	Aggregated only regression modelling	Top-down: Totex regression modelling	Bottom-up: activity level regression modelling
Opex			
Work management	Pooled in aggregated Totex regression *or* Pooled in aggregated Opex regression	Pooled Totex regression	Activity regression
Emergency			Activity regression
Repairs			Activity regression
Maintenance			Activity regression
Other direct			Non-regression
Business support			Non-regression
Shrinkage	Non-regression	Non-regression	Non-regression
Xoserve			
SIUs			
Holder decommissioning			
Holder site remediation			
Loss of metering			
Tier 2/3 survey costs			
MOBs surveys			
Interruptible contracts			
Smart metering set up			
Capex			
Connections	Pooled in aggregated Totex regression *or* Pooled in aggregated Capex regression	Pooled Totex regression	Activity regression
Mains reinforcement			Activity regression
LTS & storage			Non-regression
Governors			Non-regression
Other operational capex			Non-regression
Holder decommissioning			Non-regression
Fuel poor extensions	Non-regression	Non-regression	Non-regression
Repex			
Tier 1 – mains & services	Pooled in aggregated Totex regression *or* Pooled in aggregated Repex regression	Pooled Totex regression	Pooled Repex activity regression
Tier 2 – above threshold			
Tier 2 – below threshold			
Tier 3 – mains & services			
Other repex			
Other costs			
Street-works	Non-regression	Non-regression	Non-regression

Source: CEPA

Table B.2: Illustration of possible cost pooling under ***Options 3 & 4***

Cost activity	Option 3		Option 4
	Top-down: Totex regression modelling	Bottom-up: Opex Plus regression modelling	Opex Plus only regression modelling
Opex			
Work management	Pooled Totex regression	Pooled Opex regression *Plus* Capex or Repex items identified as complementary for pooling from listed activities below	Pooled Opex regression *Plus* Capex or Repex items identified as complementary for pooling from listed activities below
Emergency			
Repairs			
Maintenance			
Other direct			
Business support			
Shrinkage	Non-regression	Non-regression	Non-regression
Xoserve			
SIUs			
Holder decommissioning			
Holder site remediation			
Loss of metering			
Tier 2/3 survey costs			
MOBs surveys			
Interruptible contracts			
Smart metering set up			
Capex			
Connections	Pooled Totex regression	Non-regression unless added as a complementary item to the Opex regression	Non-regression unless added as a complementary item to the Opex regression
Mains reinforcement			
LTS & storage			
Governors			
Other operational capex			
Holder decommissioning			
Fuel poor extensions	Non-regression	Non-regression	Non-regression
Repex			
Tier 1 – mains & services	Pooled Totex regression	Non-regression unless added as a complementary item to the Opex regression	Non-regression unless added as a complementary item to the Opex regression
Tier 2 – above threshold			
Tier 2 – below threshold			
Tier 3 – mains & services			
Other repex			
Other costs			
Street-works	Non-regression	Non-regression	Non-regression

Source: CEPA

APPENDIX C ILLUSTRATIVE REGRESSION MODEL OUTPUTS

In this appendix we present a number of illustrative regression model outputs in light of a number of the issues discussed in the main report. All models have been run using a dataset provided by Ofgem's cost assessment team. For avoidance of doubt, CEPA have not quality assured the data provided, and all model results should be considered as illustrative only. We note that in contrast to the model specifications used by Ofgem at GD1, we include a constant / intercept term in our regressions.

C.1. TOTEX MODELS USING A LINEAR TIME TREND VERSUS TIME DUMMIES

To provide an illustration of econometric modelling with a linear time trend and linear time dummies, we regressed totex on a the GD1 totex CSV[152] for the period between 2008/09 to 2017/18 under three model specifications: (1) The model includes no time controls at all. (2) The model includes a linear time trend parameter. (3) The model includes a set of time dummy variables.

Under all model specifications, the totex CSV cost driver and the dependent variable are regressed in logarithmic form and cluster robust standard errors are used. The econometric output from each model specification is presented below.

