Review of Ofgem's proposed approach to cost assessment at GD3

—

Prepared for Wales & West Utilities

22 November 2024

Contents

Oxera Consulting LLP is a limited liability partnership registered in England no. OC392464, registered office: Park Central, 40/41 Park End Street, Oxford OX1 1JD, UK with an additional office in London located at 200 Aldersgate, 14th Floor, London EC1A 4HD, UK; in Belgium, no. 0651 990 151, branch office: Spectrum, Boulevard Bischoffsheim 12–21, 1000 Brussels, Belgium; and in Italy, REA no. RM - 1530473, branch office: Rome located at Via delle Quattro Fontane 15, 00184 Rome, Italy with an additional office in Milan located at Piazzale Biancamano, 8 20121 Milan, Italy. Oxera Consulting (France) LLP, a French branch, registered in Nanterre RCS no. 844 900 407 00025, registered office: 60 Avenue Charles de Gaulle, CS 60016, 92573 Neuilly-sur-Seine, France with an additional office located at 25 Rue du 4 Septembre, 75002 Paris, France. Oxera Consulting (Netherlands) LLP, a Dutch branch, registered in Amsterdam, KvK no. 72446218, registered office: Strawinskylaan 3051, 1077 ZX Amsterdam, The Netherlands. Oxera Consulting GmbH is registered in Germany, no. HRB 148781 B (Local Court of Charlottenburg), registered office: Rahel-Hirsch-Straße 10, Berlin 10557, Germany, with an additional office in Hamburg located at Alter Wall 32, Hamburg 20457, Germany.

Executive summary

WWU has a strong track record as a cost-efficient company that delivers on its service commitments. However, benchmarking based on Ofgem's GD2 main cost assessment model would suggest that WWU's cost-efficiency rank relative to other GDNs has been deteriorating (from second to fifth over the first three years of GD2).

In this report we assess the main drivers of this seeming deterioration, and whether it is more likely to be a relative increase in inefficiency or due to other factors (e.g. mismatches in profiling of GDN expenditures, new/additional exogenous cost pressures not accounted for in Ofgem's GD2 modelling suite, and/or potential cost savings at other GDNs not meeting their minimum service standards). We also review Ofgem's broader modelling considerations for GD3.

We identify the increased complexity of remaining mains replacement (REPEX) workloads, and also increased IT and cyber security requirements (if not excluded and separately assessed), as the main drivers of WWU's step change in costs. These activities are exogenous and affect all GDNs (though the timing of the impacts on costs may differ), as they are driven by the HSE's mains replacement policies, increased cyber security requirements on key national infrastructures and the increased digitalisation required from modern utility operators respectively.

It is therefore likely that the whole industry will face a similar step change in terms of the value and technical complexity of their underlying workloads. It will thus be important to assess this step change with the inclusion of GD3 forecast data (i.e. using structural break tests), and where appropriate, account for this in the GD3 modelling approach.

Note that WWU may have undertaken a relatively greater share of these more complex GD2 workloads over 2022–24 (e.g. for REPEX) than other GDNs, and is thus experiencing this step change earlier than other GDNs—explaining its seeming deterioration in the short term. This can be addressed by selecting a sufficiently long benchmark period, e.g. over the full RIIO-GD3 period, to smooth out profiling mismatches.

The GD3 cost assessment framework also needs to account for increased workload complexity. Otherwise, the cost assessment framework risks underfunding GDNs that undertake more complex, mandated workloads to meet customer commitments (and conversely awards others that avoid undertaking these workloads). For example, several GDNs have failed their minimum service standards for REPEX and emergency response times over the first three years of GD2.

Based on our assessment, we identify the following main considerations (and implications) for Ofgem's GD3 approach.

- **Time period used**: it will be important to test for a structural break in cost–cost driver relationships. Contingent on the results thereof (for each relevant cost assessment category), Ofgem may need to reconsider the relative weighting and/or treatment of historical and forecast data in its benchmarking.
- **REPEX complexity**: as WWU and Ofgem's consultants also noted at GD2 determinations,[1] the REPEX synthetic cost driver (and accompanying regional factor normalisations) will need to account for additional complexity drivers such as the technique, ground surface, pipe material[2] and sparsity of workloads.
- **Increased IT and cyber requirements**: the simplest way to deal with the step change in these costs may be a separate assessment (as the underlying costs are unlikely to be sufficiently captured by the cost drivers in the main regression). Alternatively, specific activity drivers for IT, cyber and related costs causing the step change would need to be used in the modelling.

We have also reviewed the following related areas, considered by Ofgem's SSMD and/or raised by companies in their response to the consultation.

- **Level of aggregation**: Given reporting inconsistency concerns[3] (and potential implications for operational trade-offs and cherry-picking of efficiency benchmarks), a TOTEX approach is likely to remain the most appropriate for determinations at GD3.
- **Choice of benchmark**: The choice of benchmark will depend on the robustness of Ofgem's models for GD3 (e.g. the statistical

[1] For example, see business plan document WWU (2019), 'Appendix 9D – Mains Replacement Performance RIIO-GD1', pp. 14–15 and CEPA (2020), 'RIIO-GD2: Synthetic Unit Costs Update', 27 February, p. 7.
[2] Accounting for types of iron mains (ductile vs spun or cast iron), and the differences in costs between them, at a more disaggregated level.
[3] That is, that cost allocations and capitalisation rates differ between GDNs, and within GDNs over time. See Ofgem (2024), 'RIIO-3 Sector Specific Methodology Decision – GD Annex', 18 July, paras 5.25–5.28 and WWU (2023), 'RIIO-3 Sector Specific Methodology Consultation (SSMC) – Wales & West Utilities (WWU) response', 6 March, GDQ50 and GDQ53, pp. 66–72.

precision of estimates and the reliability of the underlying data). The current five-year forecast benchmark period is likely to remain the most appropriate, in terms of both length (to smooth out differential expenditure profiles) and capturing expected cost pressures. Ofgem should also ensure that the benchmark is not influenced by GDNs that are underinvesting and/or not meeting minimum service standards.

- **Economies of scale—accounting for shared group costs:** Ofgem's GD2 GDN-level assessment of BSCs and other shared group-level costs is not consistent with operational or economic rationale (or Ofgem's precedent on the treatment of shared costs elsewhere). Shared costs that benefit from company-level economies of scale should be benchmarked at the group level.
- **MEAV as scale driver:** The alternative scale drivers being considered by Ofgem (e.g. customer numbers, throughput) will remain stable and eventually decline over time—and are thus inappropriate in the context of step changes in costs. That is, they would not account for increased operational costs per unit of scale. An asset value metric (such as MEAV), while still not capturing the step change immediately, is more appropriate as it at least incorporates increased workload complexity (and cost) steadily over time. As Ofgem notes, MEAV performs well at the TOTEX level, and when adding CAPEX to the cost pool.[4] As noted above, a more appropriate solution may thus be to find alternative activity drivers for, or separately assess, the elements of business support costs and work management that are not as well explained by MEAV. Note that a greater weighting to customer numbers or throughput, as alternative scale drivers, would also affect the level of pre-modelling regional factor adjustments required (given the correlation between customers/demand and sparsity/density).[5]

[4] Ofgem (2024), 'RIIO-3 Sector Specific Methodology Decision – GD Annex', 18 July, para. 5.34.
[5] As discussed in Oxera (2024), 'Regional factors for RIIO-GD3: Sparsity', report prepared for Wales & West Utilities, November.

1 Introduction

Ofgem's recent RIIO-3 Sector Specific Methodology Decision (SSMD) for gas distribution (GD) summarises the regulator's view on model development for GD3. The SSMD is informed by Ofgem's initial model testing and sector consultation through cost assessment working groups (CAWGs) and Sector Specific Methodology Consultation (SSMC).[6] This builds on Ofgem's approach in GD2.

Ofgem's overall GD2 cost assessment framework is summarised in Figure 1.1 below, which consisted of the following.[7]

- **A single top-down, TOTEX regression model** used to assess 86% of forecast costs (based on costs modelled over historical and forecast data from financial years 2014–26[8]).
- **Separate assessments** for the remaining forecast costs, using (i) non-regression techniques for the 8% of costs not well explained by the cost drivers in the TOTEX model, and (ii) technical, bottom-up assessments for the remaining 6% of costs relating to discrete, atypical and/or bespoke investments.

Before GDN costs were benchmarked in the main regression, Ofgem applied a range of pre-modelling normalisations and adjustments to submitted costs—with the aim of comparing costs on a like-for-like basis. These included:

- **reclassifications**, where activities costs have been reported incorrectly and/or inconsistent with other GDNs;
- **regional and company-specific factors**, to account for contextual factors that may cause the efficient costs to be higher in certain regions (and is not captured in the modelling);
- **workload (forecast) adjustments**, where the forecast volumes of work are considered inefficient or poorly justified;
- **exclusions**, where costs are not explained by the cost drivers used, or where there is a substantial change in the nature of the activity being undertaken (e.g. cyber security).