Table C.1: Totex models using a linear time trend versus time dummies

Totex models using a linear time trend vs. time dummies	No time controls	Linear time trend	Time dummies
GD1 Totex CSV	0.799***	0.758***	0.758***
Time Trend		-0.019***	
Dummy_2010			0.002
Dummy_2011			-0.007
Dummy_2012			-0.014
Dummy_2013			0.03
Dummy_2014			-0.051
Dummy_2015			-0.058
Dummy_2016			-0.112***
Dummy_2017			-0.119***
Dummy_2018			-0.166***
Constant	-0.233	37.549***	0.099
Observations	80	80	80
R-squared	0.883	0.924	0.930

*Source: CEPA analysis. *** p<0.01, ** p<0.05, * p<0.1*

[152] The RIIO-GD1 totex CSV is specified as:

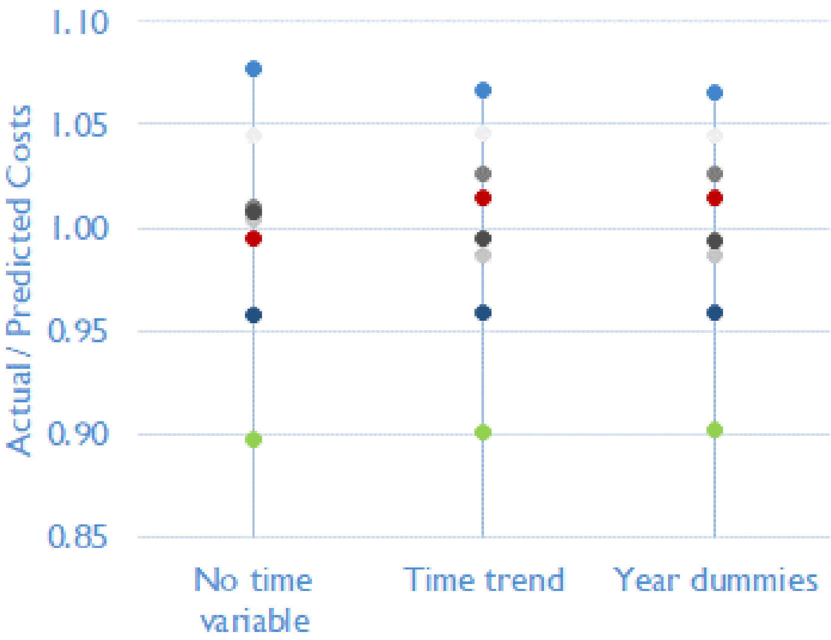
$$CSV = meav^{0.37} repex\ driver^{0.42} connections\ driver^{0.02} mains\ driver^{0.02}$$
$$external\_conditions\_reports^{0.06} maintenance\_meav^{0.06} emergency\_CSV^{0.06}$$

Under each model specification, the coefficient on the totex CSV is positive and significant.

The time trend in the second model specification is negative and significant. Three out of nine time dummies are statistically significant in the third model specification. Overall, model 2 (time trend) and model 3 (time dummies) produce very similar results.

The totex average efficiency scores of the GDNs over the period 2009 to 2018 across each model specification are shown below. The range of efficiency scores are slightly narrower when a time trend or time dummies are included but overall the results are very similar across all three models.

Figure C.1: GDN totex efficiency scores with no time variable, a time trend or time dummies, 2009-2018 average.



Source: CEPA analysis

C.2. REPEX MODELS USING A LINEAR TIME TREND VERSUS TIME DUMMIES

To provide an illustration of econometric modelling with a linear time trend and linear time dummies, we regressed repex on the GD1 repex CSV for the period between 2008/09 to 2017/18 under the same three model specifications:

- The model includes no time controls at all.
- The model includes a linear time trend parameter.
- The model includes a set of time dummy variables.

Under all model specifications the cost driver and dependent variable are regressed in logarithmic form and cluster robust standard errors are used. The econometric output from each model specification is presented below.

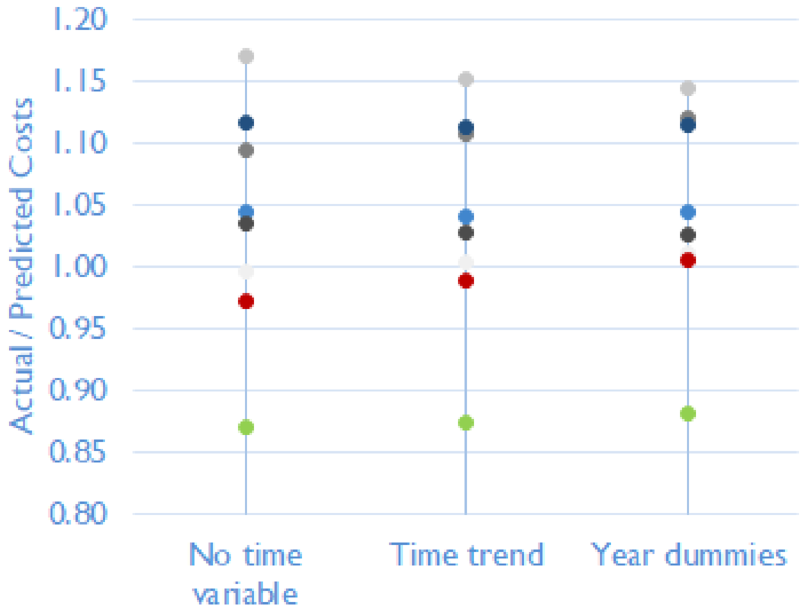
Table C.2: Repex models using a linear time trend versus time dummies

Repex models using a linear time trend vs. time dummies	No time controls	Linear time trend	Time dummies
Repex CSV	0.872***	0.838***	0.815***
Time Trend		-0.011*	
Dummy_2010			0.056*
Dummy_2011			0.023
Dummy_2012			0.004
Dummy_2013			0.150**
Dummy_2014			-0.065*
Dummy_2015			-0.04
Dummy_2016			-0.034
Dummy_2017			-0.049
Dummy_2018			-0.077
Constant	0.681*	23.842*	0.927*
Observations	80	80	80
R-squared	0.876	0.883	0.900

*Source: CEPA analysis. *** $p<0.01$, ** $p<0.05$, * $p<0.1$*

The coefficient on the repex CSV is positive and significant under all model specifications. The magnitude of the coefficient falls when a linear time trend is included relative to no time controls. The magnitude of the CSV coefficient falls again when the linear time trend is substituted by the set of time dummy variables. As under the totex modelling, the model fit is highest when the set of time dummy variables are used. The repex efficiency scores of the GDNs across the modelling period are shown below.

Figure C.2: GDN totex efficiency scores with no time variable, a time trend or time dummies, 2009-2018 average.



Source: CEPA analysis

C.3. TOTEX MODELS WITH SCALE VARIABLES

To provide an illustration of econometric totex benchmarking with alternative scale variables, we regressed the following models for the period 2008/09 to 2017/18:

- Totex on the GD1 Totex CSV, which is defined as:
    - Totex CSV= (MEAV ^ 0.37) * (Repex Driver ^ 0.42) * (Connections Driver ^ 0.02) * (Mains Driver ^ 0.02) * (External Condition Reports ^ 0.06) * (Maintenance MEAV ^ 0.06) * (Emergency CSV ^ 0.06)
- Totex on CSV1, which is defined as:
    - CSV1 = (Customers ^ 0.25) * (Mains ^ 0.25) * (Throughput ^ 0.25) * (MEAV ^ 0.25)
- Totex on CSV2, which is defined as:
    - CSV2 = (Customers ^ 0.33) * (Mains ^ 0.33) * (Throughput ^ 0.33)
- Totex on MEAV

A linear time-trend is included in all specifications. In all models, the cost driver and the dependent variable are in logarithmic form and cluster robust standard errors are used. The econometric output from each model specification is presented below.