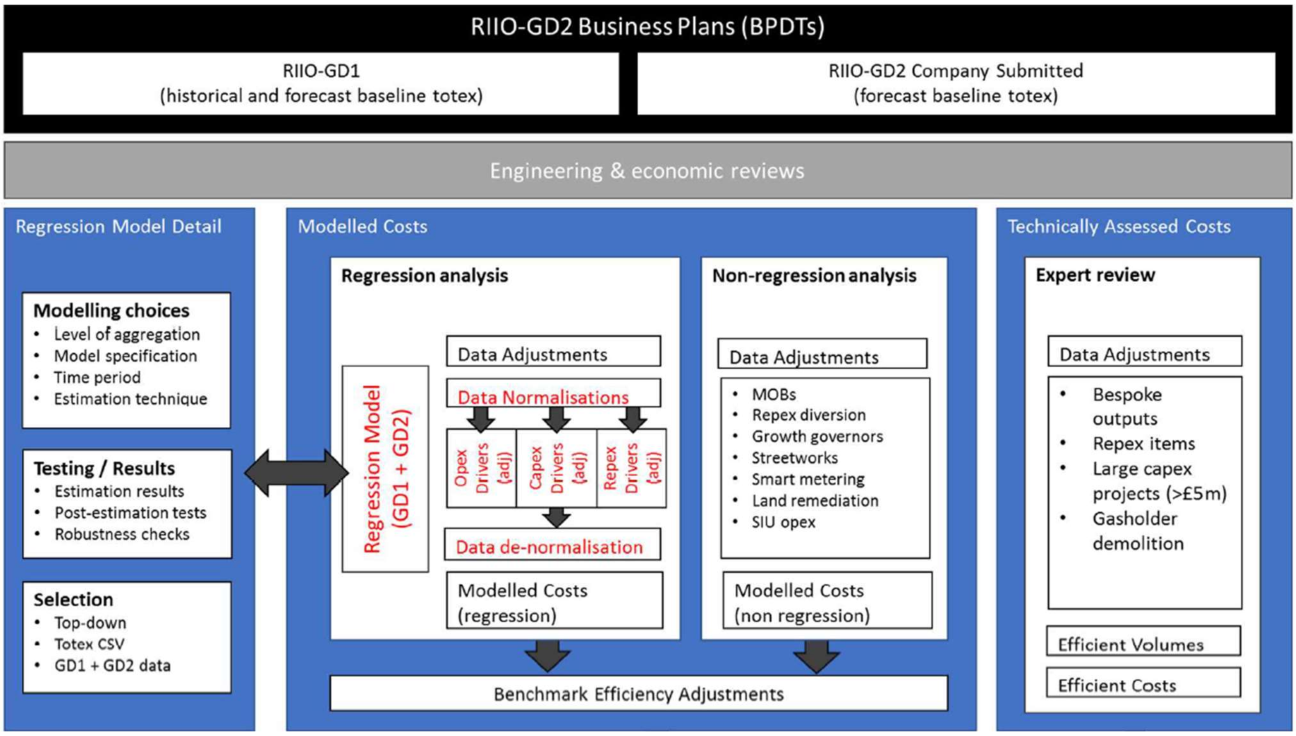
[6] Ofgem (2024), 'RIIO-3 Sector Specific Methodology Decision – GD Annex', 18 July.
[7] Details of the cost assessment approach in Ofgem (2021), 'Decision - RIIO-2 Final Determinations – GD Sector Annex (REVISED), 3 February, section 3.
[8] For the remainder of this report, we refer to financial years based on the year in which they end (e.g. 2013/14 is referred to as 2014), unless stated otherwise.

Ofgem applied two types of efficiency challenge to GDNs' predicted costs, post modelling/assessment:

- **a static catch-up efficiency** challenge for all GDNs whose costs are assessed to be less efficient than the benchmark, or frontier, GDN (which Ofgem refers to as 'benchmarking efficiency'). This challenge applies to both regression and non-regression 'modelled' costs (but not technically assessed costs);
- **a frontier shift challenge** applied to all GDNs' costs (including those technically assessed), to capture the frontier GDN's expected ongoing efficiency improvements over time.[9]

In GD2, the catch-up benchmark was based on the main TOTEX model and an assessment of the benchmark GDN's forecast efficiency for the five-year regulatory period (2022–26).[10] A 75th percentile benchmark was selected for the first year, followed by a glide path to the 85th percentile (the latter applicable from the last two years).

Figure 1.1 Ofgem's GD2 overall cost assessment process



Source: Ofgem (2020), 'RIIO-GD2 Final Determinations: Step-by-Step Guide to Cost Assessment', figure 1.

[9] Ofgem (2021), 'Decision - RIIO-2 Final Determinations – GD Sector Annex (REVISED), 3 February, para. 3.8.
[10] Efficiency is defined as GDNs' forecast costs over model predicted costs.

Ofgem has not made definitive decisions on its modelling approach for GD3. Instead, it has indicated that the GD2 framework will largely form the basis for the GD3 approach, and that it will continue model testing with GD3 business plan forecast data. Nevertheless, Ofgem's SSMD (read alongside its CAWG presentations) does provide an indication of the direction of travel and potential changes for GD3—summarised below.

- **Pre-model normalisations:** to recalibrate regional wage indices and the activities they cover (e.g. noting that London wages have increased more slowly than the rest of the UK), as well as testing a density (squared) variable within-model to account for both urbanity and sparsity effects (not our focus in this report[11]).[12]
- **Cost drivers**: refining the GD2 TOTEX composite scale variable (CSV) by reassessing the role of modern equivalent asset value (MEAV), seemingly as it relates to Work Management and Business Support costs (BSCs) in particular; recalibrating the synthetic workload drivers for REPEX and connections (CAPEX); exploring alternative cost drivers and CSVs (including alternative CSV weightings)—discussed in section 3.1.
- **Level of aggregation**: testing multiple TOTEX model approaches and considering combining the results across robust models (discussed in section 3.2).
- **Time period**: testing alternative time periods, dummies, trends and conducting related specification tests (discussed in section 3.3).[13]
- **Group-level scale effects**: considering potential economies of scale benefits, e.g. from sharing BSCs or bulk asset purchasing.
- **Exclusions and separate assessments**: Ofgem has indicated that it will revisit the separate assessment of certain areas, although it is minded to retain as many cost areas/activities as possible within the primary regression model(s) (discussed in section 3.5).[14]

[11] We deal with the approach to normalisations, specifically regional factors, in separate reports Oxera (2024), 'Regional factors for RIIO-GD3: Sparsity', report prepared for Wales & West Utilities, November and Oxera (2024), 'Regional factors for RIIO-GD3: Regional wages', report prepared for Wales & West Utilities, November.
[12] Ofgem (2024), 'RIIO-3 Sector Specific Methodology Decision – GD Annex', 18 July, paras 5.46–5.47.
[13] See Ofgem (2024), 'RIIO-GD3 Cost Assessment Working Group 7. Totex modelling and BPDT development', 10 April, slide 7.
[14] Ofgem (2024), 'RIIO-3 Sector Specific Methodology Decision – GD Annex', 18 July, paras 5.30–5.35, 5.54, 5.67, and 5.80–5.81.

One element not discussed in the SSMD, but which Ofgem will also need to reconsider (in light of the robustness of its eventual GD3 modelling) is the efficiency benchmark choice. This is discussed in section 3.4.

In light of this, Wales & West Utilities (WWU) has commissioned Oxera to review (i) the scope of Ofgem's main modelling considerations for GD3,[15] and (ii) whether these are likely to be sufficient in scope to deal with the step change in costs and related changes in the complexity of workload that WWU is, and GDNs more generally are, expecting over the remainder of GD2 and GD3.

WWU is particularly concerned that exogenous changes in the nature and complexity of its remaining mains replacement (REPEX) and IT and cyber workloads (captured under BSCs and other CAPEX) will not be sufficiently accounted for within the models. This would risk underfunding these necessary activities (or cause shortfalls elsewhere).

The remainder of this report is structured as follows.

- **Section 2** provides an overview of the main exogenous drivers, from an operational perspective, for the step-change in costs that WWU expects going forward.
- **Section 3** discusses implications for Ofgem's cost assessment framework, both the current TOTEX regressions (sections 3.1–3.4) and exclusions and broader separate assessment (section 3.5).
- **Section 4** concludes.

[15] Specially as it relates to cost aggregation, model specification and cost drivers for the main TOTEX regression, and potential interactions with separately assessed costs.

2 Context: expected step change in efficient costs

In this section we discuss the following.

- In section 2.1 we assess WWU's relative efficiency at the TOTEX level over time and investigate what the drivers of WWU's seeming decrease in relative position are (based on middle-up models)
- Section 2.2 assesses what the drivers of WWU's relative increase in costs are (based on descriptive evidence from WWU's annual regulatory reporting). We also assess whether the latter are outside of management control (i.e. exogenous), and, if so, whether they are sufficiently accounted for in Ofgem's broader cost assessment framework.
- Section 2.3 investigates WWU's relative service performance .

2.1 WWU's relative cost efficiency

Historically, WWU has performed at or better than what Ofgem has considered to be the efficiency benchmark. Both at the time of the GD2 final determinations (FD) and over the entire GD1 outturn period (2014–21[16]), WWU has been a cost-efficient company—ranking second out of the eight GDNs (as shown in Table 2.1 below).[17]

However, WWU's relative position has deteriorated when focusing on only the first three years of GD2—ranking fifth (see similar results in Ofgem's CAWG modelling tests[18]). Note that while three years is too short a time period over which to consider WWU's performance relative to the benchmark, it does provide an indicative snapshot.

[16] Regulatory years run in parallel to financial years (i.e. July to June of the following year). However, for simplicity, we refer to the calendar year in which the regulatory year ends, e.g. 2020/21 is referred to as 2021.
[17] Based on Ofgem's current 85th percentile cost efficiency benchmark, the top two companies are efficient.
[18] Ofgem (2024), 'RIIO-GD3 Cost Assessment Working Group 7. Totex modelling and BPDT development', 10 April, slide 11.

Table 2.1 WWU's cost efficiency to date

Assessment	Model period	Benchmark period	WWU rank
Ofgem FD[1]	2014–26	2022–26	2
GD1 outturn[2]	2014–21	2017–21	2
GD2 outturn (to date)[3]	2014–24	2022–24	5

Sources: [1] Ofgem (2020), '[7] CostAssessment_File', FD model; [2] Oxera analysis based on Ofgem's RRP23 updated data as at November 2023 (correcting MEAV data pull errors[19]); [3] Oxera analysis based on RRP24 data, removing re-openers, major capital projects (>£5m) and using Ofgem's 2024 updated regional factor indices.

The reason for what appears to be a deterioration in position over 2022–24 may be due to real deterioration in efficiency. However, it may also be explained by other factors (especially given the narrow benchmarking window), including the following.

- **Mismatches in profiling of expenditure:** While Ofgem uses (a more appropriate) five-year benchmarking period, one would expect relative efficiency trends over too narrow a time window to be volatile due to profiling mismatches. For example, if WWU has front-loaded its GD2 planned activities and expenditures relative to other GDNs, this may not be reflected immediately and sufficiently in Ofgem's TOTEX CSV (because major elements, such as MEAV, are fairly stable over time), and/or in other GDNs' cost bases. While WWU would then appear inefficient *temporarily*, profiling mismatches would smooth out over a longer benchmarking period (such as five years).[20]
- **New/additional cost pressures not accounted for**: WWU may be experiencing exogenous cost pressures not accounted for in the current modelling—e.g. due to the increased complexity of workloads and/or additional activities not accounted for by either the cost drivers or normalisations in Ofgem's modelling suite. Unlike profiling mismatches, these issues may or may not be experienced by other GDNs (and will therefore not necessarily smooth out over time[21]). The latter is contingent on

[19] We noted several GDNs (incorrectly) had the exact same MEAV sub-component values, due to a data-pull error in the relevant Ofgem file. We corrected these before updating the analysis.
[20] WWU may also be experiencing short-term fluctuations because it has incurred large investments (e.g. new IT and cyber security systems), which other GDNs have either incurred earlier, or are still to incur (which will similarly smooth out over longer periods of assessment).
[21] This may interact with WWU's profiling of expenditures. For example, if WWU has front-loaded activities relative to other GDNs, and is experiencing higher costs due to the increased complexity of workloads and/or additional work that these activities entail, these relatively higher costs will

whether these activities, and related upward cost pressures, will eventually be common across GDNs (with WWU just being affected earlier).

The latter is of greatest concern—as profiling mismatches should be ameliorated by a longer benchmarking period—and thus our focus in this report.

By examining the middle-up models, we identify the drivers of WWU's seeming decrease in relative position. Figure 2.1 shows GDN rankings from middle-up modelling over the GD1 and GD2 outturn period to date (2014–24). Note that we remove re-openers (e.g. cyber security and net zero expenditure) and major CAPEX projects across companies, assuming that these costs will remain separately assessed at GD3.

Based on our analysis, it appears that WWU has had a drop in efficiency at the TOTEX level which is being driven by REPEX—WWU's ranking has dropped from second over GD1 to eighth over GD2 to date. Otherwise, WWU's position has remained unchanged (CAPEX) or improved (OPEX).

only start to materialise for other GDNs if and when they incur similar expenditures at later points in time.

Figure 2.1 Middle-up efficiency rankings over time

REPEX

GDN	GD1 outturn	2021/22	2022/23	2023/24*	GD2 outturn
EoE	3	7	7	6	7
Lon	5	6	6	5	5
NW	4	5	5	4	4
WM	7	8	4	3	3
NGN	1	1	3	1	1
Sc	6	2	1	2	2
So	8	4	2	8	6
WWU	2	3	8	7	8

OPEX

GDN	GD1 outturn	2021/22	2022/23	2023/24*	GD2 outturn
EoE	6	7	6	4	6
Lon	8	8	8	7	8
NW	7	6	7	5	7
WM	5	4	3	3	3
NGN	3	1	2	6	4
Sc	1	2	1	2	1
So	2	3	5	8	5
WWU	4	5	4	1	2

CAPEX

GDN	GD1 outturn	2021/22	2022/23	2023/24*	GD2 outturn
EoE	7	5	3	8	6
Lon	2	2	5	1	2
NW	3	7	6	7	7
WM	5	8	8	6	8
NGN	8	4	2	3	1
Sc	6	3	4	2	3
So	1	1	7	4	5
WWU	4	6	1	5	4

Source: Oxera analysis based on Ofgem Nov. 2023 model update and 2024 RRP data.

2.2 WWU's cost pressures

In this section, we provide more detail on the drivers of the increase in WWU's costs. First, in section 2.2.1, we examine REPEX, then, in section 2.2.2, we examine IT and cyber security costs (as part of BSCs and other CAPEX).

In recent annual reporting, CAWG inputs and SSMC response, WWU consistently references two categories of cost pressure, resulting in the higher-than-initially-expected TOTEX requirements:

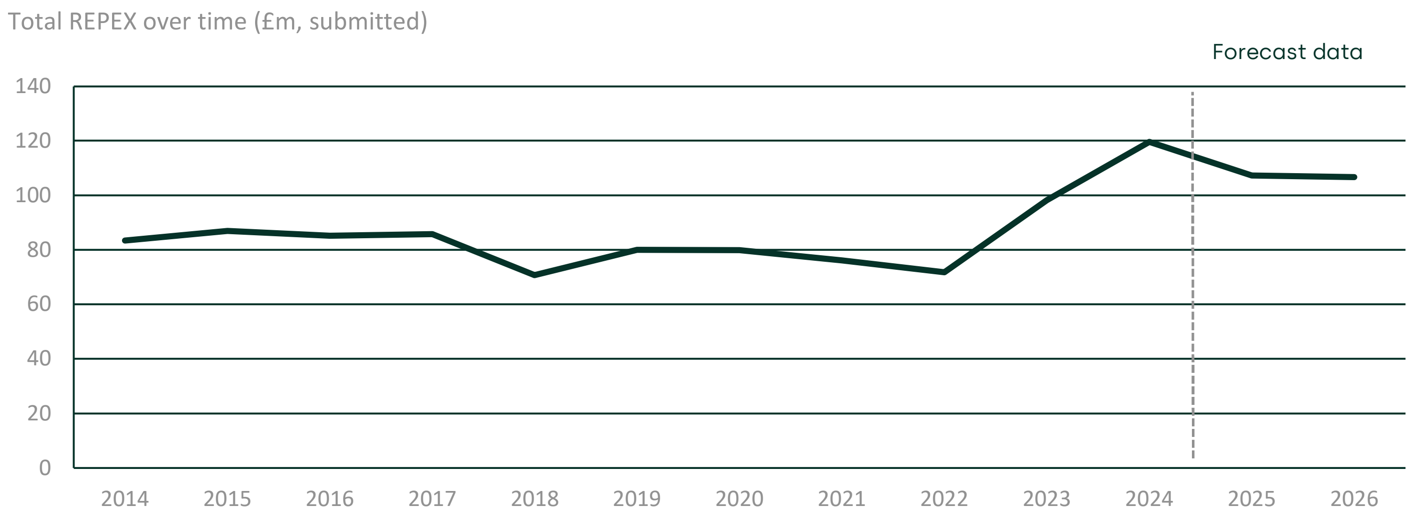
- **REPEX**: cost pressures resulting from more complex remaining workloads, contract labour and broader supply-chain shortages (WWU has reacted to the latter by insourcing a significant portion of the workload);
- **strategic IT investments and cyber security**: under the Network and Information Systems Regulations (NIS-R), and changes therein, GDNs are required to take appropriate and proportionate cyber security measures to manage the risks and be resilient against the increased threat of cyber-attacks.[22]

This is consistent with WWU's disaggregated cost trends.

2.2.1 REPEX

Figure 2.2 shows a clear step change in WWU's REPEX costs over 2023 and 2024, which is expected to be sustained over the remainder of the period (based on WWU's latest RRP 2024 forecasts).

Figure 2.2 WWU step change in REPEX



Note: Submitted costs, 2018/19 prices.

[22] As discussed in WWU (2024), 'RIIO-GD2 Year Three Strategic Performance Overview', July, pp. 13 and 18–19; WWU (2023), 'RIIO-GD2 Year Two Strategic Performance Overview'', July, pp. 3 and 9; WWU (2023), 'RIIO-3 Sector Specific Methodology Consultation (SSMC) – Wales & West Utilities (WWU) response', 6 March, GDQ50 and GDQ51, pp. 66–70.

Source: WWU 2024 RRP forecasts.

As Ofgem notes, the Health and Safety Executive's (HSE) regulations are the primary driver of REPEX.[23] That is, the sequencing of workloads is determined by the HSE's three-tier, risk-based approach under the Iron Mains Risk Reduction Programme (IMRRP) (and changes therein over time)—with the current tiering system summarised in Table 2.2.

Table 2.2 Iron mains categories in the HSE's IMRRP

Tier	Characteristics of mains	Action on GDNs
Tier 1	Less than or equal to eight inches in diameter.	Must be decommissioned by 2032.
Tier 2A	Greater than eight inches to less than 18 inches in diameter, which breaches a risk-action threshold.	Must be decommissioned or remediated over the period of the GDN's Approved Programme.
Tier 2B	Greater than eight inches to less than 18 inches in diameter, which is below a risk-action threshold.	Mains can remain operational but decommissioning can be funded if supported by cost–benefit analysis (CBA).
Tier 3	Greater than 18 inches in diameter.	Mains can remain operational but decommissioning can be funded if supported by CBA.

Source: Ofgem (2023), 'RIIO-3 Sector Specific Methodology Consultation – GD Annex', table 4.

The IMRRP has restricted the ability of GDNs to choose which mains to replace, requiring them to decommission those classified as greatest risk first. The nature of GDNs' remaining workloads are thus largely outside of management control, predetermined by the decisions and direction of the HSE's policy (and developments therein) to date.

These highest-risk mains were, however, also the lowest cost to replace: typically lower diameter, cast- or spun-iron mains within close proximity to buildings (and thus often under simpler surface areas). In contrast, WWU indicates that the remaining workload complexity has increased

[23] Ofgem (2023), 'RIIO-3 Sector Specific Methodology Consultation – GD Annex', 13 December, para. 3.6.

over GD2 (and will remain similarly so for GD3),[24] driven by the factors outlined below.