Table C.3: Totex models with alternative scale variables

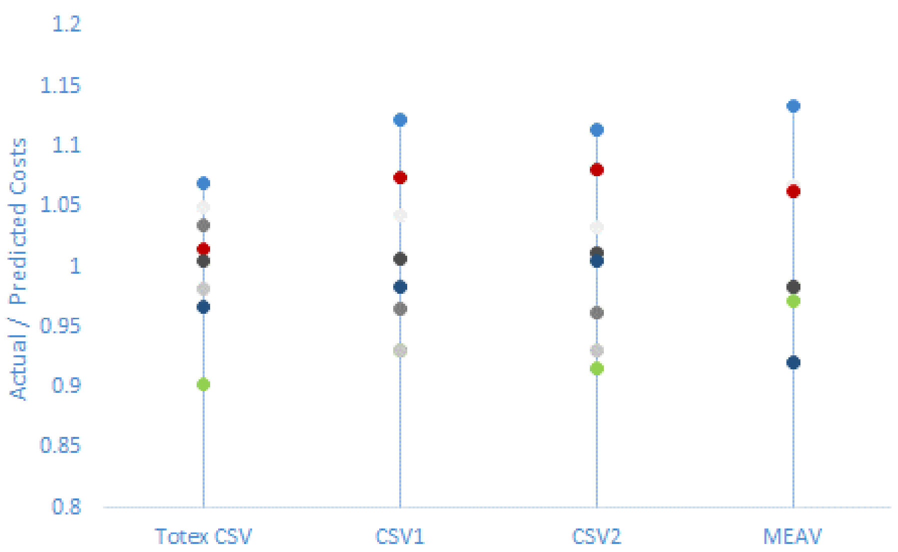
Totex models with alternative scale variables	GD1 Totex CSV	CSV1	CSV2	MEAV
GD1 Totex CSV	0.758***			
CSV1		0.760***		
CSV2			0.757***	
MEAV				0.742***
Time Trend	-0.019***	-0.033***	-0.032***	-0.034***
Constant	37.549***	62.382***	60.761***	67.711***
Observations	80	80	80	80
R-squared	0.924	0.881	0.880	0.860

*Source: CEPA analysis. *** p<0.01, ** p<0.05, * p<0.1*

The coefficient on the cost driver under both model specifications is positive and significant. The estimated coefficient on the scale variable is significantly less than one in all four models, which indicates the presence of significant economies of scale.

However, the model fit when CSV1, CSV2 or MEAV is used as the regressor is lower than when the GD1 CSV driver is used. This is reflected in the chart below, which presents the average efficiency scores over the period 2009 to 2018, and shows that the efficiency score range is slightly narrower with the totex CSV than the models that include CSV1, CSV2 or MEAV,

Figure C.3: GDN totex efficiency scores when the RIIO-GD1 totex CSV and scale-variable CSV is used, 2009-2018 average.



Source: CEPA analysis

C.4. RESTRICTED VERSUS UNRESTRICTED TOTEX MODELS

We have regressed totex on a number of different CSVs ('restricted' models) as well on the individual explanatory variables within each CSV ('unrestricted' models) to illustrate the effect of aggregating cost drivers into a CSV. A linear time-trend is included in all specifications. In both models, the cost drivers and the dependent variable are in logarithmic form and cluster robust standard errors are used. The econometric output from each model specification is presented below.

Table C.4: Totex models with alternative scale variables

Totex models with restricted and unrestricted CSV	CSV1		CSV2		GD1 Totex CSV	
	Restricted	Unrestricted	Restricted	Unrestricted	Restricted	Unrestricted
CSV1	0.760***					
CSV2			0.757***			
Customers		0.756***		0.815***		
Network Length		-0.167		-0.109*		
Throughput		0.086		0.071		
MEAV		0.108				-0.367
GD1 Totex CSV					0.758***	
Repex driver						0.282***
Connections driver						-0.002
Mains driver						0.005
External condition reports						-0.121
Maintenance MEAV						0.219
Emergency CSV						0.744**
Time Trend	-0.033***	-0.036***	-0.032***	-0.036***	-0.019***	-0.022***
Constant	62.382***	65.905***	60.761***	66.959***	37.549***	40.481***
Observations	80	80	80	80	80	80
R-squared	0.881	0.920	0.880	0.921	0.924	0.949

*Source: CEPA analysis. *** p<0.01, ** p<0.05, * p<0.1*

As expected, model fit (as indicated by adjusted R-squared) is higher in the unrestricted models compared to the restricted models. However, not all explanatory variables are statistically significant at the 10 percent level as a result of high multicollinearity.

The table below presents the correlation matrix for all the variables contained within CSV1 and CSV2, which shows that all correlation coefficients are greater than 0.9 (i.e. very highly correlated). It may therefore be sensible to estimate multiple regressions with different scale explanatory variables and triangulate across models than place restrictions on the underlying relationships within a CSV.