- **Increased diameter of the pipes to be replaced:** larger diameter pipes are more expensive and increase workload complexity (e.g. requiring larger holes, more specialised skills, and are often located under more challenging road surfaces).
- **The material of pipes**: a greater share of the remaining mains are ductile iron—which are more costly replacements (requiring more time and specialist tools to cut, relative to simpler cracking for cast- and spun-iron mains).
- **More challenging ground surfaces**: a greater share of mains are under roadways, which have higher costs relative to footpaths or the verges due to implications for traffic management and reinstatement costs.
- **Required mains replacement technique**: a greater share of the remaining projects require more costly and complex open cuts (relative to simpler insertions).
- **Increased workload sparsity**: the remaining workloads are in more sparse areas, increasing transport, labour and logistics costs—see Oxera (2024).[25]
- **The smaller size of the remaining projects**: given the fixed mobilisation (transport and logistics) cost per project, total costs increase as the length of mains replaced per project decreases.

This is consistent with WWU's increased simple REPEX unit costs shown in Figure 2.3,[26] which have increased significantly from 2021 and over the first three years of GD2. From GD1 to GD2 to date, WWU's average unit cost has increased from £198.34 to £241.34 per meter replaced in aggregate (c. 22%), and from £171.57 to £215.30 per meter for Tier 1 mains (c. 25%)—with Tier 1 mains the predominant REPEX activity.[27]

[24] WWU (2023), 'BPDT & RRP feedback, CAWG meeting 2', 16 November and WWU (2024), 'Repex', CAWG meeting 5, 27 February and WWU (2023), 'RIIO-3 Sector Specific Methodology Consultation (SSMC) – Wales & West Utilities (WWU) response', 6 March, GDQ50 and GDQ51, pp. 66–70.
[25] Oxera (2024), 'Regional factors for RIIO-GD3: Sparsity', report prepared for Wales & West Utilities, November.
[26] Replacement and related service costs per meter of mains replaced.
[27] For WWU, Tier 1 mains represent 88% of volumes (km) and 76% of costs (REPEX) over 2014–24.

Figure 2.3 Increase in WWU's REPEX unit cost (per meter)

Note: Submitted costs (including mains replaced and related services), 2018/19 prices.
Source: Oxera analysis based on Ofgem data (GD1) and WWU 2024 RRP (GD2).

2.2.2 IT and cyber security (under BSCs and other CAPEX)

Figure 2.4 shows a similar step change in WWU's BSCs, driven by IT and cyber security investments and the resulting higher operational costs (which are expected to peak over 2025–26). Of WWU's £190.7m planned BSCs over GD2, almost half (48%) are IT and cyber security, with the annual share of these categories growing from 21% in 2022 to 55% forecast by 2026—as shown in Figure 2.4.

Note that this also affects the 'Other CAPEX' category (not shown here), where the corresponding IT and cyber investment expenditure is captured.

Some of the short-term cost increases are funded through re-openers, mostly for cyber-operational and information technologies (OT and IT) and some non-operational IT—collectively 19% of total GD2 BSCs.

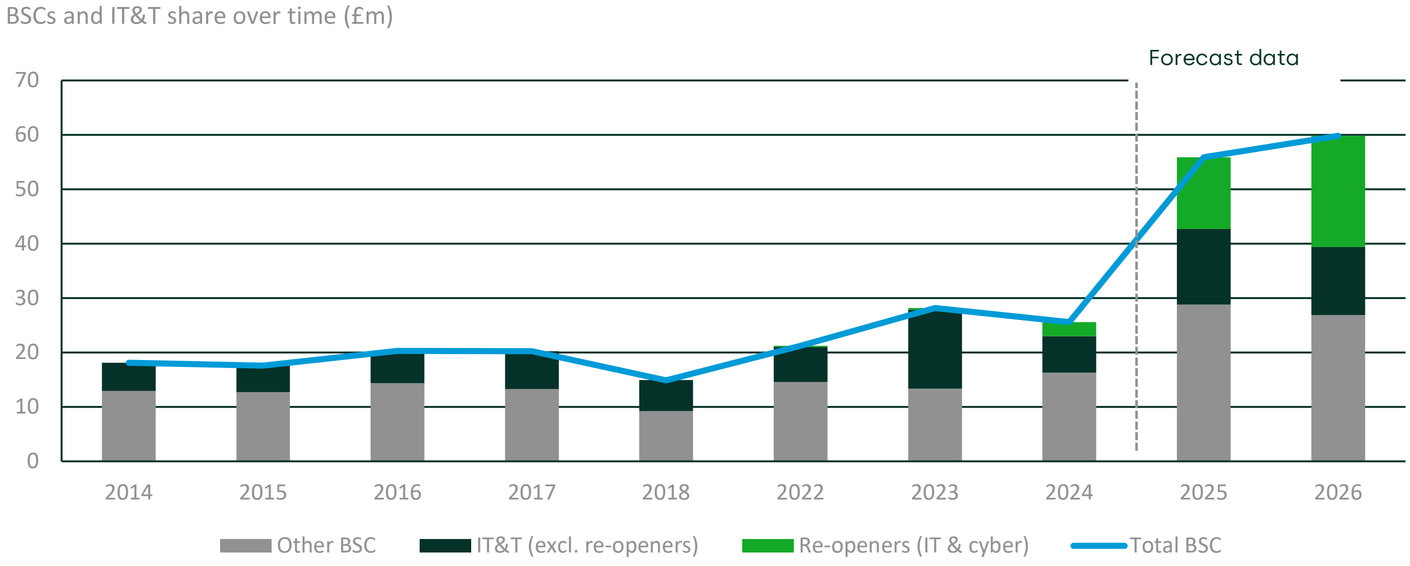
Cyber security was excluded from the main regression at GD2 (similar to Ofgem's ED2 approach, where cyber costs were technically assessed[28]). We would thus not expect these costs to be assessed under the main

[28] Ofgem (2022), 'RIIO ED2 Final Determinations Core Methodology Document', 30 November, table 16, p. 222.

regression modelling at GD3. However, Ofgem has not yet stated whether it will continue to apply exclusions and separate assessments for cyber security, nor how the other IT re-openers will be treated.

Further, we note that even excluding cyber re-openers, WWU's annual BSCs are expected to be 48% higher on average compared with its BSCs at GD1.[29] WWU expects that the sector will have to maintain a higher base operating BSC, following current IT and cyber investments (alongside other new net zero and energy-transition-related spend).

Figure 2.4 WWU's IT driven step change in BSCs



Note: Submitted costs, 2018/19 prices.
Source: Oxera analysis based on WWU 2024 RRP.

2.3 WWU's relative service performance

At the end of GD1 (2021), Ofgem concluded that GDNs had, in general, delivered on their output targets—bar some connection and supply interruption targets affected by COVID-19 (and which GDNs committed to delivering during GD2 without additional allowances).[30] Ofgem's summary of GDN service performance is shown in Table 2.3 below.

[29] BSCs over GD2, excluding cyber re-openers, are £28.7m p.a.—in comparison to c. £19.4m p.a. over GD1.
[30] Ofgem (2022), 'RIIO-GD1 Annual Report 2020-21', 28 September.

At the time of GD2 FD, Ofgem therefore had no reason to suspect that GDNs were outperforming allowances by underinvesting or not meeting minimum service standards (i.e. reducing quality).

Table 2.3 Ofgem summary of GD1 service performance (2021)

Company	Network	Environmental	Connections	Customer Service	Social Obligation	Safety	Reliability
Cadent	EoE		*				
	Lon						*
	NW		*				
	WM		*				
NGN	NGN						
SGN	Sc						
	So						
WWU	WWU		*				

KEY	Score based on annual, cumulative performance and Ofgem's consideration of COVID-19 pandemic.
	GDNs have successfully achieved an annual output or met RIIO-GD1 target.
*	GDNs output target impacted by COVID-19 pandemic and are committed to achieving target in RIIO-2 without additional allowance.
	GDNs have not achieved an annual or cumulative target.

Source: Ofgem (2021), 'RIIO-GD1 Annual Report 2020-21', p. 6.

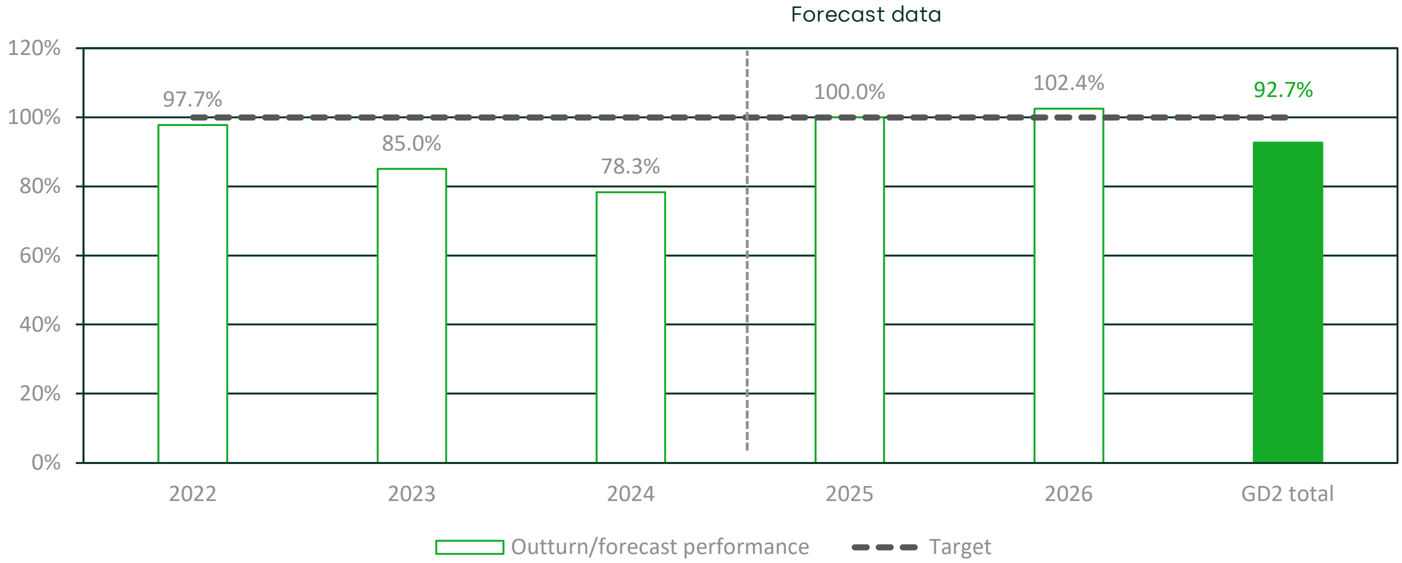
However, over GD2, while WWU is on track to continue to meet or exceed its service performance targets, this has not proved to be the case across all GDNs. The most clear example of this is minimum safety standards, where not all GDNs are meeting (i) their respective HSE primary risk removal targets for Tier 1 mains replacements, or (ii) the industry-wide emergency response standard.