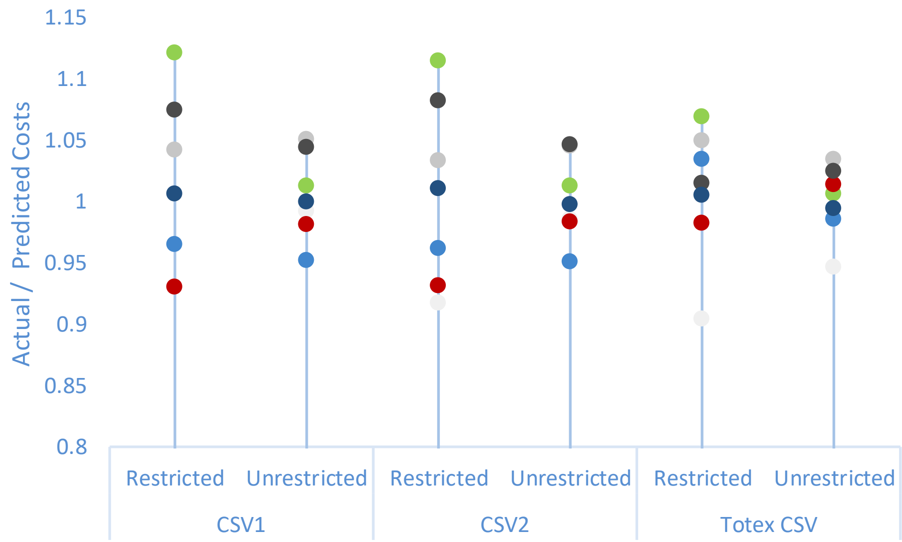
Table C.5: Correlation coefficient matrix between scale explanatory variables

	CSV1	CSV2	Totex CSV	MEAV	Customers	Network Length	Throughput
CSV1	1						
CSV2	1.00	1					
Totex CSV	0.95	0.95	1				
MEAV	0.99	0.98	0.94	1			
Customers	0.98	0.98	0.93	0.96	1		
Network Length	0.98	0.98	0.91	0.97	0.93	1	
Throughput	0.97	0.97	0.94	0.93	0.93	0.92	1

Source: CEPA analysis. Note: A correlation coefficient is the measure of linear interdependence between two variables. The value ranges from -1 to 1, with -1 indicating a perfect negative correlation and 1 indicating a perfect positive correlation.

The totex efficiency scores of the GDNs across the modelling period are shown below.

Figure C.4: GDN totex efficiency scores – restricted versus unrestricted models, 2009-2018 average.



Source: CEPA analysis

C.5. OPEX MODELS WITH RANGE OF EXPLANATORY VARIABLES

We have estimated a range of opex models using OLS over the period 2009 to 2018 to illustrate the impact of different model specifications and different levels of cost aggregation (bottom-up and top-down), as summarised in the table below.

Table C.6: Opex models

Bottom-up opex	Top-down opex
Work management	Work management + Emergency + Repairs + Maintenance
Emergency	
Repairs	
Maintenance	

Model Specification 1: MEAV + Time Trend

Model Specification 2: CSV1 + Time Trend

Model Specification 3: CSV2 + Time Trend

Source: CEPA

CSV1 and CSV2 are composite scale variables, which are calculated as follows:

- CSV1 = (Customers ^ 0.25) * (Mains ^ 0.25) * (Throughput ^ 0.25) * (MEAV ^ 0.25)
- CSV2 = (Customers ^ 0.33) * (Mains ^ 0.33) * (Throughput ^ 0.33)

The bottom-up and top-down opex bottom up model results are presented in the table below. With the exception of the time trend, all variables have been log transformed. Therefore, estimated coefficients can be interpreted as elasticities.

We find that the top-down opex model generally performs better than the bottom-up opex models from a statistical perspective across all three model specifications. All three model specifications are generally weak at explaining variations in work management and repairs expenditure. In contrast, MEAV appears to be a relatively significant driver of maintenance costs.