For example, for Tier 1 mains replacement, SGN's Southern GDN has both underperformed against its 2022–24 annualised decommissioning targets[31] and is forecasting not to deliver its cumulative GD2 period

[31] Achieving 97.7%, 85.0% and 78.3% of its baseline target workloads over each year of 2022–24.

workload target (Southern is forecasting that it will decommission only 92.7% of its c. 3,001km GD2 target—as shown in Figure 2.5).[32]

Figure 2.5 Southern's GD2 Tier 1 mains replacement performance



Note: Performance measured against annual baseline target of 600.3km of Tier 1 mains decommissioned.
Source: Oxera analysis based on Southern's 2024 RRP and GD2 FD targets.

Four GDNs have also failed to meet their annual emergency response standards over GD2—which requires GDNs to attend 97% of reported escapes within one hour for uncontrolled escapes and two hours for controlled escapes respectively (see Table 2.4)[33]—and are being investigated by Ofgem.[34] In contrast, WWU has exceeded its response standard—averaging at and above 99% for the two emergency response categories and ranking second on average.

[32] While other GDNs such as WWU and Lon have also not kept in step with their Tier 1 decommissioning targets over 2022–24 (achieving 99% and 95% of their respective target workloads to date), both GDNs are forecasting to achieve their cumulative targets by the end of GD2. WWU has also caught up by delivering most of its lagging volumes in 2024 (delivering more than 107% of its annualised workload target).
[33] This is based on GDNs not reaching annual targets. Based on average performance over the 2022–24 period, Cadent's London and SGN's Southern GDNs have also underperformed the 97% standard for controlled escapes on average, and Southern has underperformed the equivalent standard for uncontrolled escapes, on average.
[34] Ofgem, (2024), '[Investigations into Cadent Gas Limited, Scotland Gas Networks Plc and Southern Gas Networks Plc, and their compliance with their obligations under their gas transporter licence]', 24 August.

Such underperformance against minimum performance standards may affect the reliability of Ofgem's cost benchmarking—and should therefore be carefully considered. For example, if a GDN is avoiding the more complex (and costly) elements of its mandated REPEX programme by not delivering it, this would be to the detriment of the cost models by (i) making the GDN seem more efficient than it would have been (if it had undertaken the mandated work), and (ii) potentially biasing downwards the efficient cost estimates for other GDNs that did undertake these more complex workloads. Similarly, GDNs that retain a relatively greater number of engineers across their network to meet emergency standards may also be assessed to be relatively less efficient for doing so.

Table 2.4 GD2 emergency response performance

Share of uncontrolled escapes attended to in one hour (target: 97%)

GDN	2022	2023	2024	Average	Rank
EoE	98.1%	97.1%	98.9%	98.0%	5
Lon	97.6%	95.2%	98.3%	97.0%	7
NW	98.1%	96.0%	98.8%	97.6%	6
WM	99.3%	97.6%	98.4%	98.4%	3
NGN	99.7%	99.5%	99.8%	99.7%	1
Sc	98.1%	97.2%	99.5%	98.3%	4
So	97.9%	91.9%	98.4%	96.0%	8
WWU	99.0%	98.6%	99.3%	99.0%	2

Share of controlled escapes attended to in two hours (target: 97%)

GDN	2022	2023	2024	Average	Rank
EoE	98.8%	97.7%	99.3%	98.6%	5
Lon	97.5%	94.9%	98.2%	96.9%	8
NW	98.9%	96.5%	99.0%	98.1%	6
WM	99.8%	97.8%	98.8%	98.8%	3
NGN	99.9%	99.7%	99.9%	99.9%	1
Sc	99.5%	96.6%	99.8%	98.6%	4

GDN	2022	2023	2024	Average	Rank
So	98.9%	92.9%	98.9%	96.9%	7
WWU	99.9%	99.4%	99.8%	99.7%	2

Note: Underperformance against 97% standard is highlighted in red font.
Source: Oxera analysis based on GDNs 2024 RRP and GD2 FD targets

3 Implications for cost assessment

In this section, we focus on the main regression model(s) and potential exclusions (i.e. interactions with separate assessments). In particular, we examine:

- cost drivers (for areas where a step change is expected);
- the level of aggregation;
- the time period of the cost model estimation;
- the choice of benchmark;
- which costs require separate assessment.

3.1 Cost drivers

3.1.1 REPEX

For mains replacements, Ofgem's GD2 REPEX synthetic cost driver is effectively a complex unit cost measure. It combines the length of different types of mains replaced (by diameter and material) and the number of related services with each weighted by the industry average unit costs for the specific disaggregated component.[35]

However, the GD2 REPEX synthetic cost driver only accounts for a subset of the elements that contribute to WWU's increased complexity. While the REPEX synthetic cost driver accounts for the length and diameter of mains replaced, and for some differences in pipe materials (e.g. between iron and steel), it does not account for other exogenous cost differences due to the complexity of workloads referenced above (between ductile iron and spun or cast iron, ground surface, technique required, sparsity of remaining workload, etc.).

Therefore, Ofgem should reconsider the evidence for a sparsity regional factor cost adjustment, as discussed in Oxera (2024),[36] as well as the cost driver construction as follows:

- the disaggregation, where it already collects the relevant data (e.g. ductile vs spun/cast iron);
- collect additional data on the other main elements driving workload complexity and cost.

[35] See the GD2 REPEX synthetic cost driver construction discussed in the Ofgem-commissioned report: CEPA (2020), 'RIIO-GD2: Synthetic Unit Costs Update', 27 February.
[36] Oxera (2024), 'Regional factors for RIIO-GD3: Sparsity', report prepared for Wales & West Utilities, November.

For example, we understand that all GDNs should collect the GPS coordinates of their REPEX workloads for operational purposes (also historically). For other key drivers, like ground surface and technique, this data can be requested and incorporated on a forward-looking basis through the business plan data tables (with outturn data already largely reported in GDNs RRPs).

A complementary solution, or second best (and more crude) alternative, would be to place greater weight on more recent REPEX outturn and/or forecast data (as discussed in section 3.3 below). This would improve REPEX cost predictions to the extent that the more recent data better accounts for the increased complexity of the underlying workloads.

We note that these points were also raised at the time of the initial GD2 consultations. For example, in the GD2 REPEX cost driver construction report commissioned by Ofgem, Ofgem's consultants stated:

We have identified four drivers of mains replacement costs: pipe diameter, replacement technique, ground surface and pipe material. Cost and volumes data are only currently broken down by pipe diameter and pipe material, which is reflected in the updated synthetic unit costs [...]

Ofgem could consider taking into account differences in replacement technique and ground surface between GDNs by making post-modelling adjustments to repex allowances based on engineering judgement.[37]

WWU already noted the forecast cost pressures in its GD2 business plan submissions. Consistent with its outturn to date, WWU expected that a changing workload mix (specifically in terms of the increased share of ductile iron mains, wider pipe diameters, and more open cut workloads, alongside the broader geographical spread thereof) would increase the complexity and cost of workloads over GD2.[38]

3.1.2 BSCs

For BSCs (and for work management), Ofgem noted in the SSMD that MEAV may not be a particularly robust cost driver. However, Ofgem also

[37] CEPA (2020), 'RIIO-GD2: Synthetic Unit Costs Update', 27 February, p. 7.
[38] For example, see business plan document WWU (2019), 'Appendix 9D – Mains Replacement Performance RIIO-GD1', pp. 14–15. WWU also noted additional contractual and labour market challenges, which it has since sought to resolve through greater insourcing.

noted that MEAV performs well at the TOTEX level, and when adding CAPEX to the cost pool.[39]

In the GD2 TOTEX model, MEAV is used as the default driver for all OPEX and CAPEX cost categories that do not have a specific activity driver. Ofgem preferred MEAV over other measures of scale, such as network length or customer numbers, because it was deemed to better capture network complexity (and so the cost impact of scale and complexity).[40]

As indicated in section 2.2.2, WWU is expecting a step change in BSCs (driven primarily by increased IT and cyber security expenditures). There is no specific activity driver in Ofgem's models that would pick up the step change in IT and cyber costs.

In this context (and especially if IT and cyber costs remain within the model[41]), a scale driver that remains fairly stable over time (and thus does not account for increased operational costs per unit of scale) would be inappropriate. This would explain why a fairly stable MEAV metric may perform less well on more recent costs.

However, we note that among the alternative cost drivers tested by Ofgem during the CAWGs—i.e. some combination of network length, customer numbers and throughput[42]—MEAV would still be preferable.

That is, these alternative cost drivers are *all* stable and/or declining, in an environment where workload complexity and costs are increasing. Customer numbers and throughput, in particular, are expected to decline for all GDN given the expected decline in gas demand.[43]). Therefore, even in a steady-state environment (where the costs to maintain the existing network are largely fixed), declining scale drivers would be inappropriate and risk underpredicting costs for all GDNs on a forward-looking basis. This is exacerbated in the current environment where workload complexity and costs are increasing per unit of scale (be it per customer or unit of gas supplied).

In comparison, an asset value metric (like MEAV) could at least incorporate the increased complexity (and cost) steadily over time.

[39] Ofgem (2024), 'RIIO-3 Sector Specific Methodology Decision – GD Annex', 18 July, para. 5.34.
[40] Ofgem (2021), 'Decision - RIIO-2 Final Determinations – GD Sector Annex (REVISED)', 3 February, para. 3.114.
[41] Note that BSCs are still expected to increase, even when IT and cyber costs are excluded.
[42] Ofgem (2024), 'RIIO-GD3 Cost Assessment Working Group 7. Totex modelling and BPDT development', 10 April, slides 10–15.
[43] See National Energy System Operator (2024), 'Future Energy Scenarios (FES) 2024'.

Relatedly, we note that if a greater weighting were given to customer numbers or throughput as a cost driver in the TOTEX level modelling,[44] this would also require Ofgem to reconsider its pre-modelling regional factor adjustments (given the correlation between customer numbers and sparsity/urbanity). For example, the greater weighting to customers, the larger the compensating sparsity adjustment that a GDN such as WWU would require (given the relatively fewer customers to serve/demand to meet over larger land areas). Urbanity regional factor adjustments would need to be revisited for the same reason.

Therefore, instead of reconsidering the cost driver, the simplest and most appropriate way for Ofgem to deal with such a step change in these costs would be to either:

- include specific activity drivers for IT, cyber and related costs causing the step change; or
- separately assess the categories of cost that are not captured by the cost drivers in the main regression. We discuss this further in section 3.5.

3.2 Level of aggregation

In general, one would expect that the use of multiple models will make a cost assessment outcome both more robust (if all the models used are similarly robust) and transparent (as consistencies and inconsistencies between models can be investigated and understood). In this sense, the use of middle-up and disaggregated models should at least help validate and explain TOTEX, top-down results.

However, for GD3 cost determinations, TOTEX models may be preferred given two issues that would specifically affect disaggregated models.

- **Reporting inconsistencies**: GDNs have noted concerns about differences in cost allocations and capitalisation rates between GDNs, and within GDNs over time.[45] This implies that Ofgem would not have the consistent allocation of costs to activities necessary for bottom-up benchmarking, which in itself is a sufficient reason to focus on TOTEX models for allowances.

[44] That is, even if Ofgem were to use customer numbers as a cost driver, for say BSC or work management, it would still be modelled at the TOTEX level (and thus affect the cost predictions for all costs with which customers are correlated, in proportion to the weighting given to it in the TOTEX CSV).

[45] Ofgem (2024), 'RIIO-3 Sector Specific Methodology Decision – GD Annex', 18 July, paras 5.25–5.28. See also WWU (2023), 'RIIO-3 Sector Specific Methodology Consultation (SSMC) – Wales & West Utilities (WWU) response', 6 March, GDQ50 and GDQ53, pp. 66–72.

- **The need to account for operational trade-offs**: relating to the need for consistently allocated cost/activity data across GDNs (to compare activities on a like-for-like basis), disaggregated models would need to be carefully specified to account for operational trade-offs (i.e. potential substitution or complementarity between different types of spend addressing the same outcomes).[46] Given consistently reported data, these issues could notionally be addressed within the modelling framework. However, the evidence to date suggests that there is not sufficient alignment on the relevant cost pools, nor the consistency in reporting required.

In contrast, top-down TOTEX models, are not subject to the same concerns due to their level of aggregation.

Further, for a suite of disaggregated models to be considered as an alternative/complementary basis for cost determinations, these models collectively need to be deemed at least as robust as, and provide additional insight on top of, the top-down modelling. We note that the disaggregated models tested by Ofgem through the CAWG process perform significantly worse than the top-down models.[47]

Assuming the issues above can be resolved and disaggregated models are used for cost determinations, the efficiency benchmark for disaggregated models would still need to be determined at the aggregate, TOTEX, level to avoid cherry picking.[48] If the benchmark were chosen at the disaggregated activity level (where GDNs' operational focus and strategy may differ, such as relying on OPEX solutions rather than CAPEX solutions and vice versa), it would create a notional 'super-efficient GDN' aggregate benchmark. Such a benchmark would be too stringent and not based on what any GDN can achieve across all its activities in practice.[49]

[46] For example, there are likely to be differences between GDNs in whether OPEX- or CAPEX-intensive activities are the most efficient solutions for specific outcomes (e.g. asset maintenance or replacement).
[47] For example, comparing model fit and other statistical results of Ofgem's TOTEX models relative disaggregated models in Ofgem (2024), 'RIIO-GD3 Cost Assessment Working Group 7. Totex modelling and BPDT development', 10 April, slides 10–15.
[48] That is, by aggregating up individual model cost predictions first, and then determining the benchmark.
[49] That is, if benchmarks are first set for each individual activity and then aggregated to the TOTEX level, it would be for a notional 'super-efficient company' that performs at the frontier on each activity, while this performance is likely to be impossible for any single GDN to achieve in practice.

3.3 Use of historical and forward-looking data

Data-period selection is critical for both estimating the cost–cost driver relationships (for efficient cost prediction) and subsequently identifying the catch-up efficiency benchmark (discussed in section 3.4). For both the modelling and benchmark selection periods, respectively, Ofgem effectively has three options:

- to use only historical data;
- to use some combination of historical and forecast data (the modelling period used at GD2 FD);
- to use only forecast data (the benchmark period at GD2 FD).

Theoretically, it is generally advised to use as much data as possible to estimate cost–cost driver relationships, as long as data is available on a consistent basis and the underlying relationships remain stable over time. Among other reasons, this is because more data:

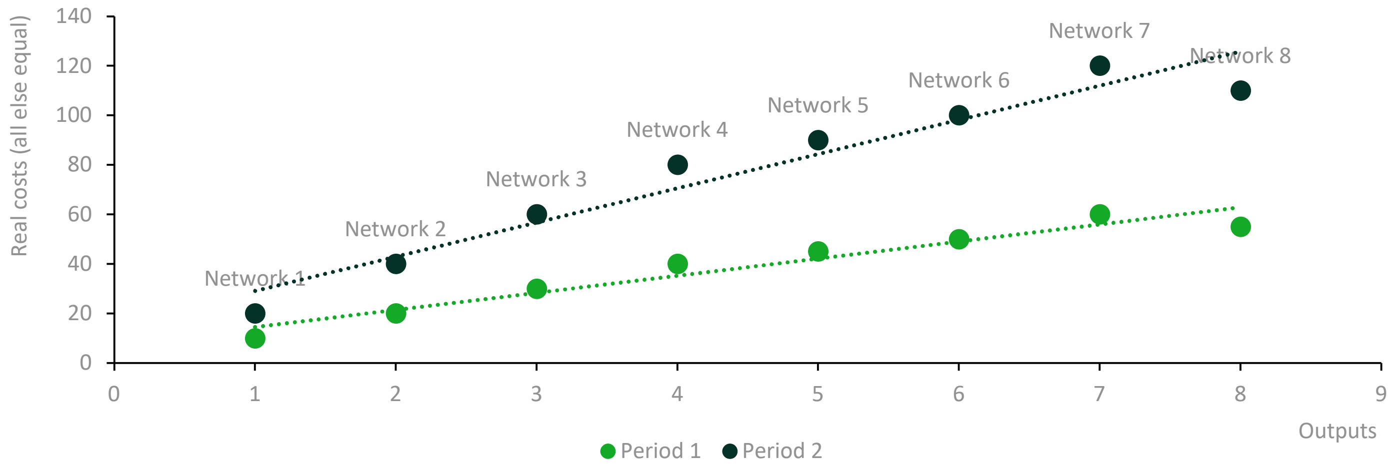
1. increases sample size and therefore also the precision of the estimated cost–cost driver relationship;
2. allows for the 'smoothing out' of the impact of cyclical and lumpy expenditures such as maintenance and capital spend (e.g. installing new IT systems);
3. allows for triangulation across information contained in both historical and forecast data.

Both historical and forecast data may contain useful information for the purposes of cost benchmarking, but may also have shortcomings that limit their reliability in a given context. The benefits and potential shortcomings of each type of data are as follows.

- **Historical data** has the benefit that the estimated cost–cost driver relationships are estimated on the basis of actual outturn (i.e. it is not subject to potential forecasting errors, information asymmetry/gaming by companies, or uncertainty about the future). However, it may be less appropriate in circumstances where the fundamental relationship between costs and cost drivers are changing in a systematic way. If this is the case, the older the data, the less relevant it is likely to be.
- **Forecast data** is inherently more uncertain, but may contain more relevant information about future cost pressures (based on networks' engineering/operational expertise, which may not be observable in historical data).

As Ofgem has noted during the CAWGs, it needs to test for structural breaks (with the inclusion of GD3 BP data),[50] given initial evidence of a step change in costs and a change in workload/activities driving it. That is, Ofgem should test for changes in cost–cost driver relationships over time (e.g. between the different regulatory periods, and between outturn and forecast data). This is illustrated in the hypothetical example in Figure 3.1. The figure shows a notional change in cost per output between two periods, where the complexity/real cost per output changes systematically over time—such that there is both a general shift upwards in costs and a steeper relationship between cost and outputs in the second period.

Figure 3.1 Notional step change between historic and future costs



Source: Oxera illustration.

If there is evidence of a structural break (i.e. a fundamental change in the estimated cost–cost driver relationships over time[51]), there are a few ways in which Ofgem could reconsider its modelling approach.