Table C.7: Opex model specification 1 results

Model Specification 1 Results: MEAV	Bottom-up Opex				Top-down Opex
	Work Management	Emergency	Repairs	Maintenance	
MEAV	0.517**	0.764**	0.737*	0.982***	0.728***
Time Trend	-0.023	-0.053***	-0.020	0.018	-0.020
Constant	45.296	102.203***	37.008	-42.567	37.510
Observations	80	80	80	80	80
R-squared	0.390	0.510	0.305	0.692	0.639

*Source: CEPA analysis. *** p<0.01, ** p<0.05, * p<0.1*

Table C.8: Opex model specification 2 results

Model Specification 2 Results: CSV1	Bottom-up Opex				Top-down Opex
	Work Management	Emergency	Repairs	Maintenance	
CSV1	0.517**	0.841***	0.836**	0.974***	0.764***
Time Trend	-0.022	-0.051***	-0.018	0.02	-0.018
Constant	41.866	95.415***	29.914	-48.909	31.864
Observations	80	80	80	80	80
R-squared	0.384	0.575	0.383	0.668	0.688

*Source: CEPA analysis. *** p<0.01, ** p<0.05, * p<0.1*

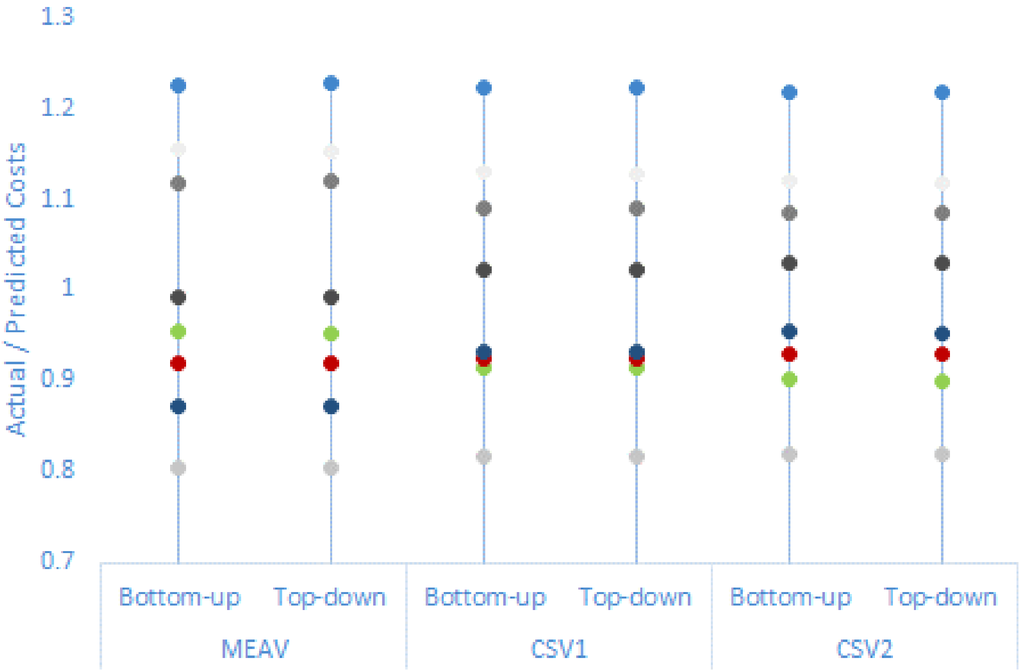
Table C.9: Opex model specification 3 results

Model Specification 3 Results: CSV3	Bottom-up Opex				Top-down Opex
	Work Management	Emergency	Repairs	Maintenance	
CSV2	0.511**	0.857***	0.859**	0.960***	0.768***
Time Trend	-0.022	-0.050***	-0.018	0.021	-0.018
Constant	40.868	93.151***	27.481	-50.726	30.083
Observations	80	80	80	80	80
R-squared	0.378	0.594	0.407	0.652	0.698

*Source: CEPA analysis. *** p<0.01, ** p<0.05, * p<0.1*

For each of the models presented above we have estimated the average efficiency score (actual costs divided by predicted costs) over the period 2009 to 2018 for each GDN. The figure below compares opex efficiency scores using the bottom-up approach versus the top-down approach for all three model specifications. As may be expected, the efficiency scores are very similar across both bottom-up and top-down approaches, and across all three model specifications.

Figure C.5: Opex model efficiency scores: bottom-up vs. top-down



Source: CEPA analysis