As discussed above, potential (at least partial) solutions include:

- adding/modifying cost drivers so that they can capture cost pressures or a changing workload mix (e.g. REPEX);

[50] Ofgem (2024), 'RIIO-GD3 Cost Assessment Working Group 7. Totex modelling and BPDT development', 10 April, slide 7.
[51] Such that separate models over the respective periods produce a better model than the combined regression over the entire period. A typical statistical test is the Chow test. Chow, G.C. (1960), 'Tests of Equality Between Sets of Coefficients in Two Linear Regressions'. *Econometrica*, **28**:3, pp. 591–605.

- separately assessing new costs and activities (such as IT, cyber security and shared group costs), if they are not readily accommodated within the main TOTEX modelling framework (discussed in section 3.5).

These are the minimum necessary considerations required if the structural break is being driven by new activities or cost pressures that are not (or insufficiently) accounted for by the model's cost drivers.

Additionally, Ofgem may also need to reconsider how much weight it places on historical and forecast data—for example by:

- placing greater weight on more recent and/or forward-looking costs by using similar models over alternative time periods (e.g. GD2 and GD3 only, or GD3 forecasts only);
- testing alternative time dummies, trends, and multiplicative terms (to capture changes in the strength of relationships over time).

Ofgem reconfigured its ED2 modelling approach based on similar considerations (see Box 3.1 below). Here Ofgem:

1. amended cost drivers (specifically, forward-looking drivers) to better capture the expected step change in costs;
2. excluded and separately assessed other costs where there was a significant changes in costs (but not captured/ accommodated in the main modelling suite);
3. placed more weight on forecast data. Ofgem has indicated that a similar differential weighting between historical and forecast data 'could be a key element in constructing a multiple model approach' at GD3.[52]

We note that there are cases where regulators may consider it appropriate to use only historical data to estimate cost–cost driver relationships, and then apply these estimates to forecast workloads to estimate future costs (e.g. Ofwat's approach to base costs). However, such an approach would not be appropriate when GDNs expect a step change in forward-looking cost pressures due to the below.

- **Modelling only historical data is only appropriate in approximately 'steady-state' environments** and requires a

[52] Ofgem (2024), 'RIIO-GD3 Cost Assessment Working Group 7 Totex modelling and BPDT development', 10 April, slide 7.

separate treatment for other costs. For example, Ofwat only considers historical cost–cost driver relationships for the 'business-as-usual', base cost element of its separate base cost-enhancement cost framework (see Box 3.2).

- **Forecast data often requires further regulatory judgement**, in the form of additional overlays to account for additional forward-looking cost pressures not accounted for in historical data (e.g. Ofwat's adjustments for mains and smart metering renewals and other factors at PR24—also in Box 3.2).

Box 3.1 Ofgem's approach to forward-looking cost drivers at ED2

At RIIO-ED2, in light of new activities associated with the energy transition (and thus the systematic changes in forward-looking cost–cost driver relationships expected), Ofgem altered its modelling framework as follows.

- Amending the cost drivers used in the models: Ofgem considered additional cost drivers for capacity released (instead of energy distributed), based on the view that it 'better controlled for the step up in reinforcement activities', as well as forward-looking cost drivers for low-carbon technology (LCT) uptake (as a proxy for incremental energy demand and thus the increased network reinforcement required in future).
- Excluding and separately assessing some categories of costs where there was a significant change in the required level of costs and/or the costs were not captured by any cost drivers in the models (e.g. cyber and physical security).
- Placing more weight on forecast data: Ofgem used three equally weighted top-down TOTEX models for cost determinations at ED2—one of which (considering LCT uptake as a driver) was based solely on forecast data.

Source: Ofgem (2022), 'RIIO ED2 Final Determinations Core Methodology Document', 30 November, section 7.

Box 3.2 Ofwat's approach to forward-looking cost drivers at PR24

Ofwat uses a split TOTEX approach, distinguishing between base costs and enhancement costs (both assessed on a TOTEX basis).

Base expenditure relates to wholesale and retail activities that include routine activities, cyclical asset maintenance and other base activities (e.g. to improve efficiency and meet legal obligations). These are modelled in a few BOTEX (base TOTEX) models for the respective service areas (wholesale water, wastewater, bioresources and retail).

The cost–cost driver relationships for base activities are estimated on historical outturn data, with these coefficients then applied to companies' cost-driver forecasts to arrive at cost predictions. Ofwat compared company explanatory variable forecasts against historical growth rates and/or independent third-party projections (e.g. ONS household projections), and in some cases adjusted company forecasts where the forecasts diverged from these without sufficient justification.[53]

In contrast to PR19, in PR24 Ofwat further applies sector-wide cost adjustments for forward-looking cost pressures not appropriately captured in its BOTEX modelling (e.g. higher mains renewal rates required, meter renewals, phosphorus removal, net zero and energy costs).[54]

However, enhancement activities are assessed separately.

Enhancement expenditure relates to activities where there is a permanent increase in the current level of service (to a new 'base' level) and/or the expanding services to new customers.[55]

[53] Ofwat (2024), 'PR24 draft determinations: Expenditure allowances', 12 July, sections 2.1.1–2.1.2.
[54] Ofwat (2024), 'PR24 draft determinations: Expenditure allowances', 12 July, section 2.2.
[55] Enhancement generally relates to environmental obligations, improving service quality and resilience (e.g. water quality, reducing leakage), and providing new solutions for water provision in drought conditions.

Enhancement costs are assessed by activity/area, using a mix of top-down econometric models, simple unit costs (ratio analysis) and engineering assessments—depending on the enhancement activity.[56] Here, forecast costs and cost drivers, at times complemented by historical outturn data, are used to benchmark costs.

Source: Ofwat (2024), 'PR24 draft determinations: Expenditure allowances', 12 July.

3.4 Benchmark choice

The choice of catch-up efficiency benchmark reflects the regulators' best *estimate* of what the efficient company cost level will be (in this case, over GD3). The choice of benchmark therefore depends on the robustness and reliability of the models, and thus on how much confidence and certainty the regulator can have in the resulting efficient cost estimates.

This is a common feature of corrected ordinary least squares (COLS) models, such as the ones used in Ofgem's GD2 cost assessment—they cannot distinguish between true differences in efficiency and noise in the data/modelling error. Therefore, the regulator needs to practise caution in setting the catch-up benchmark, proportionate to the extent to which it believes that the gap between the 'frontier' and lagging companies is indeed driven by true efficiency differences (and not statistical noise). This is why, for example, regulators often use median/average benchmarks for new and more uncertain cost assessments, but more stringent benchmarks (e.g. an upper quartile) for more robust and established cost models.[57]

In statistical terms, this is often primarily thought of in terms of model precision (i.e. level of uncertainty around the model's point estimates), alongside other measures of model performance and the reliability of the underlying data used (discussed below). For example, precision was one of the key metrics considered by the Competition and Markets Authority (CMA) during recent disputes relating to benchmark choice for

[56] Ofwat (2024), 'PR24 draft determinations: Expenditure allowances', 12 July, section 3.2.
[57] For example, Ofwat mostly uses a median/average benchmark as the catch-up efficiency challenge in (what are fundamentally less certain) enhancement models, but an upper quartile benchmark for the base costs, which benefits from long-established, robust models and more predictable costs. See Ofwat (2024), 'PR24 draft determinations: Expenditure allowances', 12 July, sections 2 and 3.

setting cost allowances in the energy and water sectors.[58] It is also considered by Ofwat when it assesses the potential impact of mergers on its ability to do robust comparative benchmarking (e.g. the recent acquisition of SES Water[59]).

At GD2, Ofgem set a glide path from the upper quartile (75th percentile) at the start to a more stringent 85th percentile benchmark by the end of the period. This was challenged during the GD2 appeals, with the CMA eventually considering that Ofgem was justified in doing so. In making its decision, the CMA gave weight to Ofgem's regulatory judgement and the fact that improved data quality and additional years of data were expected to improve the reliability of the modelling (at least to some extent).[60] However, the CMA noted that there was limited statistical evidence to support Ofgem's view that the precision of the models had improved (given comparability issues between the GD1 and GD2 modelling suites, which were not taken into account).[61]

The use of an 85th percentile benchmark for GD3 may not be justified if the robustness of the modelling suite decreases at GD3. This is something that needs to be tested empirically.

Once Ofgem resumes its model testing with the inclusion of GD3 forecast data, it will be important to consider the following factors (and their underlying measures) in deciding on the appropriate catch-up efficiency benchmark.

- **Statistical precision**: which can be measured through (i) the confidence intervals around GDN cost predictions, (ii) the resulting efficiency score distribution (and narrowness therein[62]), or (iii) conducting stochastic frontier analysis (providing a data-driven assessment of the amount of noise in the models[63]).

[58] CMA (2021), 'Cadent Gas Limited, National Grid Electricity Transmission plc, National Grid Gas plc, Northern Gas Networks Limited, Scottish Hydro Electric Transmission plc, Southern Gas Networks plc and Scotland Gas Networks plc, SP Transmission plc, Wales & West Utilities Limited vs the Gas and Electricity Markets Authority. Final determination Volume 3: Individual Grounds' [from here, Final determination Volume 3: Individual Grounds], 28 October, paras 12.135 and 12.139–12.140; and CMA (2021), 'Anglian Water Services Limited, Bristol Water plc, Northumbrian Water Limited and Yorkshire Water Services Limited price determinations. Final report', March, paras 4.493–4.494.
[59] Ofwat (2024), 'Ofwat's Opinion on Pennon's acquisition of SES Water', p. 11.
[60] CMA (2021), 'Final determination Volume 3: Individual Grounds', 28 October, paras 12.138–12.140.
[61] CMA (2021), 'Final determination Volume 3: Individual Grounds', 28 October, para. 12.135(c).
[62] The narrower the confidence intervals or efficiency score distribution, the more precise the model.
[63] Data-driven approaches like SFA, in contrast, allow one to start to disentangle how much of the assumed efficiency gap between companies and the benchmark estimate is due to statistical noise, as opposed to true inefficiency.

- **Other indicators of model performance**, including model fit, the sign and significance of the respective cost driver coefficients (i.e. the direction and magnitude should align with operational and economic insight), and other statistical tests (like sensitivity to the removal of specific observations).
- **The reliability (and comparability) of the underlying data**, which may be more of a concern if Ofgem relies to a greater extent on disaggregated analysis (without addressing reporting inconsistencies and operational trade-offs). This is also dependent on the quality of company forecasts—which Ofgem incentivises through its business planning (so-called 'truth-telling') incentives.

Also note that, given the relatively small dataset available with a limited number of independent observations (eight networks and four groups), the robustness of any statistical model in the current context will be limited. This is not necessarily improved by taking into account more years of data, given that:

- observations for a particular company are likely to be similar over time (and, in particular, many cost drivers are relatively stable over time);
- historical data may be less relevant if the nature of workloads or cost drivers changes significantly in GD3 (see section 3.3).

The five-year GD3 forecast period is likely to remain most appropriate to establish the benchmark. This ensures that the benchmark is based on the most recent and relevant data to reflect both current efficiency levels, and the expected cost pressures over GD3 faced by the sector. A longer benchmark period also helps to smooth out differential expenditure profiles across GDNs (e.g. lumpy investments or some GDNs planning to take on more complex/costly workloads earlier than others).

However, to ensure the catch-up challenge is not determined on the basis of an artificially stringent benchmark, minimum service delivery performance and differences in time profiles of capital spend across the relevant GDNs should be examined. That is, it will be important to ensure that benchmark GDNs are not merely appearing more efficient because they are underinvesting or not meeting minimum service standards, and that efficient cost predictions are not downwardly biased by certain GDNs avoiding more complex and costly mandated workloads.

For example, at GD1 some GDNs influencing the cost benchmark failed their emergency standards—so creating an inappropriate efficiency challenge for other GDNs meeting these commitments. In response, Ofgem made an upward cost adjustment for these GDNs' historical costs, to reflect the additional costs that would notionally have been required to meet the standard.[64] As discussed in section 2.3, some GDNs are not meeting their emergency (and also REPEX) minimum service standards in GD2—which would require similar consideration.

Ofgem may also want to conduct a similar assessment to that undertaken by Ofwat at PR24. Ofwat assessed whether companies at or above the benchmark were in a maintenance trough (i.e. had uncharacteristically low capital maintenance spend) over the relevant five-year period[65]—to assess whether the benchmark might have been set at an artificially too stringent level.

3.5 Separate assessment and group economies of scale

At GD1, Ofgem assessed BSCs via non-regression methods. It conducted expert reviews and bottom-up benchmarking using activity drivers. Notably, this assessment was conducted at the group (not GDN) level.[66] At GD2, Ofgem included BSCs in the main TOTEX regression, because it considered 'that costs for these activities should be stable over time'.[67]

This no longer appears to be the case. As discussed in sections 2.2.2 and 3.1.2, WWU is expecting a substantial change in the nature of the activities being undertaken (and therefore a step change in related costs). Therefore, the most appropriate way to deal with a step change to BSCs (or subcategories thereof) may be to exclude them from the main regression model and undertake a separate assessment.[68]

Furthermore, in gas distribution, as for electricity distribution, BSCs (and other shared OPEX costs) consist of cost categories that one would expect to benefit from economies of scale at the group level. For example, larger companies (groups), such as Cadent, can share CEO and group management, HR, finance and regulatory costs across four GDNs, while WWU has to absorb all these costs within one GDN. Table 3.1 summarises the broader categories of costs that WWU indicates would benefit from group-level economies of scale.

[64] Ofgem (2012), 'RIIO-GD1: Final Proposals - Supporting document - Cost efficiency', 17 December, para. 6.8.
[65] Ofwat (2024), 'PR24 draft determinations: Expenditure allowances', 12 July, pp. 24–27.
[66] Ofgem (2012), 'RIIO-GD1: Final Proposals - Supporting document - Cost efficiency.', 17 December, appendix 6.
[67] Ofgem (2020), 'RIIO-2 Draft Determinations – GD Annex.', 9 July, para. 3.78.
[68] That is, if appropriate cost drivers to pick up the change in the nature and complexity of the activities, and thus the step change in costs, cannot be accommodated within the TOTEX-level modelling.

Table 3.1 Categories of costs that may benefit from group-level scale effects

Category of cost	Subcategory
BSCs	IT & Telecoms
	Property Management
	HR & Non-Operational Training
	Audit, Finance & Regulation
	Insurance
	Procurement
	CEO & Group Management
Work management	System control
	Asset management (esp. strategy and integrity)
Other OPEX (including uncertainty mechanisms)	Net Zero (e.g. studies)
	Physical and Cyber security
	Safety and Assurance
	Environment (reporting and compliance)
	Transport

Source: Ofgem cost categorisation in GD2 normalisation files.

Benchmarking shared group costs at the GDN level may also introduce measurement error, as individual GDN costs may be influenced by reallocations of shared-group-level indirect expenditures across the relevant networks.[69]

It will therefore be important to test for group-level economies-of-scale effects with the inclusion of GD3 data (e.g. by including group-level scale drivers to GDN-level models, comparing results of modelling at the group versus network level). However, given the relatively small sample (only eight GDNs and four ownership groups), more weight should be placed on industry operational insight than on (imprecisely estimated) statistical results. For example, if one or two of the GDNs in larger groups are very inefficient (and/or singleton GDNs are very efficient, vice versa), the models may not be able to pick up the underlying economies-of-scale group-level benefits.

[69] For example, we note that group-level cost allocation data is already collected and reported for shared BSCs in sheet 4.02 of GDN RRPs.

We note that a group-level assessment of BSCs would be more consistent with Ofgem's approach at ED2 (and GD1), where core BSCs were assessed at the group level. At ED2 Ofgem used a group-level, bottom-up regression of core BSCs with MEAV as the explanatory variable.[70]

Precedent from ED2 also suggests that specific subcategories of costs may require their own separate assessments: at ED2, IT costs (CAPEX and OPEX combined) and property management had distinct bottom-up assessments, with MEAV as the cost driver,[71] while cyber security was technically assessed.[72]

Similarly, the appropriate distinct assessments at GD3 may include the following.

- **Cyber security** solutions could be technically assessed, based on the same approach that Ofgem pursued at ED2. Here, it excluded cyber security from the main regressions because of a '[s]ignificant change in the equivalent level of costs between the RIIO-ED1 and RIIO-ED2 periods'.[73] The same should be expected for gas infrastructure, as GDNs are required to make large upfront investments to install the relevant technologies and maintain higher-base-level BSCs thereafter.
- **IT costs** could be based on a separate assessment similar to ED2, using a median benchmark based on ratio analysis (i.e. unit cost benchmarking). At ED2, Ofgem based the unit cost on a subset of MEAV as the cost driver. Ofgem also used only ED2 forecast data, 'in recognition of the step change in IT costs when compared to RIIO-ED1'.[74] Similarly, if there is evidence of a structural break in the cost–cost driver relationship at GD2, it would be appropriate for Ofgem to base the assessment on more recent/forecast data only.

[70] Ofgem (2022), 'RIIO ED2 Final Determinations Core Methodology Document', 30 November, paras 7.509–7.513.
[71] Ofgem (2022), 'RIIO ED2 Final Determinations Core Methodology Document', 30 November, paras 7.291–7.293 and 7.391–7.395.
[72] Ofgem (2022), 'RIIO ED2 Final Determinations Core Methodology Document', 30 November, paras 7.22–7.23, 7.47 and 7.58
[73] Ofgem (2022), 'RIIO ED2 Final Determinations Core Methodology Document', 30 November, p. 229.
[74] Ofgem (2022), 'RIIO ED2 Final Determinations Core Methodology Document', 30 November, para. 7.295.

4 Concluding thoughts

WWU's position seems to be deteriorating over GD2, based on initial remodelling of outturn data to date. However, this does not seem to relate to a relative increase in inefficiency, but rather an exogenous increase in the complexity of its mains replacement, and potentially also IT and cyber security workloads (if not excluded from the top-down modelling). These cost pressures, resulting from a change in the nature of the underlying activities, would not be accounted for under Ofgem's current GD2 modelling suite.

This has several modelling implications for Ofgem's GD3 modelling suite.

- **Time period used**: as the step change in costs is expected across the entire sector, it will be important to conduct tests for a structural change in cost–cost driver relationships. Contingent on the results thereof (for each relevant cost assessment category), there may be a need to reconsider the relative weighting/treatment of historical and forecast data in its benchmarking.
- **REPEX complexity**: the REPEX synthetic cost driver (and accompanying regional factor normalisations) will need to be reconsidered, to account for additional workload complexity drivers such as the technique, ground surface, pipe material (ductile vs cast/spun iron) and sparsity of workloads.
- **IT, cyber and BSCs**: The cost drivers for these costs would also need to be reconsidered and/or these costs should potentially be separately assessed to account for the step change in requirements. This is driven primarily by increased IT and cyber security requirements, but also more generally, it would be important to account for group-level economies of scale in these shared costs (as Ofgem has done at ED2).

Finally, we note that the GD2 (TOTEX) level of aggregation and use of the full RIIO-GD3 period forecast data as the benchmarking period is likely to remain appropriate for cost determinations (to both smooth out differential expenditure profiles and incorporate GDN latest views on cost pressures). However, this depends on the quality of the underlying data (especially cost allocation concerns) and robustness of the models used at GD3. It will also be important to take care that the benchmark is not influenced by GDNs that underinvesting or not meeting minimum service standards.

Contact

Alan Horncastle
Partner
+44 (0) 1865 253015
alan.horncastle@oxera.com

oxera.com